

Evaluating market risk in a portfolio with
heavy-tailed risk factors using Monte Carlo
methods

Alexander Ivarsson Hannes Sternbeck Fryxell

December 21, 2012

Abstract

In this master thesis we study and implement a model for market risk in a portfolio consisting of both stock and bond indices. It is based on an article by Glasserman, Heidelberger and Shahabuddin from year 2002. The model uses the assumption that the joint distribution of the losses follows a multivariate t-distribution. The model also uses Monte Carlo simulations with importance sampling in order to improve the performance of the simulations, which is necessary to achieve statistical certainty when working with high percentiles of the losses. We focus mainly on Value at Risk, but we will also mention Expected Shortfall. We test the certainty of our model in numerous different ways.

Acknowledgements

We would like to thank our supervisors Björn Laurenzatto and Simon Eriksson at Handelsbanken Liv for all the help and guidance we have received throughout this thesis, as well as for all the data.

Furthermore, we also like to thank Magnus Wiktorsson at Lund University for the mathematical guidance.

Contents

| | | |
|-----------|--|-----------|
| 1 | Introduction | 6 |
| 1.1 | Background | 6 |
| 1.2 | Purpose | 7 |
| 1.3 | Outline | 7 |
| 1.4 | Target group | 7 |
| I | Mathematical background, methods and modelling | 8 |
| 2 | Methodology | 9 |
| 2.1 | Fitting the distribution | 9 |
| 2.1.1 | Multivariate t-distribution | 11 |
| 2.2 | Value at Risk | 11 |
| 2.3 | Expected Shortfall | 12 |
| 2.4 | Monte Carlo methods | 13 |
| 2.4.1 | Importance Sampling | 13 |
| 2.5 | Finding VaR_α using Importance Sampling | 13 |
| 2.6 | The theory behind copulas | 14 |
| 3 | Modelling | 15 |
| 3.1 | Linear approximation | 16 |
| 3.2 | Necessary steps in order to be able to find $P(L > x)$ | 16 |
| 3.3 | Implementing the importance sampling | 17 |
| 4 | Data | 19 |
| 4.1 | Our choice of data sample | 20 |
| 4.2 | Data handling | 21 |
| II | Application, analysis and results | 24 |
| 5 | Result | 25 |
| 5.1 | Stock indices v.s. Bond indices | 25 |
| 5.2 | How accurate are the simulations in the main model? | 26 |
| 5.2.1 | Different probabilities given to VaR | 28 |
| 5.3 | Simulated data and testing | 29 |
| 5.4 | Comparing VaR with ES | 30 |
| 6 | Discussion | 30 |
| 6.1 | Overall succession | 30 |
| 6.2 | The difference given by window sizes | 31 |
| 6.3 | Different thresholds of VaR | 31 |
| 6.4 | Summary | 32 |
| 6.5 | Conclusion | 32 |
| 6.6 | Possible improvements | 32 |
| 7 | References | 34 |

| | |
|--|-----------|
| A Appendix | 35 |
| A.1 Translation of the Bloomberg tickers | 35 |

1 Introduction

1.1 Background

Managing risk has always been of interest by financial institutions, and especially after the beginning of this last financial crisis. Mervyn King, British governor of the Bank of England, said the following in October 2008:

"Not since the beginning of the First World War has our banking system been so close to collapse. (...) The past few weeks have been somewhat too exciting. So let me extend an invitation to the banking industry to join me in promoting the idea that a little more boredom would be no bad thing. The long march back to boredom and stability starts tonight." (King, 2008)

This is an example of a request which has made the financial sector much more rigorous when it comes to risks. The possibility of clustering of risks has made modelling of extreme cases into a priority.

The commonly mentioned risks are Credit risk, Market risk, Liquidity risk and Operational risk. The crisis has put extra stress on the banks' risk levels, especially credit and market risk. In this thesis we will focus on the market risk. The market risk describes the risk of losses in assets do to movements in market prices, caused by different fluctuations of interests, stocks or stock indices. (Wissén, 2011)

A lot of the risk handling today is because of the new international requirements, such as the Basel Accord 'Basel II' which was gradually set into effect in the beginning of the 21st century. Since the set of regulations showed poor results during the financial crisis the capital requirements has become more strict. Having high capital requirements is expensive and therefore is something the banks tries to minimize. In other words; models which measure market risk is something that all financial institutions need and many continuously develop further.

In collaboration with the Quantitative analyst group at Handelsbanken Liv we are to investigate the possibilities of making a Value at Risk-model on events that only occur in extreme cases on a fictitious but plausible globally diverse portfolio. The requested horizon is over a 5-day period.

Using Value at Risk in combination with an extreme outcome is a easy way to quantify the risk in a monetary value at a "worst-case-scenario". It is an effective model to match the risk with a needed interconnected capital reserve and it plays an important role in the regulatory Basel Accords.

The crisis in within the European Union has this last year created a greater interest in what could happen if extreme scenarios become reality. The financial sector has raised their own requirements to take stronger measures not to be caught red handed if an extreme scenario would occur. Such an occurrence could

be created if a country, i.e. Greece, Spain or Italy, would default. This would most likely create a shockwave of financial instability throughout the global economy, and definitely effect the internationally dependent Swedish economy.

For the last few decades Paul Glasserman, Philip Heidelberger and Perwez Shahabuddin has been researching variance reduction techniques for Value at Risk estimations in different situations. Their research has been the basis of our thesis. The model by Glasserman, Heidelberger and Shahabuddin published in 2002 (from now on referred to as (GHS, 2002) or the GHS-model) is an efficient way to compute portfolio Value at Risk, when the underlying risk factors have heavy-tailed distribution.

1.2 Purpose

The main purpose of this thesis is to examine the market risk of an internationally distributed portfolio. We will evaluate the risk with a model which is highly influenced by the models of Glasserman et al. In other words we want to investigate the practical possibilities of the theoretical result of the GHS-model. The final goal of the method is to be able to use it, or a version of it, in a real scenario.

1.3 Outline

The thesis is divided into two main parts, first a theoretical mathematical background, and second an implementation of the models including results and analysis.

