FINANCIAL MATHEMATICS TEAM CHALLENGE

A collection of the five reports from the 2017 Financial Mathematics Team Challenge.

AIFMRM

AFRICAN INSTITUTE OF FINANCIAL MARKETS AND RISK MANAGEMENT

Preamble

One of the key aims of the FMTC is for South African postgraduate students in Financial and Insurance Mathematics to have the opportunity to focus on a topical, industry-relevant research project, while simultaneously developing links with international students and academics in the field. An allied purpose is to bring a variety of international researchers to South Africa to give them a glimpse of the dynamic environment that is developing at UCT in the African Institute of Financial Markets and Risk Management. The primary goal, however, is for students to learn to work in diverse teams and to be exposed to a healthy dose of fair competition.

The Fourth Financial Mathematics Team Challenge was held from the 18th to the 28th of July 2017. The challenge brought together five teams of Masters and PhD students from Australia, Germany, South Africa and the UK to pursue intensive research in Financial Mathematics. Each team worked on a separate research problem during the twelve days. Professional and academic experts from Brazil, Switzerland, Australia, Austria, Canada, South Africa, and the UK individually mentored the teams; fostering teamwork and providing guidance. As they have in the past, the students applied themselves with remarkable commitment and energy.

This year's research included topical projects on (a) portfolio risk diversification, (b) early-warning systems for financial crises and long-term asset management, (c) model calibration via neural networks, (d) the management of parameter estimation risk in a Mean-Variance portfolio optimisation exercise, and (e) rough volatility modelling paradigms. These were either proposed directly by our industry partners or chosen from areas of current relevance to the finance and insurance industry. In order to prepare the teams, guidance and preliminary reading was given to them a month before the meeting in Cape Town. During the final two days of the challenge, the teams presented their conclusions and solutions in extended seminar talks. The team whose research findings were adjudged to be the best was awarded a floating trophy. Each team wrote a report containing a critical analysis of their research problem and the results that they obtained. This volume contains these five reports, and will be available to future FMTC participants. It may also be of use and inspiration to Masters and PhD students in Financial and Insurance Mathematics.

FMTC IV was a great success, so 2018 and FMTC V are already in the pipeline!

***David Taylor**, University of Cape Town*
***Andrea Macrina**, University College London & University of Cape Town*

Contents

[1] Winning team of the fourth Financial Mathematics Team Challenge

Realistic Risk Parity

Team 1

DANIELLE CRAWFORD, University of Cape Town
CIAN CULLINAN, University of Cape Town
YANN GUGUEN, University College London
RALPH RUDD, University of Cape Town

Supervisor:
RODRIGO TARGINO, Fundação Getulio Vargas

African Institute of Financial Markets and Risk Management

Contents

1 Introduction

An important concept in portfolio construction is *diversification*. The central hypothesis of diversification is that some asset classes should perform well when others perform badly, thus the profit and loss of a diversified portfolio should remain relatively stable when there is a downturn in a single asset class.

Since 2008, portfolios that are based on *risk* diversfication have become more popular. These strategies are said to have something from US $ 150 billion to US $ 1.5 trillion invested in them[1]. The idea of such risk parity portfolios is to diversify the risk of a portfolio rather than its financial exposure.

However, through May 2013 and late 2015, there was a downturn in several asset classes, and portfolios that were (supposedly) based on risk parity principles performed poorly during this period[2]. The aim of this project is to investigate why this might be the case. The central research question is:

> "How do risk parity portfolios perform under realistic assumptions for the movements of the underlying asset returns?"

The goal is to determine whether the downturn in risk parity portfolios was due to the concept of risk parity being flawed, or rather whether there are pitfalls in the standard modelling approach.

The construction of risk parity portfolios when standard deviation is used as the risk measure is neatly reviewed by Maillard et al. (2010) and generalised to other risk measures by Tasche (2007). An investigation into the real-world performance for a very small sample of assets is performed in Stefanovits (2010).

This projects proceeds in five stages. First, a simulation environment is established for varying levels of model tractability and realism. Then, the construction of risk parity portfolios is examined. Third, a numerical investigation is performed where the simulation environment is used to test the performance of the risk parity portfolios. After the numerical investigation is complete, the most realistic model framework is fitted to 31 assets in a real-world data set. Finally, the performance of risk parity portfolios is backtested against a limited sample of historical data.

[1] https://finance.yahoo.com/news/quants-fire-back-paul-tudor-172047543.html

[2] http://www.salientindices.com/risk-parity.html

Figure 1: Four fifths of Team 1 - The Psychic Pony Skeletons.

2 Simulation

For the purpose of this investigation, we wish to be able to simulate returns from multiple assets with a complex dependency structure. This section outlines several simulation methods. First we will consider simulating from the multivariate normal (MVN) and multivariate t (MVt) distributions. Following that, we will consider a more realistic dependency model using copulas. Finally, we combine the copula dependency structure with a simple time-series model. This last case follows recommendations set forward by Nystrom and Skoglund (2002).

2.1 Multivariate Distributions

Simulating from the MVN distribution is relatively straightforward. Let us assume we wish to simulate n returns of N assets. We wish to simulate a matrix of returns $R \in \mathbb{R}^{N \times n}$ given a column vector of means $\mu \in \mathbb{R}^{N \times 1}$ and a covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$. We first compute the lower triangular matrix, L, given by the Cholesky decomposition of the covariance matrix, i.e., $\Sigma = LL^{\top}$. Then given a matrix of

random standard normal variates $Z \in \mathbb{R}^{N\times n}$;

$$R = \mu + LZ$$

where Z is simulated in MATLAB using the command `randn`(N,n). As we are simulating returns, μ is taken as a vector of zeros.

While the MVN distribution is a good starting point, it fails to capture many of the stylized facts of asset returns. One of these facts, the presence of fatter tails, is better captured by a Student's t distribution. To simulate $R \in \mathbb{R}^{N\times n}$ from a MVt distribution, we consider a similar procedure as with the MVN distribution and introduce a degrees of freedom parameter ν. The returns matrix is given by

$$R = \mu + LY,$$

where $Y \in \mathbb{R}^{N\times n}$ is a matrix of random Standard Student t variates with a common degree of freedom ν. It is important to note that in this case Σ is actually not the covariance matrix of the outputted returns, but rather a scaled version of it. The true covariance of the returns is given by

$$\mathrm{Cov}(R) = L\mathrm{Cov}(R)L^T = \frac{\nu}{\nu-2}\Sigma.$$

Example 1. Consider the two asset (bivariate) case. Given a correlation ρ and fixed degrees of freedom $\nu = 3$, we can create contour plots for changing ρ. We notice how in both instances, as the correlation becomes large positive or negative the data groups closer together. The impact of the Student's t distribution is also displayed in Figure 2 as we can see a wider spread in the contours, due to the heavier tails, when compared to the MVN distribution.

These two methods of simulation give us returns with distributions which are tractable, however we would like to consider a more realistic model for returns and their underlying dependency structures. The next section introduces the concept of copulas and how to simulate returns data using them.

2.2 Copulas

When modelling the returns of multiple assets, we wish to capture both the characteristics of each individual asset and the dependence structure between the as-

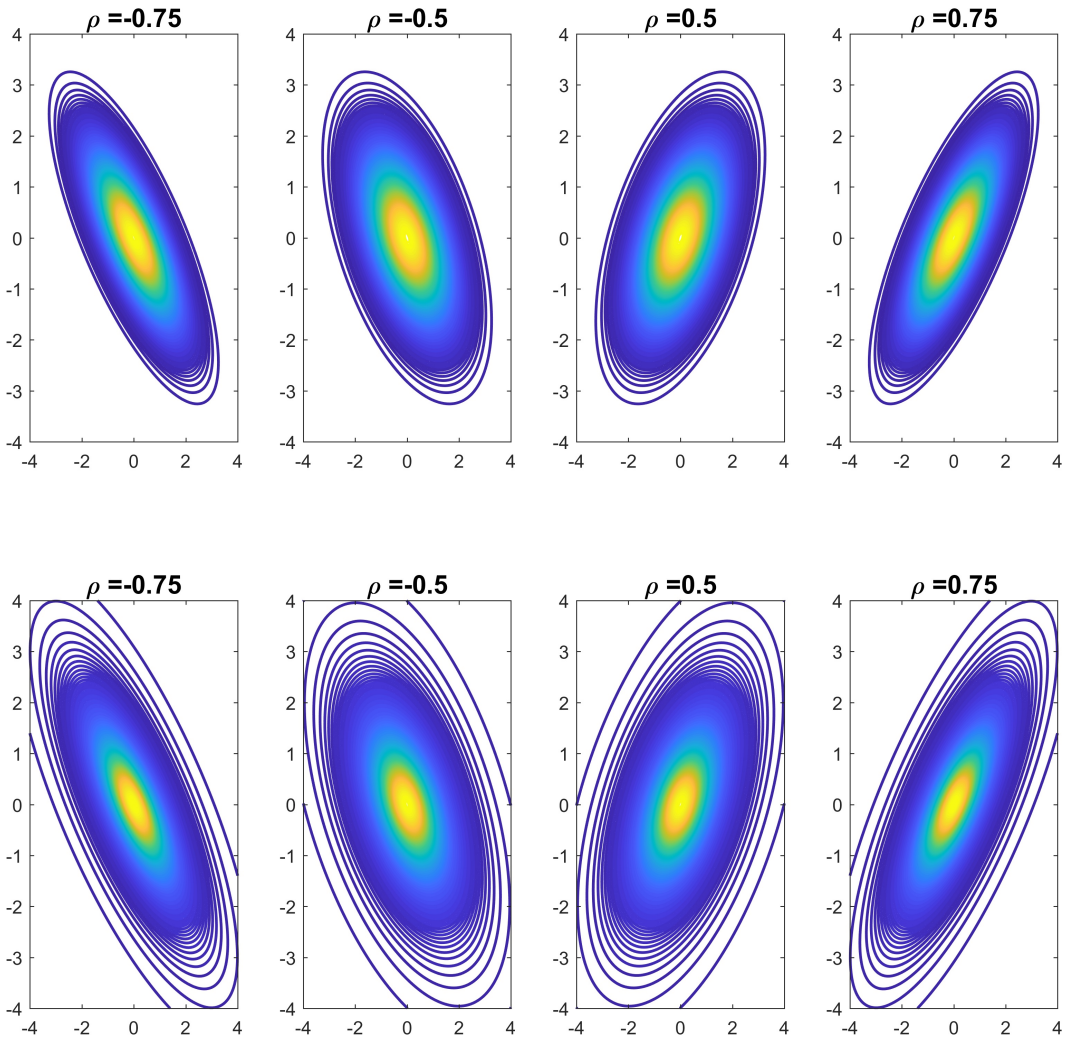


Figure 2: The bivariate Gaussian (MVN) distribution with varying correlation.

sets as best as possible. When using a multivariate distribution (such as the MVN or Mvt), we describe both of these using a joint distribution. Copulas allow us to divide up the construction of this joint distribution such that the individual asset's characteristics are described by marginal distributions while their dependence structure is described by the copula (Nystrom and Skoglund, 2002).

The fundamental theorem underlying copula analysis is Sklar's theorem, see Sklar (1959).

Theorem 2 (Sklar's Theorem)**.** Let H be an n-dimensional distribution function with marginals $F_1, \ldots, F_n$. Then there exists an n-copula C such that $\forall \mathbf{x} \in \mathbb{R}^n$,

$$H(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)).$$

and if $F_1, \ldots, F_n$ are all continuous, then C is unique. Conversely, given a copula $C : [0,1]^d \to [0,1]$ and margins $F_i(x)$ then $C\left(F_1(x_1), \ldots, F_d(x_d)\right)$ defines a d-dimensional cumulative distribution function.

Now, if we let F_i be a continuous and invertible univariate distribution function for $i = 1, \ldots, n$, we have the following corollary from Nelsen (1999) that allows us to specify a copula in terms of distribution functions.

Corollary 3. Let H be an n-dimensional distribution function with continuous marginals $F_1, \ldots, F_n$ and copula C. Then for any $\mathbf{u} \in [0,1]^n$,

$$C(u_1, \ldots, u_n) = H(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)).$$

We will consider two elliptical copulas, the Gaussian and Student's t, as well as the Clayton Archimedean copula.

2.2.1 Measures of dependence

Copulas are used to describe the dependence structure existing between random variables. It is therefore useful to have some measure of this dependence. The most common measure of dependence between two random variables, X and Y, is Pearson's correlation

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}.$$

Pearson's correlation is a linear measure of dependence, which is suitable for elliptical multivariate distributions, such as the MVN and MVt distributions, but is not appropriate for the non-elliptical joint distributions which arise when using copulas, see (Embrechts et al., 2001). More suitable measures of dependence are Kendall's tau and Spearman's rho, which are invariant under non-linear transforms. We shall consider Kendall's tau for this paper.

Definition 4 (Kendall's tau)**.** Consider two random variables X and Y, then Kendall's tau is defined as

$$\tau(X, Y) = \mathbb{P}[(X - \bar{X})(Y - \bar{Y}) > 0] - \mathbb{P}[(X - \bar{X})(Y - \bar{Y}) < 0]$$

where (X, Y) and $(\bar{X}, \bar{Y})$ are i.i.d.

Embrechts et al. (2001) provide us with a simple closed form solution for the Kendall's tau when considering a copula C with two continuous random variables X and Y.

$$\tau(X, Y) = 4\mathbb{E}[C(U, V)] - 1$$

where $U, V \sim U(0, 1)$. We shall revisit Kendall's tau when we discuss Archimedean copulas.

2.2.2 The Gaussian copula

When considering a copula we are uninterested in the marginals and the characteristics of each asset, and only consider the dependency between them. We can therefore fully specify the Gaussian copula by its correlation matrix. It is important to note here that adding a mean vector to the Gaussian copula has no effect and therefore we will always assume zero mean when simulating.

Definition 5 (Gaussian copula)**.** Let Σ_ρ be a correlation matrix with $\mathrm{diag}(\Sigma_\rho) = 1$ and Φ_ρ the standardised Multivariate Normal distribution function with correlation matrix Σ_ρ. The Gaussian copula is then defined as:

$$C(u_1, \ldots, u_n; \rho) = \Phi_\rho(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n))$$

where $\mathbf{u} \in [0, 1]^n$.

Let us return to our problem of simulating n returns from N assets in the form of the returns matrix $R \in \mathbb{R}^{N \times n}$. Let us assume the we know the form and the

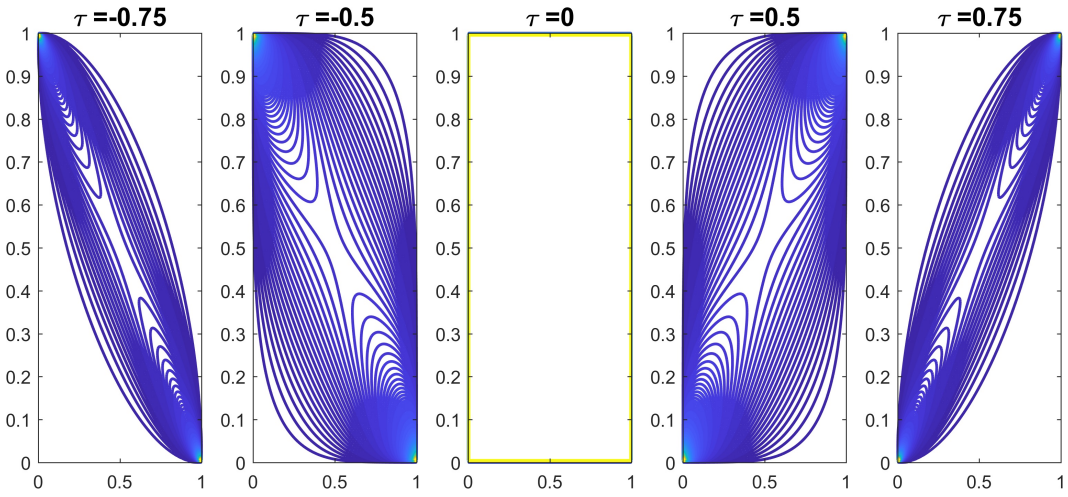


Figure 3: The bivariate Gaussian copula with varying Kendall's tau.

parameters of the marginal distributions $F_1, \ldots, F_N$ and we want to simulate from a Gaussian copula with correlation matrix Σ_ρ.

First we simulate a matrix $X \in \mathbb{R}^{N \times n}$ of MVN random numbers using the the method outlined previously. Since the means are irrelevant we can ignore them (i.e. $X = LZ$). We then transform numbers to lie on $[0, 1]$ by inputting them into the univariate standard normal cumulative distribution function, Φ. This captures the dependence structure of the assets in a unit hypercube. Applying the inverses of the marginal distribution functions $F_1^{-1}, \ldots, F_N^{-1}$ to each row of this matrix gives R with the desired returns structure.

Example 6. Consider the bivariate case of the Gaussian copula. In this case, the dependence structure is completely specified by Kendall's tau. In Figure 2.2.2, we plot contours of the generated unit square for varying Kendall's tau. The $\tau = 0$ case implies a uniform copula, with no dependence between the returns.

2.2.3 The Student's t copula

Definition 7 (Student's t copula)**.** Let Σ_ρ be a correlation matrix with $\mathrm{diag}(\rho) = 1$ and $T_{\rho,\nu}$ the standardised Multivariate Student's t distribution function with correlation matrix Σ_ρ and ν degrees of freedom. The Student's t copula is then defined as:

$$C(u_1, \ldots, u_n; \Sigma_\rho, \nu) = T_{\rho,\nu}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_n))$$

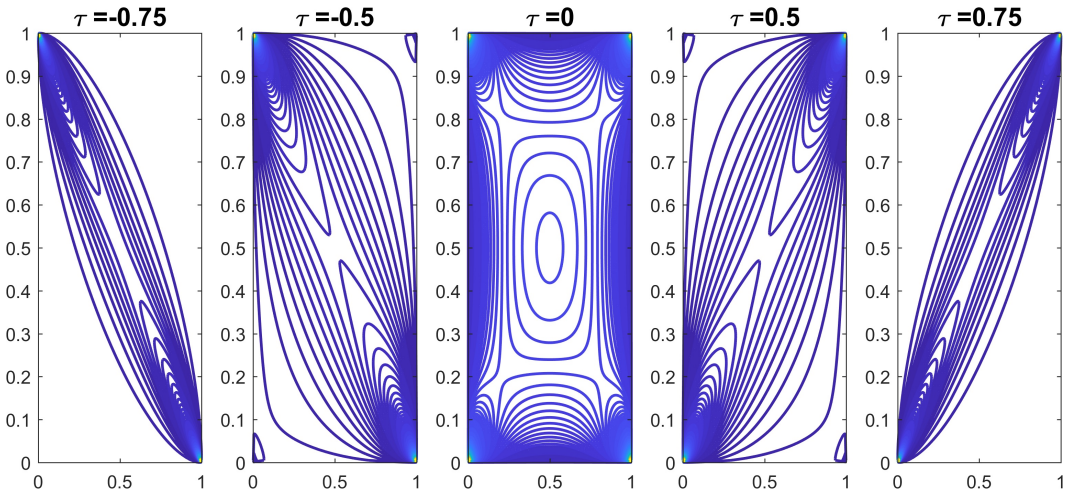


Figure 4: The bivariate Student's t copula with varying Kendall's tau.

where $\mathbf{u} \in [0,1]^n$ and t_ν^{-1} is the inverse univariate Student's t distribution with ν degrees of freedom.

To simulate from the Student's t copula with known parameters Σ_ρ and ν and known marginal distribution functions $F_1, \ldots, F_N$ we follow a similar method to the Gaussian copula. We first simulate random numbers from a MVt distribution with parameters ρ and ν using the method described previously. We then transform these realisations onto $[0,1]^N$ by applying the inverse of the Student's t distribution with ν degrees of freedom. Lastly, applying the inverses of the marginal distribution functions $F_1^{-1}, \ldots, F_N^{-1}$ to each row of this matrix gives R with the desired returns structure.

Example 8. Consider the bivariate case of the Student's t copula and set the degrees of freedom, $\nu = 3$. The dependence structure is now specified by the Kendall's tau and the degrees of freedom. In Figure 2.2.3, we plot contours of the generated unit square for varying Kendall's tau. In contrast to the Gaussian copula, we see that there is dependence structure when the Kendall's tau is zero. This is due to the structure imparted by the degrees of freedom, resulting in a non-uniform copula. In the other plots, we notice that there is stronger dependence towards the tails than with the Gaussian copula. This is due to the thicker tails of the Student's t distribution.

2.2.4 Archimedean copulas

Definition 9 (Archimedean Copula)**.** We define an Archimedean copula by its generator $\varphi(u)$, where:

$$C(u_1, \ldots, u_n) = \begin{cases} \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_n)) & \text{if } \sum_{n=1}^{N} \varphi(u_n) \leq \varphi(0) \\ 0 & \text{otherwise} \end{cases}$$

and $\varphi(u)$ is a continuous, strictly decreasing and convex function with $\varphi(1) = 0, \varphi'(u) < 0$ and $\varphi''(u) > 0$ for all $0 \geq u \geq 1$.

Archimedean copulas often present us with several desirable properties and can simplify the mathematics (Bouyé et al., 2000). One such such simplification is a closed-form solution for Kendall's tau in the bivariate case,

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(u)}{\varphi'(u)} \mathrm{d}u.$$

This is useful when we wish estimate the parameters of the copula.

For this paper, we will only consider the Clayton Archimedean copula. This copula is appropriate to financial applications since it has asymmetric dependence (unlike the elliptical copulas) skewed towards the lower tail. This fits into what we witness in financial markets as assets often move together more in crashes than in booms.

The d-dimensional Clayton copula has a generator of the form

$$\varphi_\theta(x) = (1 + \theta)_+^{-\frac{1}{\theta}}$$

where $\theta \geq \frac{-1}{d-1}$ (McNeil and Nešlehová, 2009). For the bivariate case, we can get an estimate for θ directly from an estimate for Kendall's tau, while for higher dimensions we need to calculate the Kendall's tau function described by McNeil and Nešlehová (2009). Once θ has been estimated, simulating from d-dimensional Clayton copulas is relatively simple.

McNeil and Nešlehová (2009) give a detailed overview of the mathematics underlying the simulation procedure, but we will just summarise the necessary results. We use the Marshall Olkin's simulation algorithm which is described by

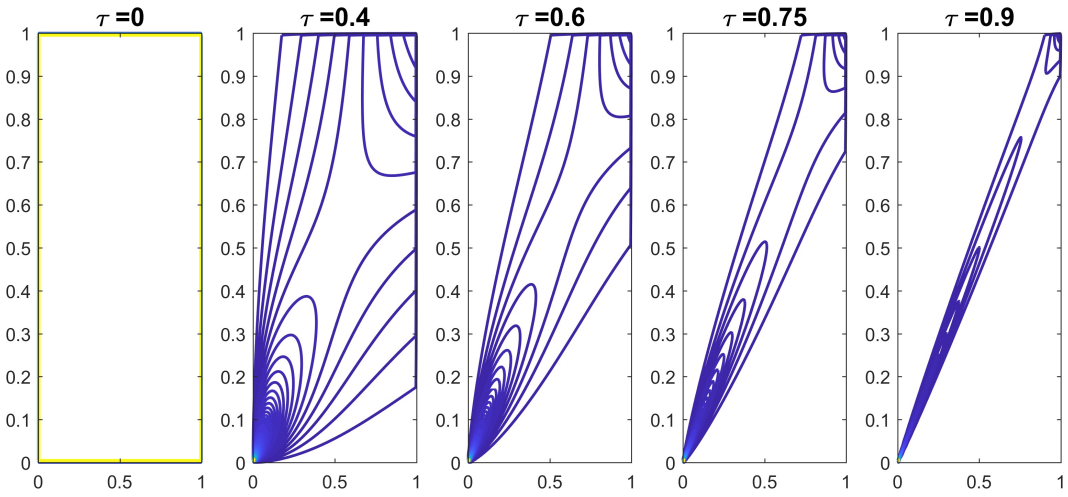


Figure 5: The bivariate Clayton copula for varying values of Kendall's tau.

Hofert (2008). This method relies on knowing the form of the radial distribution, F_V, of the copula. For the Clayton copula, this is Gamma distribution with shape parameter $\frac{1}{\theta}$ and scale parameter 1. Given θ, the algorithm for simulating 1 sample, $(U_1, \ldots, U_d)$ from a d-dimensional Clayton copula is;

1. Sample $V \sim \text{Gamma}(\frac{1}{\theta}, 1)$.
2. Sample i.i.d. $X_i \sim U[0,1], i \in \{1, \ldots, d\}$.
3. $U_i = \varphi_\theta^{-1}(-\frac{log(X_i)}{V})$.

This allows us to generate a sample from the Clayton copula via sampling from the Gamma and Uniform distributions.

Example 10. Consider a bivariate Clayton copula. The copula is fully specified by the parameter θ which can be derived directly from the Kendall's tau. We now consider a range of only positive Kendall's taus, and plot the resultant contours of the unit square in figure 10. As with the Gaussian copula, when the Kendall's tau is zero, there is no visible structure. As the Kendall's tau increases so does the co-depedence, however unlike the elliptical distributions, the dependence is skewed to the lower tail.

2.3 Time-series

Nystrom and Skoglund (2002) argue that the combination of time-series models with copulas is sufficiently general to model financial returns and manage risk. A time-series model is fitted to each underlying asset and a copula is used to introduce interdependence between the assets. This dependence is, in general, more complex than can be captured by a single variable, like correlation. In the simplest case, they suggest a GARCH(1,1) time-series model, estimated from each underlying along with a t-copula to model their dependence. In this section, we briefly discuss how to simulate from such a model.

Consider a series of GARCH(1,1) models, each estimated from an underlying asset. Each individual return is then described by

$$r_t^i = \mu^i + \epsilon_t^i, \tag{1}$$

$$\epsilon_t^i = \zeta_t \sigma_t^i, \tag{2}$$

for $i = 1, \ldots, N$, where $\zeta_t \overset{iid}{\sim} [0, 1]$. Thus, $\epsilon_t^i \sim [0, (\sigma_t^i)^2]$. The conditional variance is given by

$$(\sigma_t^i)^2 = \omega^i + \alpha_1^i \epsilon_{t-1}^2 + \beta_1^i (\sigma_{t-1}^i)^2. \tag{3}$$

As we are simulating returns we assume $\mu_i = 0$ for every i. Furthermore, we assume the residuals, ϵ_t^i, are Gaussian.

Simulating from the above is fairly straightforward: given initial values, Gaussian random variables can be generated and used to advance the returns process through time. To introduce the copula dependance, the Gaussian random variables must be inferred from random samples of the copula.

Thus, to advance the model one step, N random realisations are generated from the copula. These realisations must then be standardised such that they have zero mean and unit variance (this step is *essential* when the copula is not Gaussian). The inverse Gaussian cumulative distribution function is then applied to these random variables, which will be distributed on the uniform hypercube of dimension N, to obtain the Gaussian variates required to drive the process.

The procedure is illustrated in the numerical investigation of Section 4.6.

3 Portfolio construction

In this section, we are interested in constructing risk parity portfolios for different risk measures. Risk parity portfolios are portfolios such that the risk contribution to the selected risk measure from each component is equal. Intuitively, the risk contribution of a component is equal to the product of its weight in the portfolio and its *marginal risk contribution*. This section follows Tasche (2007).

3.1 Marginal risk contribution for different risk measures

3.1.1 General formulation

We rely on Euler's theorem for homogenous functions in order to decompose the risk of the portfolio into the risk contributed by each of its components.

Definition 11 (Homogeneous function)**.** A function $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is called homogeneous of degree 1 if for all $h > 0$:

$$f(hu) = hf(u)$$

.

When the function f is differentiable, we can use Euler's theorem on homogeneous functions so as to decompose the function on each of its component.

Theorem 12 (Euler's theorem on homogeneous functions)**.** Let $f : U \subset \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. Then f is homogeneous of degree 1 if and only if

$$f(u) = \sum_{i=1}^{n} u_i \frac{\partial f(u)}{\partial u_i}$$

where $u = (u_1, ..., u_n) \in U, h > 0$.

Now that we have seen Euler's theorem, we can use it to construct our risk parity portfolio. First, consider n real valued random variables $X_1, ..., X_n$ that will represent the return of the n assets in our portfolio for some time horizon and $w = (w_1, ..., w_n)$ the vector of weights in our portfolio. The return of our portfolio is then

$$X(w) = \sum_{i=1}^{n} w_i X_i$$

for the considered horizon. Let ρ be our risk measure, such that the economic capital required by our portfolio is $\rho(X(w))$. We are interested in evaluating the risk contribution of asset i, denoted $\rho(X_i \mid X)$ to the total risk of the portfolio, $\rho(X(w))$. Assuming our risk measure is homogeneous of degree 1 and continuously differentiable, we can use Euler's theorem to compute the risk contributions of each component:,

$$\rho(X_i \mid X) = w_i \frac{\partial \rho(X(w))}{\partial w_i},$$

and

$$\rho(X) = \sum_{i=1}^{n} \rho(X_i \mid X) = \sum_{i=1}^{n} w_i \frac{\partial \rho(X(w))}{\partial w_i}.$$

We now consider three different commonly used risk measures: standard deviation (σ), Value-at-Risk (VaR) and Expected Shortfall (ES). All of those risk measures are homogeneous of degree 1 and under smoothness conditions for value at risk and expected shortfall (see (Tasche, 1999)) we can derive the risk contributions for each of the risk measures. The relevant computational formulae are given by:

$$\sigma(X_i \mid X) = w_i \frac{(\Sigma w)_i}{\sqrt{w' \Sigma w}},$$
$$\mathrm{VaR}_\alpha(X_i \mid X) = -w_i E[X_i \mid X = -\mathrm{VaR}_\alpha(X)],$$

and

$$\mathrm{ES}_\alpha(X_i \mid X) = -w_i \frac{1}{\alpha} E[X_i \mid X \leq -\mathrm{VaR}_\alpha(X)].$$

The details of the derivation can be found in Stefanovits (2010).

3.1.2 Closed-form formulas

The marginal risk contributions for standard deviation are always available in closed-form as long as we have an estimate for the covariance matrix of the returns. It does *not* depend on the underlying distributions that gave rise to the covariance estimate. For Value-at-Risk and Expected Shortfall, we can derive analytical solutions for the case when the returns are multivariate normal or multivariate

Student's t, as in Section 2.1. For the MVN case,

$$\mathrm{VaR}_\alpha(X_i \mid X) = -w_i \left(\frac{(\Sigma w)_i}{\sqrt{w'\Sigma w}} \Phi^{-1}(\alpha) - \mu_i \right),$$

and

$$\mathrm{ES}_\alpha(X_i \mid X) = -w_i \left(\frac{(\Sigma w)_i}{\alpha\sqrt{w'\Sigma w}} \phi(\Phi^{-1}(\alpha)) - \mu_i \right),$$

whereas for the MVt case

$$\mathrm{VaR}_\alpha(X_i \mid X) = -w_i \left(\frac{(\Sigma w)_i}{\sqrt{w'\Sigma w}} T_\nu^{-1}(\alpha) - \mu_i \right)$$

and

$$\mathrm{ES}_\alpha(X_i \mid X) = -w_i \left(\frac{\nu}{1-\nu} \frac{(\Sigma w)_i}{\alpha\sqrt{w'\Sigma w}} \left(1 + (\frac{T_\nu^{-1}(\alpha))^2}{\nu}) \right) t_\nu(T_\nu^{-1}(\alpha)) - \mu_i \right).$$

3.1.3 Monte Carlo methods

When the underlying dependency structure is complex, closed-form expressions for the marginal risk contributions are unavailable. To construct risk parity portfolios, we turn to Monte Carlo simulation to estimate the required quantities. We focus only on estimating the marginal risks for expected shortfall.

If we assume that the distribution of the returns is continuous, the general formula for the marginal risk is

$$\frac{\partial \mathrm{ES}_\alpha(X)}{\partial w_i} = -\frac{1}{\alpha} E\left[X_i \mid X \leq -\mathrm{VaR}_\alpha\left(\sum_{i=1}^{n} w_i X_i \right) \right].$$

To estimate this, we generate N simulations of our n assets, either using model generated data or empirical distributions. Given a portfolio $w = (w_1, ..., w_n)$, we can compute an estimate of the necessary profit and loss distribution.

In order to determine the marginal risk contribution of asset *i*, we sort the array of portfolio returns. Based on this sorted array, the returns of asset *i* are also sorted. To compute the marginal risk, we sum the first αN sorted values of the returns of asset *i*, and we divide it by $-\frac{1}{\alpha N}$. However, sorting an array costs a significant

amount of execution time. Instead of sorting the array of the portfolio returns, we can simply sum the returns of asset i whenever the return of the portfolio is less than its α-quantile and then divide the sum by $-\frac{1}{\alpha N}$. We thus obtain the marginal risk contribution of asset i. Using this vectorised technique instead of sorting an array is approximately ten times faster when implemented in `Matlab`.

3.2 Optimisation Problem

So far, we have seen how to compute marginal risk contributions either using the closed-form solution or Monte Carlo simulation. We are now interested in building risk parity portfolios using one of these risk measures. In another words, we want to build a portfolio $w = (w_1, ..., w_n)$ such that for all i, j:

$$\rho(X_i \mid X) = \rho(X_j \mid X) \Leftrightarrow w_i \frac{\partial \rho(X(w))}{\partial w_i} = w_j \frac{\partial \rho(X(w))}{\partial w_j}.$$

We are not considering portfolios where one can short assets.

This can be restated as the following optimisation problem:

$$w^* = \operatorname{argmin} f(w)$$
$$\text{s.t.} \sum_{i=1}^{n} w_i = 1 \qquad \text{and} \qquad 0 \leq w_i \leq 1 \forall i,$$

where

$$f(w) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(w_i \frac{\partial \rho(X(w))}{\partial w_i} - w_j \frac{\partial \rho(X(w))}{\partial w_j} \right)^2$$

The idea is to minimise f in order to equalise all the risk contributions of the portfolio. At each iteration of the algorithm, the programme computes the score function f with new weights and we find a solution when $f(w^*) = 0$ (i.e, all the risk contributions are equal for this portfolio).

The above optimisation problem can be solved using Sequential Quadratic Programming (SQP), or Cyclical Coordinate Descent (CCD), see Griveau-Billion et al. (2013).

3.3 Framework

We now have an overview of the different parts we need in order to build our risk parity portfolios. Let us see how these different parts interact with each other so as to build a risk parity portfolio for standard deviation and expected shortfall using Monte Carlo simulation.

3.3.1 Risk parity portfolio for standard deviation

For standard deviation, marginal risk contributions do not depend on the underlying distribution, but only on the covariance matrix:

$$\frac{\partial \sigma(w)}{\partial w_i} = \frac{(\Sigma w)_i}{\sqrt{w' \Sigma w}}$$

. Obtaining the risk parity portfolio is quite fast and simple. We only need to plug the closed form formula in the optimisation problem and we get the optimal weights after a few iterations of the SQP algorithm.

3.3.2 Risk parity portfolio for expected shortfall using Monte Carlo

Based on the underlying distributions - either obtained from model generated data or empirical distributions - we generate N samples of our n assets returns. At each iteration of the SQP algorithm, we compute the returns of the portfolio with the new weights and estimate the marginal risk contribution for each asset. Given the portfolio and the marginal risks we can compute the score function f for this iteration. After some iterations, the algorithm converges and we obtain the optimal weights.

As the number of assets and Monte Carlo simulations increase, the number of iterations required for the algorithm to converge increases too. Instead of using the equally-weighted portfolio as initial guess in our optimisation algorithm, we can compute the initial guess using the MVN assumption for the returns. This dramatically increases the rate of convergence.

4 Numerical Investigation

4.1 Overview

In this chapter, we investigate the sensitivity of the profit and loss distribution of five portfolios (equally-weighted risk contribution (ERC) -with volatility, expected shortfall and VaR as risk measures-, equally weighted (EW) and mean variance (MV)) to changes in the mean and covariances of the asset returns. There may be errors in estimating these parameters from asset returns data therefore we should have a good understanding of the effect these parameters have on the profit and loss distribution of the portfolios. With the assumption that the asset returns either have a multivariate Gaussian or multivariate Student's t distribution, we will analyse the following 5 cases:

1. $\boldsymbol{\mu} = \mathbf{0}$;
2. $\boldsymbol{\mu}$ proportional and inversely proportional to $\boldsymbol{\sigma}$;
3. sensitivity to $\boldsymbol{\mu}$;
4. sensitivity to $\boldsymbol{\Sigma}$; and
5. model misspecification.

Consider stock, currency and bond returns with volatilities 30%, 10% and 5%, respectively. We will use the following correlation matrix throughout this investigation:

$$\rho = \begin{bmatrix} 1 & -0.3 & -0.5 \\ -0.3 & 1 & 0.1 \\ -0.5 & 0.1 & 1 \end{bmatrix}$$

.

4.2 Multivariate Normal

4.2.1 Influence of μ

$\boldsymbol{\mu} = \mathbf{0}$

Firstly, we consider multivariate normal asset returns. When the mean of the returns are zero the risk parity portfolios all have the same profit and loss distribution. Since we consider a fixed covariance matrix and mean of the asset returns, the only variable to consider in calculating the profit and loss of the portfolios is the weights of each portfolio. In this case when the distribution of these portfolios are the same, the weights for the portfolios are the same. This can be shown by taking the marginal risk contribution of the risk parity portfolios as a proportion of the total risk of the portfolio (Stefanovits, 2010):

$$\frac{\text{marginal risk}}{\text{porfolio risk}} = \frac{\text{marginal risk}}{\text{sum of marginal risks}}; \tag{4}$$

$$\text{ERC}_{\sigma} := \frac{m_i \left(\Sigma \boldsymbol{m}\right)_i / \sqrt{\boldsymbol{m}'\Sigma\boldsymbol{m}}}{\sum_{i=1}^{N} m_i (\Sigma \boldsymbol{m})_i / \sqrt{\boldsymbol{m}'\Sigma\boldsymbol{m}}};$$

$$\text{ERC}_{\text{ES}} := \frac{m_i \left(\Sigma \boldsymbol{m}\right)_i / \sqrt{\boldsymbol{m}'\Sigma\boldsymbol{m}}}{\sum_{i=1}^{N} m_i (\Sigma \boldsymbol{m})_i / \sqrt{\boldsymbol{m}'\Sigma\boldsymbol{m}}};$$

$$\text{ERC}_{\text{VaR}} := \frac{m_i \left(\Sigma \boldsymbol{m}\right)_i / \sqrt{\boldsymbol{m}'\Sigma\boldsymbol{m}}}{\sum_{i=1}^{N} m_i (\Sigma \boldsymbol{m})_i / \sqrt{\boldsymbol{m}'\Sigma\boldsymbol{m}}}.$$

The variance and mean of the portfolio returns will be used to calculate the 0.01-quantile of the the profit and loss distribution using the *norminv* function in Matlab. The expected value of the portfolio returns (see Stefanovits (2010)) is

$$\sum_{i=1}^{N} m_i \mu_i. \tag{5}$$

The variance of the portfolio returns is

$$\sum_{i=1}^{N} \sum_{j=1}^{N} m_i m_j Cov(R_i, R_j) = \boldsymbol{m}'\Sigma\boldsymbol{m} \tag{6}$$

where $Cov(R_i, R_j)$ is the covariance between asset i and asset j.

It can be seen in Figure 6 that the portfolio risks are ordered in the following way Maillard et al. (2010):

$$\boldsymbol{\sigma}_{\text{MV}} \leq \boldsymbol{\sigma}_{\text{ERC}} \leq \boldsymbol{\sigma}_{\text{EW}}. \tag{7}$$

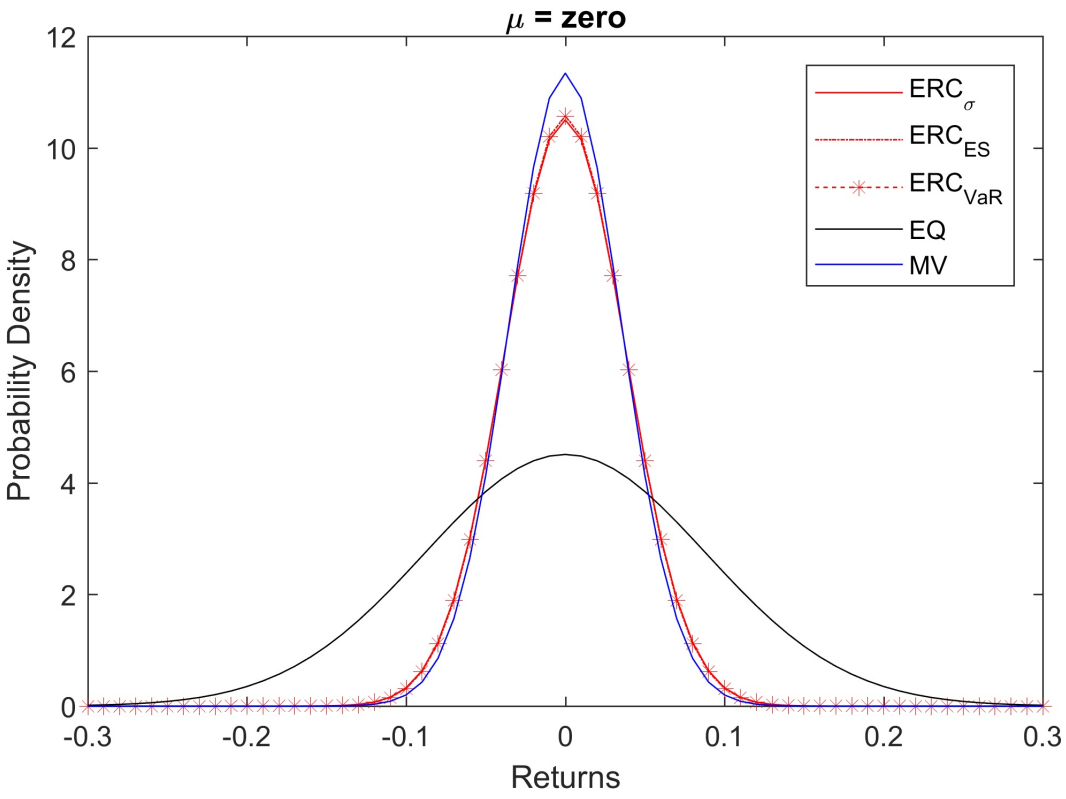


Figure 6: Profit and Loss (P&L) probability density with $\boldsymbol{\mu} = 0$.

The derivation of the relationship in equation (7) is shown in Maillard et al. (2010).

The risk parity portfolios all have the same profit and loss distribution due to the mean of the asset returns being zero. The lower 0.01-quantile is -0.0884 for the risk party portfolio, -0.2058 for the equally weighted portfolio and -0.0818 for the minimum variance portfolio. It can be seen in figure 6 that the minimum variance portfolio has the fattest tails with the largest 0.01-quantile.

μ proportional and inversely proportional to σ

The expected value of asset returns in the market is usually proportional to their volatilities i.e high profit and loss is associated with high volatility. We will consider the asset returns with means that are proportional to their volatilities and asset returns that are not proportional to their volatilities. This will give us an idea

Table 1: Portfolio mean.

Case	ERC_σ	ERC_{ES}	ERC_{VaR}	EW	MV
$\boldsymbol{\mu} \propto \boldsymbol{\sigma}$	0.0378	0.0375	0.0375	0.0567	0.0325
$\boldsymbol{\mu} \not\propto \boldsymbol{\sigma}$	0.0769	0.0865	0.0923	0.0567	0.0841

Table 2: Portfolio 0.01-quantile.

Case	ERC_σ	ERC_{ES}	ERC_{VaR}	EW	MV
$\boldsymbol{\mu} \propto \boldsymbol{\sigma}$	-0.0506	-0.0503	-0.0503	-0.1492	-0.0494
$\boldsymbol{\mu} \not\propto \boldsymbol{\sigma}$	-0.0115	0.0040	-0.0157	-0.1492	0.0023

of the effect the misspecification of the mean has on the profit and loss distribution of the portfolios. The asset mean values considered are $\boldsymbol{\mu} = [0.1 \ 0.05 \ 0.02]$ and $\boldsymbol{\mu} = [0.02 \ 0.05 \ 0.1]$.

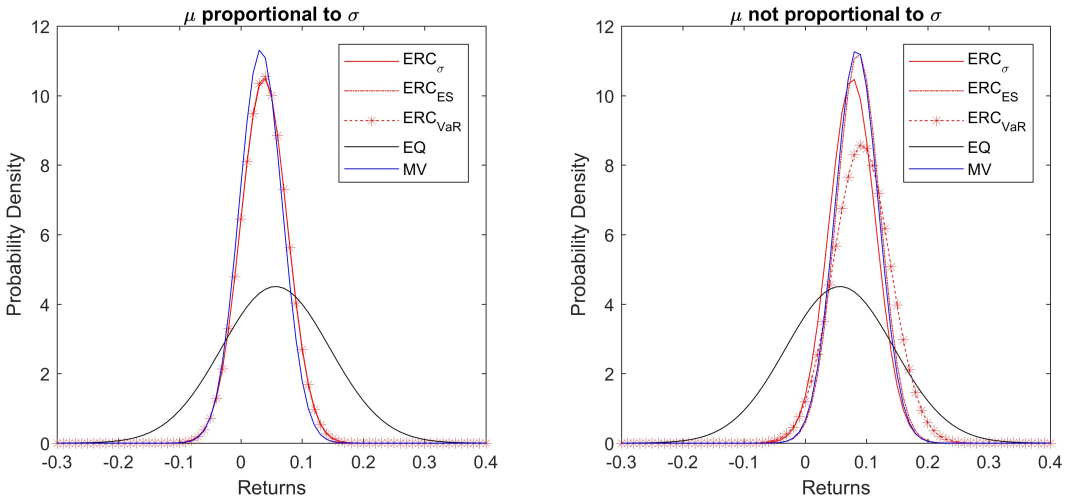


Figure 7: P&L probability density.

When the means of asset returns are non-zero and proportional to their volatilities the relationship in equation ((7)) holds. The higher the mean of the portfolio returns the fatter the tails of the distribution. We are concerned about the tails of the profit and loss distribution, in particular the 0.01-quantile. The equally weighted portfolio has the fattest tails with the largest 0.01-quantile and the highest portfolio mean; this is illustrated in figure 4.2.1. The minimum variance portfolio has the

Table 3: ERC_{es} varying $\boldsymbol{\mu}$.

	μ_{est} 7% error	μ_{est} 20% error	μ_{est} 33% error	μ_{est} 47% error	μ_{est}60%
0.01-quantile	-0.0477	-0.0427	-0.0378	-0.0328	-0.0278
mean	0.0402	0.0452	0.0501	0.0551	0.0601

Table 4: ERC_{VaR} varying $\boldsymbol{\mu}$.

	μ_{est} 7% error	μ_{est} 20% error	μ_{est} 33% error	μ_{est} 47% error	μ_{est} 60% error
0.01-quantile	-0.0477	-0.0427	-0.0378	-0.0328	-0.0278
mean	0.0402	0.0452	0.0501	0.0551	0.0601

smallest portfolio mean and 0.01-quantile. When the assets means are not proportional to their volatilities we see that the equally weighted portfolio performs the worst with the fattest tails and the lowest mean. The minimum variance portfolio has smallest 0.01-quantile with relativity high portfolio returns which is due to the fact that we have an asset that has a high mean and low variance.

4.2.2 Misspecification of μ and σ

Sensitivity to μ

We will now vary the mean of the asset returns as a proportion of itself to investigate the sensitivity of the profit and loss distribution of the portfolio with misspecification of μ. We only consider the ERC expected shortfall and the Value-at-Risk portfolios as the asset allocations for these portfolios depend on the mean of assets.

The expected shortfall portfolio is sensitive to the misspecification of $\boldsymbol{\mu}$. Making an error of 50% in estimating the means of the assets leads to a large change in the mean and 0.01-quantiles of the portfolios. This is a problem as it is difficult to estimate the mean of the asset returns accurately.

The ERC Value-at-Risk portfolio is sensitive to the misspecification of $\boldsymbol{\mu}$. If we make an error of 50% in estimating the means of the asset returns the mean and 0.01-quantile of the portfolio changes. We can see that the minimum variance and expected shortfall portfolio have same dependence on the mean; this is due to the fact that the asset returns distribution is multivariate normal which is symmetric.

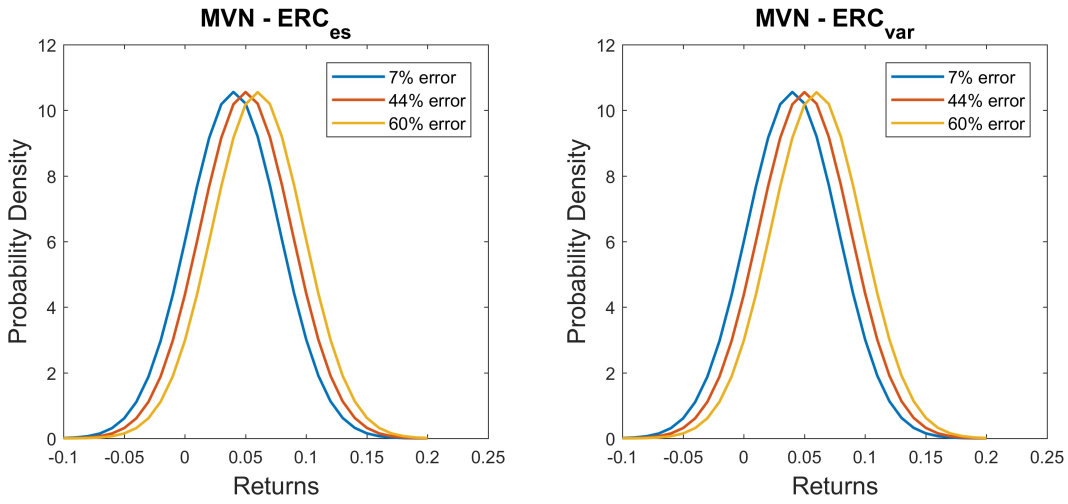


Figure 8: P&L probability density with varying $\boldsymbol{\mu}$.