The first part describes the methods we use, the theory behind them, their drawbacks and the reason why we chose them. It also brings up our reasoning behind using Student's t-distribution. We will also give the definitions and basic understandings needed to comprehend Value at Risk, Expected Shortfall, the Monte Carlo-method Importance Sampling and the usage of Copulas.

The second part is the next natural step after the first, and is main the part. Here we asses the data, its drawbacks and limitations. We also discuss the results from our models and the limitations of our analysis. We also discuss the different risks and their results on our portfolio.

1.4 Target group

This thesis is written for people who take interest in, or needs to, implement a risk measurement model for a sizeable stock portfolio.

To fully utilize the content of this thesis, one should have quite good knowledge within the area of mathematical statistics and preferably some basic knowledge of finance.

Part I
**Mathematical background,
methods and modelling**

- Distribution, Value at Risk, Expected Shortfall and
Importance Sampling

2 Methodology

When we got the assignment the main object was to measure the risk of a portfolio's tail events. We were suggested to do so with Value-at-Risk and Monte Carlo-methods, but we were free to argue for other methods. After studying several scientific articles we were convinced that this way also was the best approach to tackle our problem. *VaR* was a given because of its wide operating range. To get an effective and accurate model our focus was now on data handling and in our case Monte Carlo-methods. The tools we chose to do this is defined and described in the next few sections to come.

2.1 Fitting the distribution

In order to approach our problem, we first had to look at what the data looked like. The data (e.g. the portfolio) consisted of both stocks and bonds (why these two securities is something we will get back to later on). By plotting histograms of the data series' 5-day returns and fitting distributions to them we will see which pattern they have in their fluctuations. In Figure 1 we see the result of the normal distribution fitted to each individual asset (explanations of the Bloomberg index abbreviations are found in the appendix).

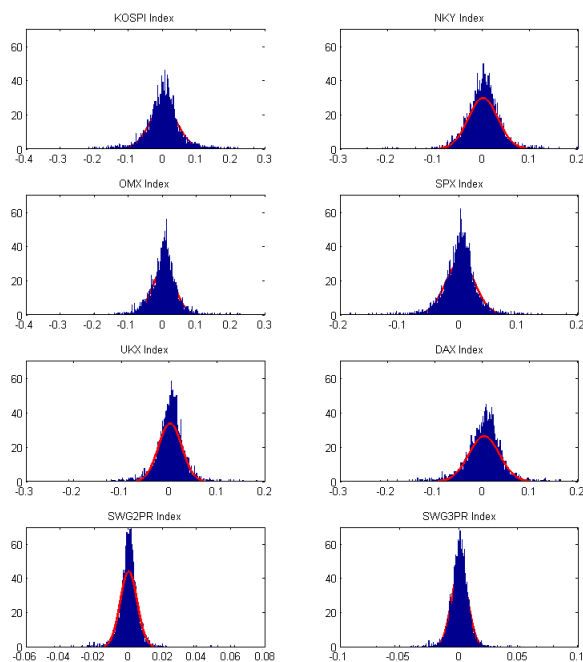


Figure 1: Normal distributions applied to our data

Seeing the result we realized that a standard normal distribution would not be

able to incorporate all risk observed in the data. The data has much heavier tails. The natural next step was to try fitting a more distribution with heavier tails to the data. The Student's t -distribution seemed to fit the data very well (seen in Figure 2), so we decided to implement a model incorporating a multivariate t -distribution. It can be added that the level of the fitted distribution made by Matlab is actually better than it looks like in the plots. Because of the low resolution in the plots the fitted distributions give the appearance of being smaller than the actual distribution, but when we are able to look closely, it is clear that the fitted distribution follows the data very well. As can be seen though, there are a few outliers in the data which are significant distances away from the distribution. Since they are so few, it is really hard to construct a distribution which will compensate for them. We feel that the fitted Student's t -distribution will be as good as we can manage, and decide to stick with it for the rest of the thesis. However, we will have in our mind that this might result in our model underestimating the risk a bit. Especially for the really high VaR levels. We will come back to this in the discussion.

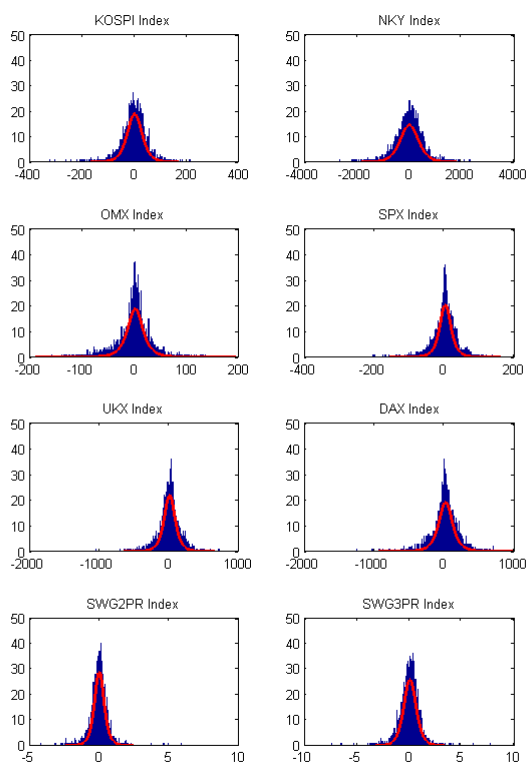


Figure 2: Student's t -distributions applied to our data.

2.1.1 Multivariate t-distribution

Since our problem is multidimensional, we will have to use a multivariate t-distribution. When we say we use the multivariate t-distribution, we refer to:

$$f_{v,\Sigma}(x) = \frac{\Gamma(\frac{m+v}{2})}{(v\pi)^{m/2}\Gamma(v/2)\|\Sigma\|^{1/2}}(1 + \frac{1}{v}x'\Sigma^{-1}x)^{-\frac{1}{2}(m+v)}, x \in \mathfrak{R}^m. \quad (1)$$

(GHS, 2002) where m is the number of underlying assets, v is the degrees of freedom and Σ is the distributions covariance matrix. We will abbreviate the multivariate t-distribution by mvt from now on.

However, the mvt is tricky to implement, since many of the straight-forward methods associated with the normal distribution has to be discarded and/or modified. This has mainly to do with the fact that the mvt has polynomial tails, and therefore does not have a moment generating function. We never explicitly find the characteristic function of the mvt-variables, which may be intractable. Instead, we use an indirect transform analysis through which we are able to compute the distribution of interest. (Glasserman, 2002)

We will combine the mvt-approximation of the data with Monte Carlo simulation methods in order to calculate loss probabilities. These are closely related to calculating Value at Risk, which is what we ultimately aim for.

2.2 Value at Risk

Since the late 1990s Value at Risk, VaR , has gained a widespread acceptance and is now probably the most common and widely used model to measure risk by financial institutions. It is also one of the models in the global regulatory standard Basel II, which demands of the banks to calculate their exposure towards market risk. More precisely than a risk measure, it rather calculates a capital adequacy for a certain risk. One of its greatest advantages is that VaR is both easy to grasp theoretically, and relatively easy to implement practically. (McAleer et al., 2009)

The basic idea lies within the name of the model; Value at Risk. Within a certain time horizon, at an arbitrary probability, one calculates the value at risk for a chosen asset or portfolio of assets. VaR is defined here by Embrechts et al. (Embrechts et al., 2005)

Definition 1 *Given some confidence level $\alpha \in (0, 1)$. The VaR of our portfolio at the confidence level α is given by the smallest number l such that the probability that the loss L exceeds l is no larger than $(1-\alpha)$. Formally,*

$$VaR_\alpha = \inf\{l \in \mathfrak{R} : P(L > l) \leq 1 - \alpha\} = \inf\{l \in \mathfrak{R} : F_L(l) \geq \alpha\},$$

α is usually in the range between 95 % and 99.9 % level of confidence. Our primary focus is the 99.5 % level, but we will show some results for lower levels as well.

VaR has encountered some criticism throughout the years, for instance Artzner, Delbaen, Eber and Heath has criticised the model for not diversifying risk in all cases. When combining two or more portfolios, the combined VaR can be greater than the sum of the individual $VaRs$, which contradicts basic theory of portfolio diversification. In other words, VaR is not subadditive in all cases, unless the underlying distribution is normal, which makes the following and otherwise logical statement not true:

$$VaR_{\alpha}(L_1 + L_2) \leq VaR_{\alpha}(L_1) + VaR_{\alpha}(L_2).$$

Another disadvantage with VaR is that it does not give any information about the losses greater VaR_{α} . We will only know that they all surpasses VaR_{α} . This can be dangerous, especially during a time of crisis and high market volatilities. (Dowd, 2002)

Even though Value at Risk has several drawbacks it is still considered to be a highly useful tool when handling financial risk.

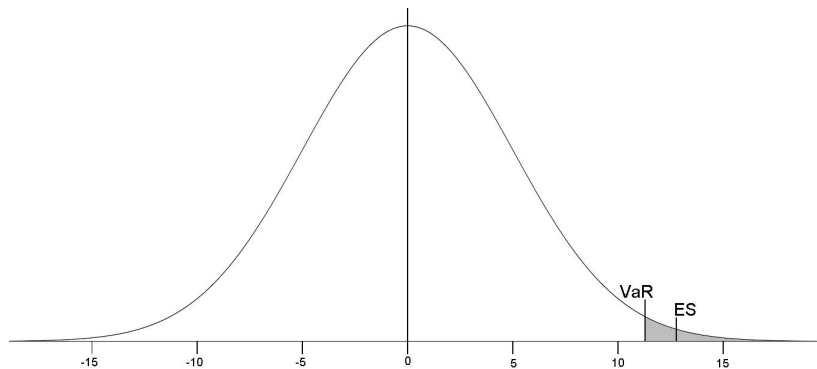


Figure 3: Value at Risk and Expected Shortfall

2.3 Expected Shortfall

Expected Shortfall, or ES , is a close relative and a complement to VaR . It was developed after VaR was getting criticised for not being able to handle extreme events. Given a certain probability, VaR only gives a threshold-loss to where an investment will not over step. In the corresponding case, ES calculates the expected value och the loss *when* it oversteps that threshold. (Acerbi et al., 2008)

According to (Embrechts, 2005) Expected Shortfall is defined mathematically as follows:

Definition 2

$$ES_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 VaR_u(L) du,$$

where α is the probability that the tail event occurs, L is the loss and VaR is the Value at Risk. If the loss distribution is continuous, the definition of ES can be written in the following, more understandable form:

$$ES_\alpha(L) = E[L : L > VaR_\alpha(L)].$$

One great advantage of ES compared to VaR is that it fulfils the properties needed for subadditivity.

2.4 Monte Carlo methods

In order to reduce the variance from a standard simulation, one can use Monte Carlo methods. The basic mechanism is to approximate the random phenomenon of the problem by a mathematical model. From this model we can obtain a large amount of independent samples, in order to eventually use these samples with the law of large numbers, which would help us to get the expected value. Since time and computer performance are our only limits, we can generate a large amount of samples using this method. This allows for a significant variance reduction.

Because our problem is complex and of high dimension, achieving an analytical solution did not seem possible. Instead we found the Monte Carlo methods satisfactory in order to receive low variance results. An introduction to Monte Carlo methods can be found in Sköld's Computer Intensive Statistical Methods. (Sköld, 2006)

2.4.1 Importance Sampling

In our case we are only interested in events which occur with low probabilities, since only the tail events on the losses side will affect our results. Because of this, we will have to throw away the vast majority of samples obtained from the standard Monte Carlo simulation. To cope with this, we introduce Importance sampling. It is a method that makes simulations of the important events more likely, while maintaining the simulations independent and unbiased. Since we can use many more samples, the variance will be greatly reduced.

The basic idea is that instead of sampling from the instrumental distribution $f_L(l)$ we sample from an alternative distribution $g_L(l)$ which has a much higher probability of producing an important sample. Then we multiply all our samples point wise by the so called likelihood function $\omega_L(l)$, which is equal to $f_L(l)/g_L(l)$. The drawback of this method lies in the difficulties of choosing and implementing a good likelihood function, which we will come back to later.