Table 5: Risk parity portfolio weights.

	Real	Σ_{est} 25% error	Σ_{est} 50% error
w1	0.1219	0.1093	0.1007
w2	0.2668	0.2902	0.3012
w3	0.6113	0.6004	0.5981

Sensitivity to Σ

We will now vary Σ as a proportion of itself to analyse the sensitivity of the portfolios to the misspecification of the covariance matrix of asset returns. We consider the case when $\boldsymbol{\mu} = 0$:

The weights of the risk parity portfolio are not very sensitive to a change in the covariance matrix. However the profit and loss distribution of the portfolio is sensitive it the misspecification of Σ. The tails of the profit and loss distribution vary significantly when Σ is incorrectly estimated. This can be seen in table 6. Because the portfolio P&L distribution is very sensitive to errors in estimating Σ,

Table 6: Risk parity portfolio 0.01-quantiles.

	Real	Σ_{est} 25% error	Σ_{est} 50% error
0.01-quantile	-0.0884	-0.1635	-0.2099

Table 7: EW portfolio 0.01-quantiles.

	Real	Σ_{est} 25% error	Σ_{est} 50% error
0.01-quantile	-0.2059	-0.2865	-0.3490

Table 8: MV Portfolio 0.01-quantiles.

	Real	Σ_{est} 25% error	Σ_{est} 50% error
0.01-quantile	-0.0818	-0.1163	-0.1163

it imported to have a good estimate for the covariance matrix of the multivariate normal asset returns.

The equally weighted portfolio is not very sensitive to the misspecification of the covariance matrix of the asset returns. We can see in table 7 that a 50% error in the estimation of the covariance matrix does not lead to a large change in the 0.01-quantile of the portfolio.

The minimum variance portfolio is sensitive to the misspecification of the covariance matrix. We can see that a 50% estimation error of the asset returns' covariance matrix does lead to a large change in the 0.01-quantile of the portfolio. All five portfolios are sensitive to errors in the estimation of the covariance matrix. Therefore, it is important to be able to get a good estimate of the covariance matrix to ensure the error in estimating the P&L of the portfolios is small.

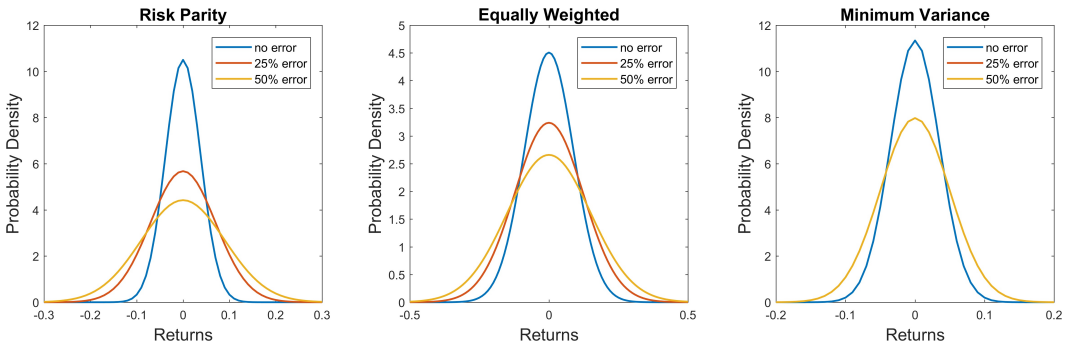


Figure 9: P&L probability density with varying Σ.

Table 9: ERC_{es} 0.01-quantiles.

	Misspecified Model	Real model
0.01-quantile	-0.1580	-0.1576

Table 10: ERC.

	Misspecified Model	Real model
0.01-quantile	-0.2066	-0.1583

4.2.3 Misspecification of the model

We now investigate the influence the misspecification of the model of asset returns has on the P&L distribution of the portfolio. This investigation will be done on the ERC expected shortfall and Value-at-Risk portfolios as these two portfolios depend on the chosen returns model to calculate the portfolio weights. We will assume the underlying distribution of the asset returns is multivariate normal, calculate the portfolio weights and then use these weights to calculate the P&L assuming the asset returns were multivariate Student-t distribution. This will be compared to the calculation of the portfolio weights and P&L assuming the asset returns were distributed multivariate Student's t distribution.

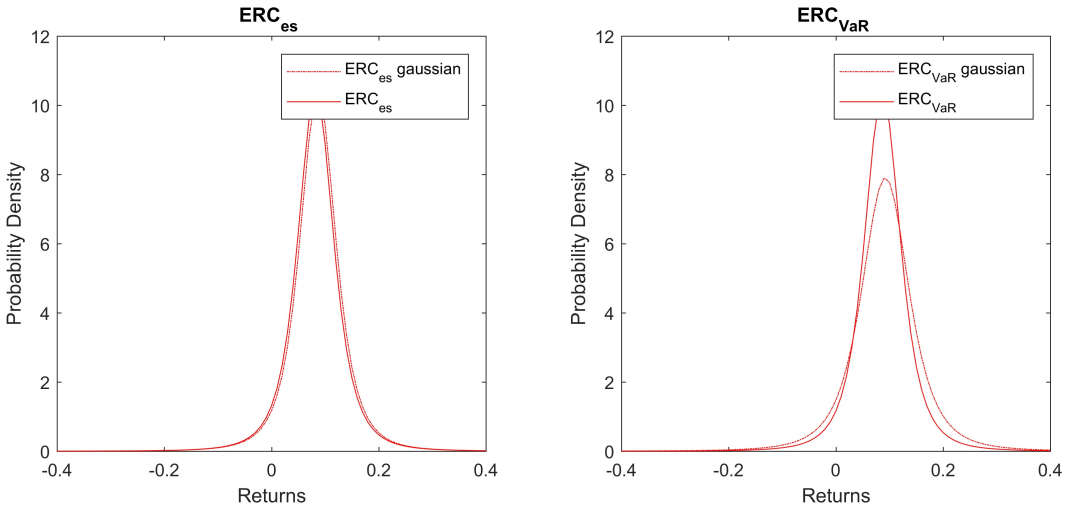


Figure 10: P&L probability density with misspecified model.

Firstly, consider the ERC expected shortfall portfolio with $\boldsymbol{\mu} = [0.02\ 0.05\ 0.1]$. Refer to table 9 for the 0.01-quantiles. Next we consider the ERC Value-at-Risk portfolio with $\boldsymbol{\mu} = [0.02\ 0.05\ 0.1]$. Figure 4.2.3 illustrates that a model misspecification for the ERC_{es} portfolio does not change the distribution of the P&L. The 0.01-quantile does not significantly change. The model misspecification for the ERC_{VaR} portfolio does change the results of the P&L distribution. The $\boldsymbol{\mu}$ that was chosen is the $\boldsymbol{\mu}$ that is not proportional to the volatilities of the asset returns. If the $\boldsymbol{\mu}$ is chosen to be proportional to the asset volatilities then the misspecification of the model for the ERC_{VaR} portfolio does not change the distribution of the P&L.

4.3 Multivariate Student-t

We will now consider the case where we assume the asset returns are distributed multivariate Student-t. We have chosen this distribution as it is a more realistic model of asset returns as it has fatter tails than the multivariate Gaussian distribution. We will investigate the sensitivity of the P&L distributions of the portfolios to $\boldsymbol{\mu}$ and Σ.

4.3.1 Influence of μ

$\mu = 0$

When the mean of the asset returns are zero, then the risk parity portfolios have equal P&L distributions. This can be shown analytically by taking marginal risk contribution as a proportion of the total risk of the portfolio in each case (see Stefanovits (2010)) using equation (4).

The risk parity portfolios all have the same P&L distributions with the 0.01-quantile being -0.1714. The 0.01-quantiles are -0.4019 for the equally weighted and -0.1598 for the minimum variance portfolio. The relationship in equation (7) holds in this case. The minimum variance portfolio has the fattest tails with the largest portfolio mean.

μ proportional and inversely proportional to σ

We will now investigate the effect the means of the asset returns has on the P&L distribution of the portfolio when the means are proportional, and not proportional

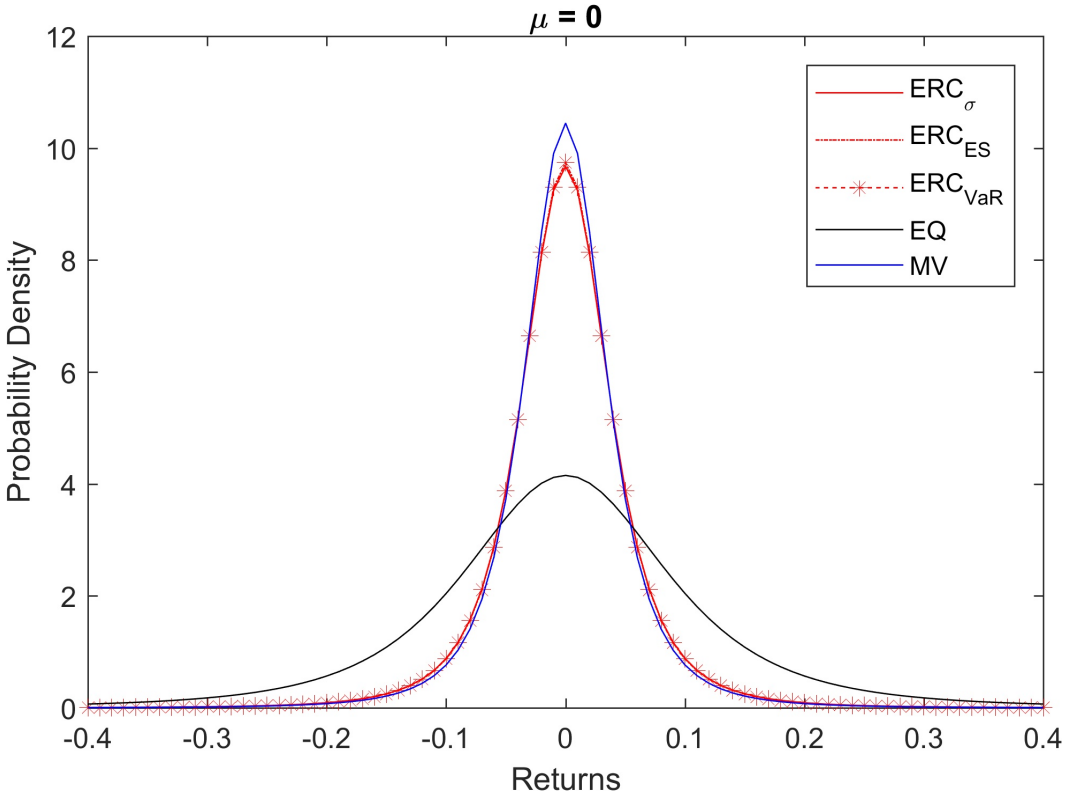


Figure 11: P&L probability density with $\boldsymbol{\mu} = 0$.

to the asset volatilities. The asset mean values considered are $\boldsymbol{\mu} = [0.1 \ 0.05 \ 0.02]$ and $\boldsymbol{\mu} = [0.02 \ 0.05 \ 0.1]$.

When the means of the assets are proportional to their volatilities the equally weighted portfolio has the fattest tails followed by the risk parity portfolios. When the means of the assets are not proportional to their volatilities the equally weighted portfolio performs the worst: it has the fattest tails and the smallest portfolio mean. We are concerned about the tails of the portfolio P&L, in particular the 0.01-quantile. The 0.01-quantiles can be seen in table 11. The minimum variance portfolio has the smallest 0.01-quantile. The 0.01-quantiles do not change drastically when $\boldsymbol{\mu}$ is not proportional to the asset volatilities.

Table 11: Portfolio quantile.

Case	ERC_σ	ERC_{ES}	ERC_{VaR}	EW	MV
$\boldsymbol{\mu} \propto \boldsymbol{\sigma}$	-0.1711	-0.1700	-0.1701	-0.3969	-0.1586
$\boldsymbol{\mu} \not\propto \boldsymbol{\sigma}$	-0.1696	-0.1576	-0.1583	-0.3969	-0.1568

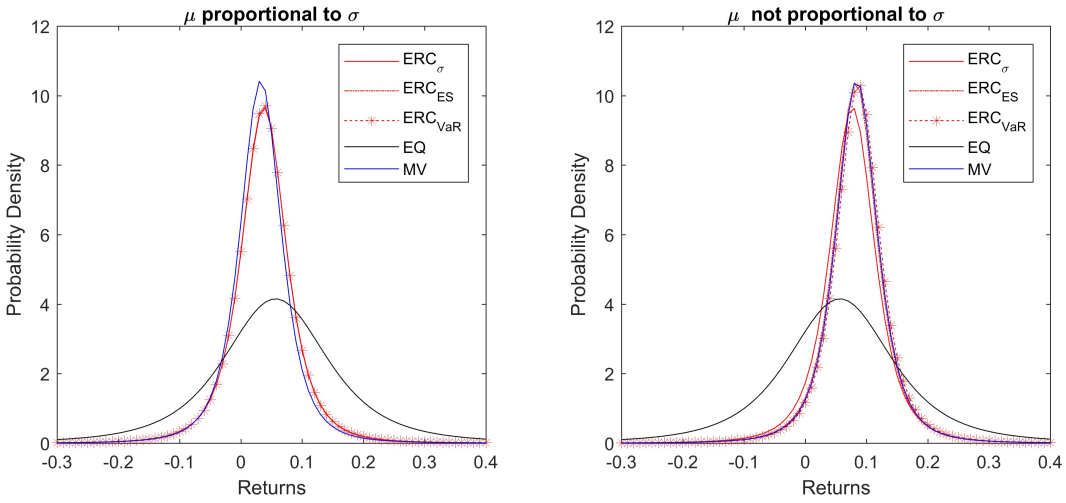


Figure 12: P&L probability density.

4.3.2 Misspecification of μ and σ

Sensitivity to μ

We will now investigate the sensitivity to of the P&L distribution of the portfolios to the misspecification of $\boldsymbol{\mu}$. This is done by varying $\boldsymbol{\mu}$ as a proportion of itself. We will consider the two portfolios that depend on $\boldsymbol{\mu}$.

Consider the sensitivity of the ERC_{es} portfolio to varying $\boldsymbol{\mu}$. It can be seen in table **??** that the values of the 0.01-quantile are very similar when $\boldsymbol{\mu}$ is incorrectly estimated. This means that the multivariate Student-t distribution is not very sen-

Table 12: ERC.

	μ_{est} 7% error	μ_{est} 20% error	μ_{est} 33% error	μ_{est} 47% error	μ_{est}60%
0.01-quantile	-0.1699	-0.1698	-0.1696	-0.1694	-0.1693

Table 13: μ portfolio varying.

	μ_{est} 7% error	μ_{est} 20% error	μ_{est} 33% error	μ_{est} 47% error	μ_{est} 60% error
0.01-quantile	-0.1700	-0.1698	-0.1697	-0.1695	-0.1694

sitive to the misspecification of $\boldsymbol{\mu}$. This is also illustrated in figure 13.

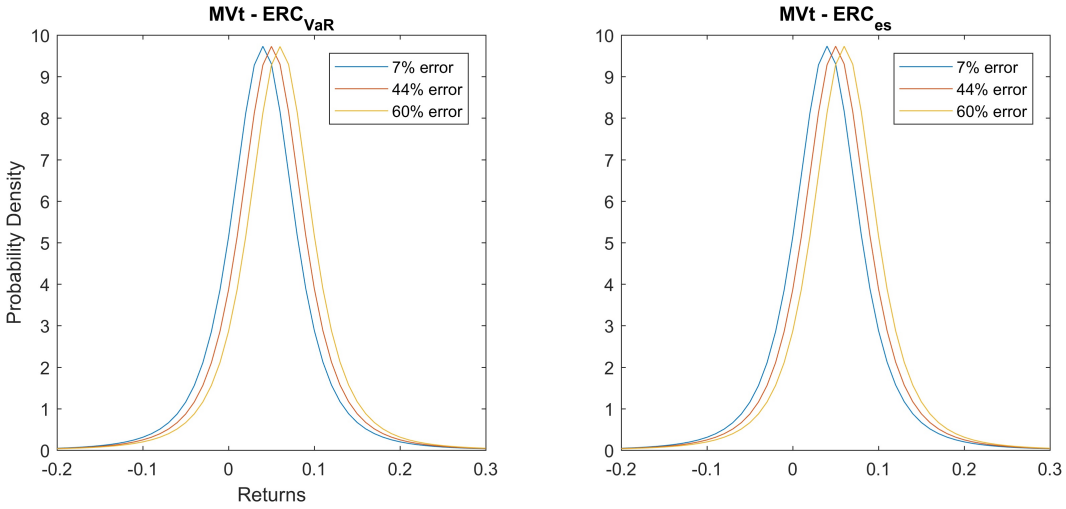


Figure 13: P&L probability density with varying $\boldsymbol{\mu}$.

Consider sensitivity of the ERC_{VaR} portfolio to varying $\boldsymbol{\mu}$. The values of the 0.01-quantiles for the ERC_{VaR} portfolio do not change drastically when $\boldsymbol{\mu}$ is incorrectly estimated. Refer to table **??** for these results. We can see that the ERC_{VaR} and ERC_{es} portfolio P&L distributions are not sensitive to the misspecification of $\boldsymbol{\mu}$.

Sensitivity to Σ

We investigate the sensitivity of the portfolio's P&L distribution to the misspecification of the asset returns covariance matrix. We set the $\boldsymbol{\mu} = 0$.

We will now vary Σ as a proportion of itself for risky parity portfolio. The change in the portfolio weights is very small but the change in the 0.01-quantiles of the P&L is large. The risk parity portfolio is very sensitive to errors in the estimation of the covariance matrix of the asset returns. These results can be seen in table 14.

Table 14: Risk parity portfolio weights.

	Real	Σ_{est} 25% error	Σ_{est} 50% error
w1	0.1219	0.1093	0.1007
w2	0.2668	0.2902	0.3012
w3	0.6113	0.6004	0.5981

Table 15: Risk Parity Portfolio 0.01-quantiles.

	Real	Σ_{est} 25% error	Σ_{est} 50% error
0.01-quantile	-0.1725	-0.3190	-0.4097

We vary Σ as a proportion of itself for the equally weighted portfolio. The change in the 0.01-quantile of the P&L of the portfolio is relatively small when the covariance matrix is incorrectly estimated. This means that the equally weighted portfolio is not very sensitive to the misspecification in the asset returns covariance matrix. These results can be seen in table 16.

We will now vary Σ as a proportion of itself for minimum variance portfolio. The change in the 0.01-quantile of the portfolio is relatively small when the covariance matrix is incorrectly estimated. This means that the equally weighted portfolio is not very sensitive to the misspecification in the covariance matrix. These results can be seen in table 17. The risk parity portfolio is the most sensitive to incorrect estimation of the covariance matrix. It is therefore important that we have a good estimate for the covariance matrix when calculating the portfolio weights of a risk parity portfolio that has multivariate Student-t asset returns.

4.3.3 Conclusion

The P&L of the portfolios that have multivariate Student-t asset returns are less sensitive to the misspecification of $\boldsymbol{\mu}$ than the multivariate normal asset returns. This may be due to the the way the marginal risk contributions are calculated. In

Table 16: EW Portfolio 0.01-quantiles.

	Real	Σ_{est} 25% error	Σ_{est} 50% error
0.01-quantile	-0.4019	-0.5592	-0.6811

Table 17: MV Portfolio quantiles.

	Real	Σ_{est} 25% error	Σ_{est} 50% error
0.01-quantile	-0.1598	-0.2270	-0.2270

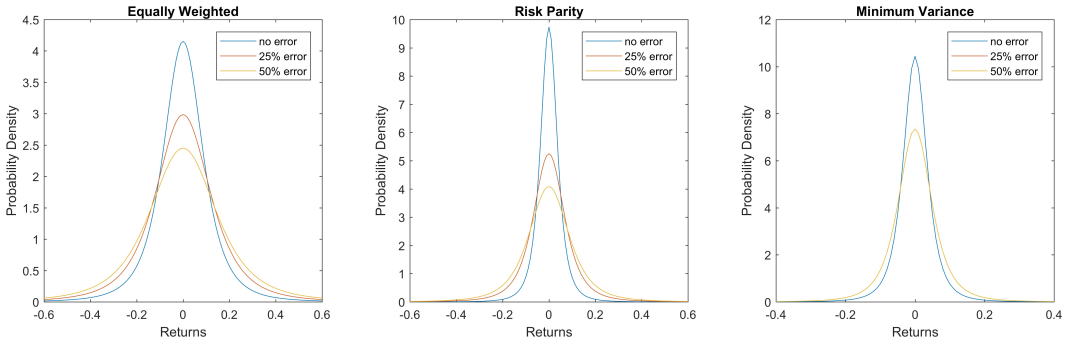


Figure 14: P&L probability density with varying Σ.

the case of the multivariate Student-t ERC_{VaR} portfolio, the marginal risk contributions are calculated with the inverse student-t cumulative distribution function (cdf) which is always larger than the inverse normal cdf due to the fatter tails. This means that the influence of $\boldsymbol{\mu}$ on the marginal risk contribution will be less for the multivariate Student-t than the multivariate normal asset returns. The sensitivity of the asset returns covariances on the P&L distributions is generally high. This means that we should have a good estimator for the covariance matrix. The misspecification of the distribution of the multivariate returns does not highly influence the P&L distribution of the portfolios.

Remark 13 (Two-asset case)**.** In our investigation, the two-asset case is not the most interesting. Indeed when we have two assets, the marginal risk contributions do not depend on the covariance between the assets for all three risk measures. We are going to prove it for the standard deviation, but the proof is the same for the other risk measures and for Gaussian and Student-t distributions. In the portfolio,

we want to equate the risk contributions. Thus,

$$\begin{aligned} w_1 \frac{\partial \rho(X(w))}{\partial w_1} &= w_2 \frac{\partial \rho(X(w))}{\partial w_2}, \\ w_1 \frac{(\Sigma w)_1}{\sqrt{w' \Sigma w}} &= w_2 \frac{(\Sigma w)_2}{\sqrt{w' \Sigma w}}, \\ w_1^2 \sigma_1^2 + w_1 w_2 \sigma_{1,2} &= w_2^2 \sigma_2^2 + w_1 w_2 \sigma_{1,2}, \\ w_1^2 \sigma_1^2 &= w_2^2 \sigma_2^2. \end{aligned}$$

As we can see, equating the risk contributions does not depend on the covariance $\sigma_{1,2}$ for the assets.

4.4 Clayton Copula with Gaussian Marginals

In this section, we study the Clayton copula with Gaussian marginals. We are going to use three assets corresponding to a stock, a currency and a bond, with means $\mu = (0.1, 0.05, 0.02)$ and volatilities $\sigma = (0.3, 0.1, 0.05)$ respectively. The Clayton copula takes a single parameter, θ, which drives the correlation between the assets.

We compare the performance of risk parity portfolios (using expected shortfall and standard deviation as risk measures) against the classical mean variance (MV) portfolio and a constant equally weighted (EW) portfolio. For each of our experiences, we generate three sets of data from the Clayton copula: the first to estimate the covariance matrix of the assets, the second to estimate marginal risk contributions for expected shortfall, and the last one to generate the profit and loss distributions. Unless stated otherwise, $\theta = 0.5$, which gives the following correlation matrix between the assets:

$$\Sigma_\rho = \begin{pmatrix} 1.0000 & 0.3171 & 0.3180 \\ 0.3171 & 1.0000 & 0.3176 \\ 0.3180 & 0.3176 & 1.0000 \end{pmatrix}.$$

4.4.1 Influence of Mean

Firstly, we consider the case where all assets have zero mean, i.e $\mu_i = 0$ for all i, but still have the volatility defined above. The resulting portfolio properties are summarised in Table 18 and the profit and loss distributions are displayed in Figure 15.

Table 18: Portfolio properties when the mean return of all the assets is equal to 0.

	ERC ES	ERC σ	Mean Variance	Equally Weighted
0.01-quantile	-0.177	-0.178	-0.12	-0.311
mean	$-1.30 * 10^{-4}$	$-1.30 * 10^{-4}$	$-0.97 * 10^{-4}$	$-1.90 * 10^{-4}$
Max drawdown	-0.38	-0.38	-0.25	-0.65

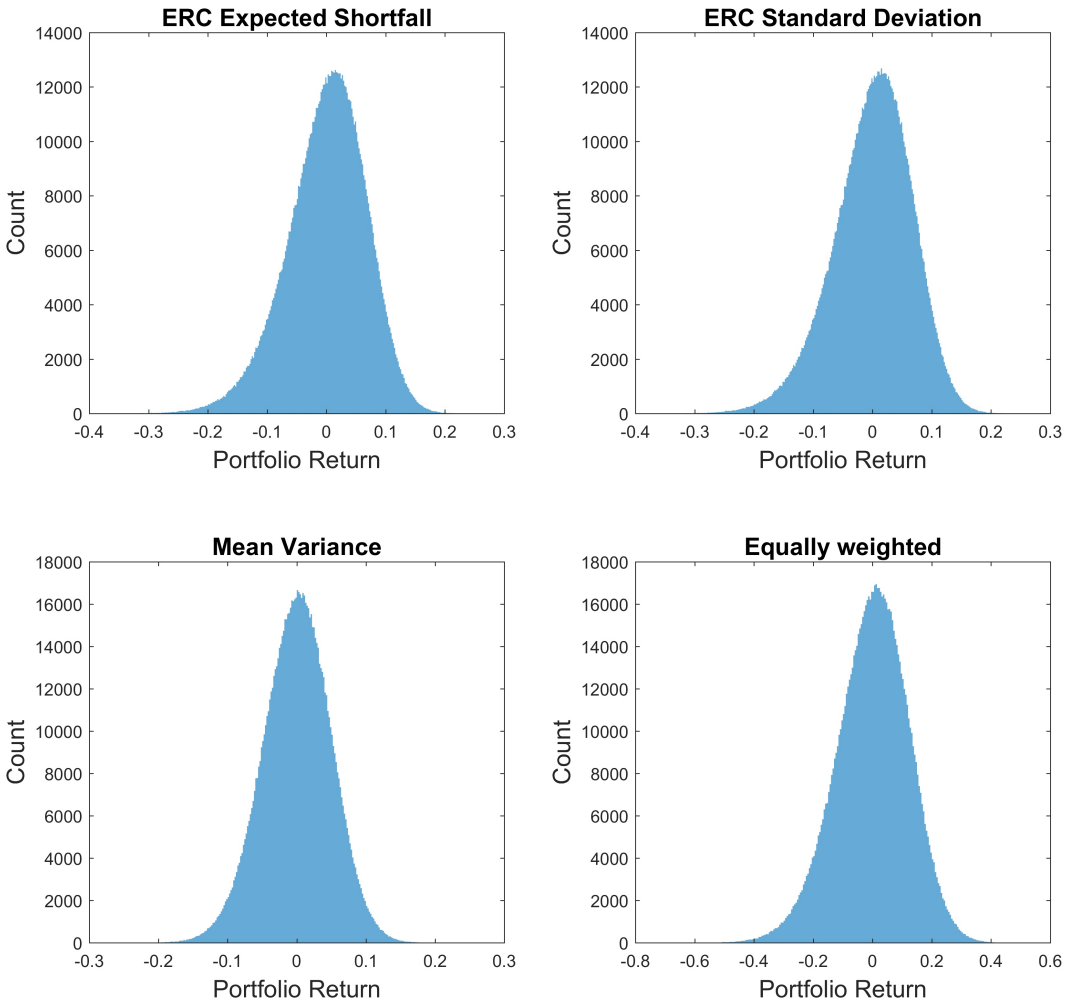


Figure 15: Profit and loss distributions when the mean of the returns is equal to 0 for all assets.

Table 19: Portfolio properties when the mean return of all the assets is proportional to their volatility.

	ERC ES	ERC σ	Mean Variance	Equally Weighted
0.01-quantile	-0.139	-0.141	-0.097	-0.254
mean	0.037	0.037	0.023	0.057
Max drawdown	-0.38	-0.39	-0.24	-0.69

Table 20: Portfolio properties when the mean return of all the assets is inversely proportional to their volatility.

	ERC ES	ERC σ	Mean Variance	Equally Weighted
0.01-quantile	-0.046	-0.101	-0.025	-0.623
mean	0.090	0.077	0.095	0.057
Max drawdown	-0.22	-0.33	-0.17	-0.62

We observe that both risk parity portfolios are performing exactly the same. In terms of risks, they are situated between the mean variance portfolio and the equally weighted portfolio. We can also see that the returns generated by the Clayton copula are not symmetrical: we can indeed note that the left tail is thicker than the right tail. This is interesting because empirical distribution exhibits the same behaviour.

Now we turn to the case where the mean is proportional to the volatility for each asset, $\mu = (0.1, 0.05, 0.02)$. The results appear in Table 19

Now that the mean of the returns are not zero, we can see that both risk parity portfolios are in-between mean variance and equally weighted portfolios. They have lower risks than the equally weighted portfolio but lower returns, and higher risks than the mean variance portfolio but better returns.

Lastly, we consider the case when the mean is inversely proportional to the standard deviation, $\mu = (0.02, 0.05, 0.1)$. The results are summarised in Table 20 and Figure 16.

We can see here that the ES portfolio behaves better than the SD one. Indeed, it has higher returns and at the same time a lower risk (better quantile and better max drawdown). This is probably due to the fact that all the assets have the same correlation and are positively correlated. We can note that the mean variance portfolio outperforms all the other portfolios in terms of returns and risks, which is normal since we have an asset with low variance and high returns. On the other hand, the

Table 21: Portfolio properties and estimation.

	Real μ	$\mu_{est}7.4\%$ error	$\mu_{est}18.9\%$ error	$\mu_{est}24.1\%$ error	$\mu_{est}50.2\%$ error
0.01-quantile	-0.14	-0.138	-0.136	-0.144	-0.141
mean	0.0370	0.0366	0.0362	0.0380	0.0373
Max drawdown	-0.354	-0.346	-0.343	-0.363	- 0.357

equally weighted portfolio behaves really badly: it has low returns and high risks. The ES, SD and MV portfolios invest mostly in the asset with low variance and high return. Since the assets are positively correlated, in case of a crash, all assets tend to go down together and the equally weighted portfolio has a non-negligible part invested in the asset with high variance, which explains its bad performance.

4.4.2 Misspecification of μ and Σ

In this section, we investigate what happens if we estimate badly either μ or Σ.

Sensitivity to mean

We add a Gaussian noise to the vector of means in order to measure the influence of a bad estimation of μ. Since the standard deviation portfolio does not depend on the returns of the assets, we only investigate the impact of the bad estimation of μ on the expected shortfall portfolio.

We can observe that the portfolio is really robust against bad estimation of μ in the case of a Clayton copula with positive correlations. Indeed, even if we make a relative error of 50% when estimating μ, the portfolio remains very close to the one when μ is known. The changes in the portfolio are very small even when the error in estimating μ is very important.

Sensitivity to covariance

Here, we add noise to the covariance matrix Σ and investigate what happens. We also consider what happens when $\mu = 0$ and $\mu \neq 0$.

When there is noise in the covariance matrix - coming from estimation for instance - we observe that the weights of the SD portfolio are very close to each other and that the performances of the portfolios are roughly the same. We can conclude that even if we have a rather bad estimation of our covariance portfolio, the resulting portfolio will still be very close to the real one.

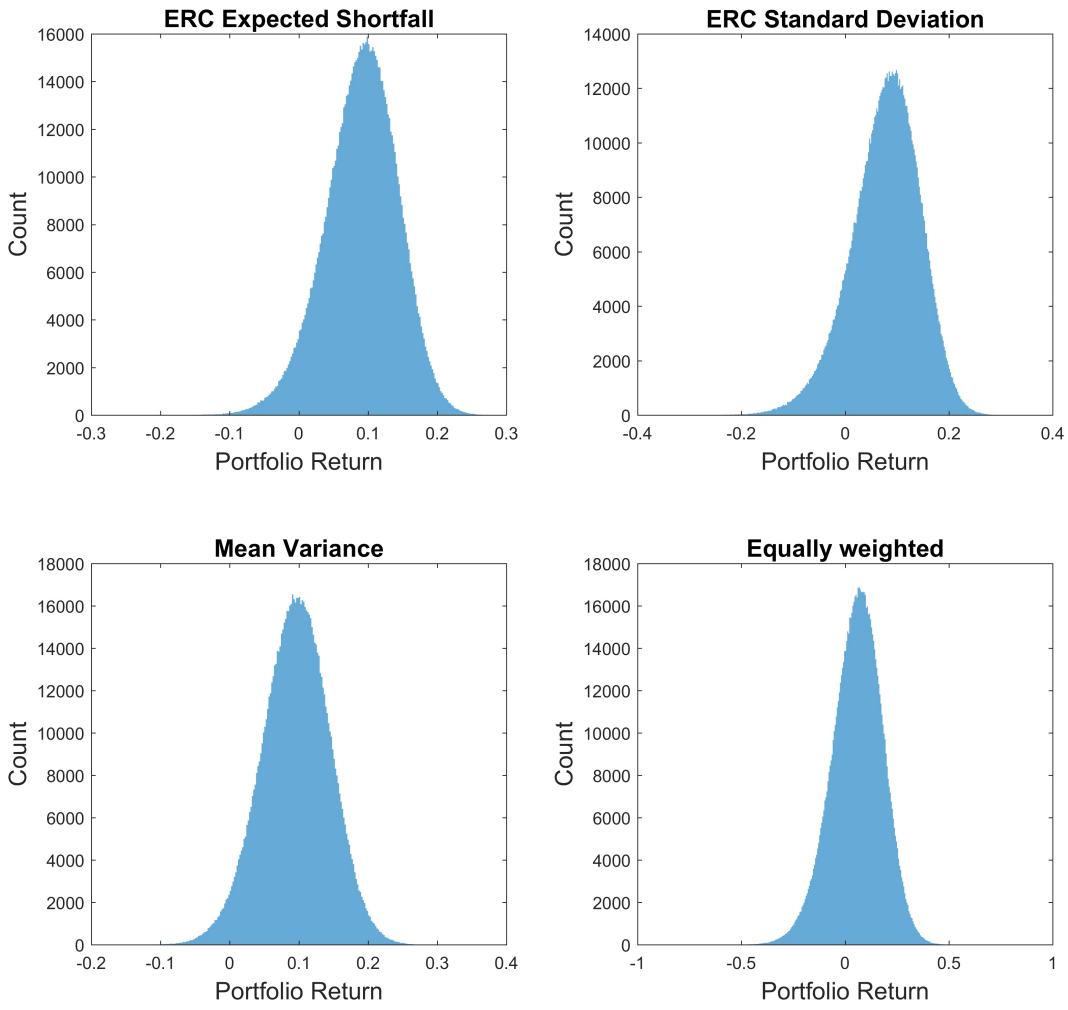


Figure 16: Profit and loss distributions for μ inversely proportional to σ.

Table 22: Portfolio weights for standard deviation risk measure when there is estimation error in the covariance matrix.

	Real Σ	$\Sigma_{est}20\%$ error	$\Sigma_{est}40\%$ error
w_1	0.1003	0.0889	0.1113
w_2	0.3008	0.3441	0.2988
w_3	0.5989	0.5670	0.5900

Table 23: Portfolio properties for zero mean and estimation error in the covariance matrix.

	Real Σ	$\Sigma_{est} 20\%$ error	$\Sigma_{est} 40\%$ error
0.01-quantile	-0.178	-0.177	-0.183
Mean	$-1.30 * 10^{-4}$	$-1.29 * 10^{-4}$	$-1.33 * 10^{-4}$
Max drawdown	-0.382	-0.381	-0.394

Table 24: Portfolio properties for non-zero mean and estimation error in the covariance matrix.

	Real Σ	$\Sigma_{est} 20\%$ error	$\Sigma_{est} 40\%$ error
0.01-quantile	-0.141	-0.139	-0.145
Mean	0.037	0.037	0.038
Max drawdown	-0.388	-0.383	-0.401

4.4.3 Copula influence

Here, we try to understand the influence of the θ in our Clayton copula. The first thing we observe is that, if we increase θ, all the correlation coefficients will increase in the same way. For instance, if all the correlation coefficients are equal to $\rho = 0.21$ and we increase the θ by 0.3, all the coefficients will become $\rho = 0.36$.

Second, we observed something very important: if we increase the θ coefficient, the weights in the risk parity portfolios do not change! Indeed, when we increase θ, we increase the correlation in the same way for all assets, and the tails are moving uniformly in all directions, which does not affect the expected shortfall.

Lastly, we were not able to get a negative correlation coefficient with the Clayton copula, which can be problematic since some assets can be negatively corre-

Table 25: Parameters.

θ	Corr	w_{ES}	w_{σ}
0.1	$\begin{pmatrix} 1 & & \\ 0.076 & 1 & \\ 0.078 & 0.077 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.097 \\ 0.311 \\ 0.592 \end{pmatrix}$	$\begin{pmatrix} 0.102 \\ 0.302 \\ 0.596 \end{pmatrix}$
2	$\begin{pmatrix} 1 & & \\ 0.684 & 1 & \\ 0.685 & 0.685 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.096 \\ 0.310 \\ 0.594 \end{pmatrix}$	$\begin{pmatrix} 0.101 \\ 0.302 \\ 0.597 \end{pmatrix}$

lated - for instance stocks and bonds. Furthermore, if we could have a correlation coefficient that moves in opposite directions, we would observe a change in the weights of the portfolio.

4.4.4 Misspecification of the model

In this section we investigate what happens when, if generating data from the Clayton copula, we assume that the returns are normally distributed.

In order to check the implications of a model misspecification, we generate N samples from the Clayton copula and estimate the covariance matrix. Using the covariance matrix, we can compute the risk contributions and the weights of the portfolio since we have a closed form solution for multivariate normal returns. We generate another N samples from the Clayton copula to estimate the ES portfolio using Monte Carlo.

We observe that the weights are the same. It means that the Clayton copula does not have any specific dependence between the assets. Indeed, the Clayton copula does not affect the tail dependences: the tail dependences are uniform in all directions, so the expected shortfall computed using Monte Carlo does not take into account the effect of the copula.

4.4.5 Conclusion

The Clayton copula is quite interesting to study as it is asymmetrical and has a thicker left tail. One of the major things that we have seen is, when using a Clayton copula which affects the tail dependencies in the same way all across spaces, we do not add a significant dependence in the tail. ES and SD portfolios are very close and do not evolve when the θ parameter of the Clayton copula increases or decreases. We would expect to have very different results with skewed copulas, which would have tail dependences between assets in a non-uniform way. Since the covariance matrix cannot capture this kind of tail dependence, we would expect our ES portfolios to perform better than the portfolios based on the estimated covariance matrix. The Clayton copula we used only allowed us to have the same positive correlation between assets, and thus does not allowed us to induce negative correlation between assets, which results in constant risk parity portfolios. However, the portfolio construction is quite robust to estimation errors in μ and Σ.

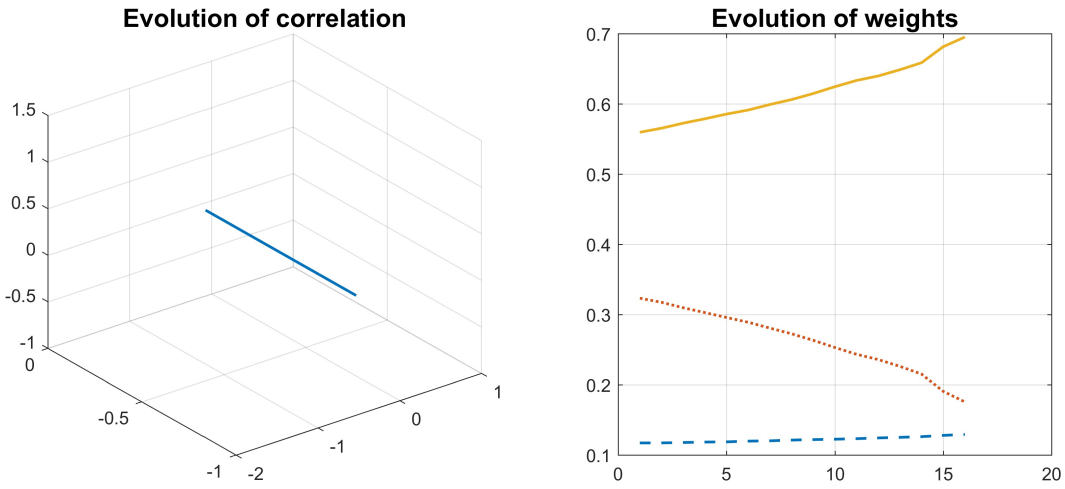


Figure 17: Evolution of correlation, and evolution of weights.

4.5 Student's t Copula with Gaussian Marginals

As we have seen that having the same correlations for all assets was an issue with the Clayton copula, we investigate what happens in the t-Copula with Gaussian marginals when we change the correlation between assets. Even if the t-Copula is symmetrical, it remains interesting because we can induce negative correlation between assets and it has thicker tails than the Gaussian copula. We investigate the influence on the SD portfolio with 3 assets when the correlation between two assets moves towards -1 or 1 and the other correlations remain the same.

4.5.1 The effect of correlation tending to 1

Since the correlation goes towards 1, the assets tend to be more and more correlated. However, the whole purpose of our risk parity portfolios is to diversify the risk. It seems logical to invest more in the asset that is not correlated to the others so as to diversify the risk, rather than holding a big proportion of two assets which are heavily correlated. We can indeed see that in the figures: the proportion of two assets are going down and the proportion of the other one is increasing.

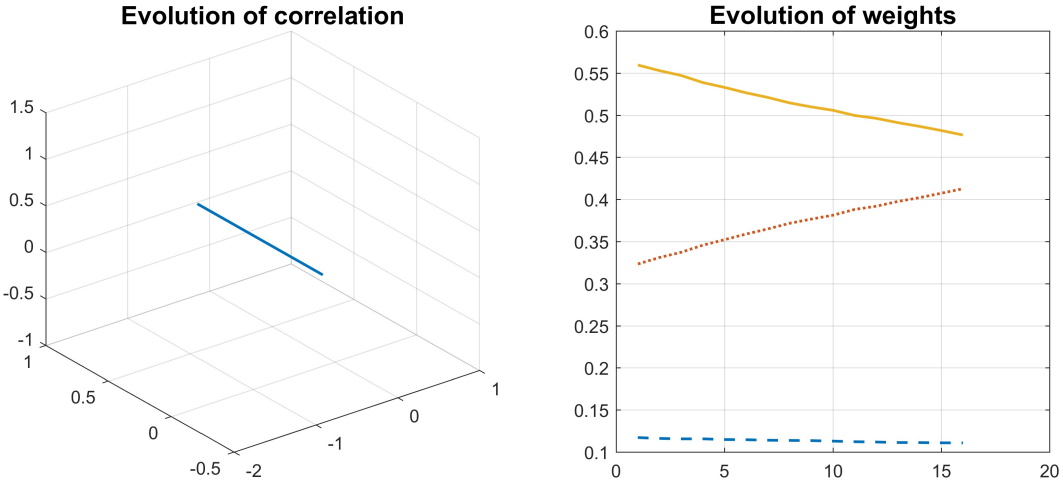


Figure 18: Evolution of correlation, and evolution of weights.

4.5.2 The effect of correlation tending to -1

Since the correlation goes towards -1, the assets tend to be less and less correlated. It seems reasonable to invest more in those two assets since they tend to move in opposite directions. This is what we observe in the graphs: the proportion of the assets that are negatively correlated increases while the proportion of the non-correlated asset decreases.

4.6 Student's t-copula with GARCH(1,1)

In this section we introduce the first simulation environment that fits into the framework suggested for risk management by Nystrom and Skoglund (2002). We model the dependence between the returns using a t-copula and for the marginal distribution of each return a GARCH(1,1) time-series model is used. The time-series model is described by (1) and (3) in Section 2.3.

Each individual return is described by

$$r_t^i = \mu^i + \epsilon_t^i$$

for $i = 1, \ldots, 3$, where $\epsilon_t^i = \zeta_t \sigma_t^i$ with $\zeta_t \overset{iid}{\sim} [0, 1]$. Thus, $\epsilon_t^i \sim [0, \sigma_t^2]$. We model σ_t^2

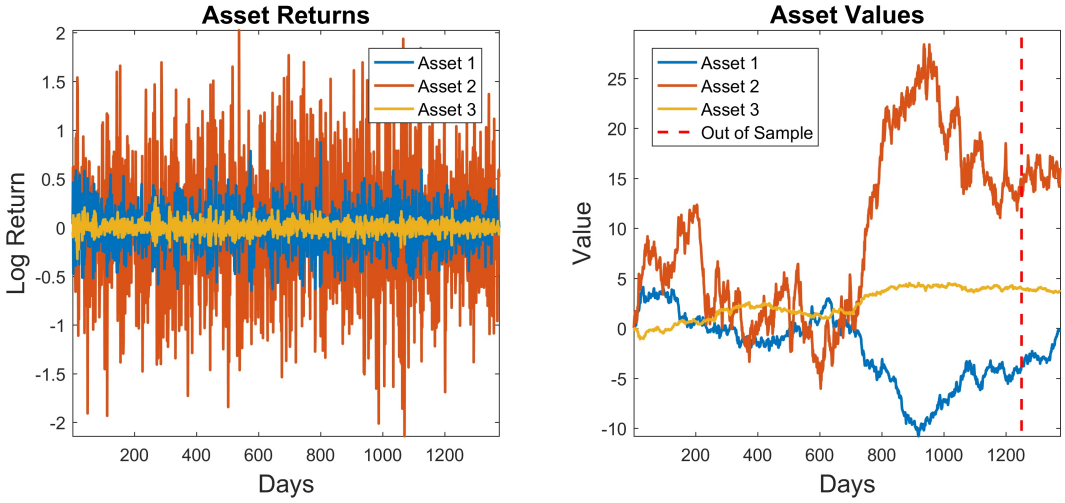


Figure 19: Sample path for three assets with t-copula dependence and GARCH(1,1) marginals.

using a GARCH(1,1), which gives a model of the form,

$$(\sigma_t^i)^2 = \omega^i + \alpha_1^i \epsilon_{t-1}^2 + \beta_1^i (\sigma_{t-1}^i)^2.$$

The copula dependence is introduced through the noise term, ζ_t, as described in Section 2.3. The time-series model parameters appear in Table 26. The student's t-copula has 3 degrees of freedom and its correlation matrix is

$$\Sigma_\rho = \begin{bmatrix} 1.0000 & -0.3 & -0.5 \\ -0.3 & 1.0000 & 0.1 \\ -0.5 & 0.1 & 1.0000 \end{bmatrix}.$$

This dependance structure is in line with what is estimated from the market in Section 5.2.1. The selected time-series parameters allow us to model three assets with very different volatilities, see Figure 19

Five years of historical, or in-sample, data is generated to construct the initial portfolios as well as estimate the initial return covariance. Four portfolios are then constructed and their performance is evaluated over the course of 6 months. The transition from the in-sample to out-of-sample data is indicated by the dashed red line in Figure 19.

Table 26: Time series parameters for the three asset case.

Parameter	Asset 1	Asset 2	Asset 3
ω^i	0.01	0.05	0.001
α^i	0.2	0.1	0.4
β^i	0.6	0.8	0.4

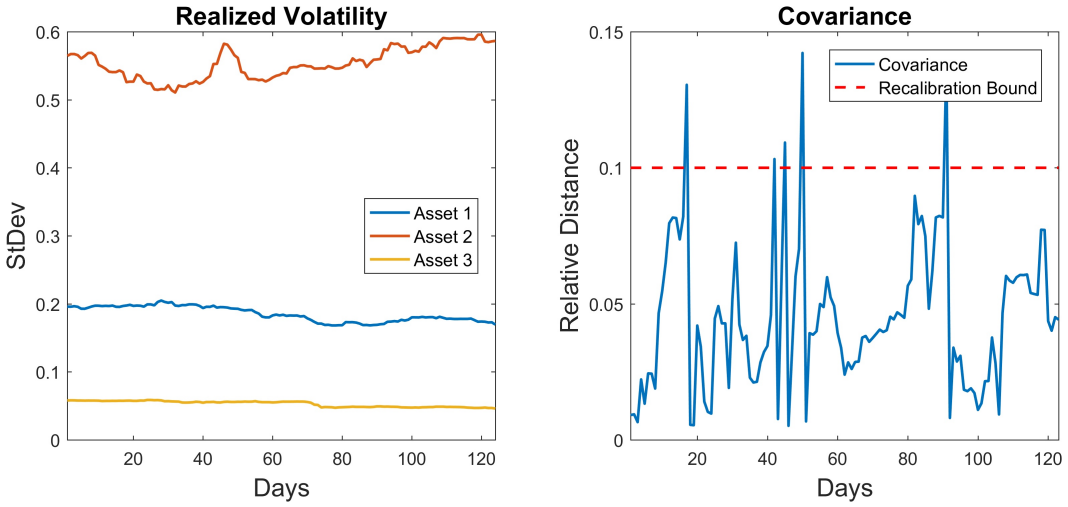


Figure 20: Realised volatility and relative change of the covariance of the simulated returns.

Portfolios are recalibrated based on the relative change in the estimated covariance matrix, i.e., when the covariance matrix has shifted more than 10% away of its previous value, all the portfolios are recalibrated. The covariance and realised volatility are estimated using a rolling 100-day window. For the illustrated path, the realised volatility of the returns as well as the relative change in their covariance is displayed in 20. Note that the relative change immediately drops after recalibration.

The considered portfolios are the classical Minimum Variance (MV) portfolio, a portfolio with constant and equal weights (EW), and two risk parity portfolios: the first based on the standard deviation risk measure, denoted ERC_σ and the second on the expected shortfall, ERC_{ES}. The asset allocation for the risk party portfolios over the one year period is displayed in Figure 21.