2.5 Finding VaR_α using Importance Sampling

Our original model has the moderate drawback that it needs the VaR_α threshold as an input parameter. We then obtain α as the output parameter. However, we

overcome this problem by implementing a binary search model. It will adjust our Var_α until it is the correct value in order to find the desired α . This increases our running time somewhat linearly with respect to the number of iterations needed. Since the original model is fast even for a high number of simulations, this really does not become a problem in practice.

2.6 The theory behind copulas

A problem with the standard GHS-algorithm is that it requires all marginal distributions to have the same degrees of freedom parameter. This means that they all have tails shaped the same way, which most often is not realistic since the marginals commonly have widely spread properties. As an example, in the histfit plot for all the data in Figure 2 the fitted distributions had degrees-of-freedom parameters ranging from 3.7 to 6.3. In order to deal with this and to allow for different degrees of freedom for the different marginals, we will use a Copula.

Copulas was first introduced in a statistical way by the mathematician Abe Sklar in 1959, but has not until these last decades been commonly used within the field of economics. It is basically used to statistically join together different multivariate distribution functions to their one dimensional marginal distribution function. (Nelsen, 2006)

We follow (Embrechts, 2002) for the definition, background and associated area of usefulness:

The dependence between the real-valued random variables X_1, \dots, X_n is completely described by their joint distribution function

$$F(x_1, \dots, x_n) = \mathcal{P}[X_1 \leq x_1, \dots, X_n \leq x_n].$$

The idea of separating F into parts, with one part describing the dependence structure, and the other parts describing the marginal behaviour, were the original thoughts which eventually led to the concept of a copula.

We could transform the random vector $X = (X_1, \dots, X_n)^t$ component-wise to have standard-uniform marginal distributions $U(0, 1)^1$. In our case, X_1, \dots, X_n are t-distributed and thus have continuous marginal distributions, called F_1, \dots, F_n . Knowing this, the transformation can be done by using the probability-integral transformation $T : \mathcal{R}^n \rightarrow \mathcal{R}^n, (x_1, \dots, x_n)^t \mapsto (F_1(x_1), \dots, F_n(x_n))^t$. The joint distribution function C of $(F_1(X_1), \dots, F_n(X_n))^t$ is then called the copula of the random vector $(X_1, \dots, X_n)^t$ or the multivariate distribution F . It follows that:

$$F(x_1, \dots, x_n) = \mathcal{P}[F_1(X_1) \leq F_1(x_1), \dots, F_n(X_n) \leq F_n(x_n)] = C(F_1(x_1), \dots, F_n(x_n)).$$

Using this, we can give two definitions of the copula.

Definition 3 A copula is the distribution function of a random vector in I^n with uniform-(0,1) marginals.

Definition 4 Alternatively, a copula is any function $C : [0, 1]^n \rightarrow [0, 1]$ which has the three properties:

1. $C(x_1, \dots, x_n)$ is increasing in each component x_i .
2. $C(1, \dots, 1, x_i, 1, \dots, 1) = x_i$ for all $i \in 1, \dots, n, x_i \in [0, 1]$.
3. For all $(a_1, \dots, a_n), (b_1, \dots, b_n) \in [0, 1]^n$ with $a_i \leq b_i$ we have:

$$\sum_{i_1=1}^2 \dots \sum_{i_n=1}^2 (-1)^{i_1+\dots+i_n} C(x_{1i_1}, \dots, x_{1i_n}) \geq 0,$$

where $x_{j1} = a_j$ and $x_{j2} = b_j$ for all $j \in \{1, \dots, n\}$.

These two alternative definitions can be shown to be equivalent.

3 Modelling

We have now stated the fundamental theoretical background needed to implement our model. We will in general follow the (GHS, 2002) model, but there will also be a number of distinctions. The most obvious is that we use a linear approximation of the relation between the change in value of the underlying assets ΔS and the portfolio value ΔV . Our portfolio consists entirely of different index funds, and hence no derivatives are included and no quadratic approximation is needed.

All of the parameters needed for the mvt-approximation are extracted from the data. We use maximum likelihood-estimates for all of them. (GHS, 2002) assume a mean of zero, as is customary in estimating risk measures over short horizons using complex models. We can, however, easily incorporate non-zero means in our model, and will do so.

The mvt-density belongs to the class of scale mixtures of normals. It can be represented as the distribution of the product of a multivariate normal random vector and a univariate random variable independent of the normal variables. Using our definition of the mvt-distribution, it can be expressed as follows: if

$$(X_1, \dots, X_m) =_d \frac{(\xi_1, \dots, \xi_m)}{\sqrt{Y/v}} \quad (2)$$

where $=_d$ denotes equality in distribution, $\xi = (\xi_1, \dots, \xi_m)$ has distribution $N(0, \Sigma)$, Y has distribution χ_v^2 (chi-square with v degrees of freedom) and ξ and Y are independent. We can see from this definition that X_i and X_j are

not independent even if they are uncorrelated, because of that they share the same Y . Small values of Y can make the underlying assets make large moves together, regardless of their correlation. From this one can also easily observe the major differences between the normal distribution and the t-distribution. (Glasserman, 2002)

3.1 Linear approximation

As mentioned before, we will use a linear approximation of the relationship between the change in the underlying risk factors ΔS and the change in the portfolio value ΔV . Our approximation ends up like $\Delta V = \delta(\Delta S) + \mu$, where δ is our portfolio weights and μ is the mean growth of the total portfolio. As it is convenient to work with losses instead of gains, this gives the approximated loss model $L \approx -\delta(\Delta S) - \mu \equiv Q - \mu$. Q is here the variable which we will simulate.

We are interested in calculating loss probabilities $P(L > x)$ when we assume equality in the previous formula, in order to solve the closely related problem of calculating VaR . This means finding a quantile x_p for which $P(L > x_p)$ where p is some small probability.

Using the non-copula model, the change in risk factors are modelled using the mvt-equation (1), using the data series covariance matrix as Σ . From equation (2) we know that ΔS has the distribution of $\xi/\sqrt{Y/v}$ with $\xi \sim N(0, \Sigma)$. If C is any matrix for which $CC' = \Sigma$, then ξ has the distribution of CZ with $Z \sim N(0, I)$. From this we see that ΔS has the distribution of CX with $X = Z/\sqrt{Y/v}$. It follows that $Q =_d (-\delta'C)X - \mu$. (Glasserman, 2002)

3.2 Necessary steps in order to be able to find $P(L > x)$

Because of that we work with an mvt-model we have to be extra careful. Though uncorrelated, X_j is not independent, as can be seen in the representation above. Hence the characteristic function does not factor as a product over j . Would it have done that, the one-dimensional transforms would still be very hard to work with. X_j is heavy tailed, and thus it does not have a moment generating function. Following (GHS, 2002), it therefore seems adequate to refer to the characteristic function of Q as intractable.

To be able to calculate the distribution $P(Q \leq x)$, we instead use an indirect approach. Using the same representation as before, we now define

$$Q_x = (Y/v)(Q - x).$$

We see that $P(Q \leq x) = P(Q_x \leq 0)$. We also observe that Q_x is not heavy-tailed, so unlike Q its moment generating function exists. This obviously makes its characteristic function more tractable. A way to find the characteristic function ϕ_x of Q_x is shown in the following theorem:

Theorem 1 *There is a matrix C satisfying $CC' = \Sigma$. Let $b = (-\delta)'C$. Then $P(Q \leq x) = F_x(0)$, where the distribution F_x has moment generating function*

$$\phi_x(\theta) = \phi_Y(\alpha(\theta)) \tag{3}$$

with

$$\alpha(\theta) = -\frac{\theta x}{v} + \frac{1}{2v} \sum_{j=1}^m \theta^2 b_j^2 \quad (4)$$

provided $\alpha(\theta) < \bar{\theta}_Y$. In the case of multivariate t_v ,

$$\phi_x(\theta) = \left(1 + \frac{2\theta x}{v} - \sum_{j=1}^m \frac{\theta^2 b_j^2}{v} \right)^{-v/2} \quad (5)$$

The characteristic function of Q_x is given by $E[e^{i\omega Q_x}] = \phi_x(i\omega)$ with $i = \sqrt{-1}$.

We leave the proof to be found in (GHS, 2002). The delta approximation to a loss probability is $P(L > x) \approx P(Q > x - a_0)$. The constant a_0 is $-\delta t(\partial V/\partial t)$. We evaluate the approximation above using $P(Q < x - a_0) = 1 - P(Q_{x-a_0} \leq 0) = 1 - F_{x-a_0}(0)$. By using the inversion integral (obtained from the standard inversion formula (Feller, 1971)) we can find F_{x-a_0} in the following way:

$$F_{x-a_0}(t) - F_{x-a_0}(t-y) = \frac{1}{\pi} \text{Re} \left(\int_0^\infty \phi_{x-a_0}(iu) \left[\frac{e^{iuy} - 1}{iu} \right] e^{-iut} \right) du, i = \sqrt{-1} \quad (6)$$

When implementing this method we have to choose a large y for which $F_{x-a_0}(t-y)$ can be approximated to be zero. Since the mean and variance of Q_{x-a_0} can be calculated without any difficulties, a y which accomplishes this can be found by using Chebychev's Inequality.

With theorem 1 we can, given the transform analysis, allow for different degrees of thickness in the distributions' tails of the different risk factors. The result extends to this case by the chain rule for differentiation and the copula mechanism. We can model the risk factor changes, ΔS , as

$$\Delta S_i = K_i(X_i), K_i(x) = \tilde{\sigma}_i G_{v_i}^{-1}(G_v(x))$$

with (X_1, \dots, X_m) having the density $f_{v,\Sigma}$. G is here the cdf of the Student's t-distribution. It leads to:

$$\frac{\partial V}{\partial X_i} = \delta_i \frac{dK_i}{dX_i},$$

where the derivatives of K are evaluated at 0. For us, this makes

$$L \approx a_0 - \tilde{\delta}' X \quad (7)$$

where $\tilde{\delta}$ now also depends on the derivatives of K . After some computations, $\frac{dK_i}{dX_i}$ evaluated at 0 simply comes down to $\tilde{\sigma}_i \frac{g(0,v)}{g(0,v_i)}$ where g is the pdf of the Student's t-distribution.

3.3 Implementing the importance sampling

When we now come back to implementing the importance sampling algorithm, we will begin by addressing an exponentially changed measure, which we do by slightly adjusting theorem 4.1 in (GHS, 2002). It turns out like this:

Theorem 2 If $\alpha(\theta) < \bar{\theta}_Y$, then

$$dP_\theta = e^{\theta Q_x - \psi_x(\theta)} dP$$

defines a probability measure and

$$P(L > y) = E_\theta \left[e^{\theta Q_x + \psi_x(\theta)} I(L > y) \right] = E_\theta \left[e^{-\theta(Y/v)(Q-x) + \psi_x(\theta)} I(L > y) \right].$$

Under P_θ , X has the distribution of $Z/\sqrt{Y/v}$ where

$$P_\theta(Y \leq u) = E \left[e^{\alpha(\theta)Y - \psi_Y(\alpha(\theta))} I(Y \leq u) \right]$$

and conditional on Y , the components of Z are the independent with

$$Z_j \sim N(\mu_j(\theta), \sigma_j^2(\theta)), \mu_j(\theta) = \theta b_j \sqrt{Y/v}, \sigma_j^2(\theta) = 1.$$

In the specific case that the distribution of X under P is multivariate t_v (i.e., the P -distribution of Y is χ_v^2), the distribution of Y under P_θ is Gamma($v/2, 2/(1 - 2\alpha(\theta))$), the gamma distribution with shape parameter $v/2$ and scale parameter $2/(1 - 2\alpha(\theta))$.

dP_θ can be seen as $g_L(l)$ in our previous notation, making $\frac{dP}{dP_\theta}$ our likelihood ratio. $\psi_x = \log \phi_x$. θ is chosen as the root of the equation $\frac{d}{d\theta} \phi_x(\theta) = 0$. This follows (GHS, 2002) and originates from the will to minimize the upper bound of the second moment

$$E[e^{-\theta Q_x + \psi_x(\theta)} I(L > y)].$$

(Fuh et al. 2011) uses an iterative algorithm in order to determine θ exactly. Because of that the computational gain ($\approx 10\%$) seemed small compared to the technical work needed in order to implement that algorithm, we decided to use the simpler method from (GHS, 2002).