For this sample path, the portfolios are recalibrated 5 times over the out-of-

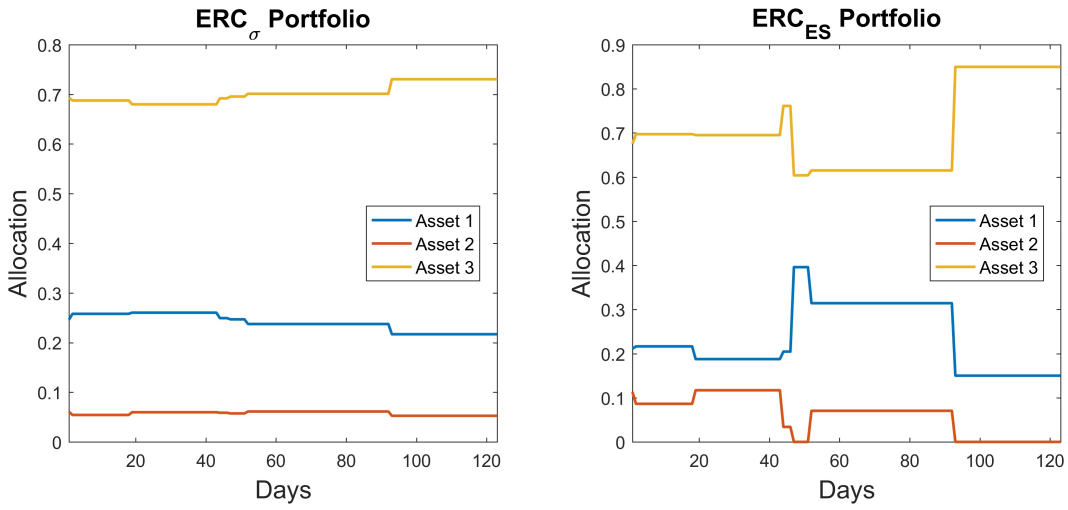


Figure 21: The asset allocation for the ERC_σ and ERC_{ES} over the 6 month out-of-sample period with a recalibration threshold of 10%.

sample period, or roughly once per month. The ERC_{ES} portfolio is significantly more dynamic (and thus more expensive) than the ERC_σ portfolio, completely closing out its position in the high-volatility asset, Asset 2, twice in this period. The first close-out corresponds to the increase in the realised volatility of Asset 2 between day 40 and day 60.

The advantage of the simulation environment is that we are not restricted to examining the results for a single path. A Monte Carlo experiment was run using $25\,000$ different realisations for the 6-month out of sample-path. This allows the portfolio profit and loss distributions to be estimated.

Both the MW and EW proft and loss are symmetric with mean approximately 0, as expected. The axis in Figure 22 are tight, so although it is not clearly visible in the depicted graph, there are actually long tails extending in both directions. The MV portfolio appears to be more exposed to high returns, although it is expected that this is a result of sample error. The depicted quantiles are for $\alpha = 0.01$.

The profit and loss distribution of the ERC_σ portfolio holds no surprises at this stage - the complex marginal used to model each underlying return has not significantly skewed the distribution. As in Section 4.2, the variance of the ERC_σ portfolio return lies between that of the MV and EW portfolios. Again, there are symmetric long tails, as evidenced by the axis limits.

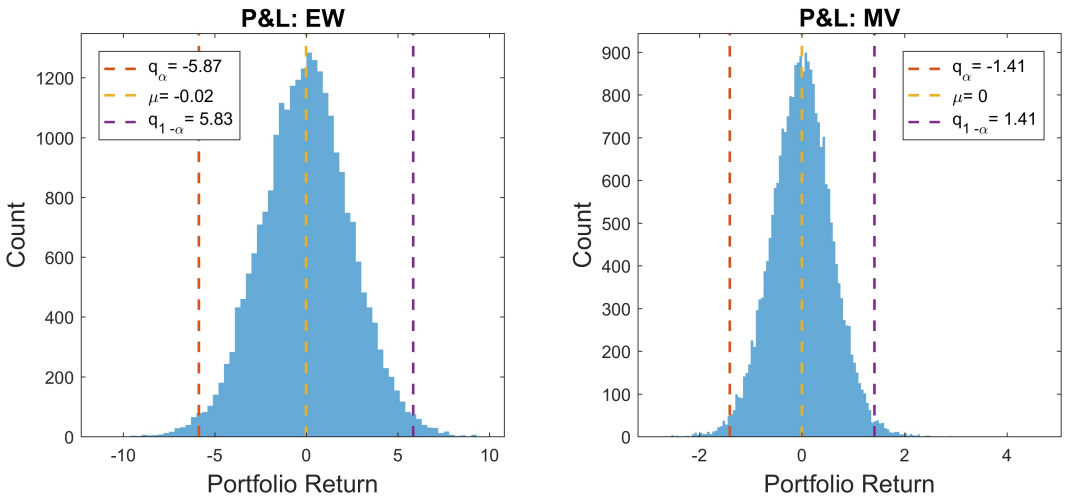


Figure 22: Profit and loss histograms for the MV and EW portfolios over the 6-month period.

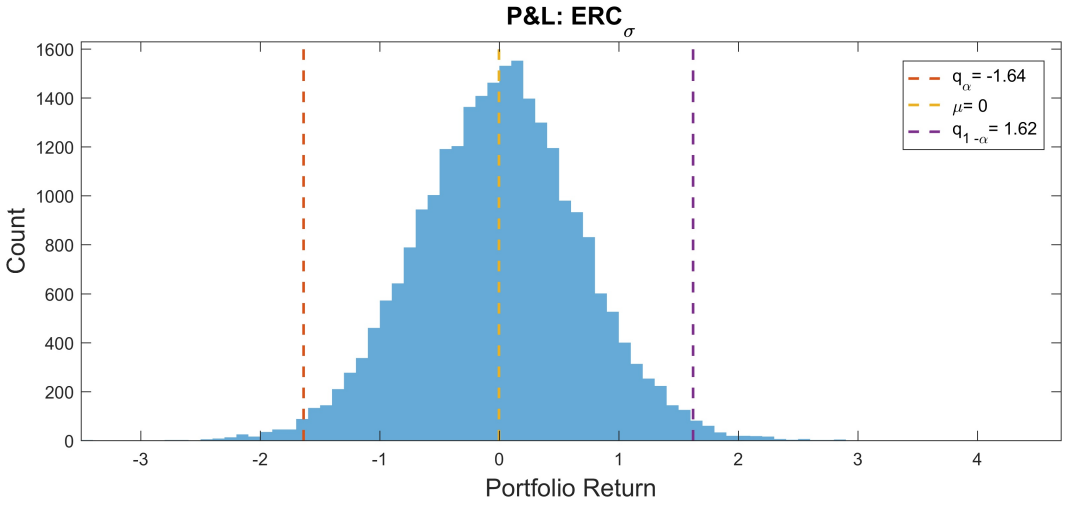


Figure 23: Profit and loss histogram ERC_σ portfolio over the 6-month period.

The investigation of the profit and loss distribution of the ERC_{ES} portfolio is computationally very expensive, as Monte Carlo methods have to be used to compute the marginal risk contributions for each asset. The result is pending.

5 Model Fitting

This part gives a description of the process followed for fitting a model to real world data and how to estimate the covariance matrix need to calculate the portfolio weights. Section 5.1 describes the process of estimation of the covariance matrix, section 5.2.1 outlines the process followed to fit time series models for the marginal distributions of the model while section 5.2.2 covers how the copula parameters were estimated.

5.1 Covariance Matrix Estimation

In this section we investigate how to estimate the covariance matrix of returns data. It is important to have a good estimate of the covariances between returns as these covariances are used in portfolio construction. Errors in your covariance matrix may lead to errors in your portfolio construction. The two methods that will be used to compute the covariance matrix are the standard estimate and the shrinkage estimate. We will assume that the returns are independent and identically distributed.

The standard estimator uses the sample mean and sample covariance. This matrix has little structure and therefore contains estimation error. This estimation error is especially prevalent when the number of assets is less than the number of realisations (Stefanovits, 2010). Suppose we have N stocks and T observations of each of these assets. The sample return and covariance are defined by:

$$\hat{\mathbf{r}} = \frac{1}{T}\sum_{i=1}^{T}\mathbf{r_i} \qquad \hat{\Sigma} = \frac{1}{T}\sum_{i=1}^{T}(\mathbf{r_i} - \hat{\mathbf{r}})(\mathbf{r_i} - \hat{\mathbf{r}})'. \tag{8}$$

The shrinkage estimator is a weighted linear combination of a structured matrix F, referred to as the shrinkage target, and the sample covariance $\hat{\Sigma}$:

$$\Sigma_{shrink} = \alpha F + (1-\alpha)\hat{\Sigma} \tag{9}$$

where α is the shrinkage constant which takes on values between 0 and 1 (Ledoit and Wolf, 2004). The shrinkage target in equation (9) is designed to have very few free parameters as well as being positive definite. This results in a shrinkage estimate that is always positive definite. All the variables that are needed to calculate the shrinkage estimate are extracted from the sample covariance matrix. Firstly the

shrinkage target is estimated using the following equations:

$$F_{ii} = \hat{\Sigma}_{ii}, \qquad F_{ij} = \hat{\chi}\sqrt{\hat{\Sigma}_{ii}\hat{\Sigma}_{jj}} \tag{10}$$

where

$$\hat{\chi} = \frac{2}{(n-1)n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\hat{\chi_{ij}}, \qquad \hat{\chi}_{ij} = \frac{\hat{\Sigma}_{ij}}{\sqrt{\hat{\Sigma}_{ii}\hat{\Sigma}_{jj}}}. \tag{11}$$

It should be noted that $\hat{\chi}_{ij}$ is the correlation coefficient between assets i and j. These correlation coefficients are used to calculate $\hat{\chi}$ where $\hat{\chi}$ is the average correlation coefficient of $\hat{\Sigma}$. This results in shrinking target matrix that corresponds to all assets being equally correlated. The next step is to estimate the shrinkage constant α. The shrinkage constant should minimise the expected value of the distance between the shrinkage estimate and the actual covariance matrix. Ledoit and Wolf (2003) provides an estimate for this parameter:

$$\hat{\alpha} = \frac{1}{T}\frac{\hat{\pi} - \hat{\rho}}{\hat{\lambda}} \tag{12}$$

Ledoit and Wolf (2004) proves that $\hat{\alpha}$ can be estimated by κ, which asymptotically behaves like a constant if N is fixed and T tends to infinity:

$$\kappa = \frac{1}{T}\frac{\pi - \rho}{\lambda} \tag{13}$$

Equation (13) can be estimated by finding estimators for π, ρ and λ:

$$\hat{\pi} = \sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\pi}_{ij}, \qquad \hat{\pi}_{ij} = \frac{1}{T}\sum_{k=1}^{T}\left[\left(r_{ik} - \frac{1}{T}\sum_{l=1}^{T}r_{il}\right)\left(r_{ik} - \frac{1}{T}\sum_{l=1}^{T}r_{jl}\right) - \hat{\Sigma}_{ij}\right]^2,$$

$$\hat{\lambda} = \sum_{i=1}^{N}\hat{\pi_{ii}}, \qquad \hat{\rho} = \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{\hat{\chi}}{2}\left(\sqrt{\frac{\hat{\Sigma_{jj}}}{\hat{\Sigma_{ii}}}}\hat{\psi}_{ii,ij} + \sqrt{\frac{\hat{\Sigma_{ii}}}{\hat{\Sigma_{jj}}}}\hat{\psi}_{jj,ij}\right),$$

where

$$\hat{\psi}_{ii,ij} = \frac{1}{T}\sum_{i=1}^{T}\left[\left(\left(r_{ik} - \frac{1}{T}\sum_{l=1}^{T}r_{il}\right)^2 - \hat{\Sigma}_{ii}\right)\left(\left(r_{ik} - \frac{1}{T}\sum_{l=1}^{T}r_{il}\right)\left(r_{jk} - \frac{1}{T}\sum_{l=1}^{T}r_{jl}\right) - \hat{\Sigma}_{ij}\right)\right]$$

$$\hat{\psi}_{jj,ij} = \frac{1}{T}\sum_{j=1}^{T}\left[\left(\left(r_{jk} - \frac{1}{T}\sum_{l=1}^{T}r_{jl}\right)^2 - \hat{\Sigma}_{jj}\right)\left(\left(r_{ik} - \frac{1}{T}\sum_{l=1}^{T}r_{il}\right)\left(r_{jk} - \frac{1}{T}\sum_{l=1}^{T}r_{jl}\right) - \hat{\Sigma}_{ij}\right)\right].$$

The above results are proven by Ledoit and Wolf (2004). The first consistent estimator $\hat{\pi}$ is the scaled sum of the asymptotic variances of the variables in the sample covariance matrix. The second consistent estimator is $\hat{\lambda}$ which is the accounts for the errors in the shrinkage target. And lastly $\hat{\rho}$ is the scaled sum of the asymptotic covariances of the variables in the shrinkage target (F) with the variables in the sample covariance matrix. It should be be noted that the shrinkage target has to be between 0 and 1, therefore if it is below 0 we truncate it to 0 and if it is above 1 then we truncate it to 1.

Example 14. We test the performance of the shrinkage estimator against that of the standard estimator. This is done by generating T random normal returns for N assets with a specified volatility for each asset. These returns are correlated by selecting a correlation matrix, constructing the covariance matrix and then using the Cholesky transformation to correlate the assets. The covariance of the returns is then estimated using equation (9) and equation (8). Equation (9) is implemented in Matlab using the code provided by Ledoit and Wolf (2003). The Frobenius norm is calculated 100 times to compare the difference between the actual covariance matrix and the estimated covariance matrix. We chose 4 increasing numbers of realisations with 20 assets to illustrate how the shrinkage estimate performs against the sample estimate. A random correlation matrix and volatilities for the 20 assets were generated.

Figure 24 above show that the shrinkage estimate is a better estimate for the covariance matrix than the sample estimate. This is especially true when the number of realisations is less than the number of assets. As the number of realisations are increased then sample and and shrinkage estimate tend toward the same values.

5.2 Fitting the model

Our data consists of the returns of 31 assets. These assets consist mostly of indices and cover most of the major asset classes. We split the time frame into a training set, from 05/01/2010 to 21/04/2016, and a 250 day testing window following the training set. This section focuses on fitting model to the training data of three assets: the S&P 500 Index, the USD/EUR exchange rate and the US Treasury Bonds Index.

5.2.1 Fitting Time Series Marginal Distributions

To fit time series models as the marginals for our data, we roughly follow the Box-Jenkins method (Pankratz, 2009). This section illustrates the procedure followed for fitting a model for the S&P 500 Index. Similar procedures were followed for the remaining assets. A quick glance at the daily logarithmic returns, displayed in figure 25, shows that the returns series appear to be stationary.

The sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) plots in figure 26 illustrate that there exists some serial correlation in the return series, but this serial correlation is not persistent. To account for this, we initially fit a Autoregressive Moving Average (ARMA) Process to the return series.

The built-in MATLAB function, `estimate`, which uses Maximum Likelihood Estimation to estimate the model specifications, suggests that an ARMA(1,1) model best fits the data. Looking at the the ACF and PACF of the standardised residuals, figure 27, we can see that we have successfully removed most of the serial correlation. Finally, we check that the residuals are distributed correctly. Figure 28 contrasts the QQ-plots when using Standard Gaussian and Student's t innovations. Clearly, the assumption of Student's t residuals is much better. This is what we would expect due to the fat tails often associated with stock returns.

The ARMA(1,1) with Student's t innovations adequately compensates for the serial correlation which we witness in the returns, however we still expect the returns to display some heteroskedasticity. Looking at the ACF and PACF of the squared returns, figure 29, we see clear persistence in both cases. This suggests we need to account for heteroskedasticity using a GARCH model.

We fit an ARMA(1,1) for the conditional mean of the return series and a GARCH model for the conditional variance. The best fitting model is a GARCH(1,1). The resultant model form is,

$$r_t = \mu + a_1 r_{t-1} + \epsilon_t + b_1 \epsilon_{t-1}$$

where $\epsilon_t = \zeta_t \sigma_t$ with $\zeta_t \overset{iid}{\sim} [0,1] \Rightarrow \epsilon_t \sim [0, \sigma_t^2]$. We model σ_t^2 using a GARCH(1,1), which gives a model of the form,

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Table 27: Time series marginals parameter estimates for the three asset case.

Parameter	S & P 500	USDEUR	US Treasury Bonds Index
μ	7.83925×10^{-5}	0.000171222	0.000143469
a_1	0.916668	-0.706299	0.119305
b_1	-0.957224	0.682714	-0.201991
ω	5.05448×10^{-6}	2×10^{-7}	4.86642×10^{-7}
α_1	0.152863	0.0410829	0.112685
β_1	0.815075	0.955154	0.826055
Degrees of Freedom	4.69221	6.55816	10

The parameters for the time series marginals when for the three asset case are summarised in Table 27

The resulting standardised residuals display almost no significant serial correlation and the assumption of Student's t innovations holds. Applying the same methodology to each of our assets gives us the required marginal distributions and their parameters.

5.2.2 Fitting a Student's t Copula

We introduce dependence into our model via the innovations of the time series marginals. We know these innovations follow a Student's t distribution each with their respective degrees of freedom. We can transform the standardised residuals to uniform variates through the inverse Student's t cdf. We assume a t copula, as suggested by Nystrom and Skoglund (2002), and use MATLAB's `copulafit` function to estimate the degrees of freedom and the correlation matrix, Σ_ρ via Maximum Likelihood Estimation. The resultant estimates are $\hat{\nu} = 3.1887$ for the degrees of freedom and

$$\hat{\Sigma_\rho} = \begin{bmatrix} 1.0000 & -0.3603 & -0.5269 \\ -0.3603 & 1.0000 & 0.1288 \\ -0.5269 & 0.1288 & 1.0000 \end{bmatrix}$$

for the correlation matrix. We now have a fully specified model from which we can simulate from.

6 Realistic Risk Parity

6.1 Backtesting

In this section we investigate the performances of risk parity portfolios against real data. We want to see how our risk parity portfolios perform against the MV and EW portfolios in terms of returns and risk. Specifically, we expect our ES risk parity portfolio to have the lowest maximum drawdown since its the only portfolio that takes the tail dependences into account.

We use a t-Copula ARMA-GARCH marginals model. We estimate the parameters using 6 years of data and run the portfolios for 1 years. At each step in time, we compute the covariance matrix over a hundred days window. We compare it to our previous covariance matrix: if it changes more than a fixed percentage, we update the covariance matrix and recalibrate the t-Copula (we do not recalibrate the marginals). Then, we compute our new risk parity portfolios. For standard deviation one, we use the new estimate of the covariance matrix to compute it, and for the expected shortfall one, we do a Monte Carlo simulation using 100 000 samples generated using our newly estimated t-Copula.

The data contains a cross-section of returns from variety of asset classes from $05/01/2010$ until $21/04/2016$ plus an additional 250 days. The contents are listed in Table 28.

6.1.1 Three-asset case

In this section, we are only going to use three assets corresponding to a stock, a currency and a bond: SPXIndex, USDEURCurncy, BUSYIndex.

Impact of the sensitivity to changes of Σ

Whenever our covariance matrix changes by a certain amount, we recompute the weights for the ES and SD portfolio. We study what is the impact on asset allocation, P&L and maximum drawdown when the barrier for the changes in the covariance matrix is 5% or 10%. For the SD portfolio, the weight allocations are very close for $\delta\Sigma = 5\%$ or 10% and the resulting P&L are respectively 4.22 and 4.12. But for the ES portfolio, we get quite different results that we will see now.

As expected, when the sensitivity in the changes of Σ increases, the number of rebalancing in the portfolio increases as well. For a 5% sensitivity, it results in a

Table 28: Historical data description.

Asset	Class	Description
SPX Index	Equity	US - SP500
RAY Index	Equity	US - Russell 3000
DAX Index	Equity	Germany - DAX
UKX Index	Equity	UK - FTSE 100
NKY Index	Equity	Japan - Nikkei 225
EEM US Equity	Equity	Emerging Markets - iShares MSCI Emerging Markets ETF
USDEUR Curncy	FX	USD/EUR
USDGBP Curncy	FX	USD/GBP
USDJPY Curncy	FX	USD/JPY
USDCNY Curncy	FX	USD/CNY
FXJPEMCI Index	FX	EM FX
LUATTRUU Index	Bonds	US - Bloomberg Barclays US Treasury Total Return Unhedged USD
BUSY Index	Bonds	US - Bloomberg US Treasury Bond Index
SPBDUBIT Index	Bonds	US - S&P U.S. Treasury Bill Total Return Index
FTFIBGA Index	Bonds	UK - FTSE Actuaries UK Conventional Gilts All Stocks Index
GEDL Index	Bonds	Germany - BofA Merrill Lynch Diversified Germany Bond Index
BGER Index	Bonds	Germany - Bloomberg Germany Sovereign Bond Index
BRIT Index	Bonds	UK - Bloomberg U.K. Sovereign Bond Index
BGSV Index	Bonds	Bloomberg Global Developed Sovereign Bond Index
BEMS Index	Bonds	Bloomberg USD Emerging Market Sovereign Bond Index
BEUR Index	Bonds	Bloomberg Eurozone Sovereign Bond Index
BUSG Index	Bonds	Bloomberg US Government Bond Index
BPJN Index	Bonds	Bloomberg Japan Sovereign Bond Index
BLCSV Index	Bonds	EM Local Markets
HYG US Equity	Bonds	US High Yield
BJPN Index	Bonds	Bloomberg JPN
LUMSTRUU Index	Mortgage	US Mortgage
XAU Curncy	Commodities	Gold
OIL US Equity	Commodities	Oil
BCOM Index	Commodities	Bloomberg Commodities Index
DJCI Index	Commodities	Dow Jones Commodity Index
SPGSCI Index	Commodities	S&P GSCI (Goldman Sachs Commodity Index)

Table 29: Influence of the sensitivity of the relative change in σ.

	$\delta\Sigma = 5\%$	$\delta\Sigma = 10\%$
Number of rebalancing	47	24
P&L	3.063	0.0775
Max drawdown	-1.08	-1.02

Table 30: Performance of three asset portfolio.

	ERC_{ES}	ERC_{σ}	Mean Variance	Equally Weighted
P&L	3.15	4.12	2.87	5.80
Max drawdown	-1.08	-0.68	-0.66	-0.85

rebalancing of the portfolio roughly every week. We also observe that it has a big impact on the P&L: the P&L drastically improves when the sensitivity goes to 5%. Concerning the maximum drawdown, the impact is not very significant. We also observe that the weight allocations are different between the ES and SD portfolios: it means that those two risk measures are not capturing the same things. Indeed, we expect the ES portfolio to be grasping the tail dependence between the assets while the SD portfolio will not.

Change in the parameters of the copula

Each time $\delta\Sigma$ hits the barrier, we re-estimate the parameters of the copula. What we can see is that, when we recalibrate, the changes in the correlation matrix of the copula are between 1% and 16%, and the degree of freedom moves between 3.5 and 17.5.

Performance Comparison

We compare the performances of the portfolio when we rebalance the portfolios when $\delta\Sigma \geq 5\%$

On the one hand, the SD performs very well against all the other portfolios: its P&L is between the MV and the EW portfolio, and its maximum drawdown is very similar to the MV portfolio and much better than the EW one.

On the other hand, the ES portfolio has a P&L in between the MV and the EW portfolio, but its maximum drawdown is much worse. With this portfolio, we were expecting to capture the tail dependence between the assets and specifically avoid this kind of drastic losses. One explanation could be that our t-Copula is not capturing the tail dependence between the assets. Indeed, the t-Copula is symmetrical - which means booms and crashes will have the same tail dependence - and only has one degree of freedom for all assets: this may be a problem if we want to have different tail dependences between the assets.

Table 31: Influence of the sensitivity of the relative change in σ.

	$\delta\Sigma = 5\%$	$\delta\Sigma = 10\%$
Number of rebalancing	56	26
P&L	0.026	0.022
Max drawdown	-0.0075	-0.0056

Table 32: Performance of 31 asset portfolio.

	ERC_{ES}	ERC_{σ}	Mean Variance	Equally Weighted
P&L	0.026	0.033	0.022	0.030
Max drawdown	-0.0075	-0.0057	-0.0035	-0.0159

6.1.2 All asset case

We now consider the case when we have 31 assets from different markets (commodities, FX, stocks, bonds).

Impact of the sensitivity to changes of Σ

Here, we consider the impact on asset allocation, P&L and maximum drawdown when the barrier for changes in the covariance matrix is 5% or 10%.

Again, as expected, when the sensitivity in Σ increases, the number of rebalancing increases too. The P&L are are quite close, but the maximum drawdown is worse when the sensitivity increases.

Change in the parameters of the copula

Performance Comparison

We compare the performances of the portfolio when we rebalance the portfolios when $\delta\Sigma \geq 5\%$

The risk parity portfolio we have built using ES is in between the MV and EW portfolios in terms of performance (regarding P&L and maximum drawdown). The SD portfolio outperforms the EW in both returns and risk. However, we would have expected the ES portfolio to have a much lower maximum drawdown compared to the SD portfolio. Indeed, the SD portfolio only takes into account the covariance matrix of the returns in its construction, but the covariance matrix does not capture the tail dependence between assets. On the other hand, the ES portfolio construction is supposed to take into account the tails when we compute it.

One of the possible explanation is that the t-Copula that we used to model the dependences between assets does not really capture the different tail dependences between the assets.

6.1.3 Conclusion

In this section, we wanted to see how our ES and SD risk parity portfolios performed when we run them on real data. Our dataset was composed of different categories of assets such as stocks, bonds, commodities or currencies. We assumed that the underlying copula was a t-Copula with ARIMA GARCH and we estimated the parameters of the copula and the marginals using the first five years of the data. Then, we used the last year to see how the portfolios behaved. Every time there was a significant change in the covariance matrix of the returns, we re-estimated the weight allocation in the portfolios and the parameters of the copula.

What we observed is that the performance of the SD portfolio is much better compared to the EW portfolio (it has higher returns and lower risk). In comparison to the MV portfolio, it has higher returns but lower risk. Since the SD portfolio is only based on the covariance matrix, we expected such results - except the fact that the SD outperforms the EW portfolio in both returns and risk. However, the performances of the ES portfolio are a bit different. In term of returns, it is between the EW and the MV portfolios. But we were expecting this risk portfolio to capture the tail dependence between the assets, and thus avoid drastic losses. This is not what we observed. Indeed, in the case of three assets, it is the portfolio which has the maximum drawdown. In the case of 31 assets, even though its maximum drawdown is better than the EW portfolio, it is worse than the SD one. A possible cause to this is that the ES portfolio does not capture the tail dependence between the assets. This may be because the copula we are using is not fit to capture the real tail dependence. It would be interesting, for instance, to test this portfolio when we use a skewed t-copula with different degrees of freedom for each assets to model the dependence.

7 Conclusion

In this project we investigated the performance of risk parity portfolios for various risk measures, most notably standard deviation and expected shortfall, in an

attempt to explain their real-world performance. Several interesting features were discovered.

In the simulation environment it is clearly shown how the standard deviation based risk parity portfolio is fairly robust - indicating that it would be ideally suited for a buy-and-hold strategy. The expected shortfall portfolio is more dynamic, requiring frequent trades and often completing closing out positions. This seems to indicate that a buy-and-hold strategy would also be inappropriate for risk parity portfolios based on other risk measures more complex than the standard deviation, and could provide an explanation for the poor performance of these portfolios in the real-world stress period.

The performance of the expected shortfall and standard deviation based risk parity portfolios in the simulation environment, even in the most realistic case, did not match the performance when using the real-world data. The changing correlation and degrees of freedom of the estimated t-copula could indicate that a time-varying model of the return interdependence is needed. And, as shown, the current copula framework does not allow for changing tail dependence, which would definitely affect the performance of the expected shortfall portfolio.

Finally, although the backtesting indicated that our most realistic simulation environment is not realistic enough, this in itself actually provides the most explanatory power for the performance of risk parity portfolios in the real-world: correctly modelling the changing interdependence of the underlying returns is very complex and computationally demanding.

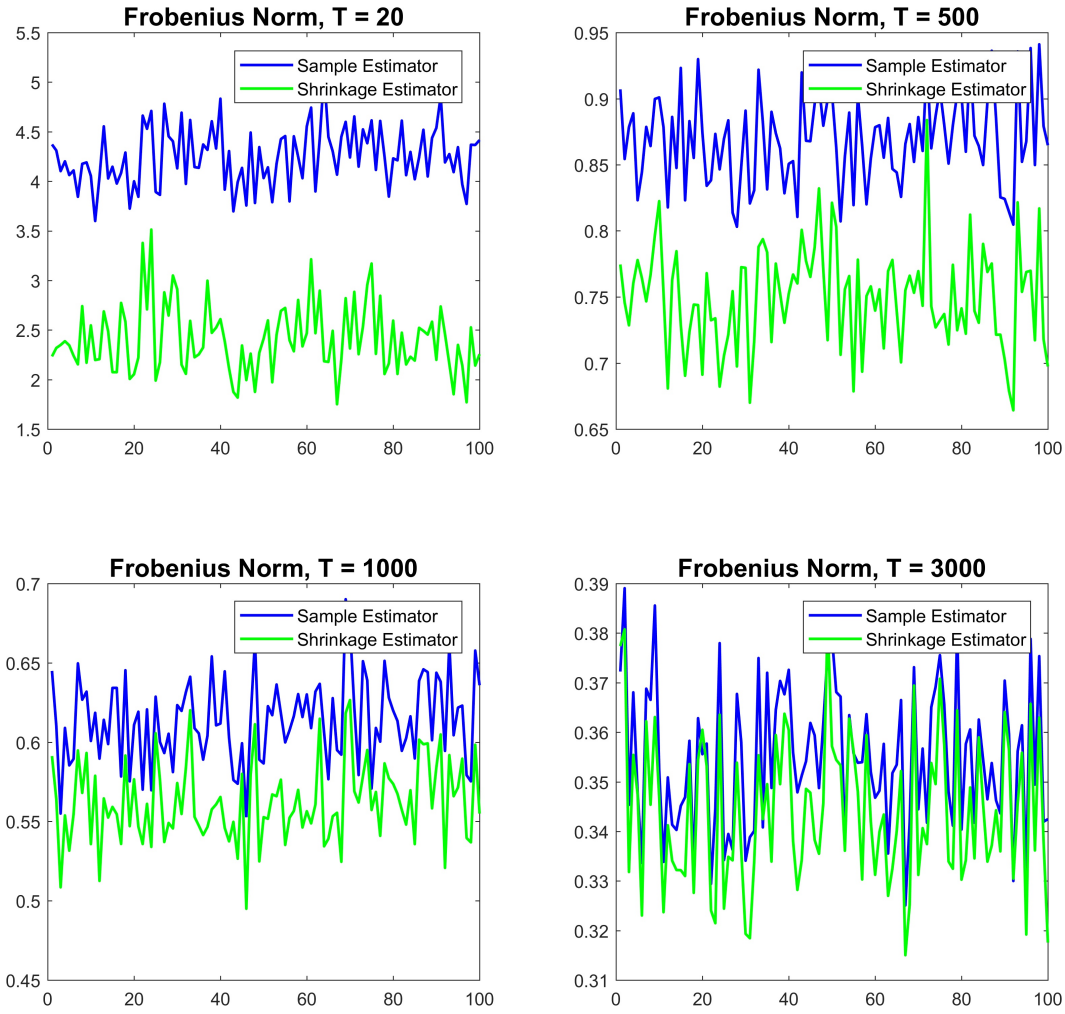


Figure 24: Shrinkage estimate vs. sample estimate illustration

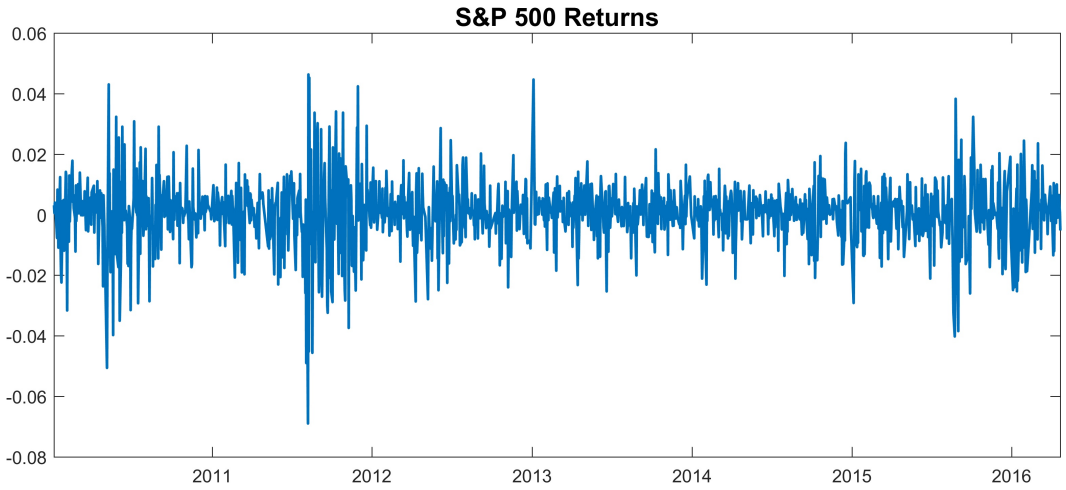


Figure 25: Daily Logarithmic Returns.

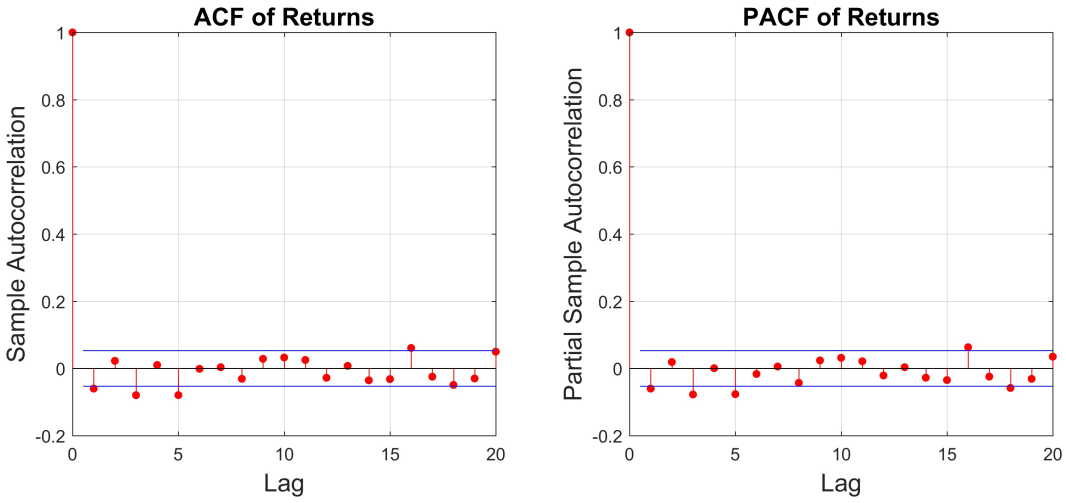


Figure 26: Sample ACF and PACF of Returns.

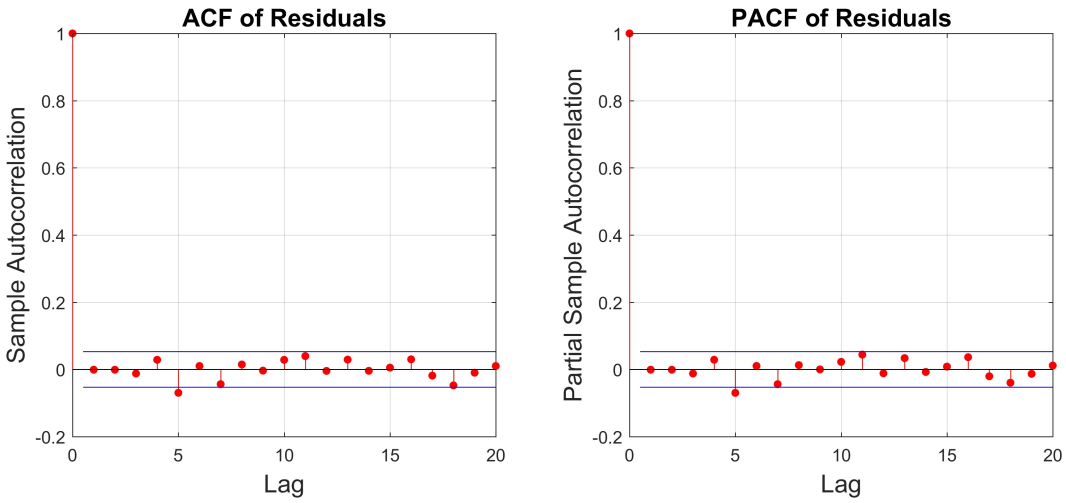


Figure 27: Sample ACF and PACF of residuals of ARMA(1,1) model.

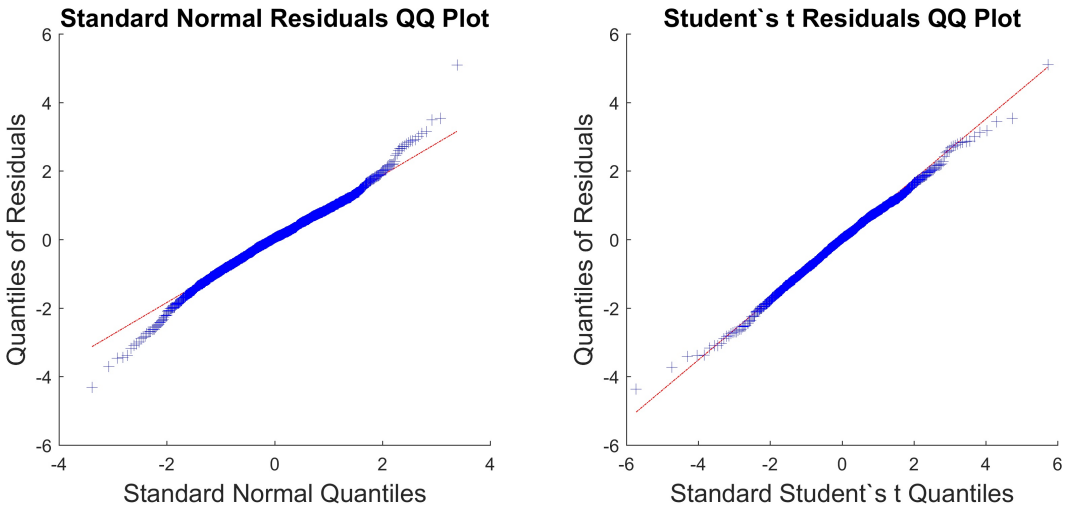


Figure 28: Quantile-Quantile plots comparing the assumptions of Standard Gaussian and Student's t distributed residuals.

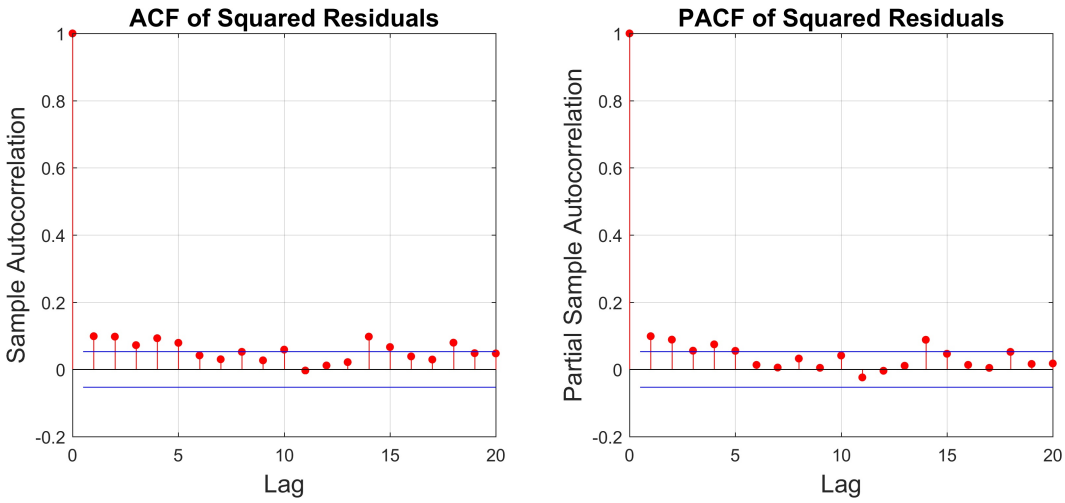


Figure 29: Sample ACF and PACF of the squared residuals.

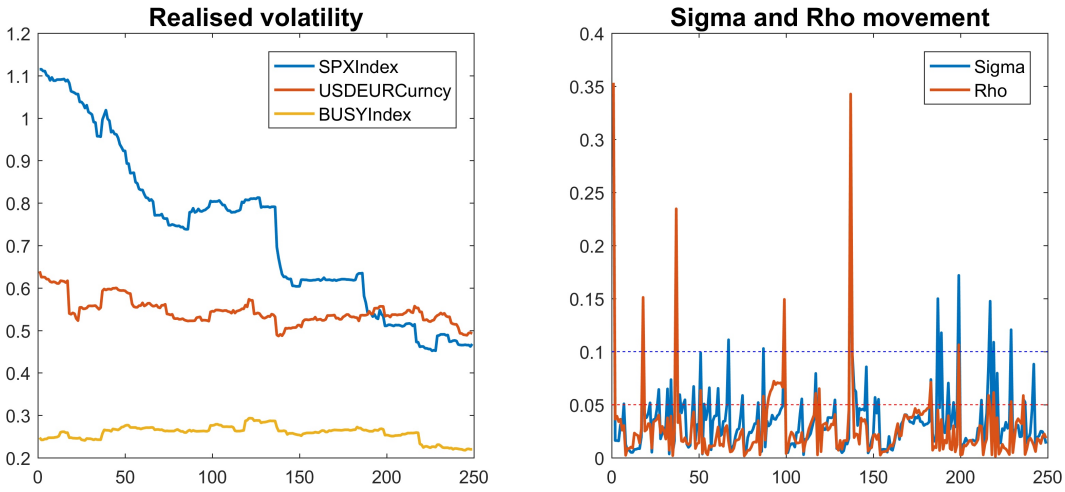


Figure 30: Realised volatility, changes in Sigma and Rho.

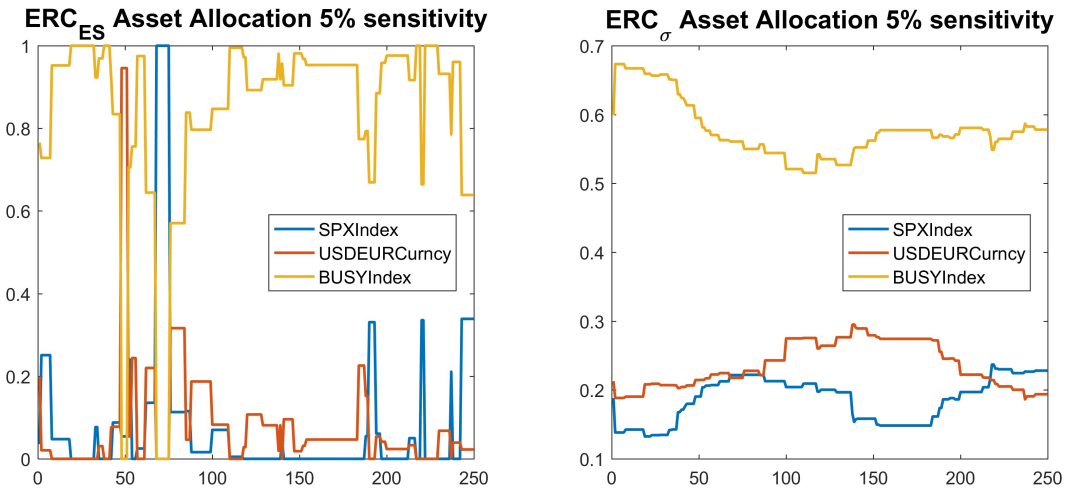


Figure 31: Asset allocation through time with 5% sensitivity.

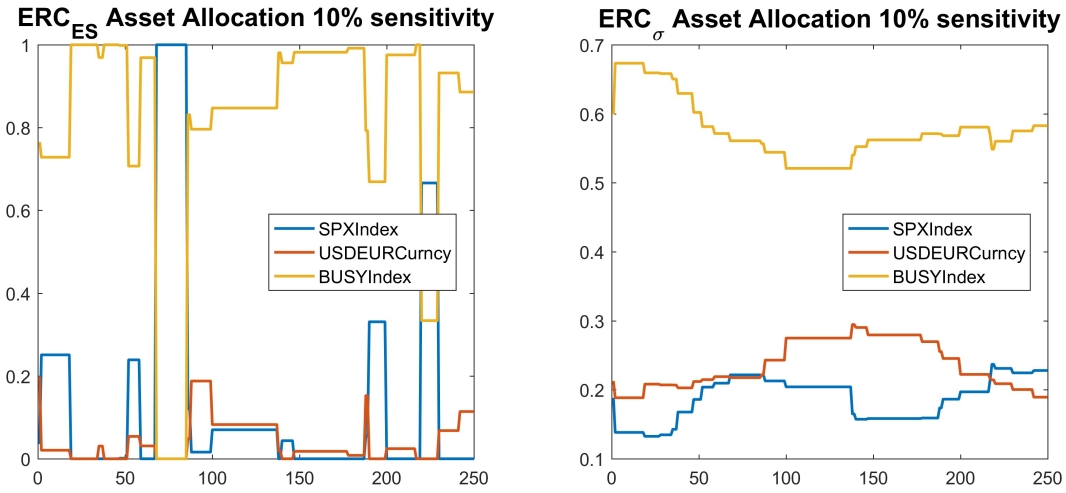


Figure 32: Asset allocation through time with 10% sensitivity.

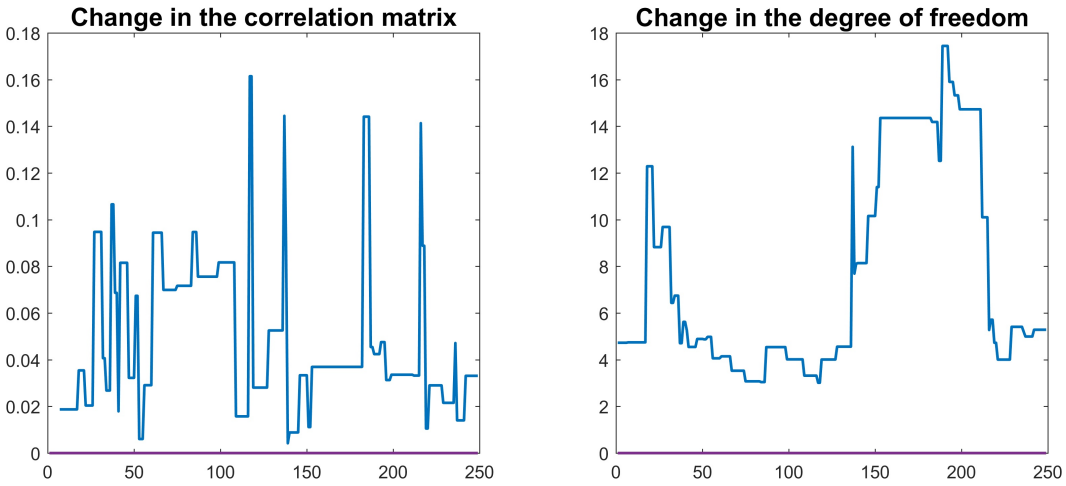


Figure 33: Parameters of copula through time.

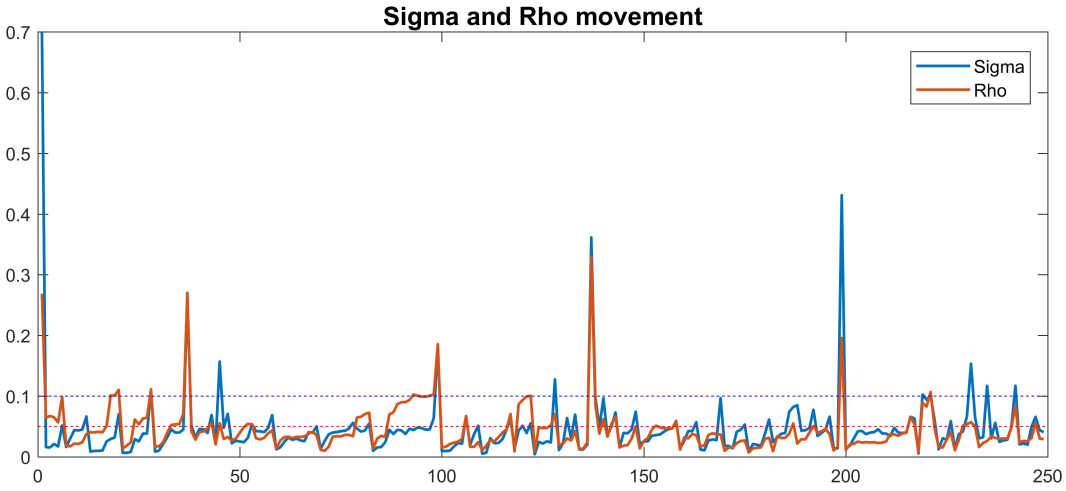


Figure 34: Realised volatility, changes in Sigma and Rho.

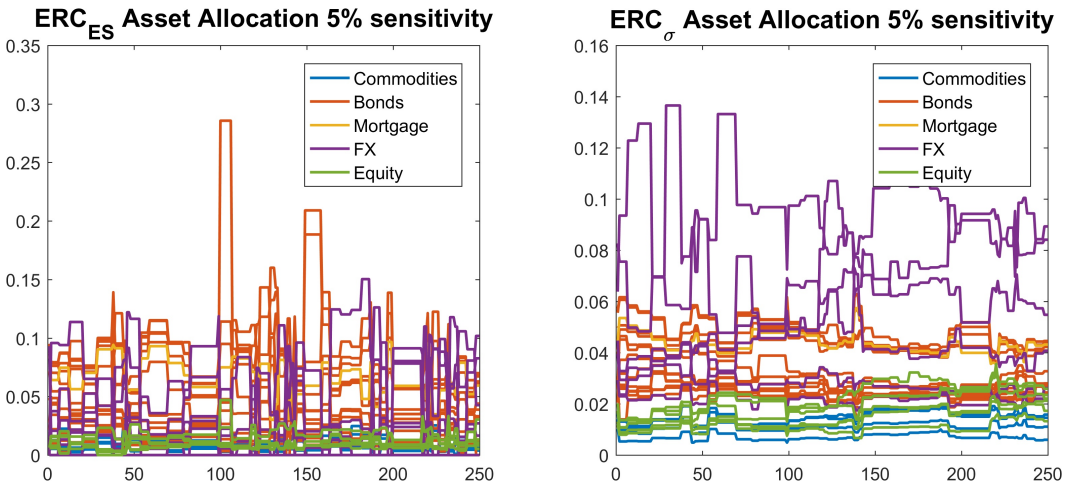


Figure 35: Asset allocation through time with 5% sensitivity.

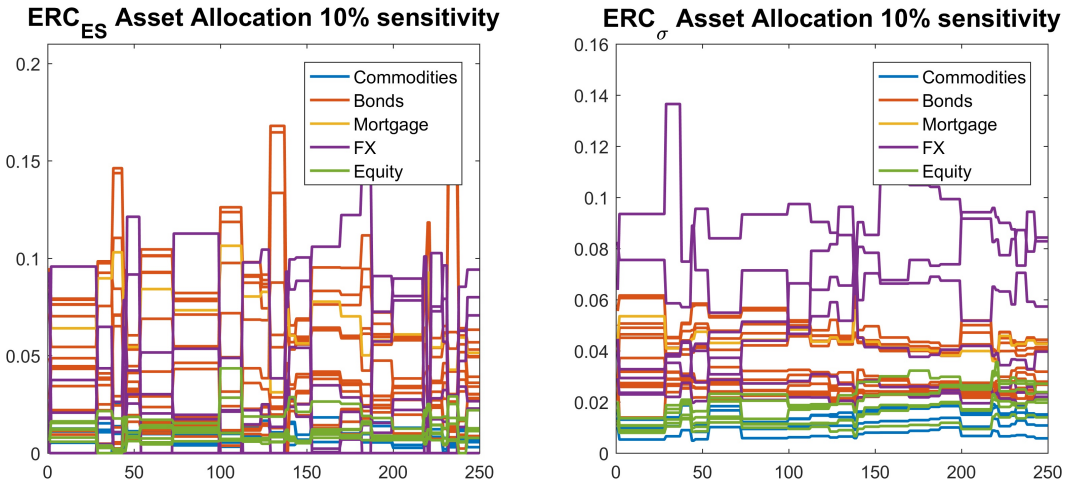


Figure 36: Asset allocation through time with 10% sensitivity.

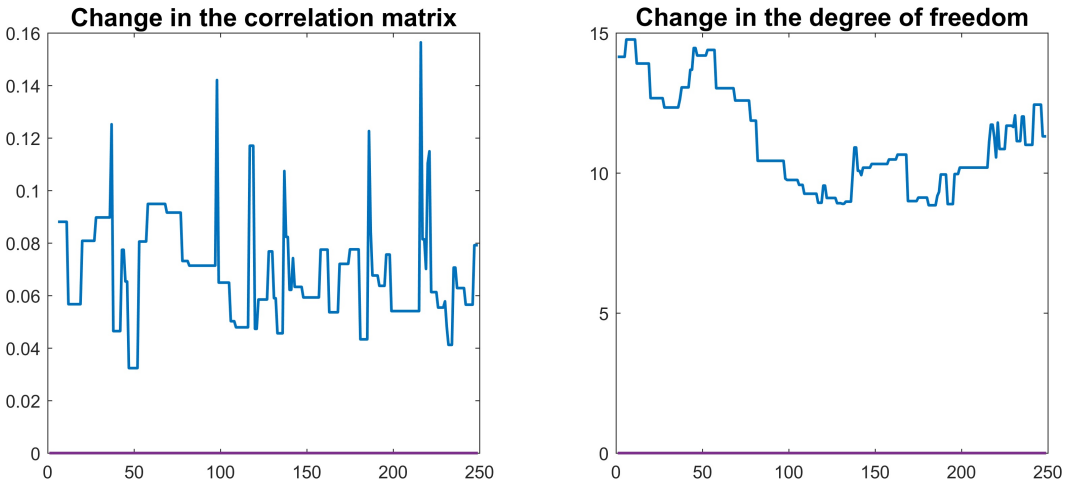


Figure 37: Parameters of copula through time.

Bibliography

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., Roncalli, T., 2000. Copulas for finance-a reading guide and some applications.

Embrechts, P., Lindskog, F., McNeil, A., 2001. Modelling dependence with copulas. Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich.

Griveau-Billion, T., Richard, J.-C., Roncalli, T., 2013. A fast algorithm for computing high-dimensional risk parity portfolios.

Hofert, M., 2008. Sampling archimedean copulas. Computational Statistics & Data Analysis 52 (12), 5163–5174.

Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of empirical finance 10 (5), 603–621.

Ledoit, O., Wolf, M., 2004. Honey, i shrunk the sample covariance matrix. The Journal of Portfolio Management 30 (4), 110–119.

Maillard, S., Roncalli, T., Teïletche, J., 2010. The properties of equally weighted risk contribution portfolios. The Journal of Portfolio Management 36 (4), 60–70.

McNeil, A. J., Nešlehová, J., 2009. Multivariate archimedean copulas, d-monotone functions and ??-norm symmetric distributions. The Annals of Statistics, 3059–3097.

Nelsen, R. B., 1999. Introduction. In: An Introduction to Copulas. Springer, pp. 1–4.

Nystrom, K., Skoglund, J., 2002. A framework for scenario-based risk management. Preprint, Swedbank.

Pankratz, A., 2009. Forecasting with univariate Box-Jenkins models: Concepts and cases. Vol. 224. John Wiley & Sons.

Sklar, M., 1959. Fonctions de repartition an dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris 8, 229–231.

Stefanovits, D., 2010. Equal contributions to risk and portfolio construction. Masters thesis, ETH Zürich.

Tasche, D., 1999. Risk contributions and performance measurement. Report of the Lehrstuhl für mathematische Statistik, TU München.

Tasche, D., 2007. Capital allocation to business units and sub-portfolios: the euler principle. arXiv preprint arXiv:0708.2542.

An Early Warning System for Financial Crises and Long-term Asset Management

Team 2

FEROZ BHAMANI, University of Cape Town
NICOLAI HAUSSAMER, University of Cape Town
YUPENG JIANG, University College London
THABANG MATHONSI, University of Cape Town

Supervisor:
MATHEUS GRASSELLI, McMaster University

African Institute of Financial Markets and Risk Management

Contents

1 Introduction

A financial crisis is an event during which a given system becomes unstable, and is subject to abrupt and considerable change. Crises place stress on a system, and are often characterised by the detrimental effects they can have; indeed, those with sizeable exposure to a system experiencing a crisis may be temporarily or permanently harmed as a result. These considerations are particularly relevant to financial systems and the corresponding markets. The three primary types of crisis in finance are banking crises, currency crises, and stock market crises – each of which may impact the values of assets in the money markets, bond markets, and equity markets. In light of this, it would be of great benefit to a long-term investor to have an early warning system whereby they can reliably anticipate a crisis, and which would allow them to adapt their investment strategy appropriately.

The prior research on this topic by Grasselli (2013) and Vermersch (2013) has yielded an early warning system comprising a set of crisis indices. These indices gauge the possibility of a future crisis within a defined period. As the index values change over time, they can be used to adjust the asset allocation of an investor's portfolio. Typically, the goal of using early warning systems to adjust the asset allocation is to maximise the gains (minimise the losses) in the value of the portfolio, although it is possible to apply different goals within the same framework. The central premise underlying the development and use of an early warning system as part of an investment strategy is that crises in financial markets are intrinsically linked to the behaviour of economic variables, see Grasselli (2013) and Vermersch (2013). Moreover, the behaviour of economic variables in the present has some non-trivial *predictive* ability regarding the emergence of financial crises in future. For example, a higher lending rate is an indicator for a potential financial crisis, as this may lead to funding problems for companies. We can observe from Figure 1 that the level of lending rate in the United States issues some accurate signal for its actual crises.

Demirgüç-Kunt and Detragiache (1998) lists several determinants indicators for banking crises. Kaminsky et al. (1998) proposed the *signals approach* as a way to predict (or, at least, anticipate the possibility of) a crisis. Under this approach, several economic variables serve as crisis *indicators*. When they exceed certain thresholds, the signals issued by indicators suggest the possibility of future crises. The optimal

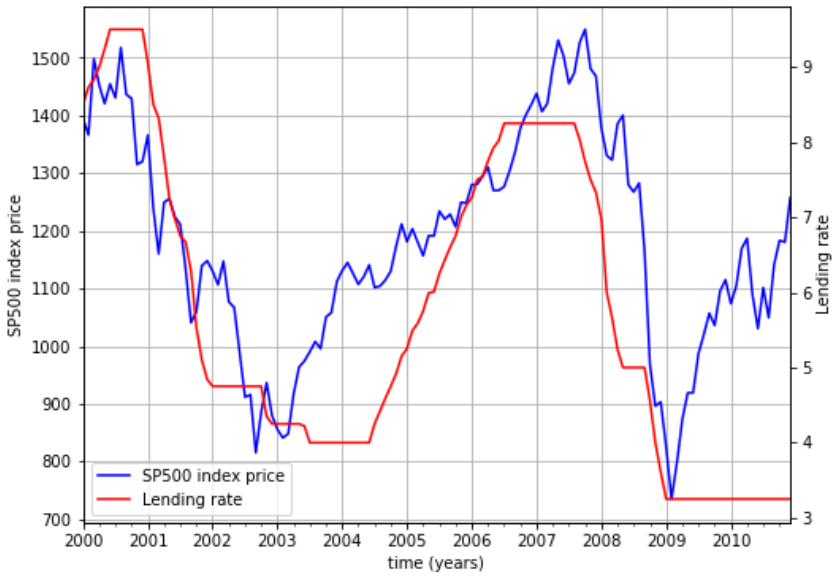


Figure 1: The monthly average of the SP500 index price (blue) and the US lending rate (red) from 2000 to 2010.

levels for these thresholds are found by ensuring that false signals (Type I and Type II errors) are as infrequent as possible, and that other measures of the quality of indicator performance, such as the average time between the first signal being issued and the subsequent crisis emerging, are relatively high. Drehmann and Juselius (2014) continued in a similar vein as Kaminsky et al. (1998) in order to learn more about banking crises in particular. Two important aspects of that research are the use of signal persistence as another measure of the quality of an indicator, as well as the conclusion that there are no unique or optimal combinations of indicators in the signals approach.

The research report by Grasselli (2013) points to several examples of economic variables that precede crises, some of which are given below for illustration:

- changes in international reserves, current account balance, or exchange rates can lead to currency crises;
- changes in real interest rates or GDP growth rates can lead to banking crises; and
- changes in lending rates, current account balance, or GDP growth rates can

lead to stock market crises.

Naturally, these variables, and many others included in the subsequent sections of this report, are amenable to use as indicators in the signals approach. However, making investment decisions on the basis of so many indicators is difficult. As such, the indicators are transformed and aggregated in order to produce *crises indices* for each type of crisis. These indices effectively summarise the information from the indicators regarding possible future crises.

The idea behind using crisis indices in investment strategy is that the possibility of a future crisis should alter how an investor would allocate their portfolio in the present. For a given initial portfolio, a strategy that integrates crisis indices involves using the observed crisis index values at each time step to change the asset allocation, with the degree of adjustment being a function of the type of crisis that may occur in future. To illustrate the benefit of using crisis indices in setting an investment strategy, we focus on three asset classes, rather than particular assets; these are cash, bonds and equities. Historical data are used to calibrate the parameters of the portfolio allocation function, based on the objective of maximising the portfolio return.

In this report, we present the application of the signals approach to a large collection of economic variables in Section 2, which includes an evaluation of each of these variables as indicators under this approach. Section 3 shows how these indicators have been used to construct crisis indices for banking, currency, and stock market crises. In Section 4, we present an implementation of investment strategies that incorporate with crisis indices, as well as the interim results obtained when optimising the parameters for the investment strategies. Finally, we include a comparison of the crisis index-based investment strategy and a benchmark investment strategy, and draw conclusions regarding the efficacy of the crisis index-based investment strategy.

2 Early signals for financial crisis from economic indicators

Early signals for financial crises can be used to predict the occurrence likelihood of economic recessions. In reality, regulators are interested in considering these early signals so as to formulate policies. Market speculators will seek these information for a potential future crisis in order to make profitable investment decisions. Asset managers will optimise their portfolio allocations by avoiding the prospective risks as well. In the literature investigating the early signals for financial crises, two major methods are presented: the logistic regression method and the signals approach, see Kaminsky et al. (1998), Grasselli (2013) and Vermersch (2013). The essence for both methods is identical, which is to find the relationship between the early running economic indicators and the financial crisis based on historical data of the indicators.

The method of logistic regression is widely utilised in the literature. However, the signals approach offers certain advantages over the logistic regression method. One of the disadvantages for the logistic regression is that it does not provide the ranking information of all the input indicators. However, a ranking can achieve for the indicators in the signals approach using different performance measures. Moreover, among the performance measures, the signals approach provides the approximate leading time before the crisis. This can be beneficial for decision making by the regulators and financial market participants. On the other hand, the logistic regression model will fail if there is a data period inconsistency for all the indicators, which may be caused by missing data. The usual method in the logistic regression model is to proceed with the interpolation algorithm in order to fill the gaps. However, the interpolation sometimes will miss some good signals or capture the wrong signals, which cause further inaccuracy for the approximation. In the signals approach, we treat the data indicator by indicator, which will solve this problem caused by missing data.

2.1 Definition of a crisis

In order to model the probability of the occurrence of a financial crisis, one has to present its definition. A financial crisis defined on the fulfilment of subjective criteria. In the following context, we define the various kinds of crises:

- The occurrence of a currency crisis is that the annual depreciation of the currency exceeds 15%;
- The occurrence of a banking crisis is that bank runs lead to the closure, merging or takeover by the public sector of one or more financial institutions;
- The definition of stock crises is often done using the running maximums of the stock indices or the returns:
    1. The `CMAX` definition of a stock crisis, which involves calculating the following:
    $$\texttt{CMAX}_t = \frac{P_t}{\max(P_{t-24}, ..., P_{t-1}, P_t)}.$$
    where P_{t-i} is the share price at time $t - i$. With the mean and standard deviation of `CMAX` series, one defines a stock crisis when the $\texttt{CMAX}_t$ drops below the n standard deviations from its mean.
    2. A alternative definition is to examine the following return series:
    $$\text{Return}_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$
    Similarly, we compute the mean and standard deviation of the returns time series. A stock crisis is defined by the fact that the returns index drops below n standard deviations from the mean.

2.2 The signals approach

As mentioned previously, the signals approach involves tracking the monthly values of a large set of economic indicators over time. It is necessary to analyse a transformed variable of the economic indicator rather than the level of the indicator itself. This is because the units of measurement for some indicators differ for different countries, and also there may be seasonality in the time series data. One of the transformations for the indicators is to perform a percentage change with its value in the same month of the previous year. For example, this can be done for quantities such as GDP, exports and imports. As an example, we plot the original level of the import value of South Africa and its percentage change in Figure 2 and 3. It may also be useful to consider the second difference percentage changes of

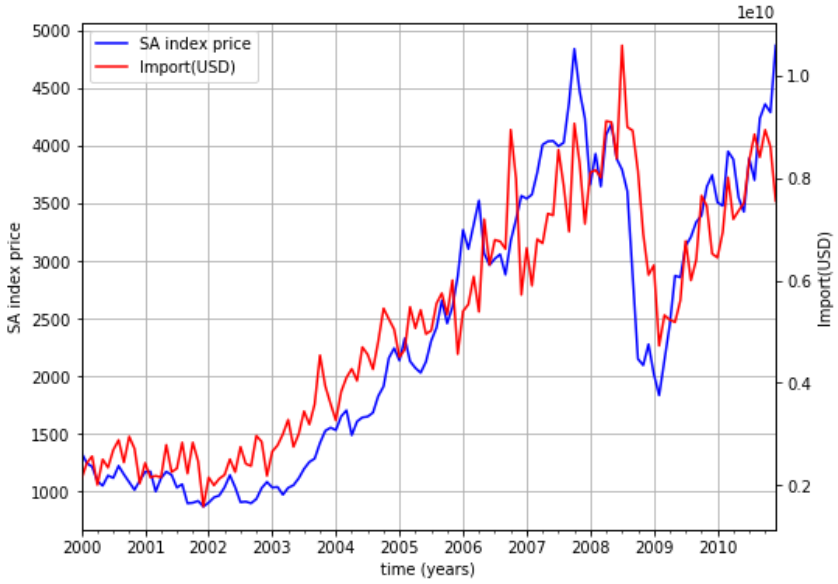


Figure 2: The monthly average of the JSE index price (blue) and the import (red) from 2000 to 2010.

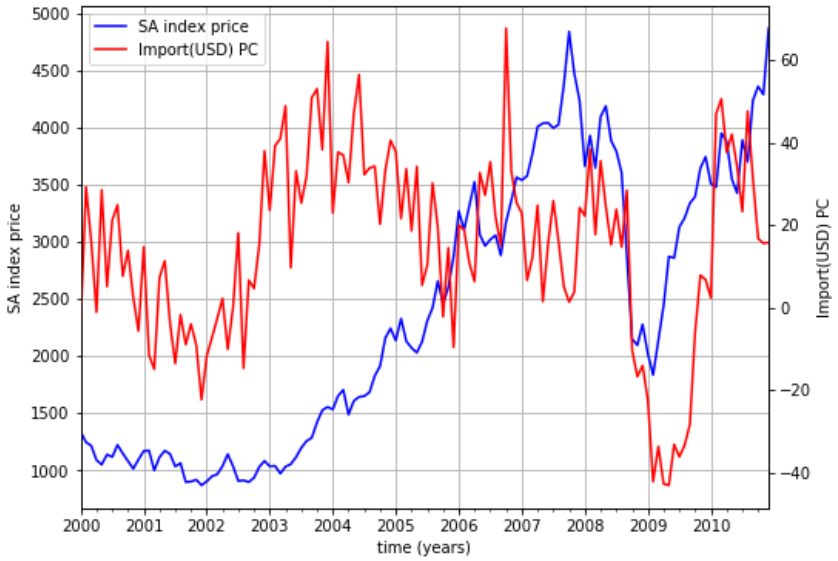


Figure 3: The monthly average of the JSE index price (blue) and the transformed import (red) from 2000 to 2010, where the transformation is the percentage change with the same month last year.

some variables. This provides the rate at which the indicator is increasing or decreasing. The second difference percentage changes are again obtained with the percentage changed data. For example, one of the variables that are considered in this way is GDP acceleration, which is the second differencing of GDP levels. In some cases, where it is also important to consider the de-trended level of the economic indicators, such as the inflation rate and the exchange rate. We list all our available indicators and their corresponding transformation in Table 1.

Table 1: The 33 indicators we consider and their corresponding transformations.

Indicator	Transformation	Indicator	Transformation
Current account	PC	Current account acc.	SD
Current account/GDP(USD)	PC	Domestic credit(NC)	PC
Domestic credit(USD)	PC	Domestic credit/GDP(USD)	PC
Deposit rate	DT	DT Deposit rate	PC
Exchange rate(SDR)	DT	DT exchange rate(SDR)	PC
Exchange rate(USD)	DT	DT exchange rate(USD)	PC
Exports/GDP(USD)	PC	GDP(USD)	PC
GDP acc.	SD	Imports/GDP (USD)	PC
Industrial production	PC	Industrial production acc.	SD
Inflation	DT	DT Inflation	PC
Lending rate	DT	DT Lending rate	PC
Lending rates/deposit rates	DT	M2(NC)	PC
M2 acc.(NC)	SD	M2/Reserves(NC)	PC
M2(USD)	PC	M2(USD) acc.	SD
M2/Reserves(USD)	PC	Real interest rate	DT
Reserves(NC)	PC	Reserves(USD)	PC
Reserves(SDR)	PC		

For some indicators, it is of interest to monitor whether the value of the indicator is above a certain threshold in the upper end of the distribution of values, while for other indicators it is of interest to monitor whether the value lies below a threshold in the lower tail of the distribution. For example, the percentage change of GDP going too low would be an indication of a crisis, while on the other hand having imports going too high could signal the possibility of a crisis.

Another parameter which needs to be chosen is the window of time over which the signals will predict whether there will be a crisis or not. This is termed as the signalling horizon, and is chosen to be a 24 month period in all our numerical ex-

amples with signals approach. The motivation for this choice is that an indicator could reasonably be able to predict a crisis to a maximum of this length of time into the future.

We now define quantities that will be important to assess how good a variable is at predicting the occurrence and the leading time of a financial crisis. If a variable exceeds its threshold level and this is followed by a crisis within 24 months, this is classified as a good signal. On the other hand, if there is no crisis in the 24 months following the signal, this is called a bad signal. We consider the following contingency table:

	Crisis (within 24 months)	No Crisis (within 24 months)
Signal Issued	A	B
No Signal Issued	C	D

where:

A : Number of months in which the variable issued a good signal,

B : Number of months in which the variable issued a bad signal,

C : Number of months in which the variable failed to issue a signal,

D : Number of months in which the variable did not issue a signal.

The best variables would be those where $B = C = 0$ and $A > 0, D > 0$. With these four quantities, we can construct statistics that summarise the performance of each variable in predicting the occurrence and average leading time of the crises:

- The ratio
$$\frac{A}{A+C}$$
gives the proportion of good signals issued out of the total number of good signals that could have been issued. The higher this quantity, the better the variable is at signalling a forthcoming crisis.

- Similarly,
$$\frac{B}{B+D}$$
gives the proportion of the bad signals issued out of the total number of bad

signals that could have been issued. The lower this quantity, the better the variable is at signalling a forthcoming crisis.

- The ratio of the previous two results gives the noise-to-signal ratio (NSR):

$$\text{NSR} = \frac{B/(B+D)}{A/(A+C)}.$$

- The percentage of crises detected (PCD) represents the proportion of the total number of crises for which there was at least one signal in the 24 months preceding the crisis.

- The probability of a crisis conditional on a signal from the variable:

$$P_{c|s} = P(\text{crisis}|\text{signal}) = \frac{A}{A+B}. \tag{1}$$

- The unconditional probability of a crisis:

$$P_c = P(\text{crisis}) = \frac{A+C}{A+B+C+D}. \tag{2}$$

  If the variable is good, then (1) would be higher than (2).

While calculating the above results, one can also calculate the following quantities:

- The average lead time (ALT), which is the average number of months before a crisis when the first signal is issued by the variable.

- The persistence of a variable, which represents the average number of signals issued in the 24 month window prior to the crisis.

2.3 The optimal percetile

The optimal percentile for each variable is chosen so as to have a balance between having too many false signals, which occurs if the threshold value is too close to the mean, and the risk of not giving a warning signal when a crisis is likely, which occurs if the threshold is too far into the tails of the distribution of the variable of interest. In Grasselli (2013) and Vermersch (2013), the optimal threshold for each indicator is obtained by finding the percentile of the values of the indicator that led to a minimum noise-to-signal ratio (NSR). We obtain the optimal percentile by

minimising the NSR and maximising the probability of detecting a crisis (PCD) simultaneously.

To achieve the objective, we rank different percentiles for each variable based on their corresponding NSR and PCD measures. To perform this ranking of the percentiles, we transform the NSR and PCD measures so that the transformed performance measure is in the similar scale range and same directionals with respect to performance. We perform the following transformations:

$$\widehat{\text{NSR}}_{\text{k}} = \frac{\max(\text{NSR}_{\text{k}}) - \text{NSR}_{\text{k}}}{\max(\text{NSR}_{\text{k}}) - \min(\text{NSR}_{\text{k}})};$$

$$\widehat{\text{PCD}}_{\text{k}} = \frac{\text{PCD}_{\text{k}} - \min(\text{PCD}_{\text{k}})}{\max(\text{PCD}_{\text{k}}) - \min(\text{PCD}_{\text{k}})},$$

where the $\max$ and $\min$ are taken over all possible percentile levels.

The optimal percentile is the percentile level with the highest values of the sum of $\widehat{\text{NSR}}$ and $\widehat{\text{PCD}}$. The optimal percentile for a specific indicator will be the same across all the countries in the sample, but the threshold value for each country will depend on the distribution of the variable for that country.

2.4 Aggregated performance measure

As the four performance measures show that some of the indicators outperforms the rest as early warning signal issuers from different perspectives, it is natural to construct a aggregated performance measure for each indicator by its transformed NSR, PCD, PER and ALT. The aggregated performance measure is constructed by the following steps

(S1) We transform our four original performance measures to re-scale their range

among $[0, 1]$ as follows, so that they are comparable.

$$\text{NSR}_k^* = \frac{\max(\text{NSR}_k) - \text{NSR}_k}{\max(\text{NSR}_k) - \min(\text{NSR}_k)}, \tag{3}$$

$$\text{PCD}_k^* = \frac{\text{PCD}_k - \min(\text{PCD})}{\max(\text{PCD}_k) - \min(\text{PCD}_k)}, \tag{4}$$

$$\text{PER}_k^* = \frac{\text{PER}_k - \min(\text{PER}_k)}{\max(\text{PER}_k) - \min(\text{PER}_k)}, \tag{5}$$

$$\text{ALT}_k^* = \frac{\text{ALT}_k - \min(\text{ALT}_k)}{\max(\text{ALT}_k) - \min(\text{ALT}_k)}, \tag{6}$$

where the $\max$ and $\min$ are taken over the set of different indicators.

(S2) We determine our aggregated performance measure by the average of NSR*, PCD*, PER* and ALT* as follows

$$\text{Performance}_k = \frac{\text{NSR}_k^* + \text{PCD}_k^* + \text{PER}_k^* + \text{ALT}_k^*}{4}. \tag{7}$$

The aggregated performance measure can be utilised to rank our indicators.

2.5 Numerical results

We present here the results of our signals approach for banking, currency and stock market crises, where the data is the monthly indicator data from 1960 to 2008. Banking and currency crises data was available for a total of 22 countries, whereas stock market crises data was available for a larger set of 46 countries. The data was obtained from Reinhart and Rogoff (2013), Reinhart and Rogoff (2009), the World-Bank Database and the IMS Database.

Table 2: Indicators for banking crises based on monthly data for 22 countries from 1960 to 2008.

	A	B	C	D	PCD	NSR	PER	ALT	Percentile	Performance
Reserves (USD)	581	1337	1694	5994	0,822	0,714	7,217	16,848	20	0,816
M2 acc. (NC)	402	1079	1738	5011	0,889	0,943	4,769	18,498	18	0,763
Reserves (SDR)	541	1406	1770	6037	0,784	0,807	6,696	16,708	20	0,761
Inflation (level)	562	1271	1713	5634	0,501	0,745	5,886	19,056	80	0,716
GDP (USD)	460	507	1845	6856	0,595	0,345	5,720	15,436	10	0,700
GDP acc.	392	1253	1912	6094	0,835	1,002	4,407	17,772	17	0,695
Exchange rate (SDR)	378	317	1933	7340	0,348	0,253	4,166	18,082	7	0,657
Exchange rate (USD)	359	336	1952	7321	0,301	0,282	3,857	17,967	7	0,619
Lending rates/deposit rates	332	529	922	2518	0,484	0,656	4,695	16,198	80	0,589
Inflation PC	300	592	1975	6109	0,679	0,670	3,127	15,649	90	0,582
Lending rate (level)	302	232	1127	3186	0,425	0,321	4,366	14,903	89	0,574
Current account	408	741	1304	3291	0,569	0,771	4,339	15,123	20	0,550
Real interest rate	226	149	1454	3525	0,404	0,301	3,655	15,053	93	0,550
Current account/GDP (USD)	408	741	1304	3291	0,563	0,771	4,310	15,130	20	0,547
M2 (USD)	288	374	1858	5802	0,485	0,451	3,392	15,106	8	0,543
Exports/GDP (USD)	340	1481	1935	5848	0,712	1,352	3,893	17,016	19	0,540
Reserves (NC)	298	952	1977	6379	0,629	0,991	3,729	15,747	13	0,529
Imports/GDP (USD)	207	454	2098	6677	0,518	0,709	2,198	14,977	93	0,459
M2/Reserves (NC)	387	1252	1723	4826	0,592	1,123	3,406	15,107	20	0,458
Current account acc.	164	462	1544	3555	0,658	1,198	1,754	16,449	11	0,452
Domestic credit (USD)	172	178	2047	6346	0,329	0,352	2,139	14,356	4	0,434
Exchange rate (USD) PC	352	1309	1959	6144	0,420	1,153	4,167	15,551	17	0,431
Deposit rate PC	265	510	1343	3044	0,408	0,871	2,983	15,266	15	0,427
Industrial production	104	240	892	3720	0,331	0,580	1,704	15,455	7	0,413
M2 (USD) acc.	90	159	2050	5967	0,583	0,617	0,926	12,874	3	0,384
Lending rate PC	187	417	1134	2917	0,389	0,884	2,135	14,814	13	0,372

Table 3: Indicators for currency crises based on monthly data for 22 countries from 1960 to 2008.

	A	B	C	D	PCD	NSR	PER	ALT	Percentile	Performance
Imports/GDP (USD)	866	642	3206	4722	0,676	0,563	5,405	17,140	84	0,866
Inflation (level)	749	446	3087	4898	0,410	0,427	5,023	18,598	87	0,810
GDP (USD)	763	305	3375	5225	0,472	0,299	5,015	14,857	11	0,758
M2 acc. (NC)	628	1017	2745	3840	0,763	1,125	3,704	18,171	20	0,733
Reserves (USD)	601	365	3520	5120	0,568	0,456	3,913	14,720	10	0,705
Reserves (SDR)	543	338	3642	5231	0,576	0,468	3,736	14,898	9	0,702
GDP acc.	709	1224	3421	4297	0,736	1,291	3,954	17,499	20	0,684
M2 (USD)	539	291	2910	4582	0,481	0,382	3,262	13,213	10	0,613
M2 (USD) acc.	265	232	3144	4625	0,640	0,614	1,746	15,282	6	0,609
Current account acc.	394	578	1895	2858	0,505	0,977	2,415	18,149	17	0,597
Exchange rate (SDR)	303	103	3972	5590	0,186	0,255	2,354	17,424	4	0,592
Inflation PC	413	397	3315	4851	0,504	0,683	2,918	14,808	91	0,589
Reserves (NC)	433	334	3688	5151	0,399	0,580	2,750	14,772	8	0,559
M2/Reserves (NC)	313	261	3089	4525	0,390	0,593	1,878	14,626	7	0,506
Industrial production	222	225	1385	3124	0,217	0,486	1,846	16,311	9	0,503
Current account	458	691	1841	2754	0,410	1,007	2,659	15,654	20	0,499
Deposit rate (level)	274	101	1971	3008	0,223	0,266	1,792	14,093	93	0,486
Exchange rate (SDR) PC	639	921	3541	4663	0,254	1,079	3,518	15,981	16	0,477
Deposit rate PC	306	364	1807	2685	0,282	0,824	1,743	16,596	13	0,465
M2/Reserves (USD)	240	168	3162	4618	0,338	0,498	1,482	13,664	95	0,458
Lending rate (level)	281	157	1359	3050	0,149	0,286	1,621	13,591	91	0,432
Lending rates/deposit rates	312	549	1109	2331	0,200	0,868	1,849	15,638	80	0,404
Exchange rate (USD)	73	29	4202	5664	0,116	0,298	0,630	14,524	1	0,392
Domestic credit (USD)	173	90	3429	5051	0,265	0,364	1,174	11,370	3	0,379
M2 (NC)	526	1140	2923	3733	0,215	1,534	2,948	16,847	20	0,368
Current account/GDP (USD)	166	237	2133	3208	0,342	0,953	0,947	14,413	7	0,364

Table 4: Indicators for stock market crises (CMAX definition) based on monthly data for 46 countries from 1960 to 2008.

	A	B	C	D	PCD	NSR	PER	ALT	Percentile	Performance
Exchange rate (SDR)	1325	1288	5487	17955	0,356	0,344	5,145	19,841	10	0,820
M2 (USD) acc.	1304	2763	4406	11878	0,812	0,826	4,814	16,937	20	0,780
Exchange rate (USD)	965	810	5847	18433	0,316	0,297	3,806	18,689	7	0,713
GDP acc.	952	2023	5238	14678	0,769	0,788	3,498	16,108	13	0,680
GDP (USD)	1286	1931	4908	14811	0,590	0,556	4,670	13,331	14	0,655
Exports/GDP (USD)	1081	3073	4942	12811	0,739	1,078	3,739	16,698	19	0,631
Lending rate (level)	1171	1881	3271	8931	0,347	0,660	4,510	16,563	80	0,620
Current account	1009	2002	3818	8228	0,508	0,936	3,803	17,595	20	0,606
Industrial production	986	1701	3442	8766	0,528	0,730	3,773	15,325	18	0,598
Current account/GDP (USD)	972	1923	3729	7858	0,492	0,951	3,689	17,651	20	0,592
Imports/GDP (USD)	701	1058	5287	14831	0,517	0,569	2,579	15,916	92	0,583
Inflation PC	741	1259	5438	14733	0,578	0,656	2,796	15,325	91	0,581
Lending rate PC	1015	1926	3278	8495	0,416	0,782	3,794	16,142	20	0,568
Real interest rate	937	2178	3475	8989	0,316	0,918	3,459	18,983	80	0,559
M2 acc. (NC)	679	1556	5032	13049	0,647	0,896	2,494	16,178	11	0,558
Current account acc.	620	1328	4193	8868	0,687	1,011	2,355	16,765	13	0,555
Reserves (USD)	663	997	5812	16364	0,432	0,561	2,343	13,546	7	0,476
Inflation (level)	800	1013	5466	15444	0,313	0,482	2,685	13,471	92	0,467
Deposit rate PC	521	930	3985	10737	0,337	0,689	2,034	15,208	9	0,440
Domestic credit (NC)	977	3336	5209	12041	0,398	1,374	3,350	16,899	20	0,421
Exchange rate (USD) PC	526	988	6057	17356	0,292	0,674	1,890	14,752	6	0,407
Reserves (SDR)	390	583	6299	17082	0,313	0,566	1,366	14,232	4	0,398
Lending rates/deposit rates	526	1279	3461	8618	0,331	0,980	1,904	16,178	87	0,390
Reserves (NC)	411	787	6064	16574	0,347	0,714	1,525	13,790	5	0,372
M2 (USD)	407	421	5365	14343	0,304	0,404	1,406	10,685	4	0,337
Domestic credit (USD)	421	653	5765	14718	0,310	0,624	1,456	12,094	5	0,329

Table 5: Indicators for stock market crises (Returns index definition) based on monthly data for 46 countries from 1960 to 2008.

	A	B	C	D	PCD	NSR	PER	ALT	Percentile	Performance
GDP acc.	1654	2692	6247	12298	0,831	0,858	4,731	17,573	19	0,786
Exchange rate (SDR)	1558	1321	7217	15959	0,368	0,431	5,194	18,747	11	0,771
M2 acc. (NC)	1475	2586	5826	10429	0,779	0,984	4,322	18,214	20	0,736
M2 (USD) acc.	1496	2571	5812	10472	0,770	0,963	4,348	17,211	20	0,714
Inflation (level)	1982	2556	5998	12187	0,473	0,698	5,781	16,227	80	0,711
M2/Reserves (USD)	1552	2439	5580	10374	0,592	0,875	4,969	17,280	80	0,701
Exchange rate (USD)	1231	804	7544	16476	0,344	0,332	4,328	16,496	8	0,689
Current account acc.	1193	1655	5085	7076	0,729	0,998	3,436	18,288	19	0,676
Exports/GDP (USD)	1449	2932	6166	11360	0,722	1,078	4,107	17,311	20	0,662
GDP (USD)	1614	2288	6298	12736	0,549	0,747	4,703	14,806	17	0,642
Deposit rate (level)	1538	1636	4752	8787	0,368	0,642	4,951	15,295	81	0,626
Exchange rate (SDR) PC	1351	2234	7338	14580	0,432	0,855	3,979	15,783	14	0,567
Domestic credit (USD)	1320	2123	6450	11664	0,480	0,906	3,442	15,046	16	0,529
M2/Reserves (NC)	1422	2569	5710	10244	0,516	1,006	3,418	15,182	20	0,521
Industrial production	1148	1827	4758	7162	0,539	1,046	3,101	15,362	20	0,510
Real interest rate	1167	1948	4714	7750	0,361	1,012	3,608	16,075	80	0,496
Domestic credit/GDP (USD)	1120	2831	5996	9806	0,356	1,423	2,923	19,412	20	0,451
M2 (NC)	1260	2851	6126	10305	0,371	1,270	3,601	16,465	20	0,450
Lending rates/deposit rates	1020	1618	4518	6728	0,402	1,053	2,520	15,591	81	0,439
Inflation PC	317	350	7541	13963	0,446	0,606	0,993	13,184	97	0,427
Imports/GDP (USD)	472	607	7117	13681	0,378	0,683	1,379	14,007	95	0,424
Domestic credit (NC)	1224	3089	6546	10704	0,383	1,422	3,214	17,325	20	0,423
Exchange rate (USD) PC	647	867	7760	15653	0,351	0,682	1,990	13,207	6	0,422
Reserves (SDR)	419	554	8054	15327	0,318	0,705	1,291	14,477	4	0,405
Reserves (NC)	481	717	7761	14877	0,330	0,788	1,473	14,637	5	0,403
Current account	477	580	5817	8183	0,373	0,873	1,332	14,936	7	0,399

3 Crisis indices – the aggregated indicators

The previous results show that some indicators outperform others. The signal approach only gives a measure to rank all the indicators, however, one can hardly justify the likelihood of a crisis based on the measure of a single indicator. Therefore, we will combine all the indicators to give a crisis index for each of the financial crisis, which can be utilised to build a early warning system for the prospective crises. We define the aggregated index for a specific crisis in country i at time t by

$$I^{(i)}(t) = \sum_{k \in \mathbb{K}} w_k s_k^{(i)}(t), \tag{8}$$

where w_k is the country and time independent weight for the $k-$th indicator, and $s_k^{(i)}(t)$ is the numerical value for the $k-$th indicator in country i at time t.

3.1 The aggregation of indicators

With the aggregated performance measures, we are able to calculate the weight of indicator k in the corresponding crisis index, which is given as

$$w_k = \frac{\text{Performance}_k}{\sum\limits_{\ell \in \mathbb{K}} \text{Performance}_\ell}. \tag{9}$$

It is obvious that $\sum_{k \in \mathbb{K}} w_k = 1$ is satisfied.

The series of numerical value is calculated in the following steps,

(L1) Generate the histogram of indicator k in country i with historical data. Calculate the optimal threshold and obtain the signal trigger level $L_k^{(i)}$.

(L2) Discretise the domain

$$\left[\inf_{t>0}\left\{P_k^{(i)}(t) : P_k^{(i)}(t) \geq L_k^{(i)}\right\}, \sup_{t>0}\left\{P_k^{(i)}(t) : P_k^{(i)}(t) \geq L_k^{(i)}\right\}\right], \tag{10}$$

into 10 uniform intervals with grid point

$$t_0 < t_1 < \cdots < t_{10}$$

where

$$t_0 = \inf_{t>0} \left\{ P_k^{(i)}(t) : P_k^{(i)}(t) \geq L_k^{(i)} \right\},$$
$$t_{10} = \sup_{t>0} \left\{ P_k^{(i)}(t) : P_k^{(i)}(t) \geq L_k^{(i)} \right\}.$$

(L3) For indicators with upper thresholds, assign

$$s_k^{(i)}(t) = \ell + 1 \quad \text{if} \quad P_k^{(i)}(t) \in [t_\ell, t_{\ell+1}), \text{ for } \ell = 0, 2, \cdots, 9. \tag{11}$$

For indicators with lower thresholds, assign

$$s_k^{(i)}(t) = 10 - \ell \quad \text{if} \quad P_k^{(i)}(t) \in [t_\ell, t_{\ell+1}), \text{ for } \ell = 0, 2, \cdots, 9. \tag{12}$$

With steps in (S1) to (S2) and (L1) to (L3), we are able to evaluate the numerical value for each indicator in a specific country over time. In essence, the numerical value can be treated as a refinement of the procedure used to define a signal: rather than assigning 1 for all the signal issuance, we distinct the strength or weakness for all the signals by scoring them between 0 and 10.
With the formula

$$I^{(i)}(t) = \sum_{k \in \mathbb{K}} w_k s_k^{(i)}(t), \tag{13}$$

one can come up with the crisis indices for banking, currency and stock crises.
We plot the crisis indices for United Kingdom as follows

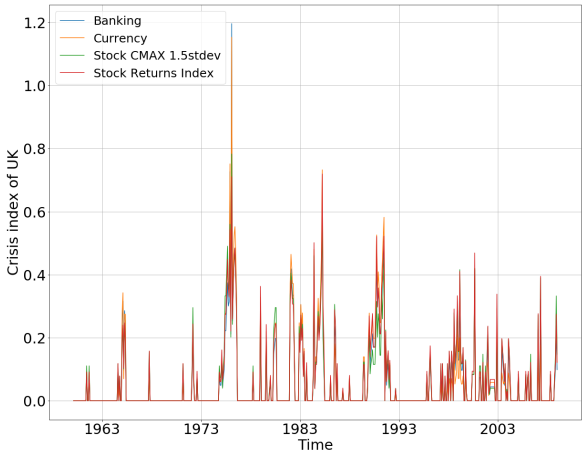


Figure 4: The monthly crisis indices for banking, currency and stock crises of United Kingdom from 1960 to 2008.

Meanwhile, we also list the major economic recessions in United Kingdom from 1960 to 2008 as follows:

Years	crisis
1973 - 1975	Mid-1970s recessions
1980 - 1981	Early 1980s recession
1990 - 1991	Early 1990s recession
2008	Great Recession

We can observe from Figure 4 that the crisis indices of United Kingdom issue useful signals for the forthcoming crisis.

3.2 Transformation of crisis indices

Apart from the United Kingdom, we do the same experiment for the United States and South Africa. The crisis indices are given in Figure 5 below.

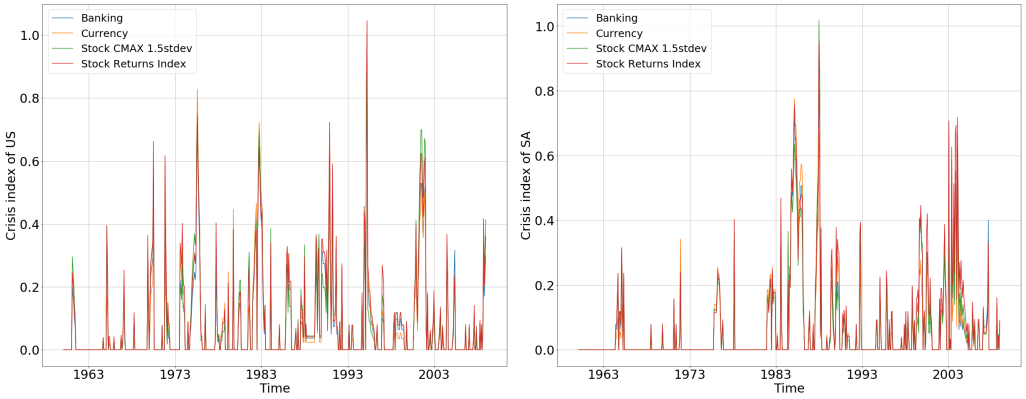


Figure 5: The monthly crisis indices for banking, currency and stock crises of the United States (left panel) and South Africa (right panel) from 1960 to 2008.

It is can be observed that the paths of our crisis indices are rough, which can hardly be used for developing trading strategies or run further simulations. Therefore, we need to filter the rough process to make the it smoother. We denote our banking, currency and stock crisis index of country i at time t by $I_b^{(i)}(t)$, $I_c^{(i)}(t)$ and $I_s^{(i)}(t)$, i.e. we have $I_x^{(i)}(t)$ where $x = b$, c or s. We proceed the filtering $I_x^{(i)}(t)$ with the following steps:

(R1) We take the moving average of $I_x^{(i)}(t)$ with rolling window $k_1^{(x)}$:

$$\tilde{I}_x^{(i)}(t) = \frac{1}{k_1^{(x)}+1} \sum_{j=0}^{k_1^{(x)}} I_x^{(i)}(t-j);$$

(R2) We then take the difference of $\tilde{I}_x^{(i)}(t)$:

$$\Delta\tilde{I}_x(t) = \tilde{I}_x^{(i)}(t) - \tilde{I}_x^{(i)}(t-1);$$

(R3) Next, we take the moving average of $I_x^{(i)}(t)$ with rolling window $k_2^{(x)}$:

$$\tilde{\Delta}_x^{(i)}(t) = \frac{1}{k_2^{(x)}+1} \sum_{j=0}^{k_2^{(x)}} I_x^{(i)}(t-j);$$

(R4) Finally, we take a weighted sum of $\tilde{\Delta}_x^{(i)}(t)$ to form the transformed crisis in-

dices $B(t)$, $C(t)$ and $S(t)$ for banking, currency and stock crisis, respectively:

$$B(t) = \frac{1}{10}\sum_{j=9}^{18}\tilde{\Delta}_b(t-j),$$
$$C(t) = \frac{1}{10}\sum_{j=9}^{18}\tilde{\Delta}_c(t-j),$$
$$S(t) = \sum_{j=0}^{24}\tilde{\Delta}_x(t-j)f(j),$$

where f is a non-negative function peaked in the middle with

$$\sum_{j=0}^{24} f(j) = 1.$$

For example, f can be the following function

$$f(j) = \begin{cases} j, & \text{for } 0 \le j \le 12, \\ 25-j, & \text{for } 13 \le j \le 25. \end{cases}$$

We take the United Kingdom as an example and observe the transformation step by step from Figure 6 to 9. It is easy to observe that the path is smoother compared with the original crisis indices.

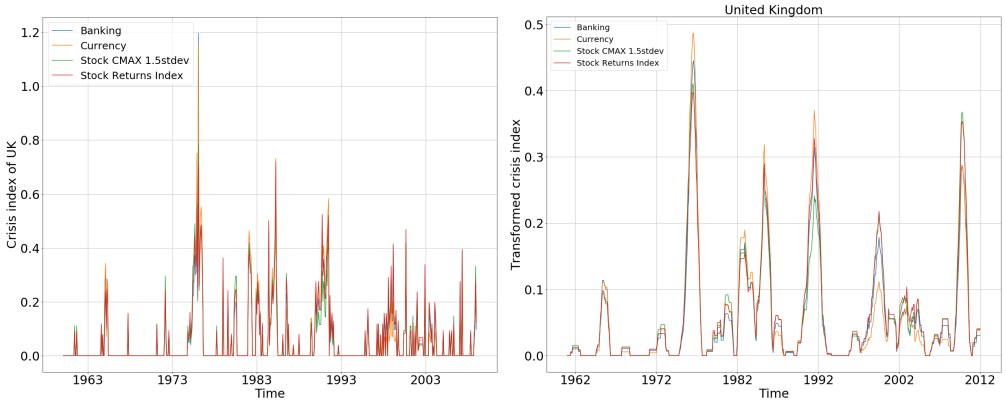


Figure 6: The transformed $\tilde{I}_x(t)$, which is taking the moving average with rolling window for $k_1 = 12$.

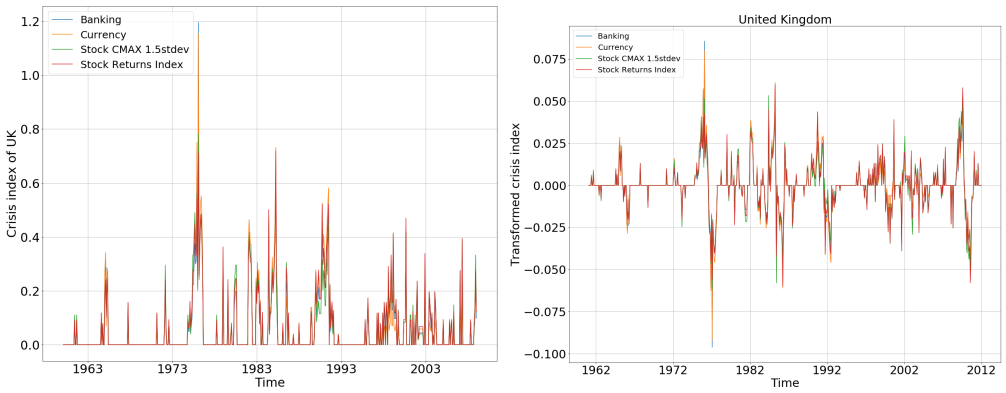


Figure 7: The transformed $\Delta \tilde{I}_x(t)$, which is taking the difference.

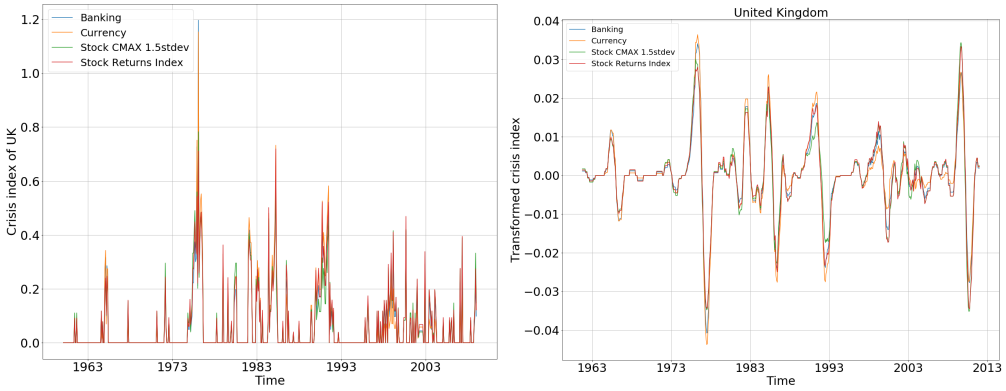


Figure 8: The transformed $\tilde{\Delta}_x(t)$, which is taking the moving average with rolling window for $k_2 = 9$.