For short, P_θ is the new measure which we will use to sample from. This measure was first developed for normally distributed risk factors in (GHS, 2000), but was extended to the mvt case in (GHS, 2002).

Once we have obtained the likelihood ratio, we have everything we need to be able to sample from Algorithm 1. For the non-copula model, we use the square matrix root of the covariance matrix of the data as C . This has the benefit of preserving all magnitudes of the data in a simple way.

When using the copula model we are not as fortunate, as it is necessary that Σ has all diagonal elements equal to 1. To cope with this, we use Σ as the correlation matrix of the copula variables. By doing this, we have to get the information lost by not using the covariance matrix back into the model once again. We do this by scaling our losses by what could be seen as the covariance matrix of the pre-scaled mvt-variables, defined as $\sqrt{D\Sigma D}$, where D is the diagonal matrix with the σ 's of the pre-scaled mvt-variables on the diagonal.

The algorithm by Glasserman, Heidelberger and Shahabuddin used for Importance sampling estimation of loss probability looks as follows.

Algorithm 1 For each of n independent replications:

1. Generate Y from the distribution $P_\theta(Y \leq u) = E[e^{\alpha(\theta)Y - \psi_Y(\alpha(\theta))} I(Y \leq u)]$
2. Given Y , generate independent normals Z_1, \dots, Z_m with parameters given by $Z_j \sim N(\mu_j(\theta), \sigma_j^2(\theta))$, where $\mu_j(\theta) = \theta \cdot b_j \sqrt{Y/\nu}$ and $\sigma_j^2(\theta) = 1$.
3. Set $X = Z/\sqrt{Y/\nu}$.
4. Set $\delta S = CX$ and calculate the resulting portfolio loss L and the quadratic approximation Q . Set $Q_x = (Y/\nu)(Q - x)$.
5. Multiply the loss indicator by the likelihood ratio to get

$$e^{-\theta Q_x + \psi_x(\theta)} I(L > y)$$

Average this product over the n independent replications.

For the copula case we change part 4, so that ΔS gets sampled from $K(\tilde{C}X)$, where \tilde{C} is the matrix square root of Σ , the correlation matrix of the copula variables. This alters some of the model parameters a bit as well, but does not require any new major computations.

4 Data

Looking at our test portfolio below, with a start at the first trading day of 1992, we see the individual indices and their respective progress. You find the most movement in the Swedish OMX, which is probably because of the financial bank crisis we found ourselves in at the time of the start of the indexation. The other parts of the world did not find themselves in the same situation and therefore had different starting positions. Three of the indices separate themselves from the rest, since the others have the same kind of movements but only differ in the degree of change. The two stable (grey and dark blue lines) are the Swedish government bonds SWAG2PR and SWAG3PR. Because they are fixed income securities they are expected to have stable and different movement patterns (compared to the stock indices). The worst outcome has come from the Japanese Nikkei 225 Index, where the country has had their own domestic financial problems for several decades.

Figure 5 shows the daily returns for each asset in our portfolio. It quite clearly shows the last decades' financial turbulence. It is easy to notice the IT-bubble in the interval 1500-3000 days, as well as the immediate reaction of the fall of Lehman Brothers around day 4500.

The plot also shows the well-known phenomenon of volatility clustering. It refers to the phenomenon that market prices tend to exhibit long consecutive periods of both high and low volatility. For example, during a crisis, the fluctuations in the financial markets tend to be large in both directions. This is a contrast to the smaller market swings during periods of stability. We had plans on incorporating volatility clustering in our model, but since we did not have enough time to incorporate it, we had to leave it out.

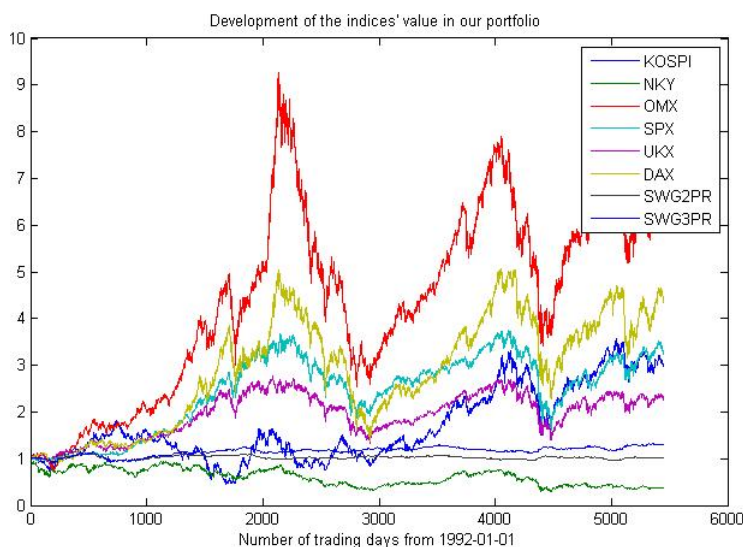


Figure 4: The outcome of the indices from 1992 until today.

4.1 Our choice of data sample

We wanted to see how well our model worked in many different applicable areas. We chose stocks and bonds as securities, which are widely and commonly used among most categories of financial institutions. Both stocks and bonds are traded in highly liquid markets and represent a great deal of all financial transactions.

We also wanted to make a realistic portfolio selection. Since the common investor is risk-averse, we felt the most natural thing to do was to try use a portfolio which could be attractive to investors. Our stocks are globally diversified, since we believe it could represent the interests of a stock broker in this financially unstable world.

Our stocks are in fact represented by stock indices for the respective countries from where the stocks are issued. For example, all stocks issued on the Swedish market will be lumped together as a combined, large position in the OMXS30. This has a couple of reasons. Since the dependence structure estimated from the daily returns requires all trading days in the data to match, the length of our usable data will be determined by the shortest data series for any of the stocks. One do not want to reduce the amount of usable data just because a few minor stocks in the portfolio only date back just a few years, whilst the majority dates back many years. This is vital in order to obtain the amount of data needed to be able to make good parameter estimations for the model.