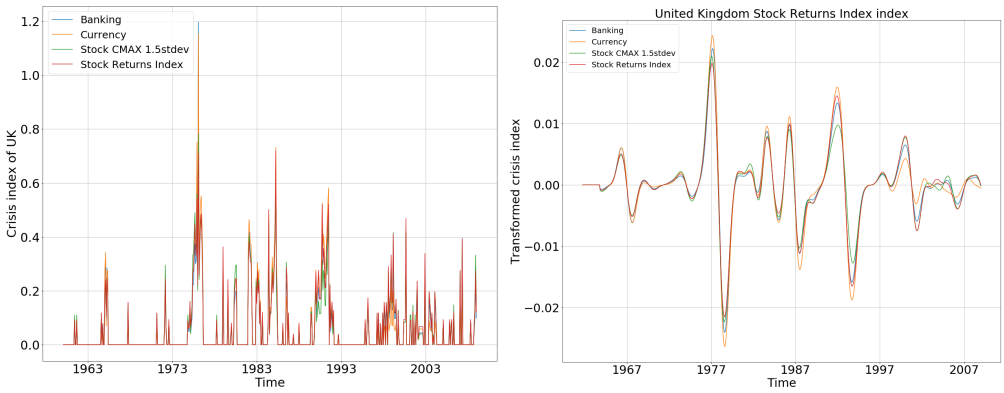


Figure 9: The transformed $B(t)$, $C(t)$ and $S(t)$.

4 Asset management based on early warning signals

The primary aim of integrating the signals approach and the use of crisis indices into an investment strategy is maximising portfolio gains (minimising portfolio losses). Here we develop a trading model that determines the monthly optimal portfolio allocations needed to achieve this aim, based on the values of the crisis indices.

Vermersch (2013) cautions against a major shortfall during the development of such a model, namely, a forward-looking model that takes future data into consideration as part of its calibration. One could, for instance, erroneously use all of the data up to and including the present day to backtest the model, as if this data was accessible at any time during the simulation. Mathematically, this is tantamount to conditioning on a filtration $\mathcal{F}_t$ to predict future values, from some present time $s < t$. In order to avoid this, we use two sets of historical data from disjoint periods to calibrate and then test our trading model.
The theoretical development of the model is outlined in section 4.1, and the numerical results obtained from implementing and testing our model are given in section 4.3.

4.1 Linear trading model

We develop a model based on a portfolio comprising three asset classes: cash, bonds, and equities. We denote the proportion of the portfolio allocated to each asset class at time t as

$$\begin{aligned} c(t): &\quad \text{cash, e.g. 3M treaury bills,} \\ b(t): &\quad \text{bonds, e.g. 10Y government bonds,} \\ e(t): &\quad \text{equities, e.g. FTSE index.} \end{aligned}$$

These proportions are subject to the following constraints:

$$\begin{aligned} c(t) + b(t) + e(t) &= 1, \\ 0 \leq c(t) &\leq c_{max}, \\ 0 \leq e(t) &\leq e_{max}, \end{aligned}$$

where we have set

$$c_{max} := 75\%,$$
$$e_{max} := 50\%.$$

The linear trading model is proposed in Grasselli (2013) and Vermersch (2013). It is specified by the change in the proportions held in each of the three asset classes, as

$$\Delta c(t) = F_1 \cdot B(t) + F_2 \cdot C(t) + F_3 \cdot S(t), \tag{14}$$
$$\Delta e(t) = G_1 \cdot B(t) + G_2 \cdot C(t) + G_3 \cdot S(t), \tag{15}$$
$$\Delta b(t) = -\Delta c(t) - \Delta e(t), \tag{16}$$

where the last equation ensures that the portfolio is self-financing. $B(t)$, $C(t)$, and $S(t)$ are, respectively, the values at time t for the banking crisis, currency crisis, and stock market crisis indices. For simplicity, we consider the case with

$$F_2 < F_3 = 0 < F_1, \tag{17}$$
$$G := G_3 < G_1 = G_2 = 0. \tag{18}$$

4.2 Trading models with risk preferences

We extend the linear trading model to include investor risk preference with respect to each asset class. This is incorporated by defining functions $f_i(x)$ and $g_i(x)$ for $i = \{1, 2, 3\}$ and setting

$$\Delta c(t) = F_1 \cdot f_1(B(t)) + F_2 \cdot f_2(C(t)) + F_3 \cdot f_3(S(t)), \tag{19}$$
$$\Delta e(t) = G_1 \cdot g_1(B(t)) + G_2 \cdot g_2(C(t)) + G_3 \cdot g_3(S(t)), \tag{20}$$
$$\Delta b(t) = -\Delta c(t) - \Delta e(t), \tag{21}$$

and retaining the simplifying conditions given in 17. The convexities and concavities of the functions $f_i(x)$ and $g_i(x)$ determine the risk preferences of the investor.

Our first variation of such a trading model is that which exhibits linear risk preference for banking-related and currency-related risks, and (relative) risk aversion for

stock-related risks. This is given by

$$\Delta c(t) = F_1 \cdot B(t) + F_2 \cdot C(t), \tag{22}$$

$$\Delta e(t) = G \cdot (e^{\alpha S(t)} - 1), \tag{23}$$

$$\Delta b(t) = -\Delta c(t) - \Delta e(t), \tag{24}$$

where α is determined by the level of risk aversion of the investor. Similarly, we define a trading model that exhibits (relative) risk-seeking behaviour for stock-related risks, given by

$$\Delta c(t) = F_1 \cdot B(t) + F_2 \cdot C(t), \tag{25}$$

$$\Delta e(t) = G \cdot log(S(t) + 1), \tag{26}$$

$$\Delta b(t) = -\Delta c(t) - \Delta e(t). \tag{27}$$

Note that the second model may be defined using a power function $g_3(x) = x^\alpha$, with α denoting the level of risk aversion, if desired.

4.3 Numerical results

We proceed by calibrating and testing our models for the United Kingdom. As representative of each asset class, we use 3-month treasury bills for cash, 10-year government bonds for bonds, and the FTSE 100 Index for equities. For all model calibration, we use January 1982 as the starting month, as this is the first month from which we have all of the required market data. All data used are monthly data. Without loss of generality, for the stock market crisis index, we use the definition by return.

For the purspose of comparison, we construct a benchmark portfolio that follows a naive diversification strategy, that is

$$c(t) = e(t) = b(t) = \frac{1}{3}, \quad \forall t. \tag{28}$$

Maintaining the above allocations requires the benchmark portfolio to be rebalanced monthly. As such, it is comparable to our trading models. Furthermore, the

initial allocation for each of our trading models is

$$c(0) = e(0) = b(0) = \frac{1}{3}. \tag{29}$$

The model parameters F_1, F_2, and G, as well as the parameters $k_1^{(x)}, k_2^{(x)} (x \in \{b, c, s\})$ in Section 3, are obtained by maximising the value at the end of the calibration period of a portfolio with initial allocation given by 29. The calibration period runs from January 1982 to December 1999.

4.4 Backtesting results

From the calibration exercise, we obtain the following optimal parameter values:

Table 6: Calibrated parameter values for three trading models with calibration period Jan. 1982 - Dec. 1999

	Banking crisis			Currency crisis			Stock market crisis		
	k_1	k_2	F_1	k_1	k_2	F_2	k_1	k_2	G
Linear	5	5	7.154	24	11	-2656.510	8	8	-3.334
Risk-averse	24	6	1301.847	19	8	-2298.340	6	1	-70.816
Risk-seeking	25	10	171.065	20	7	-871.323	10	10	-82.290

The backtest of the trading models using the calibrated parameters is shown below. The graph indicates that over the calibration period, all three trading models outperform the benchmark in terms of portfolio value. In addition, the portfolio that is traded using the relatively risk-seeking trading model provides the greatest overall return, followed by that using the linear trading model. The portfolio traded using the relatively risk-averse trading model provided the second-lowest overall return, with the benchmark portfolio generating the lowest overall return. These results are in line with our expectations, as they are taken from the backtest: since the model parameters having been optimised for each trading model, our expectation is that the risk-seeking trading model produces the highest overall return in the backtest, and so on.

4.5 Out-of-sample results

We use the period of January 2000 to December 2010. As we are investigating the use of the signals approach and crisis indices in long-term asset management,

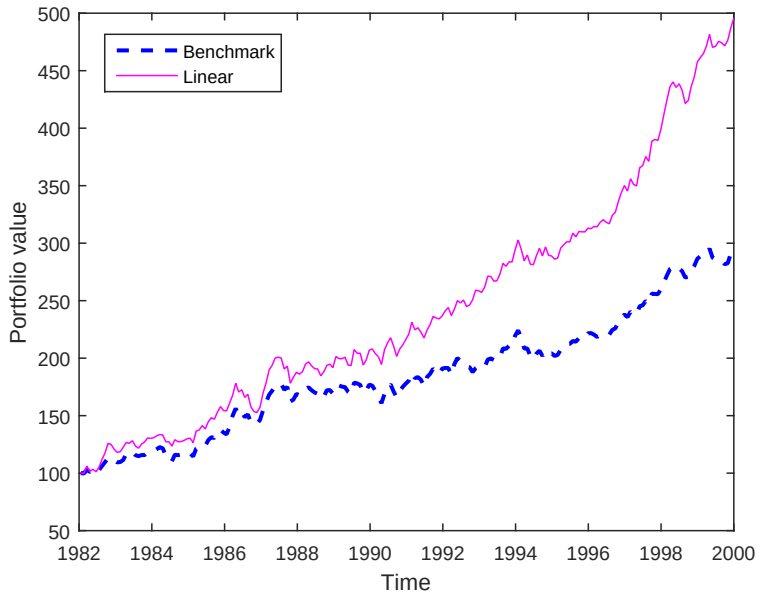


Figure 10: Backtesting performance for our linear trading model.

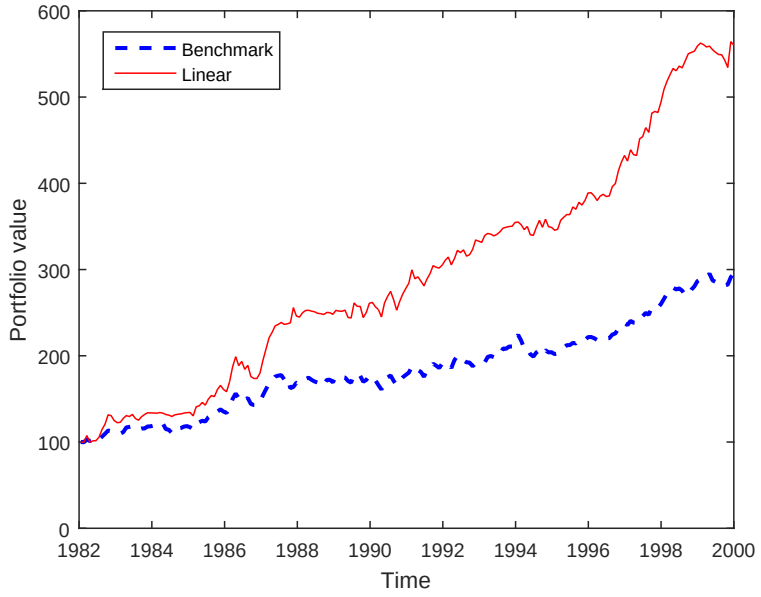


Figure 11: Backtesting performance for our risk-seeking model.

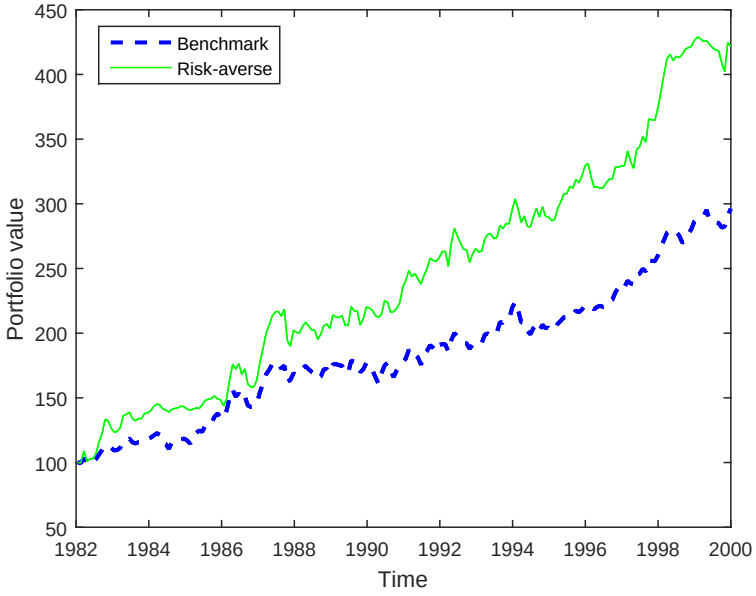


Figure 12: Backtesting performance for our risk-averse trading model.

the length of this period is of appropriate length for evaluating the efficacy of our trading models. Using the parameters derived from the calibration, the out-of-sample test yields the following results:
We find that the relatively risk-averse trading model and relatively risk-seeking trading model perform best (in that order), with the linear trading model performing similarly to the benchmark.

Up until shortly before the beginning of 2008, when most of the portfolios achieve their highest values (as seen by the vertical reference line), the relatively risk-seeking trading model produces the highest portfolio value. Moreover, it significantly outperforms the other trading models, as well as the benchmark. In the context of a bull market, this would be expected. Furthermore, the other portfolios based on our trading models perform similarly to the benchmark.

Between 2008 and 2010, all of the portfolios undergo reductions in value. As expected, the portfolio based on the relatively risk-seeking trading model experiences the largest losses, losing most of the value it previously held in excess of the benchmark portfolio. In contrast, although the portfolio based on the relatively risk-

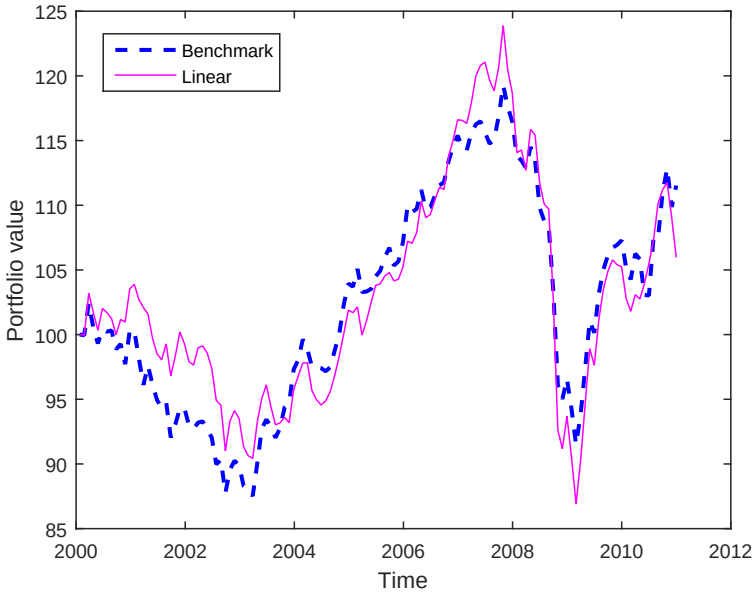


Figure 13: Out-of-sample performance for our linear trading model.

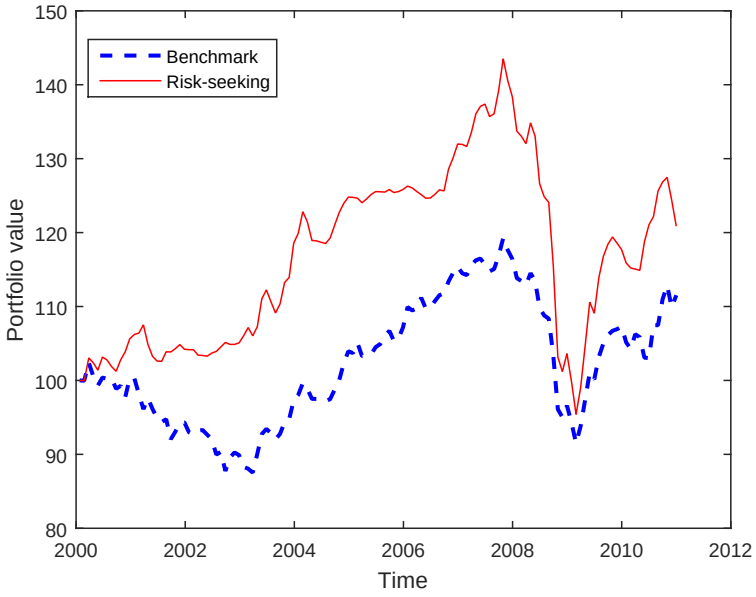


Figure 14: Out-of-sample performance for our risk-seeking trading model.

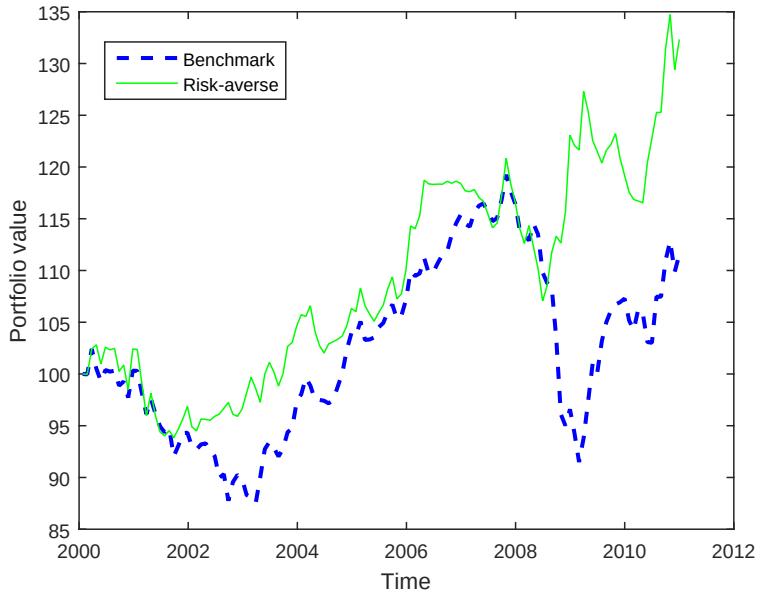


Figure 15: Out-of-sample performance for our risk-averse trading model.

averse trading model loses some value in the first part of 2008, it actually increases in value while all of the other portfolios drop to their respective troughs. This shows that the relatively risk-averse trading model performs significantly better relative to the benchmark and other trading models during bear markets; again, this is in line with our expectation. Finally, in terms of the overall result over the trading period from January 2000 to December 2010, it is the relatively risk-averse trading model that performs best, followed by the relatively risk-seeking trading model.

5 Conclusion

With the historical indicator and crisis data up to 2010, we are able to rank the leading economic indicators for banking, currency and stock crises with four performance measures. We list the top five indicators for three types of crises as follows:

Table 7: Top five indicators.

	Banking crises	Currency crises	Stock crises (CMAX)	Stock crises (Return)
1	Reserve (USD)	Import/GDP (USD)	Exchange rate (SDR)	GDP acceleration (USD)
2	M2 acceleration (NC)	Inflation	M2 acceleration (USD)	Exchange rate (SDR)
3	Reserve (SDR)	GDP (USD)	Exchange rate (USD)	M2 acceleration (NC)
4	Inflation	M2 acceleration (NC)	GDP acceleration (USD)	M2 acceleration (USD)
5	GDP (USD)	Reserve (USD)	GDP (USD)	M2

The leading indicators make sense for the corresponding crises in economics as well. For example, rising inflation and decreasing reserves can lead to prospective currency crises. Our model could provide more empirical results for the study of the relationship between the economic direction and macro-indicators.

With the four performance metrics, a early warning system by the crisis indices with economic indicators is established. Regulators can adopt such a early warning system to monitor the economic environment and establish suitable protections for potential financial crises. On the other hand, speculators or asset managers can adjust their portfolio positions based on this early warning system, which can avoid the severe loss due to the occurrence of a crisis.

We also develop our asset management strategies based on our crisis indices. With our risk-seeking trading algorithm, one can gain more profits compared with the risk-averse algorithm and linear algorithm. However, the risk-seeking algorithm will suffer the largest maximum drawdown when the economic scenario turns bad. Compared with risk-seeking strategy, our risk-aversion strategy is able to preserve our loss when the economic scenario turns bad, while it will lead to lower profitability in good economic scenarios, though it can still outperform our benchmark trading strategy. With our models, the investor can determine the choice of portfolio adjustments based on their risk profile, which offers more flexibility.

Bibliography

Demirgüç-Kunt, A., Detragiache, E., 1998. The determinants of banking crises in developing and developed countries. Staff Papers 45 (1), 81–109.

Drehmann, M., Juselius, M., 2014. Evaluating early warning indicators of banking crises: Satisfying policy requirements. International Journal of Forecasting 30 (3), 759–780.

Grasselli, M., 2013. Understanding financial crises. McMaster University working paper.

Kaminsky, G., Lizondo, S., Reinhart, C. M., 1998. Leading indicators of currency crises. Staff Papers 45 (1), 1–48.

Reinhart, C. M., Rogoff, K. S., 2009. This time is different: Eight centuries of financial folly. Princeton University Press, Princeton.

Reinhart, C. M., Rogoff, K. S., 2013. Banking crises: an equal opportunity menace. Journal of Banking & Finance 37 (11), 4557–4573.

Vermersch, M., 2013. Part ii: Long-term investments strategy with early warning signals. Field Institute working paper.

Model Calibration with Neural Networks

Team 3

ALEXIA MARR, University of Cape Town
MELUSI MAVUSO, University of Cape Town
NICOLAS MITOULIS , University of Cape Town
ADITYA SINGH, University College London

Supervisors:
JOSEF TEICHMANN, ETH Zurich
CHRISTA CUCHIERO, University of Vienna

African Institute of Financial Markets and Risk Management

Contents

1 Introduction

One of the major considerations practitioners make in choosing a pricing model is its practicality. It would be of no use to have a model that prices market instruments correctly if calculating prices with such a model takes (say) a week. These considerations have often lead practitioners to choosing simpler models instead of more complex models that could better describe the observed market prices. The first goal of the project is to solve one such practical consideration: calibration speed.

Choosing a simpler, parsimonious model also has its own disadvantages; one such disadvantage is parameter instability. Having calibrated a model at a particular time point, it is desirable that the model parameter values do not change drastically at a later date, since otherwise, the model assumptions (that the parameters are constant) would be violated (see work by Schlögl (2015) on classifying different forms of model risk). The second aim of this project is to propose a model for interest rates that is not susceptible to parameter instability while being quick to calibrate using the solution to the first problem.

Intuitively, model calibration means choosing the model parameters that 'best' reproduce the market prices of certain instruments. These instruments — often called the *calibration instruments* — are usually chosen to be highly liquid assets whose prices are easily accessible. A criteria to penalize deviations from the prices is chosen, called a *cost function*, and the parameters are chosen to minimize this cost function. This optimization procedure is usually slow, and has at times resulted in practitioners discarding some potentially good models.

To be more precise, let $Q^{mkt} = \{Q_1^{mkt}, \ldots, Q_N^{mkt}\}$ be a set of market prices for the instruments to be used for calibration. If $\mathbb{M} = \mathbb{M}(\theta)$ is a pricing model that depends on the parameters $\theta = (\theta_1, \ldots, \theta_n)$, we denote the corresponding model prices of the securities by $(Q_1^\theta, \ldots, Q_N^\theta)$. We can then view calibration as finding $\theta \in \mathbb{R}^n$ that minimizes the *cost function*

$$\theta^* \mapsto \mathrm{Cost}(Q^{mkt}, Q^{\theta^*}).$$

From an abstract viewpoint, calibration defines a mapping $\Theta : \mathbb{R}^N \longrightarrow \mathbb{R}^n$ that gives the parameters θ for each vector of market prices Q^{mkt}. We will call the mapping Θ the *calibration function*. The idea behind this project is to estimate the calibration function "offline" (i.e., before the model is used for live trading), so that during live trading, the process of calibration reduces to simple, quick function evaluation of the market prices to give the relevant parameters. Taking ideas from Hernandez (2016) we will approximate Θ using Neural networks.

2 Calibration

2.1 General calibration methodology

The calibration of a model involves determining the parameters for that model that 'best' fit the market data. Calibration can be thought of as an *inverse problem* associated with the pricing of a financial derivative, where in the theoretical situation, all prices are known and the calibration problem is solved by fitting parameters to a model so that is reproduces the known market data. Calibration thus allows for the pricing of less liquid instruments or more complex instruments.

Let $\mathbb{M} = \mathbb{M}(\theta)$ be a market model that depends on the parameters θ. Assume that the model $\mathbb{M}$ has n parameters that need to be calibrated using a set of N market instruments whose (true) market prices are given by $Q_1^{mkt}, \ldots, Q_N^{mkt}$. We will denote the corresponding model prices by $Q_1^{\theta}, \ldots, Q_N^{\theta}$. The choice of instruments to use for calibration is not trivial and arguments can often be made for the inclusion or exclusion of a particular quote (instrument) based on, for instance, liquidity concerns or avoiding over-fitting to a particular maturity region. The general principals ensure instruments are chosen to include the important maturity dates of the underlying cash flows as well as choosing actively traded instruments. At the money options are most often used as calibration data.

Previously it was stated that the model parameters are chosen to best fit the market data. The concept of 'best' is evaluated by a cost function $\mathrm{Cost}(\cdot, \cdot)$. This cost function quantifies the distance between quotes obtained from the model and market quotes. This distance can be likened to an error term. This is intuitive as we want our model to be able to price accurately. The goal of calibration, therefore, involves

finding those model parameters i.e. solving for θ such that the distance between the model and market quotes is minimised. This is done by determining a cost function $\mathrm{Cost}(\cdot,\cdot)$ and using an optimisation scheme to determine the parameters θ that minimise this cost function.

There is a choice of which cost function to implement. This is often influenced by the practical concern of reducing calibration time. The cost function $\mathrm{Cost}(\cdot,\cdot)$ usually takes the form of some weighted average of each difference between the market quote and the model quote. If calibration is done to N market quotes possible cost functions include Larsson (2015)

$$\mathrm{Cost}\left(Q^{\theta}, Q^{mrkt}\right) = \sum_{i=1}^{N} w_i \left| Q_i^{\theta} - Q_i^{mrkt} \right|^p, \tag{1}$$

$$\mathrm{Cost}\left(Q^{\theta}, Q^{mrkt}\right) = \sum_{i=1}^{N} w_i \left| \frac{Q_i^{\theta} - Q_i^{mrkt}}{Q_i^{mrkt}} \right|^p, \tag{2}$$

$$\mathrm{Cost}\left(Q^{\theta}, Q^{mrkt}\right) = \sum_{i=1}^{N} w_i \left| lnQ_i^{\theta} - lnQ_i^{mrkt} \right|^p, \tag{3}$$

where $i = 1.....N$, Q_i^{θ} is the quote obtained from the model with parameters θ and Q_i^{mrkt} is the quote obtained in the market. p must be greater than one and weights w_i are positive.

We will not argue here for the use of one cost function over another. We will use one of the common formulations being the weighted sum of squares formulation to illustrate a cost function.

$$\mathrm{Cost}\left(Q^{\theta}, Q^{mkt}\right) = \sum_{i=1}^{N} w_i \left(Q_i^{\theta} - Q_i^{mkt} \right)^2. \tag{4}$$

Solving for the parameter set θ to minimise the cost function using an optimisation technique yields

$$\theta = \underset{\theta^* \in S \subseteq \mathbb{R}^n}{\operatorname{argmin}} \ \mathrm{Cost}\left(\theta^*, \{\mathbf{Q}^{mrkt}\}\right). \tag{5}$$

Optimisation schemes that are commonly used to solve for the parameter set θ in minimising the cost function Cost are Newton's method, Bisection method and the method of Gradient Descent.

The calibration problem can now be seen as a problem with N arguments and n outputs where $\mathbf{\Theta}$ is the calibration function that maps the much larger data set N onto a smaller output vector of size n, the calibrated parameters, i.e.

$$\mathbf{\Theta} : \mathbb{R}^N \mapsto \mathbb{R}^n. \tag{6}$$

Calibration is currently done on-line alongside active trading, as N can often be quite large using optimisation techniques on-line poses challenges to the speed and accuracy of the calibration of models. This presents a good use case for a neural network where the calibration problem can be reduced to finding a neural network to approximate $\mathbf{\Theta}$, this is introduced in the next section.

2.2 Some examples of calibration

2.2.1 The Black-Scholes model

Consider the Black-Scholes model under the pricing measure. Assume that Q^{mrkt} is the price of a European call option issued on an asset S_t with time to maturity T, strike K and has a price process

$$dS_t = rS_t dt + \sigma S_t dW_t. \tag{7}$$

Here W is a *Wiener process*, r is the *interest rate* and σ is the *volatility* of S_t.
Although options are priced in terms of volatility we assume the implied volatility is the non-observable parameter in the model that needs to be calibrated i.e. $\theta = \sigma$
The model price Q^θ of the option at time t and todays price S_t is

$$Q^\theta = S_t \Phi\left(d_1\right) - Ke^{-r(T-t)}\Phi\left(d_2\right), \tag{8}$$

where

$$d_1 = \frac{\ln(S_t/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \tag{9}$$

$$d_2 = d_1 - \sigma\sqrt{.T-t} \tag{10}$$

In this simple example the parameter θ is calibrated to one option quote Q^{mrkt}. Therefore, to calibrate the Black Scholes model for one option quote we want to find the parameter θ such that the cost function $\text{Cost} = (Q_i^\theta - Q_i^{mrkt})^2$ is minimised. In

this example $N = 1$ as one option quote is used. The minimum of the cost function is zero, therefore, $Q^{\theta} = Q^{mrkt}$ or more clearly

$$S_t \Phi\left(d_1\right) - K e^{-r(T-t)} \Phi\left(d_2\right) = Q^{mrkt} \tag{11}$$

where d_1 and d_2 are defined in Eq.12 and 13. In this example, calibrating to one option quote allows Eq. 14 to be solved directly for θ which is the implied volatility σ. If calibration is done using a set of N market quotes instead, an optimisation scheme would be necessary to minimise the cost function and solve for the unknown parameter θ

$$\text{Cost} = \sum_{i=1}^{N} w_i \left(Q^{\theta}(\tau_i) - Q^{mrkt}(\tau_i)\right)^2 \tag{12}$$

$$\theta = \underset{\theta \in S \subseteq \mathbb{R}^n}{\operatorname{argmin}} \, \text{Cost}\left(\mathbf{Q}^{\theta}, \mathbf{Q}^{mrkt}\right) \tag{13}$$

$$\mathbf{\Theta} : \mathbb{R}^N \mapsto \mathbb{R}^1 \tag{14}$$

where $i = 1.....N$ and $\mathbf{Q}^{\theta}$ and $\mathbf{Q}^{mrkt}$ are vectors of model and market quotes. $\mathbf{\Theta}$ is the calibration function that maps the set of N quotes to output the unknown parameter. This is a single parameter calibration example where Eq.16 consists of one unknown parameter θ and can be solved using an optimisation scheme..

2.2.2 The Heston model

The Heston model is a stochastic volatility model. It assumes volatility is not constant as under the Black-Scholes model but allows volatility to follow a random process.

The Heston model assumes the underlying asset follows a stochastic process with stochastic variance υ_t.

The underlying asset and volatility progression is characterised by the following risk-neutral dynamics Crisostomo (2015)

$$dS_t = rS_t dt + \sqrt{\upsilon_t} S_t dW_t^1 \tag{15}$$

$$d\upsilon_t = \alpha(\bar{\upsilon} - \upsilon_t)dt + \eta\sqrt{\upsilon_t} dW_t^2 \tag{16}$$

$$dW_t^1 dW_t^2 = pdt \tag{17}$$

where

- S_t is the price of the underlying asset;
- r is the risk free rate;
- υ_t is the variance at time t;
- $\bar{\upsilon}$ is the long-term variance;
- α is the rate of variance mean-reversion;
- η is the volatility of the variance process;
- W_t^1,W_t^2 are the correlated Wiener processes with correlation coefficient ρ.

One of the main advantages of the Heston model in addition to stochastic volatility is that the price of European options can be estimated using a quasi-closed form valuation formula involving characteristic functions Crisostomo (2015).
Availability of closed-form solutions is particularly useful in calibration where many plain vanilla option repricings need to be done in order to find the optimal parameters. The Heston model has five unknown parameters $\theta = \{\upsilon_0, \bar{\upsilon}, \alpha, \eta, \rho\}$ that need to be calibrated. As stated in section 1, the goal of calibration is to find the set of parameters θ through an optimisation scheme that minimises the cost function $\mathrm{Cost}(\cdot, \cdot)$. This is now more complicated than calibrating the Black-Scholes model as there are far more parameters to calibrate. The cost function and optimisation are the same as those depicted in Equation (15) and (16), however, the calibration function is

$$\boldsymbol{\Theta} : \mathbb{R}^N \mapsto \mathbb{R}^5 \tag{18}$$

as there are now 5 unknown parameters to calibrate.
By calibrating the parameters of the Heston model an evolution for the underlying asset that is consistent with the market prices of plain vanilla options is obtained.

2.2.3 The Hull-White model

The Hull-White model is a model of interest rates. It is a no-arbitrage model that fits today's term structure of interest rates. It is able to translate the evolution of

interest rates onto a tree or lattice so interest rate derivatives such as swaptions can be valued by the model Gurrieri et al. (2013).
The model is a short rate model with dynamics

$$dr_t = (\theta(t) - \alpha r_t)dt + \sigma dW_t \tag{19}$$

where α is the mean reversion parameter and σ is the volatility. Both are constant parameters and are shared across all market quote inputs. $\theta(t)$ is chosen to replicate the current market yield curve $y(t)$. The yield curve is an exogenous factor which is the ϕ of the model while τ of the model will include properties of the options such as day-count convention and maturity.
The Hull-White model can be used to price any vanilla interest rate derivative, swaptions will be used here for illustration.
Swaption prices are quoted as implied volatilities in the market. The market price of the swaption is obtained from inputing the implied volatility into Black's Formula. This is needed in order to compare the market quote with the model quote for calibration. The Hull-White model has two parameters that need to be calibrated, $\theta = \{\alpha, \sigma\}$. In order to calibrate the Hull-White model the same principles are followed as outlined above: market quotes are obtained from the market such as swpation prices with varying maturities and tenors, the cost function is determined to quantify the distance between the market quotes and the model quotes and an optimisation scheme is chosen to minimise the cost function by solving for θ. The problem is then

$$(\alpha, \sigma) = \mathbf{\Theta}\left(\left\{Q^{mrkt}\right\}; \{\tau\}, y(t)\right). \tag{20}$$

The calibration function can be viewed as a function that maps the input market quotes of size N and outputs calibrated parameters α and σ Gurrieri et al. (2013)

$$\mathbf{\Theta} : \mathbb{R}^N \mapsto \mathbb{R}^2 \tag{21}$$

2.3 Challenges of on-line calibration

Calibration at trading desks is currently done on-line. This means calibration is done in real time while active trading is taking place, therefore, the speed at which a trader can calibrate their model and output a price is an essential considera-

tion. The speed at which parameters can be calibrated largely informs the choice of model used by practitioners. This could result in forgoing the more accurate model for a simpler model whose parameters can be calibrated faster.
Calibrating on-line decreases the time allowed for debugging certain pressure points within the calibration function, thereby, making on-line calibration more risky. Most of the pressure points involve the chosen optimization scheme and ensuring a global minimum is found. These optimization challenges can be refined using off-line calibration without affecting the speed or accuracy with which live trades can be executed.
The aim of this paper is to eliminate the calibration speed consideration from informing the choice of model by taking the calibration of parameters off-line using neural networks. This is introduced in the next section.

3 Neural networks

3.1 Basics

Neural networks aim to imitate the highly connected structure of neurons in the brain. The neural networks used in computing are more aptly called artifical neural networks (ANNs).
Neurons receive signals or information from other neurons through input branches called dendrites. The neurons choose whether or not to pass on the signal or alter it.

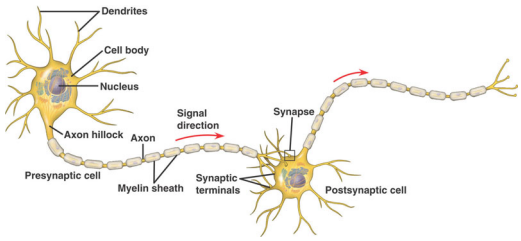


Figure 1: Structure of a neuron. Available at: http://biomedicalengineering.yolasite.com/neurons.php

The underlying motivation for neural networks is the Universal Approximation Theorem which states that any function, f, may be approximated by a weighted combination of some non-linear function, ψ, of the scalar product of f's inputs with some weights.

Theorem 3.1 (Universal Approximation Theorem)**.** Let $\psi : \mathbb{R} \to \mathbb{R}$ be a fixed activation function[1] and N be a positive integer. For every continuous function $f : [0,1]^N \longrightarrow \mathbb{R}$ and every $\epsilon > 0$, there exist a positive integer m and vectors $\{w_i : i = 1, \ldots, m\}$, β and b such that

$$\sup_{x \in [0,1]^N} \left| f(x) - f^{\mathcal{NN}}(x) \right| < \epsilon,$$

where

$$f^{\mathcal{NN}}(x) = \sum_{i=1}^{m} \beta_i \psi(w_i^T x + b_i), \quad x \in [0,1]^N.$$

3.1.1 Gradient descent algorithm

The gradient descent algorithm is a first order linear optimisation algorithm for finding the minimum of a function. It is a simple and robust algorithm which causes the parameter which is being optimised to move in the opposite direction to the sign of the gradient. It is implemented by repeating the following until convergence:

$$\omega_{k+1} = \omega_k - \eta J'(\omega_k)$$

. Where $J(\omega)$ is the function to be minimised. See figure for further intuition.

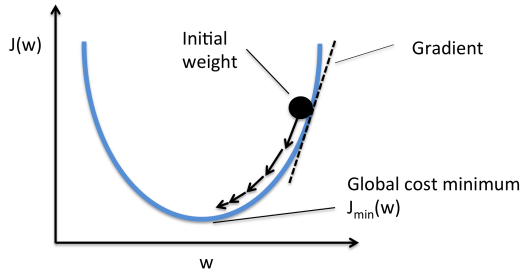


Figure 2: Sketch of gradient descent w.r.t. one parameter. [Available at: https://sebastianraschka.com/faq/docs/closed-form-vs-gd.html]

In neural networks, we use gradient descent to minimise the error or loss function while training, where the loss function is defined as a function of the difference between the neural network's output or estimate and the true output.

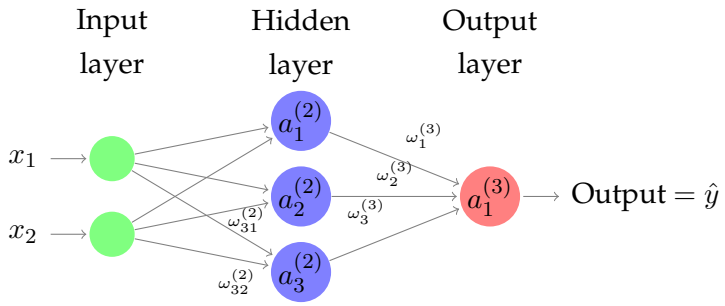
$$J(\omega) := \ell(f_\omega^{NN}(x), f(x))$$

[1] non-constant, continuous and bounded

The gradient descent algorithm is one of the most commonly used optimisation methods when training neural nets. More specifically, extensions of the gradient descent algorithm which have been created to achieve faster convergence and have higher likelihood of avoiding local minima. One extension is the stochastic gradient algorithm which iteratively updates the weights by calculating the derivative with one training example at a time. Other methods, such as Adam and Nadam, use "momentum" in a sense where learning is sped up where the gradient is consistent and slows down where it fluctuates. We used the latter which implements a jump of the parameter and calculates the gradient at that point to make a correction (Dozat (2016)).

3.1.2 Feedforward neural networks

A feedforward neural network is one where inputs move from the input layer, through the network, to the output layer with no return. This is also sometimes referred to as forward propagation. A feedforward neural network is usually one that has been trained to minimise the error on a training set.



$$z_3^{(2)} = \omega_{31}^{(2)} x_1 + \omega_{32}^{(2)} x_2$$
$$a_3^{(2)} = g(z_3^{(2)})$$

The structure of a neural network includes input and output layers of various dimensions with potentially multiple hidden layers. Above we consider a network with one hidden layer for simplicity, two input units and one output. Weights are applied to the input units and sent to the neurons in the hidden layer which receive them as a linear combination and apply an activation function, g.

This activation function stems from the Universal Approximation Theorem. This activation function is required to be differentiable to apply the backpropagation algorithm and non-linear, as if it wasn't, a network with many hidden layers with all but one linear activations could be represented by a network with one hidden layer or none if the activation functions are all linear.
A variety of activation functions exist, a common one being the sigmoid or logistic activation function and others such as ELU, ReLU and SoftPlus. As a side note, using the sigmoid function with only an input layer and output layer of one dimension equates to logistic regression. In the training of the Hull-White mixture model, a variety of activation functions were used in the same vein as Hernandez (2016).
The activated units $a^{(2)}$ are passed through the network just as the input units were until the output layer which will give us an estimate $\hat{y}$ of the function being approximated.

3.2 Training with backpropagation

3.2.1 Supervised learning

Neural networks are used to solve supervised learning problems where each observation forms a pair of input and output objects. This is opposed to unsupervised learning where there is no observable response variable, where methods such as PCA, clustering and self-organising maps are used to identify patterns and structure in the data.
Neural networks can be trained to infer the function between input and output objects, and thus do well in supervised learning problems especially with complicated functions and high dimensional data.

3.2.2 Training

In training a neural network, we use a training set of paired input and output examples such as $\{(\overrightarrow{x}_1, y_1), \ldots, (\overrightarrow{x}_m, y_m)\}$, where $\overrightarrow{x}_i$ is a vector as in the previous network.
The estimated outputs of the $\overrightarrow{x}_i$'s are $\hat{y}$ and are compared to the actual responses y_i by calculating the loss function. In the neural networks that were trained the

mean squared error was calculated and minimised as

$$J(\Omega) = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}_i - y)^2,$$

where Ω are the weights of the neural network.

3.2.3 Backpropagation

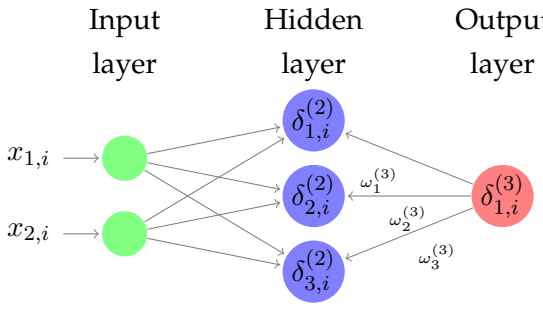
The goal is to minimise the loss function J with respect to the weights ω of the neural network, we thus need the following derivative to utilise the gradient descent algorithm and an application of the chain rule yields

$$\frac{\partial J}{\partial \omega_{jk}^{(2)}} = \frac{\partial J}{\partial z_{j,i}^{(2)}} \frac{\partial z_{j,i}^{(2)}}{\partial \omega_{jk}^{(2)}}$$

for the weights applying to the hidden layer in the simple network considered earlier. Of course we also need take the derivative of the loss with respect to the weights applying to the output layer, again using the chain rule.
The backpropagation algorithm works by first calculating the errors between the estimated and actual output object for each training example,

$$\delta_i^{(3)} = \hat{y}_i - y_i.$$

These errors are then propagated backwards along the branches by multiplying them by the weights in the same way that forward propagation was done, however without applying the activation functions. This allows us to get the errors at each node as follows.



It can be shown that,

$$\frac{\partial J}{\partial z_{j,i}^{(l)}} = \delta_{j,i}^{(l)}$$
$$\frac{\partial z_{j,i}^{(2)}}{\partial \omega_{jk}^{(2)}} = x_{k,i}$$
$$\frac{\partial z_{i}^{(3)}}{\partial \omega_{k}^{(3)}} = a_{j,i}^{(2)}$$
$$\implies \frac{\partial J}{\partial \omega_{jk}^{(2)}} = \delta_{j,i}^{(2)} x_{k,i} \quad \text{and} \quad \frac{\partial J}{\partial \omega_{k}^{(3)}} = \delta_{1,i}^{(3)} a_{j,i}^{(2)}$$

and this allows us to use back-propagation to minimise the loss function by updating the weights as per the gradient descent algorithm.

3.2.4 Hyper-parametrisation

One thing to consider is how many times to pass through the training set and update the weights. This corresponds to setting the number of epochs while training which is the number of times a forward and backwards pass of the training set is completed.
This is a so called hyper-parameter of the network, more specifically while training. There are many ways to parametrize neural networks, not only by number of neurons and layers.
Other hyper-parameters include performing forward and backward passes on subsets of the data, this is done by setting the batch size, or alternatively by specifying the number of forward and backwards pass in one epoch (number of iterations).

3.3 The Python Keras machine learning library

Keras is a high-level Python library and is rather seen as a deep-learning neural networks interface. It operates on top of either TensorFlow (C++), Theano (Python) or Microsoft Cognitive Toolkit (C++/CUDA, previously known as CNTK).
In constructing the neural networks in this report, TensorFlow was used as the backend for Keras. TensorFlow is an open-source machine learning library developed by Google and it provides an end-to-end machine learning framework for Keras to use, although it does provide its own Python API.

3.3.1 The Keras sequential model

The sequential model is a class of neural networks in Keras which is a linear stack of layers. It supports the constructing of usual ANNs as well as convolutional neural networks (basically 3-dimensional networks), amongst others.
The first step in the construction is to specify an input shape by creating the first hidden layer and specifying the input object dimensions.Layers are added sequentially to build the network, specifying the number of neurons and activation function for each layer. The model is then compiled by configuring the learning process. This is done by specifying the optimizer (e.g. Nadam) and the loss function (e.g. MSE). Once fitted on a training set, the model can be used to predict on new data. A graphical illustration of this is shown in the next section.

3.3.2 Toy example: European call option

The following is a very simplistic example where we apply a neural network to learn the calibration function between the option price and implied volatility. Options trade on volatility in the market, making this application somewhat redundant, however the point is to show how a network learns a simple calibration function before applying it to the Hull-White mixture model, as well as to show the construction using the Keras sequential model.

- Consider a vanilla call option priced at t_0 with parameters

  $T = 1$

  $K = 130$

  $S_0 = 100$

  $r = 0.0$

- The Black-Scholes model price is $C(S_t) = S_t\Phi(d_1) - Ke^{-rT}\Phi(d_2)$

- Suppose we want to calibrate the Black-Scholes model with one option price. This equates to finding the implied volatility of the option.

- In our toy example we want to train a neural network to learn the Black-Scholes formula, so that we can input a price to get the implied volatility.

- The structure will have 4 hidden layers with 64 neurons in each and the sigmoid activation function for all hidden layers.

- The loss function was specified as MSE and the optimizer used was Nadam.
- Used 100 000 training examples and 500 epochs with a cross-validation split of 90% / 10%.

The following python code is used to construct the neural network using the Keras sequential model where prices and vols used to fit the model are the 100 000 training examples.

```
x_dim = 1 #dimension of incoming data
exponent = 6
network = Sequential()
network.add(Dense(2**exponent, input_dim=x_dim))
network.add(Activation('sigmoid'))
layers = 4
for i in range(layers-1):
    network.add(Dense(2**exponent))
    network.add(Activation('sigmoid'))

network.add(Dense(1))

network.compile(loss='mean_squared_error', optimizer='Nadam')

network.fit(prices, vols, epochs = 500, validation_split = 0.1,verbose=0)
```

With a relatively simple calibration function and a large training set, we expect the neural network to approximate the function perfectly. This it does for volatilies (y-axis) less than 1 as can be seen in the following figure.
For higher prices (x-axis) and volatilities, the predicted values (green) on a test set (blue) diverge away from the actual recorded responses. This is due to no training occurring for higher volatilities and prices and thus the neural net cannot accurately predict the function for these values. This highlights a shortcoming of neural nets which is the reliance on a large training set which spans the entire parameter space.

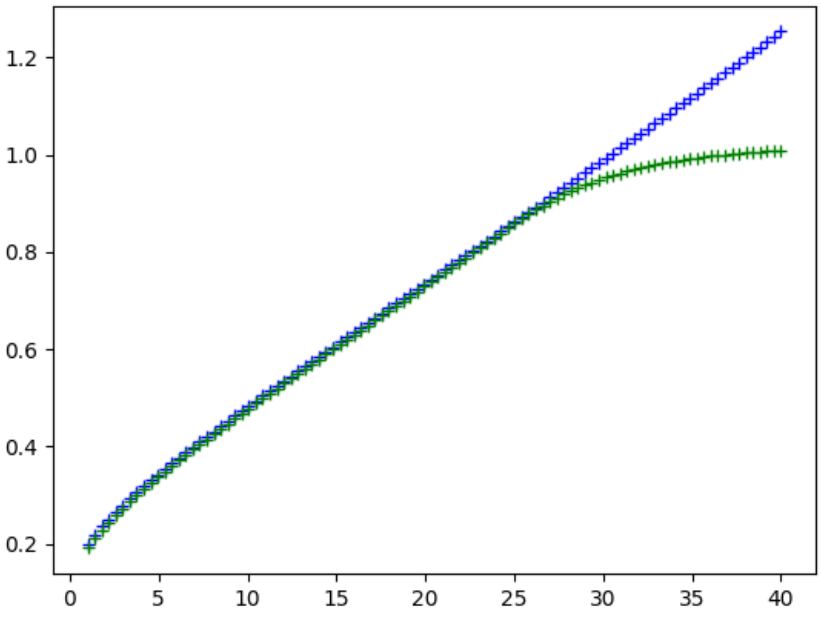


Figure 3: Performance of the network on a random test set. Estimated volatilities (green), actual implied volatilities (blue), x-axis = prices, y-axis = implied volatilities.