Our model is mainly created for banks, insurance companies and other financial institutions. These actors commonly have huge portfolios, with a myriad of different stocks and bonds. Taking this into consideration, these approximations

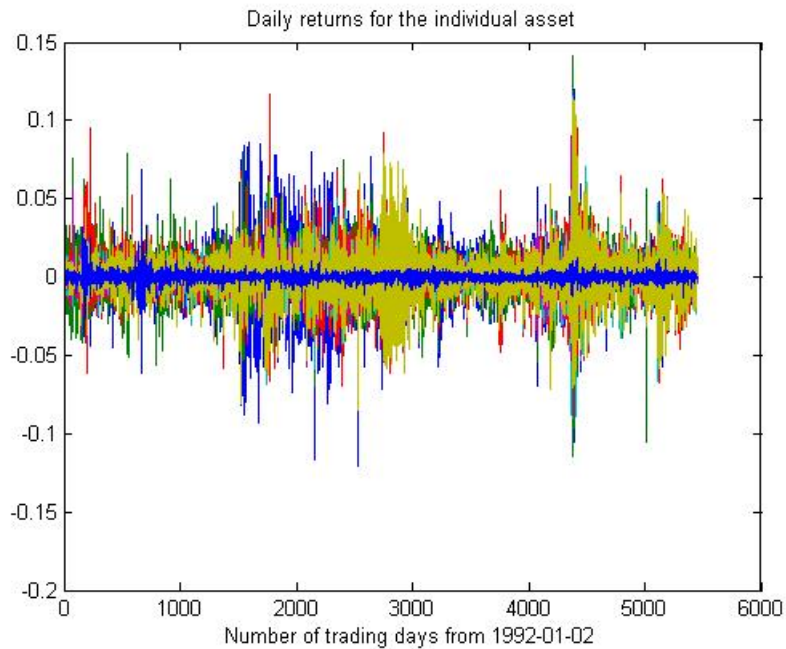


Figure 5: Shows the daily individual change of value of the six different assets of our portfolio.

will introduce relatively small errors.

When it comes to the bonds, we felt we needed data as coherent as possible. Bonds which has been continuously traded for at least a couple of decades. Swedish government bonds fulfilled our requirements. The indices we will use will represent buckets of bonds with 3-5 respective 5-7 years to maturity.

Something we have not taken into consideration is the effects of currency risk. This is mainly due to shortage of time and delimitations of this thesis. However, by having a diversified portfolio with almost all different securities being issued in different currencies, this risk is somewhat naturally minimized even though we do not deal with it in our model.

4.2 Data handling

When choosing the data we wanted it to be diversified in a way a stable and well thought through portfolio should be. We also wanted the portfolio holdings to cover the entire world market, from the United State's New York Stock Exchange, XNYS, in the west, to Japan's Tokyo Stock Exchange, XTSE, in the far east.

As with all data handling it, was problematic to format it in a way that worked for us. The securities are issued in different countries, with different cultures, which celebrate different holidays. This means while trading goes on in one

stock market, it can be closed due to local holidays in another. We simply solved this by assuming zero growth for all non homogeneous holidays. This could probably have been done in a more sophisticated way yielding slightly better results, but since this thesis's is about modelling, the data handling was time demanding enough keeping it rather simple.

However, in some Asian countries there used to be a 6-day trading weeks, which gave us bigger problems. If a country has 5 to 15 midweek holidays out of the normal 250 trading days, our zero-growth assumption, only marginally alters the data. But if we make the same assumption for Saturdays in Asia every single week it would have a much greater impact on the result. For consistency, we obviously had to remove the Saturdays as trading days. We did so, but also changed the Monday's value into the average of the Saturday's and the Monday's value. This method dampens the fluctuations imposed by the Saturdays compared to just removing it, while maintaining the correct progress of the indices.

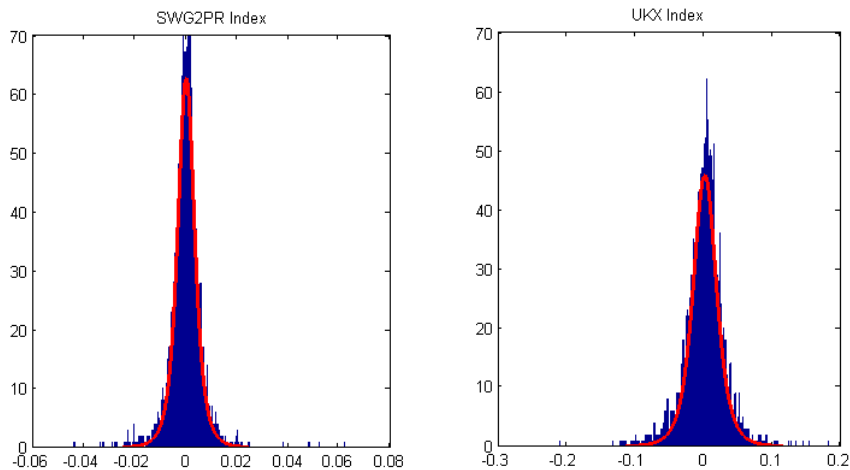


Figure 6: Swedish government bond index with maturity of 3-5 years (left) and the UKX Stock Index (right).

As mentioned, the portfolio partly consists of interest bearing securities. For our model, we treat the bond indices the same as the stock indices. The returns of the bond indices can be fitted to t-distributions in the same way as the stock indices. An example is shown in Figure 6. This gives us just as valid mathematical background for the bond as the stock indices.

There are several known time dependent effects which influence the returns on the stock market. Because of our short five day time horizon, a few of them could be taken into consideration. One such effect is the Holiday effect, which says that the market has a tendency to gain on the final days of trading just before a holiday weekend, such as Christmas or New Year (Dodd, 2011). However,

after briefly looking into this kind of data patterns, we believe that our time is better spent on evaluating and calibrating the model itself more thoroughly.