4 Applications to interest rate models

We now begin our discussion in spirit of the work done by Andres Hernandez (2016) on this particular model. As stated previously, the aim of the project is to provide a method that will perform the calibration significantly faster regardless of the model considered. This, in turn, removes calibration speed as a factor when considering a model's practicality.

4.1 Single-factor Hull-White model

The model being dealt with in this paper is the single factor Hull-White interest rate model. In its most generic formulation, it belongs to the class of no-arbitrage models that are able to fit today's term structure of interest rates. It is our goal to calibrate its parameters by means of an Artificial Neural Network.
The Hull-White model is given as :

$$dr_t = (\beta(t) - \alpha r_t)dt + \sigma dW_t \tag{22}$$

Here, the parameters of the model are θ, α and σ. Here, $\theta(t)$ represents the level of mean reversion and is a function of time determining the average direction in which *r* moves, and is chosen such that the movements in *r* are always consistent with today's zero coupon yield curve. α represents the speed of mean reversion, governing the relation between short and long rate volatilities. Finally, σ represents the annual standard deviation of the short rate (i.e. volatility) θ is calculated from the initial yield curve describing the current term structure of interest rates, whereas α and σ are assumed to be constant, and will be estimated using the Artificial Network Network.
Hence, the problem can then be given by:

$$(\alpha, \sigma) = \Theta\left(\left\{Q^{mkt}\right\}\right) \tag{23}$$

4.1.1 Basic procedure

This section aims to highlight the simple procedure used when calibrating the Hull-White model using Neural Networks.
The following points provide an overview of the steps used in Andres Hernandez's procedure:

- Obtaining the historical data of the instruments used in the calibration procedure
- Calibrating the model to set up a time series of (α, σ) and $\beta(t)$
- This data then undergoes pre-processing, and we generate many combinations of parameters
- We input all our data into a large matrix containing prices and parameters, which then serve as the input for the neural network
- Once, the neural network has been sufficiently trained, we then use out-of-sample data to test the neural network and compare to market values

4.1.2 Understanding the data set

The calibration instruments used in this setup were 156 swaptions(equally weighted). The historical data collected primarily comprised of ATM volatility quotes for GBP from January 2nd, 2013 to June 1st, 2016. The reason ATM swaptions were considered were possibly due to these options being the most liquidly traded in the market. However, In-the-Money and Out-of-Money swaptions could have also been considered. The option maturities were 1 to 10 years, plus 15 years and 20 years. The swap terms from 10 to 10 years, plus 15 years,20, and 25 years.
For the yield curve, the 6M tenor LIBOR curve was used, bootstrapped using a monotonic cubic spline interpolation of the zero curve, and built on top of the OIS (Overnight Indexed Swap) rate curve, which was also part of the data set. In order to have a consistent set of data , yield curves are always constructed using the yield rates of a set of homogeneous instruments. For bond yield curves for example, this means in particular that one always uses instruments from the same issuer or, if it is a sector curve, from issuers which belong to the same sector. But building a yield curve from classic coupon bonds would create a curve which suffers from a number of inconsistencies. Thus, for example, two bonds with the same maturity but a very different duration, will not have the same yield. Also, two identical coupons belonging to two bonds with different maturities will not be discounted at the same yield, whereas they generate the exact same cash flow. To overcome these problems, one constructs a zero-coupon yield curve from the prices of these traded instruments. As a reminder, the zero-coupon rate is the yield of an instrument that does not generate any cash flows between its date of issuance and its date of maturity. The technique used to achieve this is called bootstrapping, a term which describes a self-contained process that is supposed to proceed without external input. This method is based on the assumption that the theoretical price of a bond is equal to the sum of the cash flows discounted at the zero-coupon rate of each flow. In this particular case, only FRA's and swaps were used to bootstrap the curve. When served as an input to the neural network, the yield curve was discretely sampled at 44 point: 0, 1, 2, 7, 14 days; 1 to 24 months; 3 to 10 years; plus 12, 15, 20, 25, 30, 40 and 50 years.
The cost function considered here was the weighted average of all the differences between the market provided quote and model-provided quote. As per standard practice, a Levenberg-Marquardt local optimizer was used to minimize this cost

function.
The calibration was then done twice, with two different starting points:

- A default starting point with α = 0.1 and σ = 0.01 was chosen
- As a second starting point, the calibrated parameters from the previous day (using the default starting point) were also used.

The results from the default calibration are as shown below :

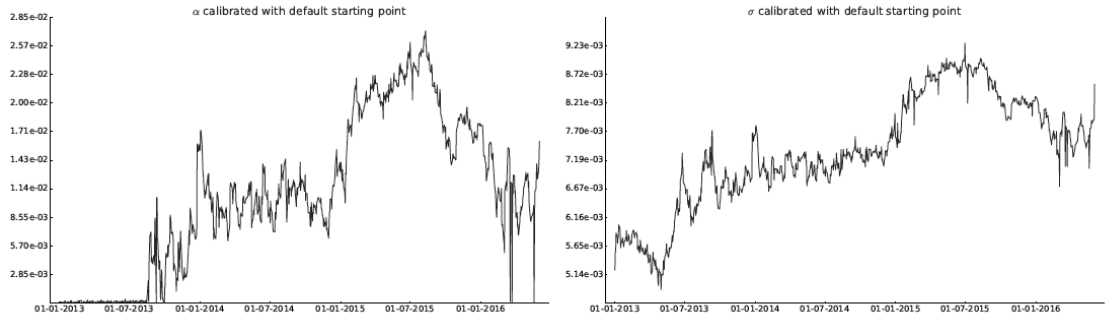


Figure 4: Parameters estimated by default calibration.

4.1.3 Training set generation

Once the calibration history is available, the training set can then be generated in the following way :

- Obtain errors for each calibration instrument for each day (i.e. *Market price - Model price*)
- Since the parameters are positive, proceed to take the natural log on the calibrated parameters. This ensures that the distance between the sizes of the parameters is significantly reduced.
- The next step involves re-scaling yield curves, parameters and errors to have a mean of 0 and a variance of 1
- Apply dimensional reduction via PCA to yield curve, while keeping parameters for given explained variance (99.5%)

- Calculate covariance of rescaled log parameters, PCA yield curve values and errors
- Generate random normally distributed vectors consistent with given covariance structure as found previously
- Then, an arbitrary reference date is selected from the set used for covariance estimation
- Finally, obtain implied volatility for all calibrations and apply random errors to the results, hence we now have a set of prices adjusted for the sampled errors

4.2 Hyper-parameter optimisation

In the context of machine learning, hyperparameters are parameters whose values are set prior to the commencement of the learning process. By contrast, the value of other parameters is derived via training. In this case, the number of neurons, the number of layers, the activation function are all examples of hyper-parameters. For the hyper-parameter optimization, the sample set was divided into three parts: 60% served as the training set, 20% served as the cross-validation set and the last 20% served as the testing set. The training was the set used during inter-epoch training, while the cross-validation set was primarily used to measure inter-epoch improvement. The testing set was then used to find the most suitable configuration, with comprised the model corresponding to the minimum error. The fitting of the hyper-parameters does not imply any problems for the actual testing done later.
The optimization was a mixture of grid and manual search, with a truncated grid search over the number of layers, neurons per layer, the learning rate for a RMSProp optimizer, and a drop-out tate, that was set equal for all layers, including the output layer.
Once the best set of hyper-parameters was selected, different activation units were tried, including a ReLU (Rectified Linear Unit) and an ELU (exponential linear unit). The ELU provided the most improvement for the model.

Having generated a sample of 150,000 samples, the hyper-parameter optimization led to the feed-forward neural network as given in the figure below:

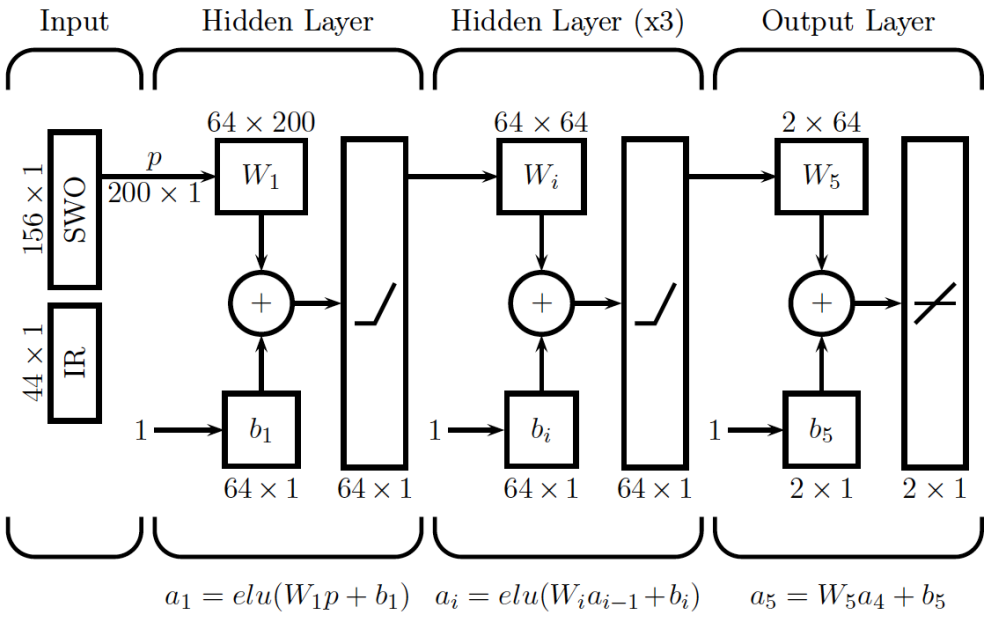


Figure 5: FNN with 4 hidden layers, each layer consisting of 64 neurons.

A standard 500 epoch was used for the purpose of training the neural network. Differnt variations of Convolutional Neural Networks were tried, however, none were a match for the best FNN. However, when the number of calibration instruments are significantly increased, CNN's could become more useful.
Two different neural networks were trained using the sample set generated with the previously estimated covariance matrix. The first sample used 40% of the historical data whereas the second sample set used 73% of the historical data. For training purposes, the sample set was split into two once again - 80% for the training set and 20% for cross-validation. The testing set was the historical data itself, which then served as backtesting for the model.

4.3 Observations and conclusions

The procedure used by Hernandez led to several interesting results. One of the major observations was that of the trained model. It was observed that it behaved well only for a period of 6 months to a year beyond the period which was used for

the covariance estimation.

As seen below, we can observe that the out-of-sample back testing shows good behaviour for a year post correlation estimation period. After that, the model produced several inaccuracies in the results.

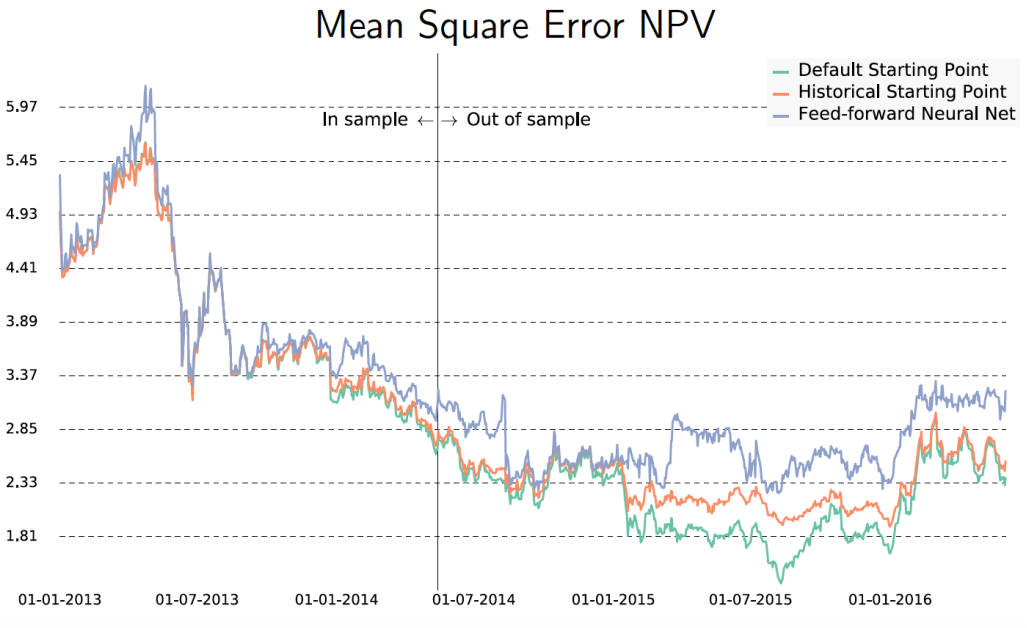


Figure 6: Correlation up to June 2014.

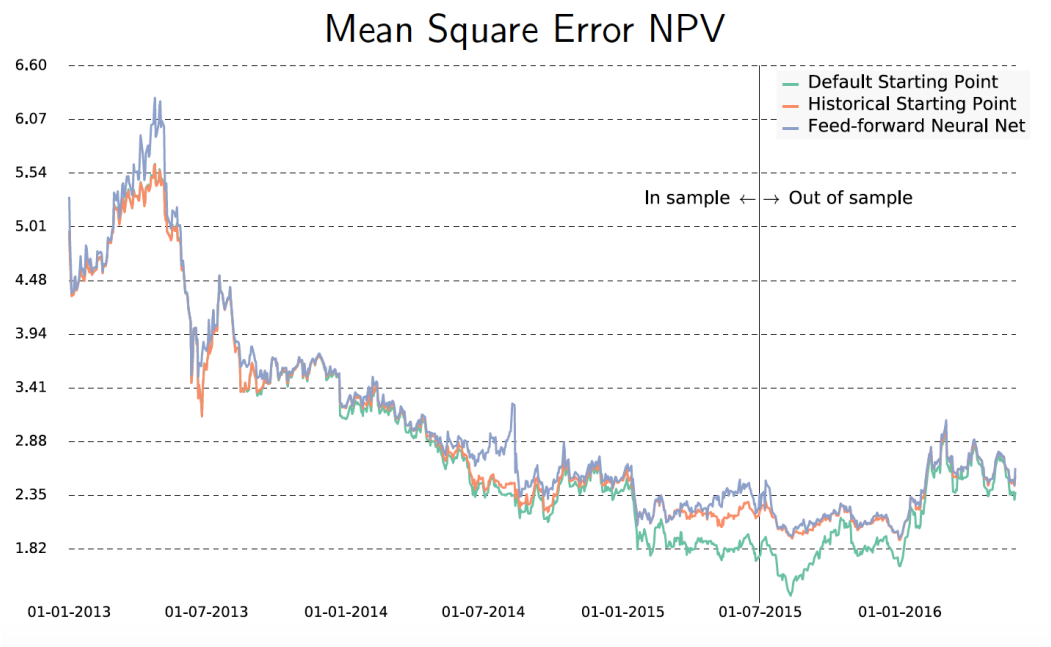


Figure 7: Correlation up to June 2015.

The performance degradation after the one year period, however, is not problematic, as one could simply re-train the model every 2-3 months, and still remain within acceptable bounds. In essence, what was learnt that sampling from a parametrized correlation structure, could extend the lifespan of a trained model.

4.4 The mixture model

As seen in the previous section, the model parameters are very unstable. As a remedy to this, we now consider a slight generalization to the standard Hull-White model. Consider the two Hull-White models (on a space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$):

$$dr_t^{(1)} = (\beta_1(t) - \alpha_1 r_t^{(1)})\, dt + \sigma_1\, dW_t$$

$$dr_t^{(2)} = (\beta_2(t) - \alpha_2 r_t^{(2)})\, dt + \sigma_2\, dW_t.$$

We will assume that the short rate r is a mixture of the two, in the sense that

$$\mathbb{P}(r_t \leq x) = \pi \mathbb{P}\left(r_t^{(1)} \leq x\right) + (1-\pi)\mathbb{P}\left(r_t^{(2)} \leq x\right),$$

where $\pi \in [0,1]$ is a fixed parameter. In particular, this implies that the price of every instrument is simply a convex combination of the prices under the two models.

To calibrate this model using standard techniques is not feasible in real time, so we used the neural network approach to estimate the calibration function Θ. We still have the same set-up (in terms of data):

- $N = 156 + 44 = 200$ input prices (swaptions + yield curve)
- $n = 44 + 4 + 1 = 49$ parameters to estimate. These are $\alpha_1, \alpha_2, \sigma_1, \sigma_2, \pi$ and $\beta_1(t)$ (or, equivalently, $\beta_2(t)$) at 44 maturities.
- Hence, the calibration function is now

$$\Theta : \mathbb{R}^{200} \longrightarrow \mathbb{R}^{49}, \quad \begin{pmatrix} SWO_1 \\ SWO_2 \\ \cdots \\ yield(0) \\ yield(1) \\ \cdots \end{pmatrix} \mapsto \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \sigma_1 \\ \sigma_2 \\ \pi \\ \beta_1(0) \\ \beta_1(1) \\ \cdots \end{pmatrix}$$

To estimate Θ using neural networks, we modified the training data that was previously used in training the calibration function for the single Hull-White model. This involved joining pairs of prices and parameters.

After training the model with 150000 training examples, we calculated (for each day) the average differences in implied volatility from the model and from the market:

$$\text{AVG volatility error} = \frac{1}{156} \sum_{i=1}^{156} |Q_i^{mkt} - Q_i^{\mathcal{NN}}|.$$

Note that the prices Q are expressed in terms of implied volatilities. Here is a diagram that shows this error, both for normal calibration using QuantLib (red) and our neural network (blue). As can be seen from the diagram, the two errors are almost identical. The time series for the estimates of π, α_1 and σ_1 are also given below.

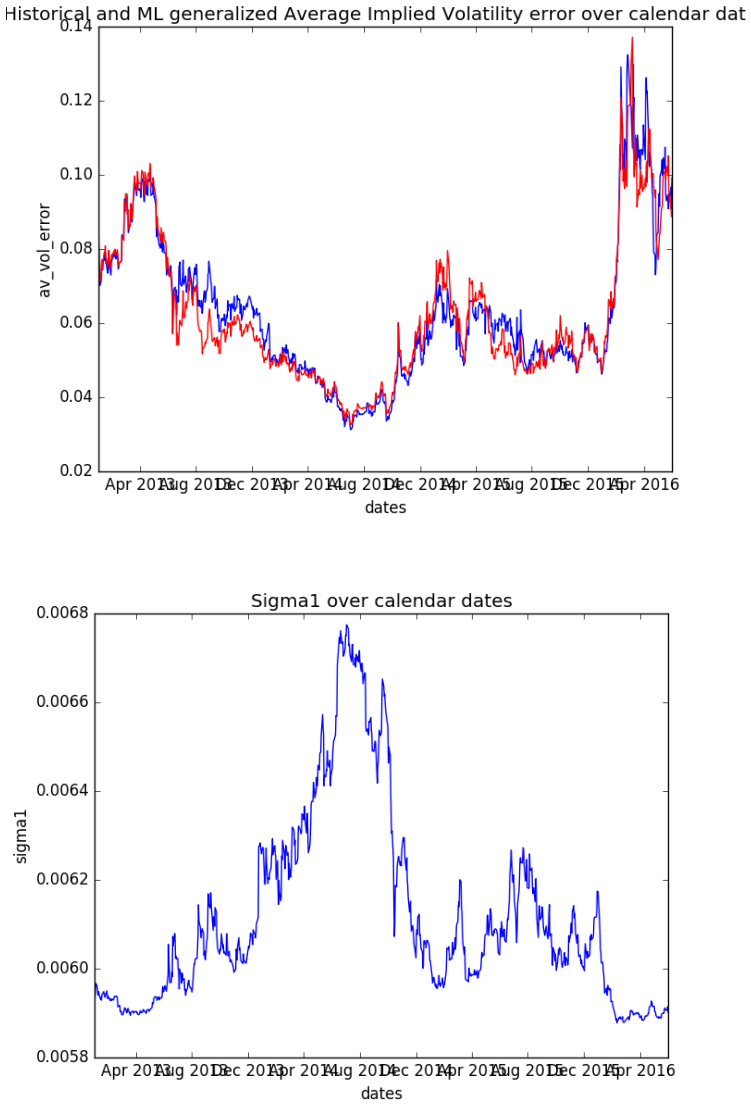


Figure 8: Parameter stability.

These graphs show parameter stability, as the changes in the values of the parameters are small (in absolute and relative terms).

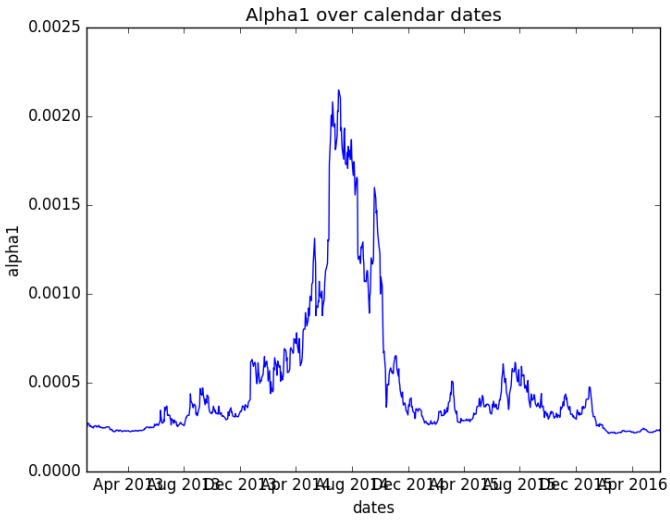


Figure 9: Parameter stability.

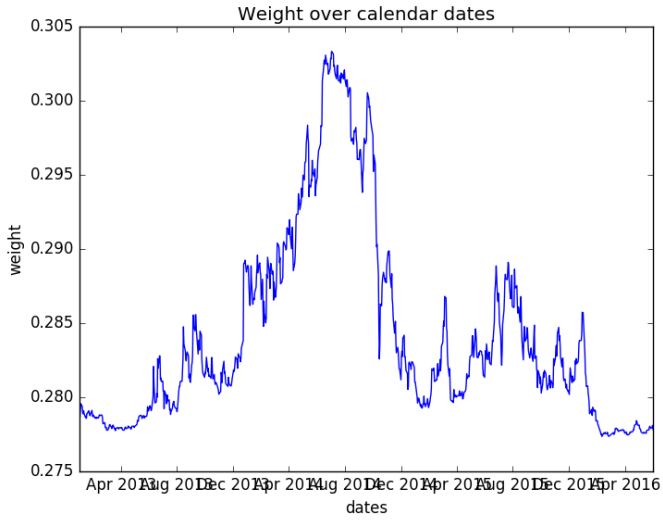


Figure 10: Parameter stability.

Bibliography

Crisostomo, R., 2015. An analysis of the heston stochastic volatility model: Implementation and calibration using matlab. arXiv preprint arXiv:1502.02963.

Dozat, T., 2016. Incorporating Nesterov Momentum into Adam. OpenReview.net.

Gurrieri, S., Nakabayashi, M., Wong, T., 2013. Calibration Methods of Hull-White Model. Risk Management Department, Tokyo.

Hernandez, A., 2016. Model Calibration with Neural Networks. Available at SSRN: https://ssrn.com/abstract=2812140.

Larsson, E., 2015. Introduction to Calibration. Uppsala University,Sweden.

Schlögl, E., 2015. Toward quantifying model risk, qFRC, University of Technology Sydney.

Managing estimation risk in Mean-Variance portfolio optimisation

Team 4

WAHID KHOSRAWI-SARDROUDI, University of Freiburg
MATTHEW GORVEN, University of Cape Town
NAILL MORLEY, University of Cape Town
VYKTA WAKANDIGARA, University of Cape Town

Supervisor:
ERIK SCHLÖGL, University of Technology Sydney

African Institute of Financial Markets and Risk Management

Contents

1 Introduction

Mean-variance portfolio optimisation, pioneered by the work of Markowitz (1959) and further extensively studied by others has for a long time been one of the most active areas of research in the intersection of finance and mathematics. In particular, its relevance in practice for portfolio managers makes a precise understanding crucial.

One of the most used frameworks in which mean-variance optimisation is performed is the CAPM and extensions like Fama-French. By imposing structural assumptions on asset returns, these models allow to estimate the high dimensional covariance matrix by performing a linear regression on the time series of asset returns. However, what CAPM and related models have in common with other mean-variance portfolio optimisation models is that they take expected returns and covariance matrices as given input, not incorporating parameter estimation uncertainty. In addition, these models make assumptions on underlying distributions and time homogeneity, both which can be in contradiction to market observations.

These issues lead to the notion of model uncertainty as discussed (among others) in Schlögl (2016) where different types of these uncertainties are presented (see further down for a more comprehensive overview). In an ideal world, where model assumptions are not violated by empirical data, the covariance matrix does not change with time and hence an optimal portfolio once computed would stay optimal as time progresses. Empirical data however rejects these homogeneity assumptions which makes it necessary to re-balance the portfolio after a period of time.

A novel idea to address these points was made in Glasserman and Xu (2014) where relative entropy is introduced to unify uncertainty - multivariate in nature when considering portfolios of multiple assets - in a single number. This allows one to develop a unifying framework to quantify model risk. One of the applications of this approach was mean-variance optimisation itself. As Glasserman and Xu show, in the context of mean-variance optimisation, the concept of relative entropy has a preserving property regarding normality assumptions.

In the context of this project we seek to address the following points, building on the results of Glasserman and Xu:

- Given historical data, what would be a reasonable method to choose the level of uncertainty formulated by relative entropy?
- After what time should one re-balance portfolios given that the used model assumes homogeneity?
- How do different estimators perform and how do they influence the level of uncertainty needed to protect against model changes in the data?

We begin by shortly presenting the classical case and stating the CAPM framework used for the estimation of the covariance matrix as an alternative to the sample covariance matrix estimator. We continue with presenting the mathematical framework of the robust version of mean-variance optimisation. As this project is mainly empirical in nature, we proceed by presenting the methodology we developed to investigate the questions raised above. In addition we attempt to explore the nature of estimated relative entropies of asset return distributions to each other. This is implemented in order to answer the question of which portion of the uncertainty level comes from parameter estimation error that in theory can be eliminated by considering large enough amounts of data and which portion comes from model risk beyond parameter estimation risk. This we complete by stating and discussing the numerical results we got from the implementation of our approach using actual market data.

2 Background

2.1 Introductory background

A common approach in asset management is Mean-Variance portfolio optimisation, also known as Modern Portfolio Theory (MPT). This approach was first laid out by Markowitz (1959) as an optimisation problem in the following form. Consider a random vector $X \in \mathbb{R}^N$ (e.g. yearly relative returns per asset) and a measurable function family $V(X)_a \in \mathbb{R}$ indexed by sum $a \in \mathcal{A}$ where $\mathcal{A}$ is the set of admissible parameters. Mathematically, we can formulate our problem as solving

$$\max_{a \in \mathcal{A}} \mathbb{E}\left[V_a(X)\right].$$

In the context of the Mean-Variance portfolio optimisation, the above problem is further specified by setting

$$\mathcal{A} := \left\{ a \in \mathbb{R}^N : \quad \sum_i a_i = 1 \right\},$$

and $V_a(x) := a^\top x - \frac{\lambda}{2} a^\top x x^\top a$. As one can easily see this translates into the maximisation problem

$$\max_{a \in \mathcal{A}} \left(a^\top \mu - \frac{\lambda}{2} a^\top \Sigma a \right),$$

where a is a $N \times 1$ vector of portfolio weights which sum to one allowing for negative entries corresponding to short selling, μ is the $N \times 1$ vector of expected portfolio returns, Σ is the $N \times N$ covariance matrix of returns, and the parameter λ denotes the risk aversion preference of an investor. This equation expresses the trade-off between expected portfolio return and portfolio variance. A maximisation of this equation therefore represents the best trade off between return and variance for a given level of risk aversion for the investor.

A standard assumption associated with this methodology is that all estimates of expected returns, variances and covariances are known and exact. However, many empirical studies have shown that estimates often deviate from their exact values which makes sense by the very nature of point estimators, leading to an error in the portfolio optimisation process. This is indeed true for all statistical estimates. However, investors often consider these values to be indisputable which should be

viewed as a naive viewpoint, as estimation risk has not been appropriately considered. Causes of estimation risk include:

- inappropriate choice of estimation method;
- inappropriate choice of data, with regards to its source, frequency and time window ;
- lack of stationarity;
- random sampling error;
- inaccurate data; and
- insufficient data.

The impact of estimation risk on portfolio optimisation is well known, and attempts at incorporating this risk into portfolio selection have been made. The literature surrounding these methods is summarised as part of this report, with the associated strengths and limitations. The main focus of our report is with the robust approach to portfolio optimisation suggested by Glasserman and Xu (2014) due to its ability to adjust for incorrect estimation of the distribution of returns.

The method of choosing a relative entropy budget, or η, is not specified by Glasserman and Xu (2014). This paper suggests that the size of this budget should correspond with the confidence associated with the estimated parameters, as parameters with higher estimation risk would require a higher entropy budget. In this paper, we solely concentrate on quantifying the risk associated with estimating the covariance matrix. As Ledoit and Wolf (2003) points out, the estimation of the covariance matrix is vital to controlling the risk associated with portfolio selection. Hence, the choice of estimation method is considered in order make an informed choice of η that allows for appropriately robust portfolio optimisation.

In Glasserman and Xu (2014), the authors do not specify a method to derive a level of uncertainty against which one wants to make the considered target robust. This report therefore aims to address this point and present an implementation of the relative entropy metric proposed by Glasserman and Xu (2014) that quantifies estimation risk associated with portfolio Mean-Variance optimisation. In addition,

it suggests a Mean-Variance portfolio optimisation methodology that accounts for worst case estimation error at a given confidence level which is linked to the covariance estimation method. Finally, the suggested robust optimisation procedure is implemented using various methods to estimate the covariance matrix of returns. A similar study conducted by Nghiem (2015) on the optimisation of beta and its effect on the CAPM model implies the time period can have a significant effect on the estimated value. Therefore, each method is compared over various time windows, in an attempt to further refine the portfolio optimisation process.

2.2 Mean-Variance portfolio theory

Mean-Variance portfolio theory was developed by Markowitz (1952,1959) as a method for constructing an investment portfolio that provides a maximum return for specified level of risk. Alternatively it provides the portfolio with the minimum level of risk given a specified return. The measure of risk is defined as the variance. Mean-Variance portfolio theory relies on several assumptions according to Bodie et al. (2014):

- All expected returns, variances and covariances of assets are known.
- Investors base all their decisions on values of expected return and variance.
- Investors are non-satiated and risk-averse.
- All investors are exposed to a single-step fixed time period.
- There are no taxes or transaction costs.
- The short-selling of assets is possible, with no maximum investment limits.

The return of the portfolio r_p is calculated by summing over the returns of all individual assets r_i held in a certain proportion x_i. The proportions is subject to the constraint that the weights must sum to one, i.e. in the introductory setting, the vector $x = (x_1, \ldots, x_N)^\top$ plays the role of a. Portfolio returns are computed as

$$r_p = \sum_{i=1}^{N} x_i r_i = x^\top r, \quad r := (r_1, \ldots, r_N)^\top.$$

The expected return of the portfolio is given by:

$$E_p = \mathbb{E}[r_p] = \sum_{i=1}^{N} x_i E_i = x^\top E,$$

again with $E_i = \mathbb{E}[r_i]$ and $E = (E-1, \ldots, E_N)^\top$. The variance of the portfolio's return is computed by:

$$\sigma_p^2 = \mathbb{V}\text{ar}(r_p) = \sum_{i=1}^{N} x_i^2 c_{ii} + \sum_{i=1}^{N} \sum_{j \neq i} x_i x_j c_{ij},$$

where $c_{ii} = \sigma_i^2 = \mathbb{V}\text{ar}(r_i)$ and $c_{ij} = \rho_{ij}\sigma_i\sigma_j$ is the covariance of assets i and j.
The method of Lagrangian multipliers can be used to determine the optimal portfolio. This method seeks to optimise the following equation by taking partial derivatives of the x_i, λ and μ and setting these to zero, which can then be solved simultaneously:

$$W = \sum_i \sum_j c_{ij} x_i x_j - \lambda(\sum_i E_i x_i - E_p) - \mu(\sum_i x_i - 1)$$

Through varying the levels of expected return, an efficient frontier is produced on the E-σ plane. Along this curve is where all rational investors will invest as it contains all efficient portfolios.The chosen portfolio is then subject to the risk appetite of the investor Bodie et al. (2014).

2.3 CAPM

The Capital Asset Pricing Model was originally developed in 1967 by William Sharpe, John Lintner and Jan Mossin to further develop portfolio optimisation. The model explains the relationship observed between the expected excess return of a financial asset and the systematic risk level of the market as a whole and is considered one of the most important models in modern financial economics (Nghiem, 2015).
The total risk of a financial asset can be decomposed into two parts, systematic risk and firm-specific risk. Firm-specific risk can be avoided or removed through the process of diversification; which is to hold a wide range of assets both by sector and industry. However, the same technique cannot be undertaken for systematic risk. As a result, an asset-risk premium is required as compensation for bearing systematic risk. The risk premium or excess return can be calculated as the difference of

the return of an asset and the risk-free rate Nghiem (2015).
The CAPM model is a single risk factor model, with the world market portfolio representing the risk factor taken into consideration. The world market portfolio consists of all available financial assets in the market held in proportion to the market share/capitalisation of each financial asset. Therefore, since the world market portfolio is considered to be the most diversified portfolio, there should be a proportionality between the expected risk premium of the world market portfolio and the expected risk premium of the individual financial asset measured by the betas Nghiem (2015).
The following assumptions are considered to underlie the CAPM model according to Bodie et al. (2014):

- The market experiences perfect competition. This implies there are many investors present in the market with each investor's total wealth representing a negligible amount compared to the total wealth of all investors.
- All investors have a single fixed identical holding period.
- Investors are able to borrow or lend at a fixed, risk-free rate and can only invest in publicly traded financial assets.
- There are no taxes or transaction costs.
- All investors are considered to be rational Mean-Variance optimisers. As well as analysing securities in a consistent manner and holding an identical economic outlook of the world.

Under the CAPM framework, the expected return on an asset i is given by:

$$\mathbb{E}[r_i] = r_f + \beta_i^M(\mathbb{E}[r_M] - r_f),$$

where r_f is the risk-free rate of return, r_i is the return on asset i, r_M is the return on the world market portfolio, $\beta_i^M = \text{Cov}\left(r_i, r_M\right)/\sigma_M^2$ is the systematic risk of asset i relative to the world market portfolio and σ_M^2 is the variance of the return on the world market portfolio.
The world market portfolio has been defined as consisting of all available financial assets in the world. However this portfolio cannot be directly observed and therefore an estimate for β_i^M is required. An index I is introduced as a proxy for

the world market portfolio Bartholdy and Peare (2005). Time series regression is undertaken to obtain such an estimate:

$$r_{it} - r_{ft} = \alpha_i + \beta_i^I (r_{It} - r_{ft}) + \epsilon_{it}$$

For $t = 1, ..., t_0$ where r_{ft} is the risk-free return at time t r_{It} is the return on index I at time t,β_i^I = Cov(r_i,r_I)/σ_I^2 is the systematic risk of asset i relative to index I, σ_I^2 is the variance of the index and ϵ_{it} is a white noise error term Bartholdy and Peare (2005). If the CAPM equation always holds we consider assets are fairly priced, the α_i should be zero indicating that CAPM always holds. A positive α_i indicates an asset is undervalued and is earning a higher excess return than expected. It is considered overvalued and earning a lower excess return than expected if an α_i is negative Nghiem (2015).

Both Mean-Variance portfolio theory and CAPM assume that exact estimates of expected returns and the variance/covariance matrix of returns are available. These assumptions are challenging as it is extremely difficult and near impossible to achieve this with a high degree of accuracy Morin (1994). This project seeks to move away from the assumption that exact estimates are available and rather allow the estimates to vary, optimising the worst case outcome at a given level of confidence.

2.4 Model risk

Model risk is the risk practitioners face by using a model which produces incorrect outputs and reports and then basing their decisions on these inaccurate information. Categories of model risk include how appropriately practitioners use their models, irrespective of model choice, as well as the ability of market data to be fitted by the relevant model. To understand and manage model risk a comparison of models has to be undertaken through some notion or measure.

The sources of model risk have recently been classified within four aspects by Schlögl (2016):

1. Type 0: Parameter uncertainty is the first type of model risk. Through the statistical estimation of model parameters a statistical error bound on these parameters will be realised. The accuracy and confidence in model outputs is effected by this uncertainty. The sensitivity of the model to incorrect parameter specifications is also considered a type 0 risk.

2. Type 1: Calibration error is the second type of model risk. The calibration error is the potential for the model to be unable to fit a full set of market observations. The Black/Scholes models inability to capture a volatility smile is a simple example of Type 1 model risk. The inability of the model to reprice market observations on a given day thus contradicts the model assumptions.

3. Type 2: The third type of model risk is induced by the recalibration of the model and the resultant change in the parameters. As time passes the model needs to undergo recalibration to ensure an accurate fit as possible. An assumption of fixed model parameters will then be violated through the potential daily changes.

4. Type 3: The final type of model risk refers to the error resulting from model dynamics not matching the true empirical state variable dynamics. Type 3 model risk is thus a violation of the model assumptions.

3 Theory

This section is meant to give a short representation of the different theoretical methods necessary to implement our approach. We do not aim for a full discussion but rather restrict ourself to the most basic version of what is needed.

3.1 Relative Entropy

The first step is to give an overview of relative entropy. We then continue with establishing the relation to Mean-Variance portfolio optimisation.

3.1.1 Introduction to Relative Entropy

We start to introduce Relative Entropy, which provides a notion of distance between two probability measures. We denote f as the reference density and any other density is denoted as $\tilde{f}$. Even though relative entropy could be defined for measures that do not have a density, it suffices for our purposes to stay in the setting where densities exist.

Definition 3.1. Likelihood Ratio: Let f and $\tilde{f}$ be two density functions. We define the Likelihood Ratio of f and $\tilde{f}$ as

$$m := \frac{\tilde{f}}{f} 1_{\{f>0\}},$$

where we assume $\{f > 0\} = \{\tilde{f} > 0\}$.

In the sequel, we will without loss of generality assume $f > 0$ as one can always reduce the state space to the support. We can now prove the following lemma.

Lemma 3.2. The Likelihood Ratio m is a Radon-Nikodym derivative .

Proof. Let X be a random variable with density f and $V(X)$ be some function of X. We have that:

$$\mathbb{E}[m(X)V(X)] = \int m(x)V(x)f(x)dx = \int V(x)\tilde{f}(x)dx = \tilde{\mathbb{E}}[V(X)]$$

□

Definition 3.3. Relative Entropy: Let m be the likelihood ratio between f and $\tilde{f}$. We define the Relative Entropy $R(f, \tilde{f})$ between f and $\tilde{f}$ as

$$R(f, \tilde{f}) := \mathbb{E}[m \log m] = \int \frac{\tilde{f}(x)}{f(x)} \log \frac{\tilde{f}(x)}{f(x)} f(x) dx.$$

Though $R(f, \tilde{f})$ effectively acts as a measure of 'distance' between two densities, it is not symmetric and is therefore not a metric. However, it does have some desirable properties that are inherent to metrics. For example, $R(f, \tilde{f}) = 0$ if and only if $f = \tilde{f}$ with respect to f. Using Jensen's inequality, one can also verify that $R(f, \tilde{f}) \geq 0$.

Relative Entropy coincides with Kullback-Leibler divergence, which is used in Bayesian Statistics to quantify how much information can be gained from collecting additional data. This involves measuring the entropy between the posterior and the prior distribution. In the case of risk measurement, Relative Entropy can hence be interpreted as a measure of the amount of information needed to make a perturbed distribution preferable to the reference distribution.

3.1.2 Relative entropy in relation to Mean-Variance portfolio optimisation

Glasserman and Xu (2014) present the relative entropy criterion as a mean to quantify model risk. It is our aim to address the question of how much model risk in means of relative entropy should be considered. This method attempts to account for a misspecification of the covariance matrix for a given "distance" away from the true value, given by relative entropy. As we stated before, relative entropy is a measure of the required additional information that would make a model $\tilde{f}$ preferable to a nominal model f.

$$\mathcal{R}(f, \tilde{f}) = \mathbb{E}[m \log m] = \int \frac{\tilde{f}(x)}{f(x)} \log \frac{\tilde{f}(x)}{f(x)} f(x) dx,$$

where m is the likelihood ratio $m = \tilde{f}/f$. In order to consider a certain range of deviation away from the nominal model f, we consider a relative entropy budget given by η. In other words, any likelihood ratio that satisfies $\mathbb{E}[m \log m] < \eta$ must be considered. Glasserman and Xu (2014) convey how Mean-Variance portfolio optimisation can be written in the form:

$$\inf_a \left(-\mathbb{E}\left[a^\top X - \frac{\gamma}{2} a^\top (X - \mathbb{E}(X))(X - \mathbb{E}(X))^\top a \right] \right).$$

The uncertainty associated with the distribution of returns is constrained to the uncertainty of the covariance matrix, therefore we assume the mean returns vector is correct. The optimisation problem considering this uncertainty becomes:

$$\begin{aligned} \inf_a \sup_m \mathbb{E}\Big[m V_a(X) \Big] = \quad & \inf_a \sup_m \left(-\mathbb{E}\Big[m \Big(a^\top X - \tfrac{\gamma}{2} a^\top (X - \mathbb{E}(X))(X - \mathbb{E}(X))^\top a \Big) \Big] \right), \\ & \text{s.t. } \mathbb{E}[mX] = \mu. \end{aligned}$$

As in Glasserman and Xu (2014),the worst case likelihood ratio for some $\theta > 0$ given a constrained mean returns vector satisfies:

$$m^* \propto \exp\left(\frac{\theta\gamma}{2} a^\top (X - \mu)(X - \mu)^\top a \right).$$

The optimal a can be found under these conditions by solving for fixed $theta$ the following optimisation problem:

$$a^*(\theta) = \arg \inf_{a \in \mathcal{A}(\theta)} \frac{1}{\sqrt{\det(I - \theta\gamma a a^\top \Sigma)}} + a^\top \mu.$$

The corresponding relative entropy budget is then:

$$\eta(\theta) = \frac{1}{2} \Big(\log(\det(\Sigma \tilde{\Sigma}^{-1})) + tr(\Sigma^{-1} \tilde{\Sigma} - I) \Big).$$

This is the only proposed method that accounts for data that is unrepresentative of the true distribution of returns. An extension to this work is to account for the misspecification of the mean returns vector. This robust optimisation using different methods of estimating the covariance matrix along with a methodology of choosing η without knowing the true covariance is the focus of this paper.

3.2 Estimation methods

All the estimation methods considered impose the following assumptions:

- Stock returns are independently and identically distributed (iid);
- Weak stationarity of returns;

- The number of stocks are finite.

The assumption that stock returns are iid is problematic, but this is a common assumption among covariance estimators that cannot be avoided. The stationarity of returns is also problematic. Relatively short time periods must be considered for this to hold. However, if the time period considered is too short, there is not enough data for a good estimate, increasing estimation error. It is therefore an optimisation problem to choose the correct time window.

In order to introduce model reliant covariance estimators, we will discuss each one in the context of CAPM. This is the commonly used model for model reliant covariance estimators, and as this paper focuses on proof of concept, it is not necessary to extend this to other models.

3.2.1 Sample covariance matrix estimation

The sample covariance matrix for n observations of random variables $\{x_1, x_2, ..., x_N\}$, is defined as:

$$\Sigma = \begin{bmatrix} \sigma_{1,1} \dots \sigma_{1,N} \\ \vdots \quad \ddots \quad \vdots \\ \sigma_{N,1} \dots \sigma_{N,N} \end{bmatrix}$$

where each element can be calculated by:

$$\sigma_{i,j} = \frac{1}{n-1} \sum_{k=1}^{n} (x_{j_k} - \overline{x_j})(x_{i_k} - \overline{x_i}).$$

This is an unbiased estimator that converges to the true covariance matrix as $n \to \infty$. However, this type of estimator is known to perform poorly in high dimensions. When the dimension to sample ratio, N/n, is too big, the estimated matrix is numerically ill-conditioned. Therefore, operations such as inversion can amplify the estimation error further. Large sample sizes are required to avoid this issue. In order to counteract the ill-conditioning of the sample estimator, model constrained covariance estimation is considered.

3.2.2 Beta covariance matrix estimation

In this section we show how to calculate the estimate of the covariance matrix implied by the CAPM. The estimator requires the calculation of the β coefficients in the CAPM; which may be derived using either the method of ordinary least squares or the maximum likelihood method. We begin by describing the former.

3.2.3 Ordinary Least Squares

To estimate coefficients of the linear regression model, the Ordinary Least Square (OLS) method is implemented. The OLS methods seeks to minimise the the sum of squared residuals. This method therefore provides us the following estimates for alpha and beta:

$$\hat{\beta} = Cov(r_i - r_f, r_I - r_f)/var(r_I - r_f) = Cov(r, r_I)/\sigma_I^2$$

$$\widehat{\alpha} = r_I - r_f - \widehat{\beta}(r - r_f)$$

where the "hat" implies that the calculated value is an estimate.

To compute the OLS regression the above formula can be expressed in matrix form.

Let y be a vector containing all returns of the form r_i - r_{fi}, let X be a $2 \times n$ matrix with a column of ones and a column of r_I - r_{fi}, let θ be a column vector containing α and β and let ϵ be a column vector with two white noise terms. Thus the regression in matrix form is:

$$y = X\theta + \epsilon.$$

This provides us with an OLS estimate of:

$$\hat{\theta_{ols}} = (X^\top X)^{-1} X^\top y,$$

where $X^\top$ denotes the transpose of X. The estimate found above is the best linear unbiased estimator according to Markov's theorem if certain assumptions cannot be rejected Nghiem (2015). This method is favourable as it is simple in concept as

well as computation. However it has a disadvantage in that the method does not consider the distribution of the variables when the coefficients are estimated. This results in an assumption of the distribution being required in order to evaluate the quality of the estimations Nghiem (2015).

Ledoit and Wolf (2003) show the method of calculating the covariance matrix from the estimated β values:

$$\hat{\Sigma} = \sigma_{00}^2 \beta\beta^\top + \Delta,$$

where σ_{00}^2 is the estimated variance of market returns and Δ is a diagonal matrix containing the residual variance estimates.

This method of estimation is biased, but well structured and stable, because of the model imposed on returns.

3.2.4 Maximum Likelihood method

Maximum Likelihood estimation uses the likelihood function in conjunction with a statistical model to estimate the models parameters. Given n samples $(x_1, x_2, .., x_n)$, the likelihood function is given by:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f(x_i|\theta),$$

where $f(x|\theta)$ is probability density at x given that θ is the parameter that describes the distribution (e.e. the mean and variance of a normal distribution). This implies that a distribution must be imposed on the model in question in order to find this estimator.
Maximising this therefore gives the most likely value of θ, assuming we have drawn from a distribution of type f.

In the context of CAPM, which we will use for estimating θ as defined in the OLS section, we assume that errors are normally distributed with mean 0, and some variance σ^2. The approach for maximum likelihood estimation under these condi-

tions is given by Nghiem (2015).

$$\mathcal{L}(\theta, \sigma) = \prod_{i=1}^{n} N(y_t|x_t, \theta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \times \exp\left\{\frac{-1}{2\sigma^2}(y - X\theta)^\top (y - X\theta)\right\}.$$

This can be maximised in order to obtain an estimator of:

$$\widehat{\theta}_{mle} = (X^\top X)^{-1} X^\top y.$$

This is the same as the Ordinary Least Squares estimate under these conditions. We therefore only consider the OLS method in this paper.

3.2.5 Shrinkage estimator

Ledoit and Wolf (2003) proposed a linear combination of the sample covariance matrix and a covariance matrix structured by a model. The aim of this style of estimator is to add the structure of a model estimated covariance to the accuracy of sample covariance. The resulting estimator is more accurate than either, with a better structure than the sample covariance. In the context of CAPM, we use a weight called the shrinkage intensity, or α between 0 and 1, resulting in an estimator in the form:

$$\widehat{\Sigma}_{Shrink} = \alpha \times \widehat{\Sigma}_{CAPM} + (1 - \alpha) \times \widehat{\Sigma}_{Sample}.$$

Using the CAPM model in this context is beneficial by being a single factor model, which allows consensus when estimating the optimal α. The optimal alpha can be found through the quadratic loss function derived by considering Frobenius norm of the difference between the shrinkage estimator and the true covariance matrix (Ledoit and Wolf, 2003):

$$L(\alpha) = ||\alpha \times F + (1 - \alpha) \times S - \Sigma||^2,$$

where $F = \Sigma_{CAPM}$ and $S = \Sigma_{Sample}$ The resulting risk function can be minimised to give the optimal shrinkage intensity, α^* by:

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

where f_{ij} is the element of Σ_{CAPM} in the i^{th} row and j^{th} coloum, s_{ij} is the element of Σ_{Sample} in the i^{th} row and j^{th} coloum. This can be represented in the form:

$$\alpha^* = \frac{1}{T}\frac{\pi - \rho}{\gamma} + O\left(\frac{1}{T^2}\right)$$

where

$$\begin{aligned}
&\pi = \textstyle\sum_{i=1}^{N}\sum_{j=1}^{N} AsyVar(\sqrt{T}s_{ij});\\
&\rho = \textstyle\sum_{i=1}^{N}\sum_{j=1}^{N} AsyVar(\sqrt{T}f_{ij});\\
&\gamma = \textstyle\sum_{i=1}^{N}\sum_{j=1}^{N}(\phi_{ij} - \sigma_{ij})^2;\\
&T = \text{"samples per stock"}.
\end{aligned}$$

However true α^* depends on the true distribution of the variables, and is therefore unobservable. The estimation of the shrinkage intensity is also given by Ledoit and Wolf (2003):

$$\widehat{\alpha^*} = \frac{(p - r)c}{T},$$

where

$$p = \sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij}$$

$$p_{ij} = \frac{1}{T}\sum_{t=1}^{T}[(x_{it} - m_i)(x_{jt} - m_j) - s_{ij}]^2.$$

Here, x_{it} is the t^{th} sample of stock i and m_i is the mean of stock i. Further

$$r = \sum_{i=1}^{N}\sum_{j=1}^{N} r_{ij}.$$

Here we have for $i \neq j$:

$$r_{ij} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{s_{j0}s_{00}(x_{it} - m_i) + s_{i0}s_{00}(x_{jt} - m_j) - s_{i0}s_{j0}(x_{0t} - m_0)}{s_{00}^2}\times\right.$$
$$\left.(x_{0t} - m_0)(x_{it} - m_i)(x_{jt} - m_j) - f_{ij}s_{ij}\right),$$

where m_0 is the sample mean of market returns and s_{i0} is the sample covariance of stock i's returns with the market.

Finally, again for $i \neq j$ we have

$$r_{ii} = p_{ii},$$

$$c = \sum_{i=1}^{N} \sum_{j=1}^{N} c_{ij},$$

$$c_{ij} = (f_{ij} - s_{ij})^2.$$

This closes this section and we can now continue with presenting our concrete approach to the previously presented problem.

4 Methodology

In this section we present our approach to the question of how to quantify a reasonable entropy budget. We start by shortly describing the data we used for our numerical experiments.

4.1 Data description

The data used in this project, which has been sourced from EOD Data, focuses on stocks that are listed on the NASDAQ Stock Exchange. Our period of interest begins on the 1st of January 2010 and ends on the 17th of December 2015. Due to exchange-related activity such as listing and de-listing, the number of stocks registered on the NASDAQ changes over this window of interest. As a result, we have only considered stocks that were listed on the first day of our investigation window (1 January 2010), till the last day (17 December 2015). This decision allows us to calculate covariances on every day since there are daily returns for each day.

There are around 1300 stocks that have been listed throughout the entire window of interest. These are far too many stocks, however, for us to compute an invertible covariance matrix. To address this problem, we selected the top 38 stocks that are contained within both our data set and the NASDAQ-100 index, ranked by alphabetical order.

We use the NASDAQ Composite Index as the proxy for our market portfolio. We feel it contains a more diverse group of NASDAQ-listed stocks than the NASDAQ-100 for example; a characteristic that the market portfolio must have. Moreover, the Composite Index excludes instruments that are not listed on the NASDAQ, unlike the S&P500 Index. Though the NASDAQ Composite Index values have been dividend adjusted, we have not managed to adjust our stock data for dividends. We did not deem it necessary to make such adjustments because dividends are released relatively infrequently compared to the size of our data and most stocks do not share cum dividend dates. Since the main focus of this paper involves estimating covariances, our estimate will therefore not suffer from a damaging level of correlation arising from dividend payouts. Some data was missing from EOD Data and so to reconcile the Index's data and the EOD Data we removed excess data points relating to the index.

Finally, the Dollar-denominated LIBOR was chosen as risk-free rate. This data was downloaded from

`https://fred.stlouisfed.org/categories/33003/downloaddata`.

4.2 Relative entropy budget using time series

For any given estimator of the covariance matrix, we find an estimate, called $\widehat{\Sigma_{\phi_i}}$ using some time period ϕ_i. We then wish to see how other estimates of the covariance matrix may be different from this estimate, due to estimation risk. We also wish to asses the effect of this difference on our optimised portfolio. We may estimate, for a different time period τ_i, a second estimate of the covariance matrix. We call this $\widehat{\Sigma_{\tau_i}}$.

We may then calculate η given that we have a notion of stationarity of the covariance matrix, using the two estimates of $\widehat{\Sigma_{\phi_i}}$ and $\widehat{\Sigma_{\tau_i}}$, as proposed by Glasserman and Xu (2014). We call this relative entropy η_i, which is calculated below:

$$\eta_i = \frac{1}{2}\left(\log\left(\det(\widehat{\Sigma_{\phi_i}}\widehat{\Sigma_{\tau_i}}^{-1})\right) + tr(\widehat{\Sigma_{\phi_i}}^{-1}\widehat{\Sigma_{\tau_i}} - I)\right).$$

This therefore gives the smallest possible η that would have accounted for the deviation of $\widehat{\Sigma_{\tau_i}}$ from $\widehat{\Sigma_{\phi_i}}$. In this way, we may construct a series of $\{\eta_1, \eta_2, ..., \eta_n\}$, corresponding to the time periods $\{\tau_1, \tau_2, ..., \tau_n\}$ and $\{\phi_1, \phi_2, ..., \phi_n\}$. In order to then find the most appropriate η to use for portfolio optimisation, we take the η corresponding to the α quartile. This η would have made our optimisation robust against α deviation away from $\widehat{\Sigma_{\tau_i}}$. Now that we have chosen an η, we can apply it to the robust Mean-Variance portfolio optimisation suggested by Glasserman and Xu (2014).

4.2.1 Time window and frequency of data

In order to compute the estimates of the covariance matrix, we must first choose our data. We my consider choosing a lot of data, which would give us a stable estimate of the covariance matrix. However, this covariance matrix may no longer represent the true distribution of returns, as it includes too much past data. The

stationarty of the data therefore must be balanced with the estimation error that results from having too little data. We must also consider frequency of data that should be used. Bartholdy and Peare (2005) do an empirical study on both of these parameters in relation to estimating β's for the CAPM model. It was found that there was a significant differences in estimation for both of time window and frequency of data. They find that using a higher frequency of data (i.e. daily returns rather than monthly returns) introduces noise into the data, which reduces the efficiency of the estimates (Bartholdy and Peare, 2005). Due to limited data available in this study, we however only consider daily returns in order to allow for sufficient data for stable parameter estimation. We do however consider time window by comparing η trends for different time windows. In other words, we calculate the relative entropy budget using a time series for different lengths of time window. We may then compare the results to see how this parameter influences η estimation.

4.2.2 Gap between time periods τ_i and ϕ_i

As we are estimating the covariance matrix for two different time windows, the effect of gap between the time windows on the estimated relative entropy needs to be considered. Given an unlimited amount of data, we would discount the possibility of using overlapping time windows, as sharing data would result in a estimated relative entropy that does not account for all types of estimation risk. We consider the case that we use two adjacent time windows. This would minimise the risk due to non-stationarity of returns, while still fully incorporating the other forms of estimation risk. However, as we are estimating an η that aims to make our portfolio robust against all forms estimation risk, which includes non-stationarity, therefore we suggest this gap should correspond to the amount of data used to estimate the covariance matrix that will be used for portfolio optimisation. In other words, the gap between the first day of τ_i and the first day of ϕ_i should be proportional to the amount of days included in the estimation of the covariance matrix used for portfolio optimisation. However, due to lack of data, we will consider overlapping time periods in our estimation of η.

4.2.3 Estimated η

The resulting estimated η requires interpretation, as it is not the true value that would come from knowing the true distribution of returns. We are in essence, calculating the distance between two estimated covariance matrices in the context of Mean-Variance portfolio optimisation. The resulting η can therefore be thought of as the budget that, assuming the first estimated covariance matrix is exact, would have made the portfolio optimisation robust against a case where the second estimated covariance matrix is the real covariance matrix describing returns. Therefore, the calculated η describes a compounded estimation error from the two covariance matrices. This compounded error could cause the matrices to appear further apart than they are, resulting in a higher η then would be correct, or they could appear closer than they really are, resulting in a lower η then would be correct. We want our η to reflect a worst case budget for estimation error, therefore, we take the $\alpha = 95\%$ quantile to account for 95% of the difference between our two estimated covariance matrices. We are therefore drawing from the η's that are calculated from covariance matrices that appear further apart then they are (the estimation error has been amplified). Our estimated η therefore may be, in some sense, too large. However, since we are attempting to account for the worst case estimation error, we argue that a cautiously large η may be the correct approach. We however recommend that further studies take this amplified error into account.

4.3 Simulation approach

The approach begins with calculating an estimate for the covariance matrix, using any given estimation method, called $\widehat{\Sigma_{\phi_1}}$. This estimate is calculated over a fixed initial time period ϕ_1. This estimate will be used to determine how future estimates of the covariance matrix may differ as a result of parameter risk which is an element of estimation risk.

Based on the initial estimate $\widehat{\Sigma_{\phi_1}}$ we compute a large number $(n = 1000)$ of Monte Carlo simulations. In each case we assume the returns r_i are multivariate-normally distributed with known mean μ and variance-covariance matrix $\widehat{\Sigma_{\phi_1}}$. We simulate the same number of return observations as the length of observations in the initial time period. From these observations we determine an estimated covariance matrix $\widehat{\Sigma_{\theta_1}}$ using the same method as the original $\widehat{\Sigma_{\phi_1}}$ is estimated from.

The time window is then shifted by a specified time step length. The method then

stores the values of all previously simulated returns that also belong to the new shifted time window. To ensure an equal number of observations are contained in both time windows we simulate a number of new returns equal to the length of the time step per simulation. Based on this new set of observations, with the same dimensions as the original period, we calculate an estimate of $\widehat{\Sigma_{\theta_{2i}}}$. This covariance matrix is then used in conjunction with the covariance matrix $\widehat{\Sigma_{\theta_1}}$ to calculate a measure of relative entropy η_{1i} per each simulation.
Therefore once the simulation procedure is complete we will have n values for η from which we are able to construct a confidence interval. This confidence interval is found by determining the 95^{th} quantile of the η values.
The procedure then repeats itself shifting again the specified time step. The values of the second simulation which are to be retained are determined based on the overlapping time periods between the second and the third window and a further number of observations are simulated to ensure an equal number of observations as the initial period. The new covariance matrix $\widehat{\Sigma_{\theta_{3i}}}$ is calculated with $\widehat{\Sigma_{\theta_1}}$ and thus the relative entropy η_{2i} is calculated along with the 95^{th} quantile of the η values.
This method continues to be implemented until the final simulated observation occurs at the time point which is equivalent to the final time point as the data. When the quantiles of the η values are plotted it is expected that their values will display an increasing trend as they will share less information. At the point where they no longer share any information with the original time window it is expected that a straight line will be observed.
The simulation is able to tested at a number of differing time step lengths to determine the relative η $95th$ quantiles. This approach thus enables us to observe the portion of estimation risk that can be assigned directly to parameter risk that the investor is exposed to.

4.4 Backtesting

In order to test optimal portfolios, we require consistent metrics associated with portfolio performance and estimation risk. We therefore consider the Sharpe ratio, which measures the risk adjusted return of the portfolio. This ratio contextualises how well this portfolio performed in terms of both the important factors in Mean-Variance portfolio optimisation. We also consider portfolio variance. This provides a measure of the risk of the portfolio.

4.4.1 Portfolio calculation

Describe the data in the following form: $\{t_0, ..., t_{est}, .., t_{rb_1}, .., t_{rb_2}, ..., t_{end}\}$ where t_{est} is the time we are estimating the distribution of returns for, and t_{rb_i} is the i^{th} portfolio rebalance. We may use the numerical methods in conjunction with the robust portfolio optimisation optimisation, or a standard naive portfolio optimisation to choose a portfolio a using a portion of our data up to t_{est}. We then hold this portfolio for the period $\tau = \{t_{est}, ..., t_{rb_1}\}$. We then rebalance using the same amount of data as with our first portfolio, but with the data going up to t_{rb_1}. We continue in this way for all the rebalancing times.

4.4.2 Sharpe Ratio

Given some portfolio a calculated at t_{est}, we can see it's realised daily returns over the period $\{t_{est}, ..., t_{rb1}\}$ by multiplying the realised daily returns of the stocks by the weightings of the stocks in our portfolio and adding them together:

$$R_{portfolio} = a \times R_{stocks}.$$

We can then calculate the excess returns by:

$$R_{excess} = R_{portfolio} - R_{riskfree}.$$

We may then calculate the Sharpe ratio of the portfolio over any period by:

$$S = \frac{\overline{R_{excess}}}{\sigma_{portfolio}}$$

where $\overline{R_{excess}}$ is the average excess portfolio return and $\sigma_{portfolio}$ is the portfolio standard deviation.

4.4.3 Variance of returns

Given an optimal portfolio, we can calculate the realised portfolio returns, which in turn allows us to calculate the realised portfolio variance using a standard variance calculation. We can also calculate the theoretical portfolio variance that we would expect if the estimated μ and Σ were the true parameters of the returns by:

$$\sigma^2_{portfolio\ theo} = \quad a^\top \widehat{\Sigma_{returns}}\ a$$

$$\mu_{portfolio\ theo} = \quad a^\top \widehat{\mu_{returns}}$$

5 Results

5.1 Simulated versus estimated entropy budget time window comparison

The covariance matrix for a given length of time windows of 30, 70, 100, and 150 days is estimated. This is then used to estimate an entropy budget using the simulation method, and the time series method for a given shift between the original time window and the compared time window. We consider two styles of covariance estimation in sample covariance and CAPM covariance. This result of entropy budget for the two methods for a given shift is plotted in the figures below. The blue line represents the estimated entropy budget, and the red line represents the simulated entropy budget:

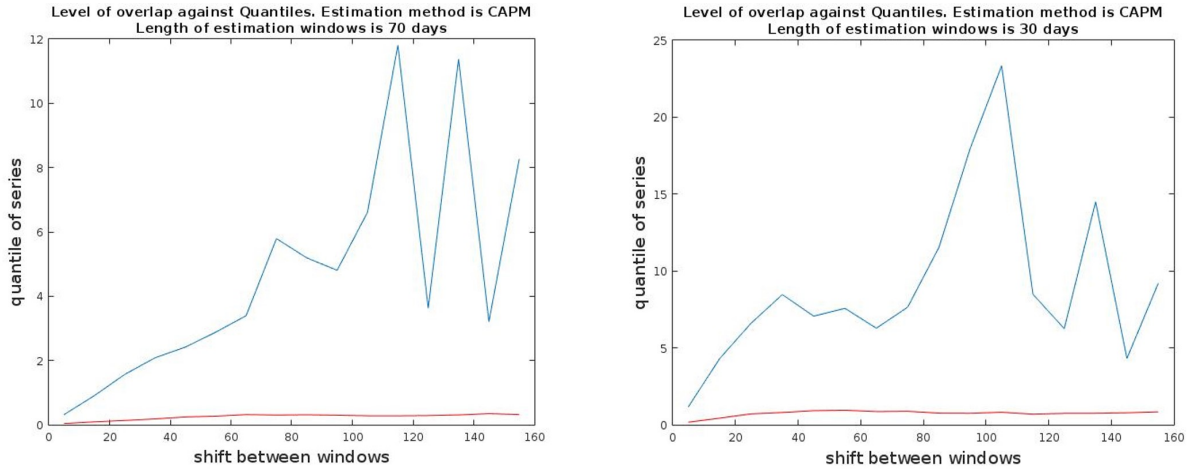


Figure 1: Simulated entropy budget vs estimated entropy budget for a time window of 30 and 70 days for CAPM.

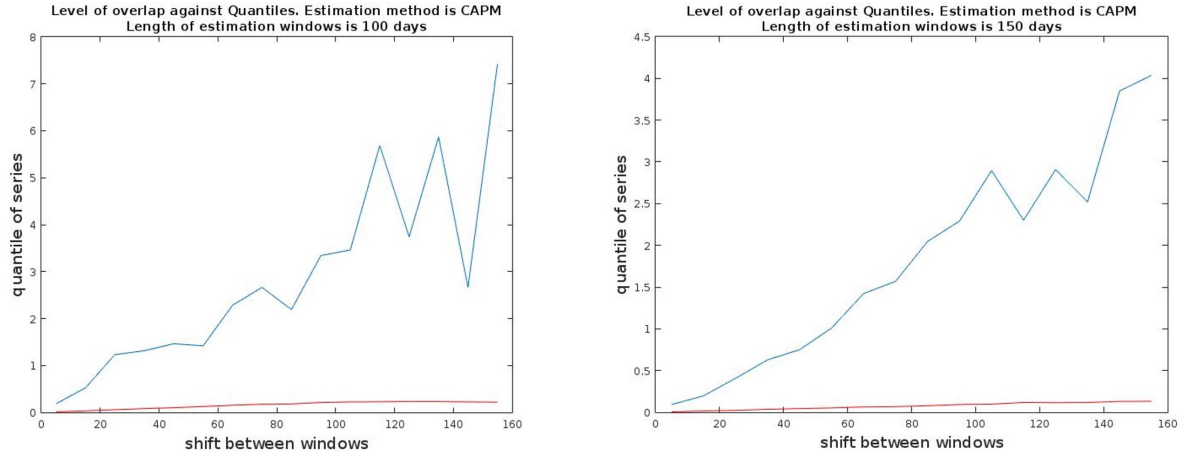


Figure 2: Simulated entropy budget vs estimated entropy budget for a time window of 100 and 150 days for CAPM.

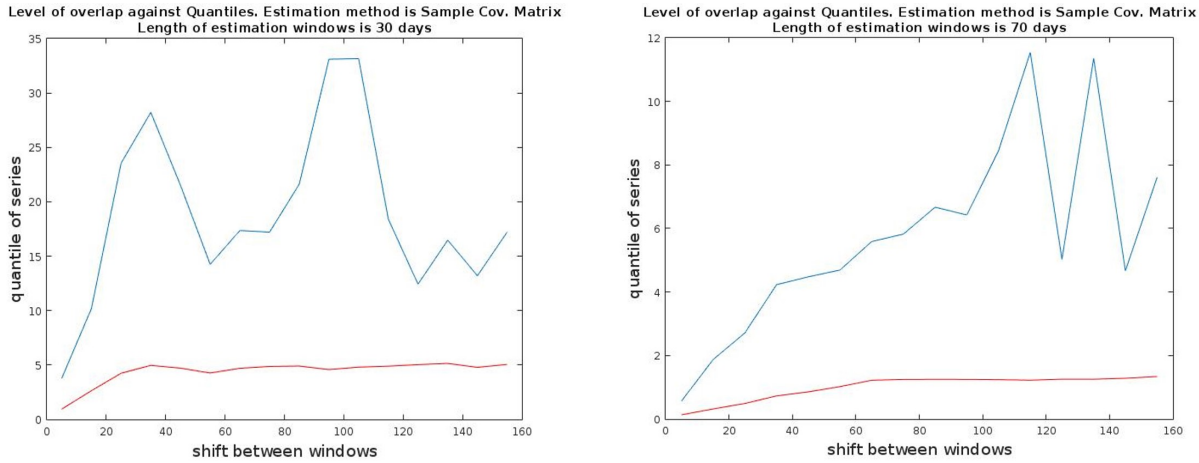


Figure 3: Simulated entropy budget vs estimated entropy budget for a time window of 30 and 70 days for the sample covariance.

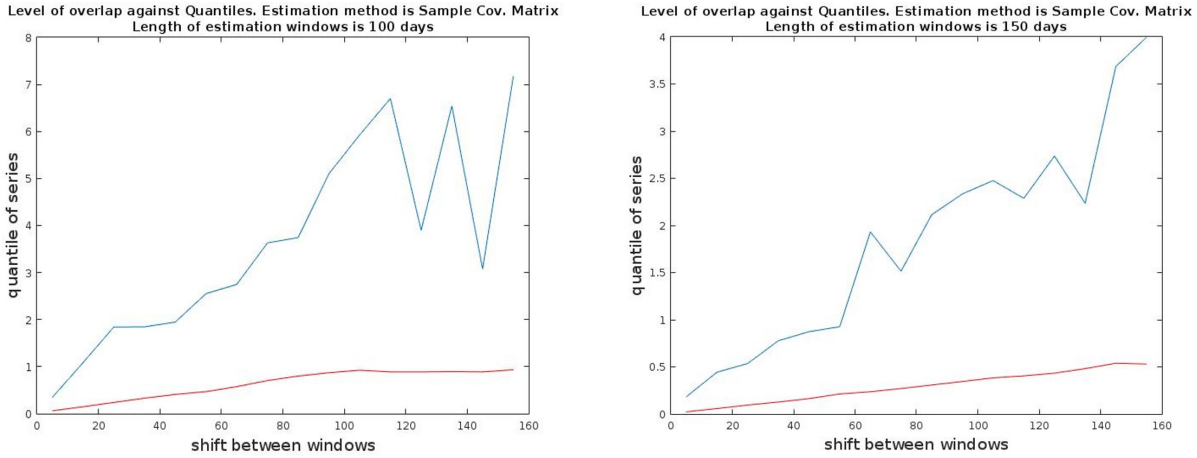


Figure 4: Simulated entropy budget vs estimated entropy budget for a time window of 100 and 150 days for the sample covariance.

5.1.1 Time window trends and general observations

The first significant trend is the times series budgets are always higher than the simulated budgets. It is also apparent that relative entropy budget decreases as the overall time window increases. Thirdly, we notice an overall increasing trend. The two covariance matrices used to estimate the entropy budget are separated when the length of the shift exceeds the length of time window. The estimated budget increases fairly linearly for the period where the two covariance estimates overlap for both the simulated and time series approaches. When the covariance matrices are separate, the simulated η's become flat. Figure 1 and figure 3 display this for both the CAPM and sample estimate cases. These figures also display high oscillation for the time series η after the covariance matrices separate.

5.1.2 Trends between sample and CAPM estimation

We can observe a significantly higher simulated budget for the sample estimation method compared to the CAPM estimation method. The time series estimation again shows a higher η for the sample estimation compared to CAPM estimation.

5.2 Backtesting Naive vs Robust

5.2.1 Realised Sharpe ratio

The Sharpe ratios for the Naive and Robust optimal portfolios were calculated.

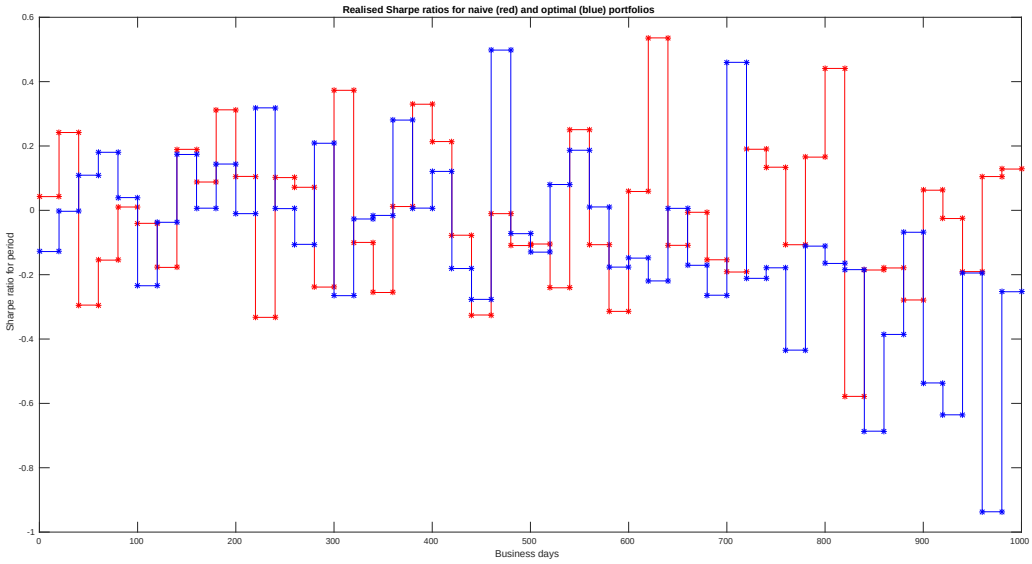


Figure 5: Naive Sharpe ratio vs Robust Sharpe ratio.

There are no observable differences between the Sharpe ratios of the Robust and Naive portfolios. Both center around 0, with a slight bias for negative values.

5.2.2 Realised returns

The realised returns for both the Naive and Robust portfolios are plotted, along with 95% confidence intervals for the theoretical variance.

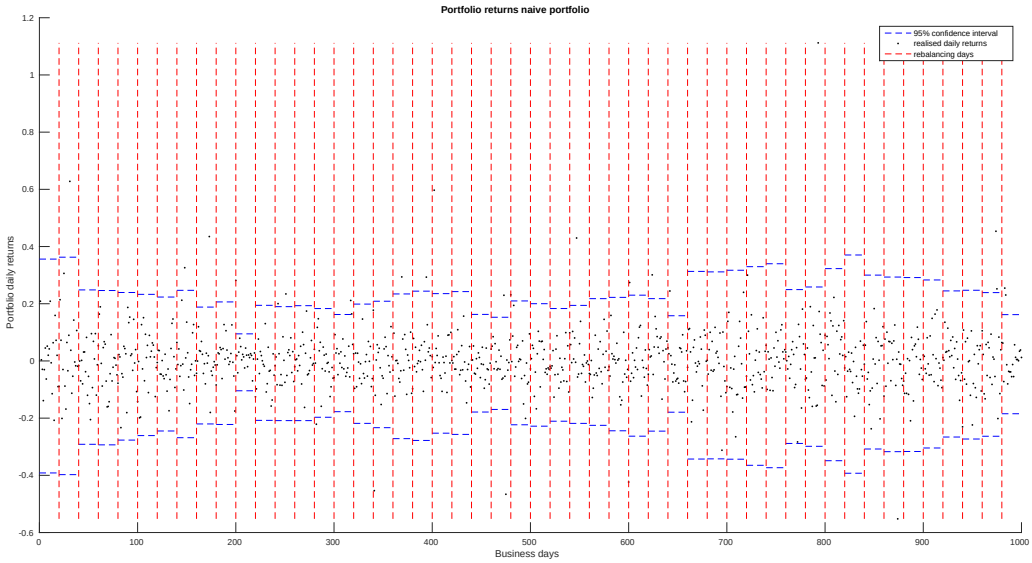


Figure 6: Naive portfolio returns plotted with the theoretical confidence intervals.

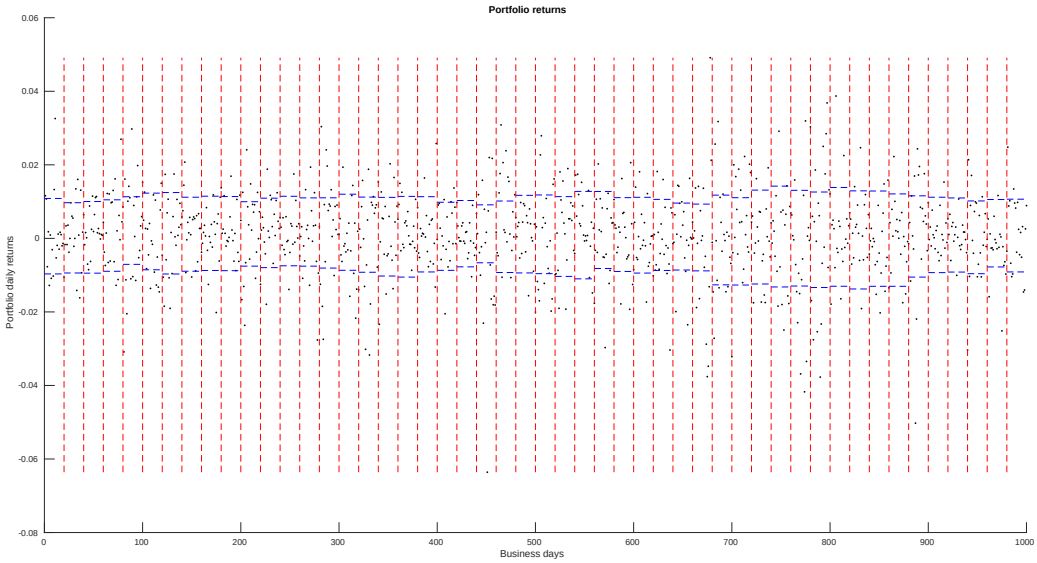


Figure 7: Robust portfolio returns plotted with the theoretical confidence intervals.

The first noticeable trend is the Naive approach results in returns that lie mostly between -0.2 and 0.2, whereas the Robust approach lies between -0.02 and 0.02. The Naive case returns are mostly contained inside the theoretical confidence intervals, whereas the Robust returns have a fair amount of outliers. We can also observe a much more stable confidence interval for the Robust portfolio.

5.2.3 Portfolio variances

The theoretical portfolio variance calculated with the estimated covariance matrix is compared to the realised portfolio variance.

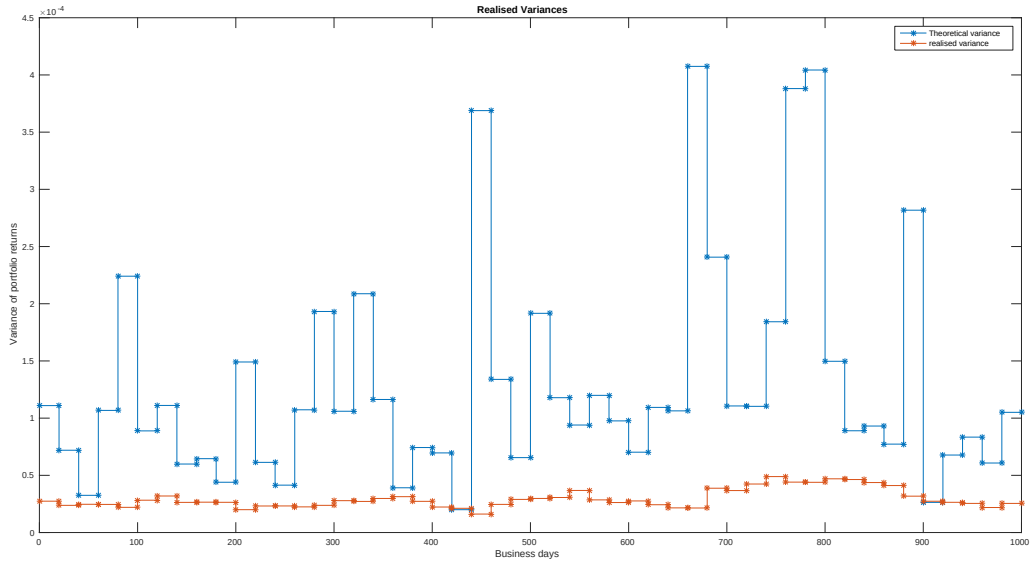


Figure 8: Realised Robust portfolio variance against theoretical portfolio variance.

The realised variance is always higher than the theoretical portfolio variance. In addition, the realised variance is highly oscillatory, whereas the theoretical variance remains fairly smooth.

6 Discussion

6.1 Quantifying random sampling error

As we fix our estimated covariance matrix and mean and use this to generate our returns, the only type of error experienced by the simulation of η's via Monte Carlo, is the error due to random sample error. The simulated η's are well below the estimated η's using the time series approach, which is to be expected, as the simulated η's only account for sample error, and the time series η's accounts for all sources of estimation risk. The required budget to account for all forms of estimation risk is therefore much higher than the budget required to account for only sampling risk. This further enforces the need for a portfolio optimisation that is robust to all forms of estimation risk.

6.2 Choice of estimator

Since the relative entropy budget for both the simulated and time series estimation methods are smaller for CAPM, we may safely conclude that the CAPM produces a more stable covariance matrix in the context of Mean-Variance portfolio optimisation. We therefore choose CAPM as our choice of estimator.

6.3 Choice of time window

Given the results of estimated η's in relation to time window, it can be concluded that a larger time window reduces the sampling risk. This is in line with expectations, as a longer time window increases the amount of information available for estimation. The optimisation procedure does not converge for high values of η, which leads us to choose budgets lower than 1. This therefore restricts us to the larger two time windows considered. In addition, our aim is to reduce estimation risk in Mean-Variance portfolio optimisation, so this choice of window aligns with our goals. We cannot choose the largest window however, as data constraints would not allow for a long enough series of η's to choose a stable 95% quantile. In addition, the largest window would have a large overlap for short re-balancing frequencies, which would miss-represent the amount of estimation risk present. We therefore choose 100 days as our time window.

6.4 Choice of re-balancing frequency

It is argued that the re-balancing frequency of the portfolio should correspond to shift between time windows, as this would give an indication as to the change in covariance matrices. Again, due to optimisation constraints, we can only choose entropy budgets smaller than 1. In conjunction with our choice for 100 days as the time window, we therefore choose a re-balancing period of 20 days, which corresponds to an η around 1.

6.5 Choice of entropy budget

We may therefore choose an entropy budget corresponding to a time window length of 100, a re-balancing period of 20, the CAPM estimator, and using the time series approach. Given our choice of entropy budget, we may now assess our robust portfolio using back testing.

6.6 Naive versus Robust

6.6.1 Implication of Sharpe ratio

There is no clear difference between the Sharpe ratios of the Robust portfolio and the Naive portfolio. Although we my not conclude that the Robust portfolio performed better than the Naive portfolio, we may conclude that there is no significant difference in risk weighted performance. However, as the Robust approach aims to protect against the worst case estimation error, rather than to optimise Sharpe ratio, this is not a negative result. We can conclude that making our portfolio Robust has not lead to reduced general performance while still possibly protecting against worse case errors.

6.6.2 Implication of Realised returns

The realised variance of the Naive portfolio is much higher than the Robust portfolio. This is clear through the range of returns for the Naive portfolio being 10 times larger than the range of the Robust portfolio. This speaks to a higher level of risk in the Naive portfolio. The large number of outliers for the Robust returns speaks to the realised variance being larger than the theoretical variance. This could also be a result of heavy tails, which is a common realisation with returns.

6.6.3 Implication of Realised variance

This again speaks to the possibility of heavy returns, or underestimated portfolio variance.

7 Further Work

7.1 Effect on estimation of mean on Mean-Variance portfolio optimisation

In order to consider making the Mean-Variance portfolio optimisation robust against estimation risk of the mean, a optimisation routine connecting relative entropy to a varying mean is needed. This is necessary since this is the other main parameter in Mean-Variance framework.

7.2 Confidence interval approach

An ideal method of estimating η would be to link it to the concept of confidence intervals associated with estimated covariance matrices. As confidence intervals of an estimator are a measure of estimation risk, a direct link between this and required model robustness could lead to an elegant choice of η.

Literature associated with the confidence of a covariance matrix estimator is sparse. A brief summary can be found below:

7.2.1 Sample covariance estimate confidence interval

Sellentin and Heavens (2015) suggest describing the sample estimate of Σ as a Wishart matrix in order to find the distribution of the estimate. If the true values of μ and Σ describing a multi-normal random variable, X, are unknown, we can estimate them using a sample, X_0, drawn from the distribution in question. We can then approximate the the distribution as $X \sim \mathcal{N}(\hat{\mu}, \hat{\Sigma})$. Using these estimated values, $P(X_0|\hat{\mu}, \hat{\Sigma}, N)$ is a likelihood function with N being the number of samples in X_0. The estimated covariance matrix is a Wishart matrix, because it is of the type $M = \sum_{i=1}^{m} Y_i Y_i^\top$, and therefore has a Wishart distribution in the form:

$$\mathcal{W}(\hat{\Sigma} \mid \Sigma/n, n) = \frac{|\hat{\Sigma}|^{\frac{n-p-1}{2}} \exp[-\frac{1}{2} n \, Tr(\Sigma^{-1}\hat{\Sigma})]}{2^{\frac{pn}{2}} |\Sigma/n|^{\frac{n}{2}} \Gamma_p(\frac{n}{2})}$$

where $n = N - 1$ is the degrees of freedom, and p is the dimension of the matrix. After a series of further calculations, given that the mean is μ, and the estimated sample covariance matrix is $\hat{\Sigma}$, the likelihood of a p-dimensional data set X_0 with N samples is:

$$P(X_0 \mid \mu, \hat{\Sigma}, N) = \frac{\overline{c}_p |\hat{\Sigma}|^{-1/2}}{\left[1 + \frac{(X_0-\mu)^\top \hat{\Sigma}^{-1}(X_0-\mu)}{N-1}\right]^{\frac{N}{2}}}$$

with

$$\overline{c}_p = \frac{\Gamma\left(\frac{N}{2}\right)}{[\pi(N-1)]^{p/2}\Gamma\left(\frac{N-p}{2}\right)}.$$

We now know the likelihood that our estimate $\hat{\Sigma}$ describes our data. This is not strictly a confidence, but it does give some indication of how reasonable our estimate is. If there is a method that describes the real confidence interval of a sample covariance matrix, it is unknown.

7.2.2 CAPM covariance estimate confidence interval

No mention of confidence interval of covariance estimate in the CAPM context was found. However, it is possible, as with the sample estimator, to describe the CAPM estimator with a Wishart matrix. We can therefore calculate the likelihood that the CAPM estimator produced the sample data X_0.

7.2.3 Shrinkage covariance estimate confidence interval

No mention of a confidence interval was made by Ledoit and Wolf (2003). It is possible that the confidence interval could be approximated by linear combination of the CAPM confidence interval and the sample confidence interval using the estimated optimal shrinkage intensity.

7.2.4 Conclusion of confidence interval approach

In order to implement this method calculation of confidence interval of a covariance matrix is required. We also require a link between confidence of the estimator and η. We leave this for future studies.

7.3 Extension to other market models

Fama French requires the estimation of several parameters, which requires large amounts of data, and introduces estimation risk into each parameter. It would therefore be interesting to test this methodology in conjunction with this model. Other commonly used model could also be considered and compared.

7.4 Analysis of worst case estimation error data

By considering data where estimation risk is highly prevalent, we are able to judge the robust portfolio's design. We use robust optimisation in order to protect against the worst case of estimation error, rather than to optimise our portfolio in all cases. Therefore, it makes sense to compare the Robust approach to the Naive using a sub-sample, that only considers cases where the realised portfolio variance is much higher than the theoretical portfolio variance, which indicates the presence of high estimation error.

8 Conclusion

We have given a proof of concept for a robust version of Mean-Variance portfolio optimisation using the relative entropy approach by Glasserman and Xu (2014) as an alternative to naive portfolio optimisation where estimation errors and violations in model-assumptions are mostly ignored. We have developed a methodology to link a multivariate point estimation, namely a high dimensional covariance matrix of asset returns to an uncertainty level described by a single number, therefore unifying the various multivariate uncertainty sources to a common concept. We have further investigated the performance of portfolios chosen by the new robust optimality criterion against naively chosen ones.

As we have implemented this approach, we have faced obstacles regarding conditioning properties such as sensitivity of η regarding θ in the interval of interest or the fact that depending on size and quality of observation data, covariance estimations might become challenging. We have linked these limitations to a choice of re-balancing time periods and the corresponding empirical estimation of relative entropy between neighboring time windows of covariance matrix estimation for re-balancing.

In addition we developed a Monte-Carlo simulation approach to estimate the portion of relative entropy between consecutive covariance estimates[1] coming purely from parameter estimation uncertainty. Finally, we have given a short discussion of future work of possible extensions to the approach presented in this report.

[1] For all our estimates, we use a normal approximation.

Bibliography

Bartholdy, J., Peare, P., 2005. Estimation of expected return: Capm vs. fama and french. International Review of Financial Analysis 14 (4), 407–427.

Bodie, Z., et al., 2014. Investments. Tata McGraw-Hill Education.

Glasserman, P., Xu, X., 2014. Robust risk measurement and model risk. Quantitative Finance 14 (1), 29–58.

Lai, T. L., Xing, H., Chen, Z., 2011. Mean-variance portfolio optimization when means and covariances are unknown. The Annals of Applied Statistics, 798–823.

Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of empirical finance 10 (5), 603–621.

Markowitz, H., 1959. Portfolio Selection, Efficent Diversification of Investments. J. Wiley.

Michaud, R. O., 1989. The markowitz optimization enigma: Is optimizedoptimal? Financial Analysts Journal 45 (1), 31–42.

Morin, R. A., 1994. Regulatory finance. Public Utilities Report, Arlington, VA.

Nghiem, L., 2015. Risk-return relationship: An empirical study of different statistical methods for estimating the capital asset pricing models (capm) and the fama-french model for large cap stocks. arXiv preprint arXiv:1511.07101.

Schlögl, E., 2016. Toward quantifying model risk, presentation.

Sellentin, E., Heavens, A. F., 2015. Parameter inference with estimated covariance matrices. Monthly Notices of the Royal Astronomical Society: Letters 456 (1), L132–L136.

9 Appendix

9.1 Related Work

9.1.1 Estimation risk minimisation via better estimators

As there are better estimators of the covariance than the sample covariance, these can be used in conjunction with the standard mean-variance optimisation process. The estimation risk associated with the optimisation is lowered because the parameters used in the optimisation are better. This will be considered in this paper.

9.1.2 Estimation risk minimisation via bootstrap

Michaud (1989) attempts to adjust for bias in the estimated optimal asset weights $\hat{a}$ through averaging bootstrap vectors:

$$\overline{a} = \frac{1}{N} \sum_{i=1}^{N} \hat{a}_i^*$$

where $\hat{a}_i^*$ is the estimated optimal asset weights calculated using the i^{th} bootstrap sample. The bootstrap samples are drawn with replacement from the full sample. This approach hopes to reduce the estimation error via reducing the effect of random sampling error. However, if the original sample is not representative of the true distribution, this approach does not reduce estimation error. In addition, if the sample is heavy tailed, this approach may not reduce sampling error, as the calculated value will oscillate rather than converging.

9.1.3 Estimation risk minimisation via prior distribution

Lai et al. (2011) adopt an approach that assumes a prior distribution for the mean and covariance matrix of returns, and formulates mean-variance portfolio optimisation as a stochastic optimisation problem. This approach aims to forecast the future distribution of returns given current returns, rather than assume stationarity of returns. The Markowitz optimisation for time $t+1$ under these conditions is described by the following:

$$\max[\mathbb{E}(a^\top X_{t+1}) - \lambda \, \mathbb{V}ar(a^\top X_{t+1})]$$

where X_{t+1} are the returns at $t+1$. This assumes μ and Σ to have Normal and Wishart priors respectively. This is solved as in the form of a maximisation problem:

$$\max_{\kappa}[\mathbb{E}(a(\kappa)^\top X_{t+1}) - \lambda\, \mathbb{V}ar(a(\kappa)^\top X_{t+1})]$$

with

$$\begin{aligned} a(\kappa) &= \tfrac{1}{C_t}\Sigma_t^{-1}1 + \tfrac{\kappa}{2\lambda}\Sigma_t^{-1}\big(\mu_t - \tfrac{A_t}{C_t}1\big) \\ A_n &= \mu_t^\top \Sigma_t^{-1}1 \\ B_n &= \mu_t^\top \Sigma_t^{-1}\mu_t \\ C_n &= 1^\top \Sigma_t^{-1}1 \end{aligned}$$

where μ_t is the estimated mean at t, Σ_t is the estimated covariance at t, and 1 is a vector of 1's.

It is also possible to combine this method with bootstrap sampling described above. This however does not account for the possibility that the true prior distribution is different from the estimated one. It may reduce the stationarity and random sampling error, but other forms of estimation error may still be unaccounted for.

Rough Volatility

Team 5

MESIAS ALFEUS, University of Technology, Sydney
FEBIN KORULA, University of Cape Town
MARCIO LOPES, University of Cape Town
ANDREW SOANE, University of Cape Town

Supervisor:
TOM MCWALTER, University of Cape Town

African Institute of Financial Markets and Risk Management

Contents

1 Introduction

As is well known, modern option pricing was initiated with the publications of Black and Scholes (1973) and Merton (1973). The asset price dynamics under the Black–Scholes framework is assumed to follow a geometric Brownian motion. The Black and Scholes model assumes that volatility is constant over time. Consequently, the model is not suitable for explaining certain phenomena observed in option price data, in particular, volatility smiles and skews[1]. To account for these features, alternative models have been proposed e.g. local volatility models; Dupire formula and stochastic volatility models; Heston model and models that incorporate jumps; Bates stochastic model with jumps. However, all these proposed models have shortcomings. For instance, local volatility models have unreasonable skew dynamics and they underestimate the volatility of volatility. Stochastic volatilty models (where volatility follows a stochastic process) e.g. the model of Heston (1993) tend to capture the long-dated skew but struggle to reproduce the near term skew. While increasingly popular, jump processes suffer too, due to the flattening of the volatility surface as time to expiry increases. For a basic introduction to volatility surface modelling we refer to the textbook by Gatheral (2006).

In volatility modelling paradigms there are two persistent stylised facts. The first one is the presence of long memory features in volatility. Long memory is exhibited in a stationary process if its covariance function decays slowly. The second stylised fact is the leverage effect: the existence of the negative correlation between price increments and volatility increments.

Empirical analysis of volatility shows that the log-volatility of historical data follows processes more like fractional processes. In their model, Comte and Renault (1998) were the first to apply a model driven by a fractional Brownian motion with Hurst parameter $H \in [\frac{1}{2}, 1)^{\dagger}$. This modelling framework is known as the Fractional Stochastic Volatility (FSV) model and is shown to reproduce the long memory property of the volatility process.

Counterintuitively, Gatheral et al. (2014) demonstrated that price and option data are more consistent with $H \in (0, \frac{1}{2})$ and with a mean reversion parameter for the volatility process kept very small, i.e. short memory. By virtue of this ob-

[1] Out of the money options are generally priced with a higher volatility.

[†] If W^H represents a fractional Brownian motion with Hurst parameter H, in this case increments of W^H satisfy a long dependence behaviour necessary for describing features such as long memory and persistence. When $H = \frac{1}{2}$, W^H is the standard Brownian motion.

servation, increments of W^H are negatively correlated which is a necessary feature for intermittency and anti-persistency; leverage effects are well-taken into account and the model gives a general interpretation of the volatility dynamics from high-frequency behaviour in the market. This modelling approach is known as the Rough Fractional Stochastic Volatility (RFSV) framework.

Even though fractional Brownian motion (fBM) seems to be an appropriate tool for modelling volatility (at least from empirical observations and statistical tests) there has been much debate about its applicability for modelling financial derivatives as fBM is not a semimartingle. For example, Rogers (1997) showed that fBM could not be used as a price process for a risky security without introducing arbitrage opportunities. As a result price processes driven by fBM do not satisfy the property of "No Free Lunch With Vanishing Risk" (NFLVR). However, Cheridito (2003) showed that fBM could still be used in a price process under certain conditions: proper restriction of the class of permissible trading strategies is necessary to eliminate arbitrage. Moreover, Jarrow et al. (2009) showed that the semimartingale property is not a necessary condition for no-arbitrage and were able to construct a class of processes which are not semimartingales but which remain arbitrage free.

The goal of this study is to investigate the claim that "Volatility is Rough" and implement the rough volatility framework of Gatheral et al. (2014). From time series data, this model is shown to account both for the short memory in volatility and the leverage effect using a fractional Brownian motion. The model provides computationally tractability, with modelling formulae for the future volatility and it is shown to perform better than classical predictors such as the Generalised Autoregressive Conditional Heteroskedasticity (GARCH) model. This area of research is still in its infancy in the financial mathematics community. For example, calibration is still an open problem (see Gatheral et al. (2014)). In this document, we introduce the area of research and then proceed to confirm some of the above recent findings.

We begin with estimation of the Hurst parameter H from time series of S&P and NASDAQ realised variance estimates. A short overview of the rough Bergomi (rBergomi) model is also considered. We then proceed to consider an example of pricing options via Fourier-based methodologies and Monte Carlo methods under the rough Heston model (or rHeston model). The pricing implementation is intensive and computationally challenging due to the non-Markovian nature of the fBM. We then compare our prices from the Fourier-based methodology e.g. Fourier

inversion, FFT and Lewis (2001) methods, to the Monte Carlo prices. We show the correctness of the implementation by verifying that the Fourier-based prices fall within the 95% Monte Carlo confidence bound.

This report is structured as follows. In Section 2.3 we estimate the Hurst parameter from realised variance data. Section 3 introduces the volatility modelling paradigms. We also discuss the extensions of the classical Heston model. We implement the Rough Heston model in Section 4. We proceed to show how to price using Monte Carlo and Fourier methods. To this end, we also show how to simulate fractional Brownian motion using two different approaches in Subsection 4.1. In Subsection 4.2 we introduce the characteristic function needed for the Fourier pricing approach. Numerical results are generated and compared in Section 4.3. We then provide a basic calibration of the model to option data on an index. We conclude the report in Section 6.

2 Fractional Brownian Motion

2.1 Definition

A fractional Brownian motion $(W_t^H)_{t\in\mathbb{R}}$ is a centred self-similar Gaussian process with stationary increments such that for all $t \in \mathbb{R}$,

$$\mathbb{E}[|W_{t+\Delta}^H - W_t^H|^q] = K_q \Delta^{qH},$$

where $\Delta \geq 0$ is a small time interval, K_q is the moment of order $q > 0$ of the absolute value of a standard Gaussian random variable and $H \in (0, 1)$ is the Hurst parameter which defines the fBM (Gatheral et al., 2014). The Hurst parameter and its estimation is given more attention in Section 2.3. A fractional Brownian motion (fBM) with Hurst parameter $H = \frac{1}{2}$ reduces to a standard Brownian motion (SBM).

While standard Brownian motion has independent increments, fBM displays autocorrelation, i.e. it does not exhibit independent increments. The covariance between an fBM process at times t and s is

$$\mathbb{E}[W_t^H W_s^H] = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}). \tag{1}$$

This covariance implies an autocorrelation function that decays slower at $H \neq \frac{1}{2}$. The autocorrelation function has the slowest decay when $\frac{1}{2} < H < 1$. This

implies that an fBM process will exhibit longer memory as H gets closer to 1. This gave rise to Comte and Renault (1998) using fBM to model volatility as a long memory process, hence their restriction of $H \in (\frac{1}{2}, 1)$. Recently, however, Gatheral et al. (2014) argues that volatility exhibits shorter memory, hence their restriction of $H \in (0, \frac{1}{2})$.

2.2 Simulating Fractional Brownian Motion

To simulate some of the models under consideration in Section 3, we need to be able to simulate fBM. We mention two methods used for exact simulation of fBM. These are methods which completely capture the covariance structure of fBM, as opposed to approximate methods which aim to reduce computation times by approximating the covariance structure.

The Cholesky Method

Let $\sigma(s,t)$ be the covariance function of a zero-mean function. Then $\sigma(s,t)$ is of the form (1) for fBM[2]. Then for a discretisation of time in N steps, $\Sigma(s,t)_{N\times N}$ defines the covariance matrix and a sample fBM path can be generated by multiplying a vector Z of *iid* standard normal variates by the Cholesky decomposition of $\Sigma(s,t)$. While very simple, the Cholesky method becomes slow (with $\mathcal{O}(N^3)$ for N points) and demanding in terms of storage (see Dieker (2004)).

The Hosking (1984) Method

This method is concerned with simulating fractional Gaussian noise (fGn). A sample fBM path can then be recovered by using a cumulative sum on the generated fGn sequence. The sequence $(X_n)_{n\in\mathbb{N}}$ of fractional Gaussian noise is computed recursively by computing the conditional distribution of X_{n+1} given $X_n, \ldots, X_0$. The required sample is found by generating a standard normal random variable X_0 and calculating the remaining X_{n+1} recursively. The conditional distribution is

[2]Note that $\sigma(s,t)$ reduces to $t-s$ for standard Brownian motion

itself Gaussian with mean

$$\mu_n = c(n)'\Sigma(n)^{-1} \begin{bmatrix} X_n \\ \vdots \\ X_1 \\ X_0 \end{bmatrix}, \tag{2}$$

and variance

$$\sigma_n^2 = 1 - c(n)'\Sigma(n)^{-1}c(n), \tag{3}$$

where $c(n)$ is an $(n+1)$-column vector with elements $c(n)_k = \sigma(m, m+k+1)$, for $k = 0, \ldots, n$. The algorithm presented by Hosking (1984) computes $\Sigma(n)^{-1}c(n)$ recursively to ensure greater efficiency. A further refined approach to the computation of $\Sigma(n)^{-1}c(n)$ is presented in Dieker (2004) and the convergence is given as $\mathcal{O}(N^2)$. More methods for exact and approximate simulation of fBM can be found in Dieker (2004).

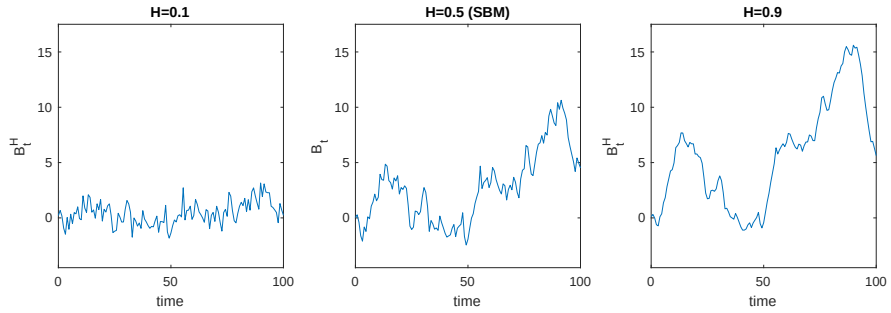


Figure 1: Three sample realisations of fBM, each with 129 points, with Hurst parameter 0.1, 0.5 (SBM), and 0.9. Each path is generated with the same seed for comparison purposes.

In Figure 1, we can see that for $H = 0.1$, W_t^H exhibits higher short-term volatility and lower long-term volatility. In contrast, when $H = 0.9$, W_t^H exhibits higher long-term volatility and lower short-term volatility.

2.3 Estimating the Hurst Index

Recall that fBM is a Gaussian process with the property that

$$\mathbb{E}[|W_{t+\Delta}^H - W_t^H|^q] = K_q \Delta^{qH}.$$

Gatheral et al. (2014) verify that the empirical distributions of log-volatility are approximately Gaussian for various time lags.

Thus, to estimate the smoothness of the volatility process, that is, H, Gatheral et al. (2014) use the following approach. Suppose that we have access to N discrete observations of the volatility process σ_k on $[0, T]$. Calculate

$$m(q, \Delta) = \frac{1}{N - \Delta} \sum_{k=1}^{N-\Delta} |\log(\sigma_{k+\Delta}) - \log(\sigma_k)|^q,$$

where $\Delta \in \mathbb{N}$ is the lag.

Note that the $m(q, \Delta)$ as specified in Gatheral et al. (2014) does not correspond to the methodology they then use to calculate its value. The definition of $m(q, \Delta)$ provided in this report describes their methodology.

Now, assuming that the log-volatility process has stationary increments, then $m(q, \Delta)$ can be seen as an estimate of

$$\mathbb{E}[|\log(\sigma_\Delta) - \log(\sigma_0)|^q] = K_q \Delta^{qH}.$$

Taking logs, we get

$$\log \mathbb{E}[|\log(\sigma_\Delta) - \log(\sigma_0)|^q] = \log K_q + qH \log \Delta.$$

We can then compute $m(q, \Delta)$ for different values of Δ for each q and regress $\log m(q, \Delta)$ against $\log \Delta$. The slope of each line of best fit is then an estimate of qH.

An estimate of the spot volatility processes needs to be used since it is not directly observable. Gatheral et al. (2014) use precomputed 5-minute daily realized variance estimates obtained from the Oxford-Man Institute of Quantitative Finance as a proxy for the S&P and NASDAQ spot variance process. They use data from 3 January, 2000 to 31 March, 2014 and obtain 3540 trading days. Using the same dates, we instead get 3552 trading days worth of realized variances data.

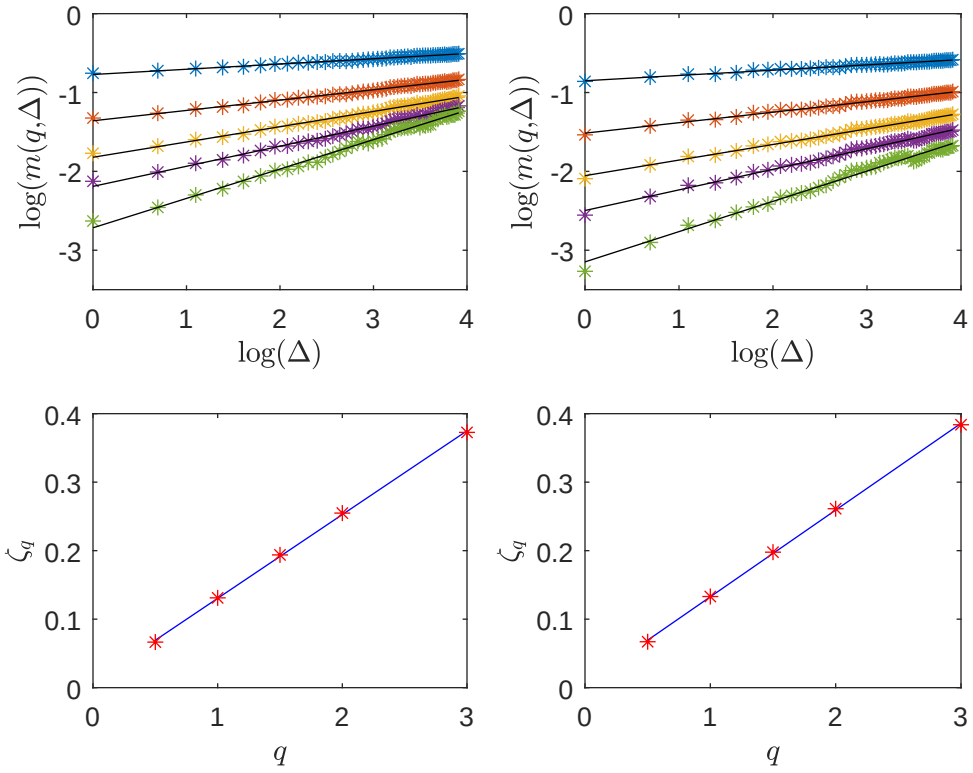


Figure 2: The left-hand column displays results for the S&P and the right-hand column is for the NASDAQ. The top row shows the results from the regressions of $\log(m(q,\Delta))$ against $\log(\Delta)$ for different q (indicated in different colours). The bottom row is a regression of ζ_q against q.

The results from the regression of $\log m(q,\Delta)$ against $\log \Delta$ are displayed in the top row of Figure 2. For a given q, we have that for both the S&P and NASDAQ, there is a linear relationship between $\log \Delta$ and $\log m(q,\Delta)$.

Let ζ_q be an estimate of Hq. By regressing ζ_q against q, we can obtain estimates of H for the different indices. The plot of ζ_q against q for both the S&P and NASDAQ indices are shown in the bottom row of Figure 2.

The estimated values for H are displayed in Table 1. The difference in the estimates between our estimates and those of Gatheral et al. (2014) could be ascribed to differences in our data sets, or perhaps to slight differences in methodology. To observe the impact of using more data points, the H-index was computed for the S&P index for various intervals of time, each starting from 3 January, 2000. The full S&P dataset until 30 June 2017 was used. The results are shown in Figure 3.

Table 1: Estimates of H for data from 3 January 2000 to 31 March 2014.

	S&P	NASDAQ
Gatheral	0.142	0.139
Regression	0.122	0.127

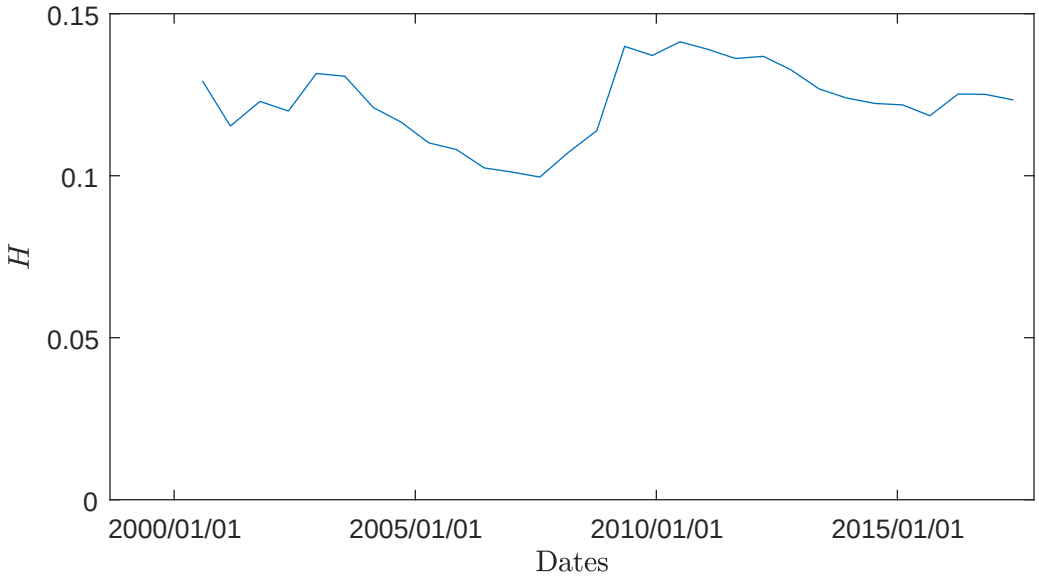


Figure 3: The change in H over time for the S&P index (calculated from January 2000).

3 Models

The following section outlines the mathematics behind the models of interest. We begin by introducing the classical Heston model, which forms the foundation of the models to follow. We then introduce FSV models of Comte and Renault (1998) and Gatheral et al. (2014), both of which modify the classical Heston model by using fBM in place of SBM. Lastly, we consider the rough Heston model of Euch and Rosenbaum (2016), which is also a modification of the classical Heston model, and uses fBM.

3.1 Classical Heston Model

The standard Heston (1993) assumes that the underlying asset price process (S_t) follows a Black-Scholes-type stochastic process but with variance modelled stochastically according to a Cox, Ingersoll, and Ross (1985) process. The model also allows correlation between variance and spot asset returns. We represent the dynamics of the forward price process $\tilde{S}_t$, as is consistent with the literature (see Euch and Rosenbaum (2016), Euch and Rosenbaum (2017), etc.), but note that other representations are possible (such as the representation of S_t under the risk-neutral measure $\mathbb{Q}$). The dynamics of the classical Heston model are

$$\begin{aligned} d\tilde{S}_t &= \sqrt{V_t}\tilde{S}_t \, dW_t \\ dV_t &= \lambda(\theta - V_t)dt + \lambda\nu\sqrt{V_t}dB_t, \end{aligned} \tag{4}$$

where

$\tilde{S}_t$ is the forward asset price;
$\tilde{S}_0 = S_0 e^{rT}$;
$\lambda > 0$ is the mean rate of reversion;
$\theta > 0$ is long-term mean of V_t;
$\nu > 0$ is the "volatility of variance", which controls the smile; and
$V_0 > 0$ the current variance.

Here, W_t and B_t are two correlated standard Brownian motions with $\langle dW_t, dB_t \rangle = \rho dt$, which controls the skew. Typically, ρ is negative pointing to the fact that a decrease in stock price is correlated with an increase in volatility (Albrecher et al., 2006). This is consistent with the market stylized fact called the leverage effect. We also require that $2\lambda\theta > \nu^2$ to ensure that V_t is always strictly positive. This is known as the Feller condition (Albrecher et al., 2006).

3.2 Fractional Stochastic Volatility (FSV) Models

Comte and Renault (1998) extended Hull and White's short-rate model and modelled the log-volatility by replacing the Wiener processes with a fractional Brownian motion (fBM). Comte and Renault (1998) restricted H to $\frac{1}{2} < H \leq 1$ which corresponds to the 'stylized fact' that volatility has long-memory. However, any non-zero α induces a process that is no longer a semimartingale (Rogers, 1997) and thus does not admit an equivalent martingale measure. Despite this, the authors

use $\mathcal{L}^2$ theory of integration for Gaussian processes to show that asset price processes maintain the semimartingale property (Comte and Renault, 1996) and thus do not admit arbitrage.

In contrast, Fukasawa (2017) shows that a stochastic volatility model where volatility is driven by fBM with Hurst index H generates an at-the-money volatility skew of the form $\psi(\tau) \sim \tau^{H-1/2}$, where τ is time to maturity. Thus, Fukasawa (2017) concludes that to generate a market-consistent volatility surface with fBM, H must be close to zero. This is in contrast to Comte and Renault's (1998) use of $H \in (\frac{1}{2}, 1)$.

Following this, Gatheral et al. (2014) went on to show that log-variance behaves essentially as an fBM with $H \approx 0.1$ at any reasonable time scale. Gatheral et al. (2014) call their model a Rough FSV (RFSV) to underline that, in contrast to FSV, $H < \frac{1}{2}$.

Regardless of H, the FSV model used by Comte and Renault (1998) and Gatheral et al. (2014) for the forward price process $\tilde{S}_t$ has dynamics

$$\begin{aligned} d\tilde{S}_t &= \sqrt{V_t}\tilde{S}_t \, dW_t \\ d\log V_t &= \lambda(\theta - \log V_t)dt + \xi dB_t^H. \end{aligned} \tag{5}$$

The parameters are as in (4), except they are interpreted to apply to $\log(V_t)$ and not V_t, and where B_t^H is fBM with Hurst index H.

3.3 Rough Volatility Models

3.3.1 Rough Bergomi Model

In Bayer et al. (2016), the authors extend the model presented in Bergomi (2005). They assume, without loss of generality, that $r = 0$. In this instance the forward price process, $\tilde{S}_t$, and the asset price process, S_t, are equivalent. The rough Bergomi model has dynamics

$$\begin{aligned} \tilde{S}_t &= \tilde{S}_0 \mathcal{E}\left(\int_0^t \sqrt{V_u} dZ_u\right) \\ V_u &= V_0 \mathcal{E}\left(\eta \tilde{W}_u\right), \end{aligned} \tag{6}$$

where η is a scaling parameter, V_0 is the variance at time 0 and

$$\tilde{W}_u = \sqrt{2H}\int_0^u \frac{1}{(u-s)^{1-\alpha}} dW_s.$$

The process $\int_0^u \frac{1}{(u-s)^{1-\alpha}} dW_s$ is known as the Volterra fractional Brownian motion with Hurst parameter $H = \alpha + \frac{1}{2}$. The symbol $\mathcal{E}(\cdot)$ denotes the Doléans exponential and Z_u and W_s are two correlated standard Brownian motions. Bayer et al. (2016) use the Cholesky decomposition $Z_t = \rho W_t + \sqrt{1-\rho^2} W_t^{\perp}$ to express Z in terms of two independent Brownian motions.

3.3.2 Rough Heston Model

Euch and Rosenbaum (2016) use an alternate representation of B_t^H and avoid simulating fBM directly. They represent the fBM B_t^H as an integral with respect to (standard) Brownian motion W_t as

$$B_t^H = \frac{1}{\Gamma(H+1/2)} \Big\{ \int_{-\infty}^{0} \Big((t-s)^{H-1/2} - (-s)^{H-1/2} \Big) \, dW_s + \int_0^t (t-s)^{H-1/2} \, dW_s \Big\}. \tag{7}$$

This representation of B_t^H is known as the Weyl fractional integral (Weyl, 1917) of B_t^H (for $H \neq \frac{1}{2}$). The kernel $(t-s)^{H-1/2}$ in (7) plays a central role in the dynamics of fBM and thus Euch and Rosenbaum (2016) introduce the kernel $(t-s)^{H-1/2}$ in a Heston-like volatility process as

$$\begin{aligned} d\tilde{S}_t &= \sqrt{V_t} \tilde{S}_t \, dW_t \\ V_t &= V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \lambda(\theta - V_s) ds + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \lambda \nu \sqrt{V_s} dB_s, \end{aligned} \tag{8}$$

where the parameters are as defined in (4).

4 Model Implementation

4.1 Monte Carlo

4.1.1 Simulation of Rough Bergomi Model

To simulate a realisation of forward asset price process of (6), we have to discretise the volatility process and the associated asset price process. The rough Bergomi

model of (6) can be rewritten as

$$\begin{aligned} \tilde{S}_t &= \tilde{S}_0 \exp\left(\int_0^t \sqrt{V_u} dZ_u - \frac{1}{2}\int_0^t V_u du\right) \\ V_u &= \exp\left(\eta \tilde{W}_u - \frac{1}{2}\eta^2 u^{2H}\right). \end{aligned} \tag{9}$$

Suppose we want to generate n simulations of the rBergomi model with m time steps. First, note that the Volterra process $\tilde{W}$ has the dependence structure, for $v > u$,

$$\mathbb{E}[\tilde{W}_v \tilde{W}_u] = u^{2H} G\left(\frac{v}{u}\right),$$

where

$$G(x) = 2H \int_0^1 \left((1-s)(x-s)\right)^{\alpha-1} ds,$$

with $x \geq 1$.

Using this dependence structure, it is possible to generate the process $\tilde{W}$ from the Brownian process W by constructing its $m \times m$ autocovariance matrix. Then compute L, where L is the lower triangular matrix obtained from the Cholesky decomposition of the covariance matrix. Generate random normal $m \times 1$ vectors W and multiply them by L to get $m \times 1$ vectors of updates of the process $\tilde{W}$.

The Brownian motion Z is simply calculated as

$$Z_t = \rho W_t + \sqrt{1-\rho^2} W_t^{\perp},$$

where W is retained from the calculation for $\tilde{W}$ and $W^{\perp}$ is an independent Brownian motion.

Since no closed-form solution for (9) exists in the literature, a simple discretisation procedure can be used to simulate S_t and V_u.

4.1.2 Simulation of Rough Heston model

We now turn our attention to the Euch and Rosenbaum (2016) approach for simulating volatility paths and subsequently forward asset price paths according to the rough Heston model of Section 3.3.2. To simulate V_t, we discretise time with $\Delta = T/10000$. For each t, V_t must be computed anew since V_t cannot be written in stochastic differential form.

The integrals in (8) are estimated using standard quadrature techniques, al-

though more reliable estimates of V_t can be obtained using, for example, predictorcorrector methods. Since fBM is not Markov, efficient Monte Carlo simulation methods remain an intricate task in the rough volatility context (see Neuenkirch and Shalaiko (2016)).

Once a realisation of V_t is computed, the forward asset price $\tilde{S}_T$, needed to calculate the payoffs of European options, is easy to compute from (8) as follows

$$\begin{aligned}
\tilde{S}_T &= \tilde{S}_{T-\Delta} + \sqrt{V_{T-\Delta}}\tilde{S}_{T-\Delta}(W_T - W_{T-\Delta}) \\
&= \tilde{S}_{T-\Delta} + \sqrt{V_{T-\Delta}}\tilde{S}_{T-\Delta}\left(\sqrt{\Delta}Z\right) \\
&= \tilde{S}_{T-\Delta}\left(1 + \sqrt{V_{T-\Delta}}\left(\sqrt{\Delta}Z\right)\right) \\
&= \tilde{S}_0 \prod_{t=1}^{T}\left(1 + \sqrt{V_{T-\Delta t}}\left(\sqrt{\Delta}Z\right)\right), \qquad (10)
\end{aligned}$$

where Δ is a small time interval, taken as $T/10000$ in our simulations. An illustration of a sample realisation of V_t and $\tilde{S}_t$ is given in Figure 4.

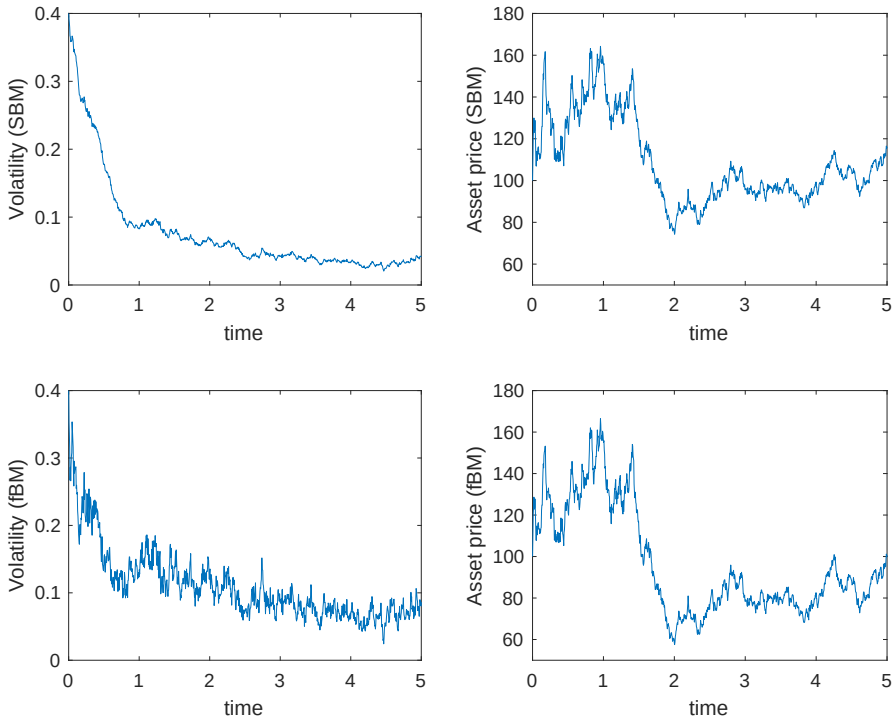


Figure 4: The top row illustrates the typical evolution of the volatility process and the forward asset price process under the standard Heston model ($H = 0.5$) while the bottom row illustrates a typical evolution under the rough Heston model ($H = 0.1$) as in (8). The parameters used are $\lambda = 2$, $V_0 = 0.4$, $\theta = 0.04$, $\nu = 0.05$, $\rho = -0.5$, $\tilde{S}_0 = 100$, $T = 5$.

4.2 Characteristic function

The characteristic function in the rough Heston is derived and presented in Euch and Rosenbaum (2016) is defined as

$$\phi_{x_T}(u) = \mathbb{E}[e^{iux_T}], \tag{11}$$

where

$$x_T = \log\left(\frac{\tilde{S}_T}{\tilde{S}_0}\right),$$

and $\tilde{S}_T$ and $\tilde{S}_0$ are the forward prices. Euch and Rosenbaum (2016) then assume an exponential form for the characteristic function with the exponent being affine

in V_0. This implies

$$\phi_{x_T}(u) = \exp(g_1(u,T) + V_0 g_2(u,T)), \tag{12}$$

where

$$g_1(u,t) = \theta\lambda \int_0^T h(u,s)ds \qquad g_2(u,t) = I^{1-\alpha}h(u,t).$$

The function $h(u,t)$ is the solution to the fractional Riccatti equation

$$D^\alpha h(u,t) = -\frac{1}{2}u(u+i) + \lambda(i\rho\nu u - 1)h(u,t) + \frac{(\lambda\nu)^2}{2}h^2(u,t), \qquad I^{\alpha-1}h(a,0) = 0,$$

where

$$D^\alpha h(u,t) = \frac{d}{dt}I^{1-\alpha}h(u,t), \text{ and } I^{1-\alpha}h(u,t) = \frac{1}{\Gamma(1-\alpha)}\int_0^t \frac{h(u,s)}{(t-s)^\alpha}ds,$$

are the Riemann–Liouville derivative and integral respectively. The fractional Riccatti equation is solved using a predictor-corrector approach. Euch and Rosenbaum (2016) also present a numerical algorithm for solving this differential equation. Similarly Diethelm et al. (2004) presents a numerical algorithm for evaluating the fractional integral for $g_2(u,t)$.

The following is a summary of these two numerical algorithms.

Algorithm 1 Predictor-Corrector Approach for Solving FDEs

$N =$ desired number of time steps = $\frac{T}{\Delta}$
$a, b =$ arrays of size N
$F(u, h(u,t)) = -\frac{1}{2}u(u+i) + \lambda(i\rho\nu u - 1)h(u,t) + \frac{(\lambda\nu)^2}{2}h^2(u,t)$
$\hat{h}(u, t_n) =$ estimated value of $h(u,t)$ for $0 \leq t_n \leq T$
$\hat{h}^P(u, t_n) =$ predictor value of $h(u,t)$
for $1 \leq k \leq N$ **do**
    $a(k) = (k+1)^{\alpha+1} - 2k^{\alpha+1} + (k-1)^{\alpha+1}$
    $b(k) = k^{\alpha} - (k-1)^{\alpha}$
end for
$\hat{h}(u, 0) = 0$
for $1 \leq j \leq N$ **do**
    $\hat{h}^P(u, t_j) = \frac{\Delta^\alpha}{\Gamma(\alpha+1)} \sum_{k=0}^{j-1} b(j-k)F(u, h(u, t_k))$
    $\hat{h}(u, t_j) = \frac{\Delta^\alpha}{\Gamma(\alpha+2)}\Big(F(u, \hat{h}^P(u, t_j)) + \big((j-1)^{\alpha+1} - (j-1-\alpha)j^{\alpha}\big)F(u, 0)$
    $+ \sum_{k=1}^{j-1} a(j-k)F(u, h(u, t_k))\Big)$
end for

Algorithm 2 Procedure for Approximating Fractional Integral

$N =$ desired number of time steps = $\frac{T}{\Delta}$
$\hat{g_2}(u, t_n) =$ estimated value of $g_2(u,t)$ for $0 \leq t_n \leq T$
$c =$ array of size $N+1$
$c(0) = (1+\alpha)N^{\alpha} - N^{\alpha+1} + (N-1)^{\alpha+1}$
$c(N) = 1$
for $1 \leq k \leq N-1$ **do**
    $c(k) = (N-k+1)^{\alpha+1} - 2(N-k)^{\alpha+1} + (N-k-1)^{\alpha+1}$
end for
$\hat{g_2}(u, t_N) = \frac{\Delta^\alpha}{\Gamma(1+\alpha)} \sum_{k=0}^{N} c(k)h(u, t_k)$

Algorithm 2, coupled with a standard quadrature integration method are used to estimate $g_2(u,t)$ and $g_1(u,t)$, respectively, used in (12). This then allows for an estimation of the characteristic function at each discrete time step between the initial time, $t_0 = 0$ and the maturity time, $t_N = T$. Using this characteristic function, prices for European call and put options can be estimated using the Gil-Pelaez Fourier Inversion formula (Wendel, 1961).

It is well know that the price of a European call option can be written as follows

$$C(K) = P_1 S_0 - K e^{-rT} P_2, \tag{13}$$

where

$$P_1 = \mathbb{Q}^s(S_T > K), \qquad P_2 = \mathbb{Q}(S_T > K).$$

Here, $\mathbb{Q}$ is the risk-neutral measure under which $S_T e^{-rT}$ is a martingale and $\mathbb{Q}^s$ is the forward measure defined by the Radon-Nikodyn derivative,

$$\frac{d\mathbb{Q}^s}{d\mathbb{Q}} = \frac{e^{rT}}{S_T/S_0}.$$

The probabilities P_1 and P_2 are then calculated using the Gil-Pelaez Fourier Inversion formula.
Let $s_T = \log(S_T)$ with corresponding characteristic function $\phi_{s_T}(u)$, then

$$P_1 = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{e^{-iu\log(K)} \phi_{s_T}(u-i)}{iu\phi_{s_T}(-i)} du,$$

$$P_2 = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{e^{-iu\log(K)} \phi_{s_T}(u)}{iu} du.$$

The characteristic function of x_T must be written in terms of the characteristic function of s_T, by first noting that

$$\begin{aligned} x_T &= \log\left(\frac{\tilde{S}_T}{\tilde{S}_0}\right) \\ &= \log(S_T) - \log(\tilde{S}_0) \\ &= s_T - s_0 - rT. \end{aligned}$$

Then,

$$P_1 = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{e^{iu(\log(S_0/K)+rT)} \phi_{x_T}(u-i)}{iu} du,$$

and

$$P_2 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{e^{iu(\log(S_T/K)+rT)}\phi_{x_T}(u)}{iu}du.$$

4.3 Results

4.3.1 rBergomi Results

Data could not be sourced to check the consistency of the rBergomi model with the SPX volatility surface, as Bayer et al. (2016) is able to do. However, Bayer et al. (2016) use the parameters $H = 0.07$, $\eta = 1.9$ and $\rho = -0.9$ to fit the SPX volatility surface for February 4, 2010. These parameters can be used in our model as a confirmation of our implementation. Computing Monte Carlo estimates of call option prices and at-the-money implied volatilities yields the results displayed in the figure below.

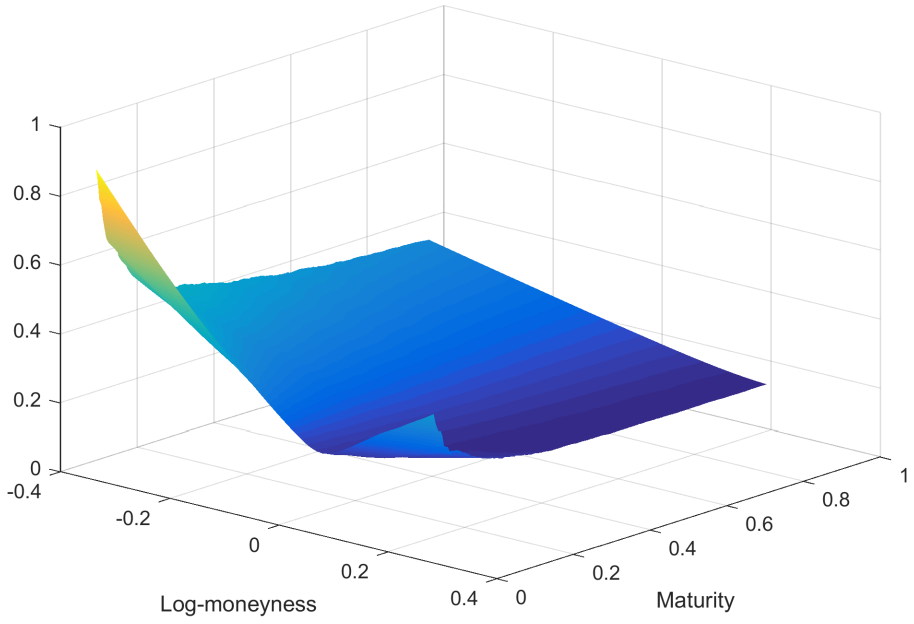


Figure 5: Volatility Surface for rBergomi model.

Taking into account that the implementation uses Monte Carlo estimates and

numerical integration dependent on the size of the time step used, the above figure could be regarded as a close approximation of Figure 4 presented by Bayer et al. (2016).

Since data to make comparisons and to confirm our model implementation was not available, we did not pursue this model any further.

4.3.2 rHeston Results

The results from the characteristic function and Monte Carlo simulation are presented with an example where a European call option is priced with the following parameters:

Table 2: Parameters used in the rHeston model.

Parameter	Value
α	0.6
λ	2
θ	0.04
ν	0.05
ρ	-0.5
T	2
V_0	0.04
$\tilde{S}_0$	100
r	0.05
Δ	0.01
K	$100e^{-rT}$

The value of the strike price is chosen such that the call option is in the money.

Figure 6 includes the integrands for P_1 and P_2 with the given set of parameters. Both integrands decay to zero relatively quickly. This implies that the bounds of integration used in the numerical computation need not be too large to still achieve accurate results. It is also noted that the integrand has no unwanted oscillations or discontinuities meaning that a standard quadrature integration technique will produce accurate estimations of the two integrals.

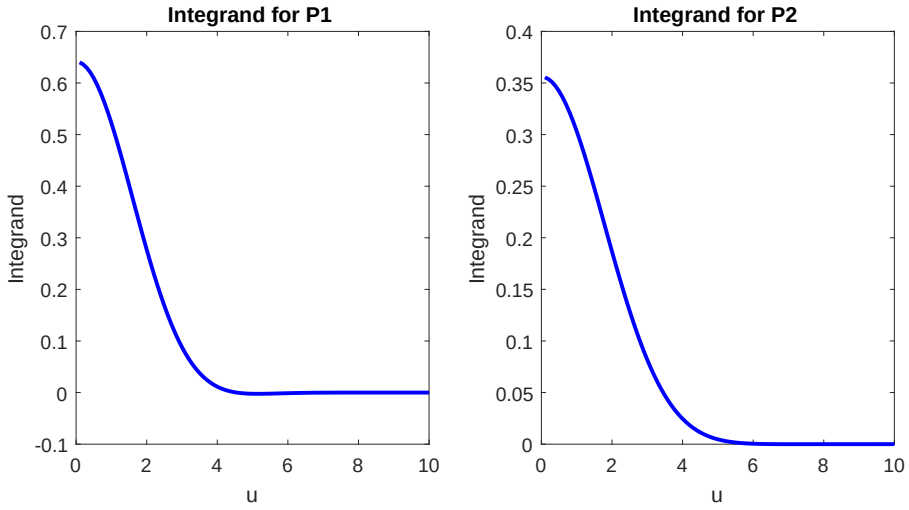


Figure 6: Integrands for P_1 and P_2 decay towards zero.

The probabilities P_1 and P_2 are then calculated using a standard quadrature technique.

Next, the option prices are calculated at different maturities with varying strike prices. These prices are then compared with the Monte Carlo prices from Section 4.1. Table 3 summarises this comparison by comparing the characteristic function pricing with the a 95% confidence interval generated from the Monte Carlo prices at $T = 5$. Figures 7 and 8 show characteristic function and Monte Carlo prices at various maturity values.

Table 3: Results at $T = 5$.

Strike (K)	**Fourier Inversion**	**Monte Carlo**		
		Lower bound	**Value**	**Upper bound**
60	37.2771	36.572	37.5901	38.6082
65	34.8865	34.3326	35.3401	36.3475
70	32.6524	31.4751	32.4256	33.376
75	30.5668	29.3683	30.3121	31.2559
80	28.6213	26.8472	27.7506	28.6541
85	26.8077	25.7911	26.7222	27.6533
90	25.1175	24.2847	25.1864	26.0881
95	23.5427	23.4445	24.3528	25.2612
100	22.0756	21.002	21.8548	22.7076

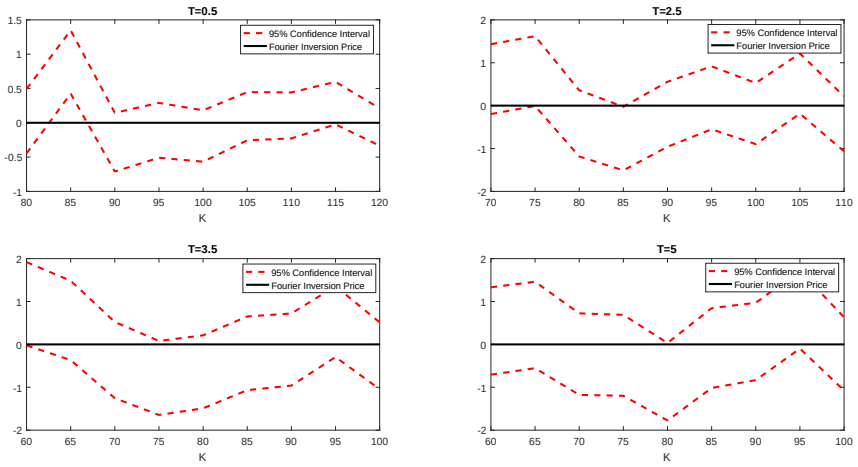


Figure 7: Call option price differences between characteristic function and Monte Carlo simulation.

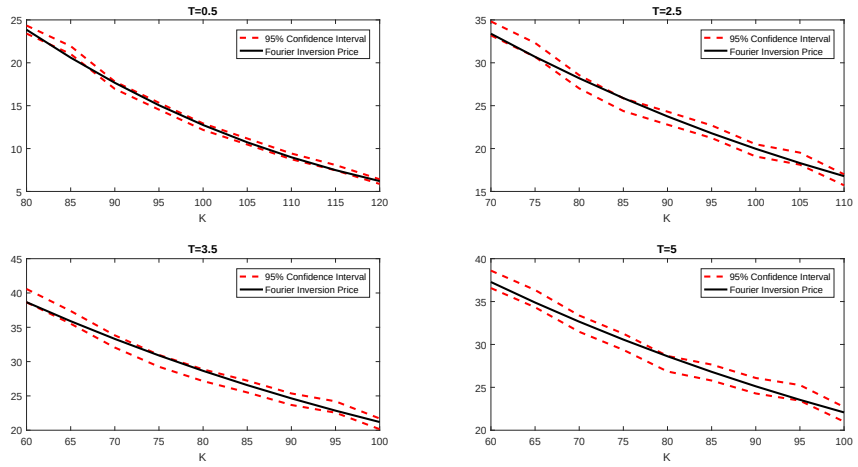


Figure 8: Call option prices for characteristic function and Monte Carlo simulation.

Figures 7 and 8 show that for almost all strike prices, the characteristic function prices lie within the Monte Carlo 95% confidence interval. It is to be expected that better results would have been achieved with larger sample sizes. Due to time constraints, the sample sizes were limited to 10 000 simulations per strike price.

These prices can be used to generate volatility surfaces using the classical Heston model as the benchmark model for calculating implied volatility. Figure 9

below shows three volatility surfaces for various Hurst indices, one surface representing the classical Heston model ($H = \frac{1}{2}$).

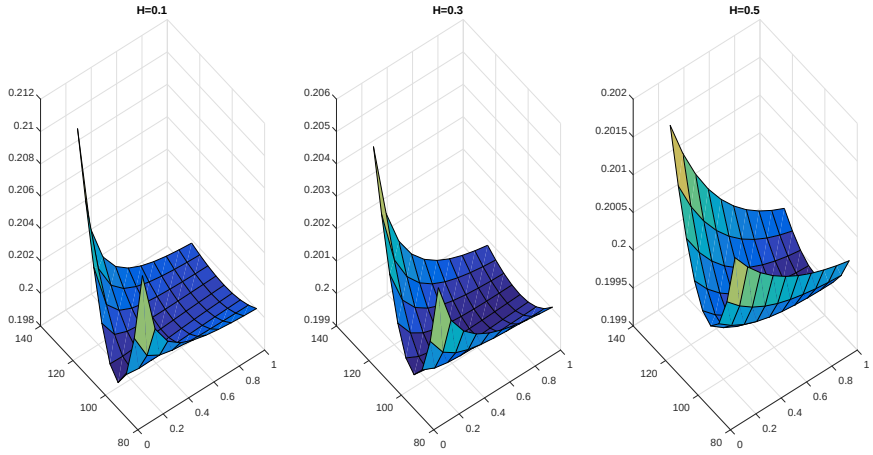


Figure 9: Volatility surfaces for various Hurst indices.

These plots show that as the Hurst index approaches zero, the volatility surface tends to fit modern volatility surfaces better. This is the key feature that motivates the modelling of stochastic volatility with fBM.

5 Model Calibration

Recall that calibration is a process of finding the model parameters such that the prices from the model match as close as possible the prices observed in the market. Efficient Rough volatility model calibration is still an open problem (see Euch and Rosenbaum (2017)). The aim of this section is to provide insight on calibration and to test the analytical tractability of the model. We take a fairly simple and straightforward approach by minimising the sum of squared differences,

$$\hat{\Phi} = \arg\min_{\Phi} \sum_{i=1}^{N} \left(\frac{C_i^{\Phi}(K_i, T_i) - C_i^{\text{mkt}}(K_i, T_i)}{C_i^{\text{mkt}}(K_i, T_i)} \right)^2, \tag{14}$$

with a set of model parameters

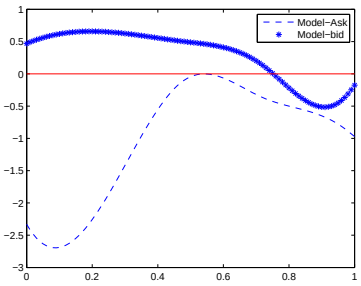
$$\Phi = \{\lambda, \theta, \nu, \rho, \nu_0\}.$$

We collected the market prices for call options written on NASDAQ and S&P500. Market data are as observed on February 14, 2014 and on February 24, 2014.

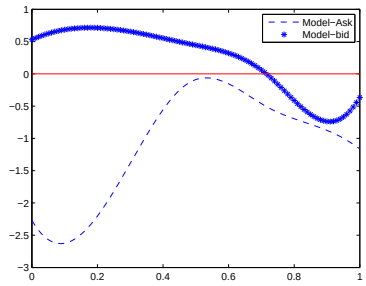
We proceed to consider a case for call options written on NASDAQ indices. Table 4 shows the calibrated Rough Heston model. We investigate cases for different values of H estimated in Section 2.3. In Figure 11, we show the calibration result.

Table 4: Calibrated parameters: 14 February 2014.

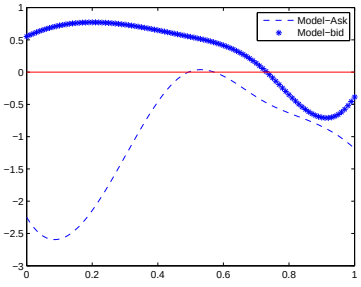
Φ/H	**0.5**	**0.3**	**0.005**
λ	0.460253	0.255436	0.194759
θ	0.388097	0.507656	0.552782
ν	0.999426	0.994158	0.488258
ρ	-0.47974	-0.61345	-0.80704
ν_0	0.103923	0.09935	0.088725



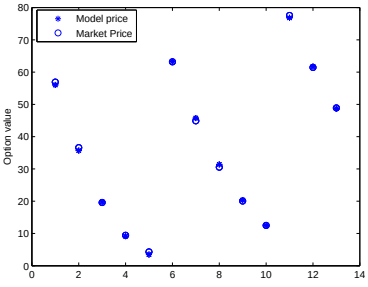
(a) $H = 0.5$



(b) $H = 0.3$



(c) $H = 0.005$

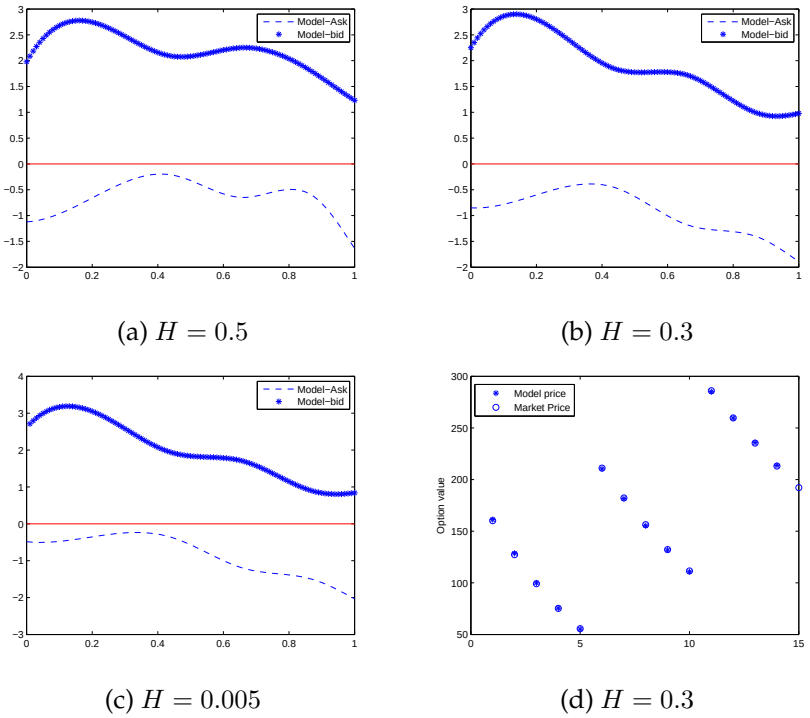


(d) $H = 0.3$

Figure 10: Calibration results for various choices of H.

Table 5: Calibrated parameters: 24 February 2014.

ϕ/H	**0.5**	**0.3**	**0.1**
λ	1	1	0.527462
θ	0.139935	0.144669	0.188619
ν	0.188603	0.164843	0.278388
ρ	-1	-1	-0.82177
ν_0	0.070764	0.064327	0.06135



(a) $H = 0.5$ (b) $H = 0.3$

(c) $H = 0.005$ (d) $H = 0.3$

Figure 11: Calibration results for various choices of H.

As mentioned before, we have used a nonlinear least square approach to fit the model price to the market price by keeping the Hurst index fixed at $H = 0.1, 0.3, 0.5$ and then iterate the other parameters such that model price is as close as possible to the market price. The fit could possibly be improved by considering a global optimisation. This work will be explored further to examine the applicability of the rough Heston model and its tractability. We have noted that the calibration is quite slow due to the intensive computation of the fractional differential Ricatti equation in the associated characteristic function using predictor-corrector schemes. An-

other improvement could be to rewrite our MATLAB implementation into C++.

6 Conclusion

The main purpose of this work was to review and implement the rough volatility framework. Inspired by claims that models where the volatility process is driven by fractional Brownian motion provide an excellent fit to historical data and volatility surfaces, we began by testing and confirming these hypotheses.

We introduced the volatility modelling paradigms and highlighted key features associated with fractional volatility models, and their rough versions such as the rough Bergomi and rough Heston models. We justified the mathematical claim that "Volatility is Rough" and we replicated most of the econometric analysis in the paper of Gatheral et al. (2014) with negligible differences, possibly from data mismatch. We implemented the pricing methodologies using the characteristic function, by using Fourier inversion[3]. We also highlighted some important concepts necessary to implement the characteristic function which are not mentioned in the literature, e.g. detailed predictor-corrector schemes. We benchmarked our Fourier pricing implementation against our Monte Carlo approach and our prices fall in the 95% confidence Monte Carlo bounds.

We went a step further to calibrate the model to observed market data. The model provided a relatively reasonable fit. The fit could possibly be improved by applying a global optimisation approach. We noted that due to the complexity of the characteristic function, with numerous function evaluations, pricing via the characteristic function is computationally expensive. All the code used for this project was implemented in MATLAB, and we shall endeavour to rewrite them in C++ to take advantage of its computational speed.

[3]We mention here that we have also looked at FFT implementation and Lewis (2001) method. Although our prices match exactly with Monte Carlo, we opted not to report these methods at this stage.

Bibliography

Albrecher, H., Mayer, P., Schoutens, W., Tistaert, J., 2006. The little heston trap.

Bayer, C., Friz, P., Gatheral, J., 2016. Pricing under rough volatility. Quantitative Finance.

Bergomi, L., 2005. Smile dynamics II. Available at SSRN 1493302.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. The Journal of Political Economy 81 (1), 637–654.

Cheridito, P., Oct 2003. Arbitrage in fractional brownian motion models. Finance and Stochastics 7 (4), 533–553.

Comte, F., Renault, E., 1996. Long memory continuous time models. Journal of Econometrics 73 (1), 101–149.

Comte, F., Renault, E., 1998. Long memory in continuous-time stochastic volatility models. Mathematical Finance 8 (4), 291–323.

Cox, J. C., Ingersoll, J. E., Ross, S. A., 1985. A theory of the term structure of interest rates. Econometrica: Journal of the Econometric Society, 385–407.

Dieker, T., 2004. Simulation of fractional Brownian motion. MSc thesis, University of Twente, Amsterdam, The Netherlands.

Diethelm, Ford, Freed, Luchko, 2004. Algorithms for the fractional calculus: A selection of numerical methods. Computer Methods in Applied Mechanics and Engineering 194 (8), 751–752.

Euch, O. E., Rosenbaum, M., 2016. The characteristic function of rough Heston models. arXiv preprint arXiv:1609.02108.

Euch, O. E., Rosenbaum, M., 2017. Perfect hedging in rough Heston models. arXiv preprint arXiv:1703.05049.

Fukasawa, M., 2017. Short-time at-the-money skew and rough fractional volatility. Quantitative Finance 17 (2), 189–198.

Gatheral, J., 2006. The volatility surface: A practitioner's guide. *John Wiley & Sons*.

Gatheral, J., Jaisson, T., Rosenbaum, T., 2014. Volatility is rough.
URL `https://arxiv.org/abs/1410.3394`

Heston, S. L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. The Review of Financial Studies 6 (2), 327–343.

Hosking, J. R. M., 1984. Modeling persistence in hydrological time series using fractional differencing. Water resources research 20 (12), 1898–1908.

Jarrow, R. A., Protter, P., Sayit, H., 2009. No arbitrage without semimartingales. Annals of Applied Probability 19 (2).

Lewis, A. L., 2001. A simple option formula for general jump-diffusion and other exponential Lévy processes. Available at SSRN 282110.

Merton, R. C., 1973. Theory of rational option pricing. Bell Journal of Economics and Management Science 1 (4).

Neuenkirch, A., Shalaiko, T., 2016. The order barrier for strong approximation of rough volatility models. arXiv preprint arXiv:1606.03854.

Rogers, L., 1997. Arbitrage with fractional Brownian motion. Mathematical Finance 7 (1), 95–105.

Wendel, J., 1961. The non-absolute convergence of Gil-Pelaez'inversion integral. The Annals of Mathematical Statistics 32 (1), 338–339.

Weyl, H., 1917. Zur gravitationstheorie. Annalen der Physik 359 (18), 117–145.