Part II
Application, analysis and results

5 Result

5.1 Stock indices v.s. Bond indices

To back up our claims that there is no great difference between stock indices and bond indices, we first looked at the two different types of distribution (seen in Figure 6). After not detecting any seemingly big difference we wanted to backtest them. Backtesting the two types gives us a clear picture of their reliability, even though we are using a quite simple method.

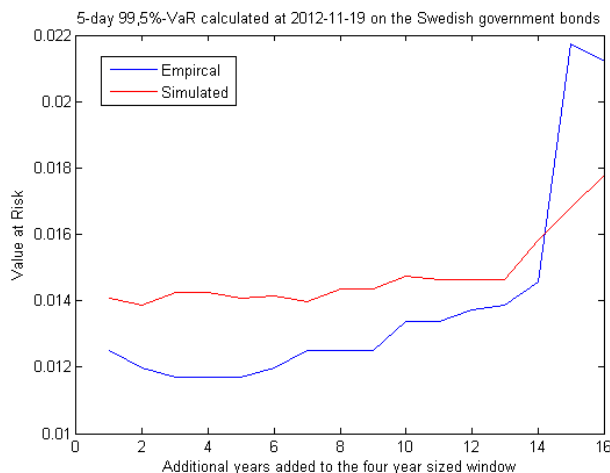


Figure 7: 5-day VaR on the bonds of our portfolio.

The sudden increase in Value at Risk is due to the bank crisis Sweden had in the early 1990's, when we could experience an interest of several hundred percent. That the model is not able to fully adapt to this is in our opinion only natural, since this was a *very* extreme event. All in all, we see that the model probably tends to overestimate the risk. Since the Student's t-distributions obviously cannot fit exactly to the empirical data, effects like these is quite expected.

When it comes to backtesting only using the bonds, we get a plot where the relationship between the simulated VaR and the empirical VaR is just the opposite. Using the same reasoning as for the bonds, we here see that the distributions have trouble fitting exactly to the data. The problem in this case leads to the distribution not being able to capture all of the really extreme values of the data. As we can see, the discrepancy diminishes as the amount of data increases. This is intuitive as we get better parameter estimations using more data, as well as a smoother behaviour of both the curves. As can be seen, the simulations still yield an underestimation compared to the empirical results though.

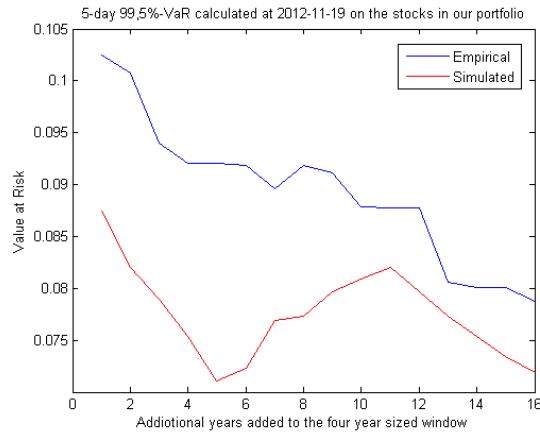


Figure 8: 5-day VaR on the stocks of our portfolio.

5.2 How accurate are the simulations in the main model?

After making this model we wanted to see how precise it actually is. To do this we created a confidence interval to show the accuracy in our simulation model.

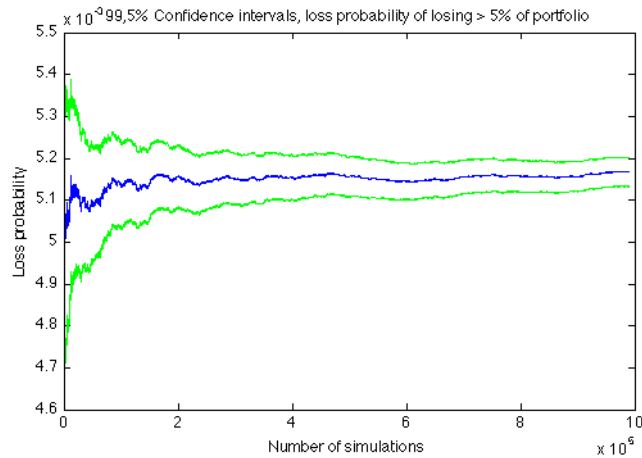


Figure 9: Confidence intervals

Here we have plotted the loss probability for losing 5% of the portfolio with 99,5% standard confidence interval against number of simulations. As we can see, already at around $N = 10^5$ the interval width gets as small as $2 * 10^{-4}$, which definitely can be seen as adequate in the context. Errors originating from other parts of the model is most likely far greater.

From this we conclude that there is no specific need to make the simulation algorithm more effective. The reason why the empirical and simulated calculations are closer to each other than in the equivalent plots seen later on is most

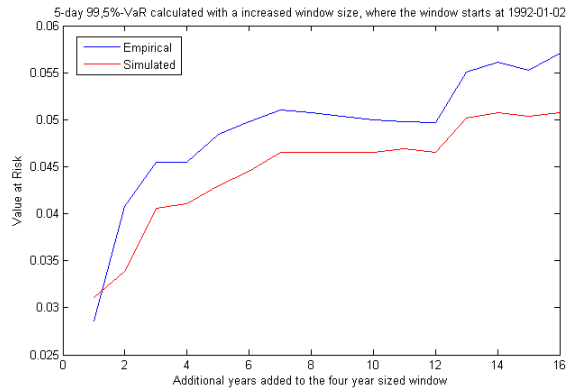


Figure 10: The VaR has here been calculated using an increasing time window. The first point represents 1992-01-02 until 1997-01-02, the second point represents 1992-01-02 until 1998-01-02 and so on until the final year of 2012.

likely because the windows in the plots of 5.2, 5.2.1 and 5.2.1 start out by creating a distribution from quite scarce data. In the financial data from 2007-2012 the outcome is scattered and it is thus hard to find a distribution which will cover all possible events, which leads to the greater differences. When we create a window which always starts in 1992 and the markets are a little bit more stable, the positive result shows in Figure 10. In the end of the plot, we find a bigger gap between the simulated and empirical results again due to the worse fit of the distribution.

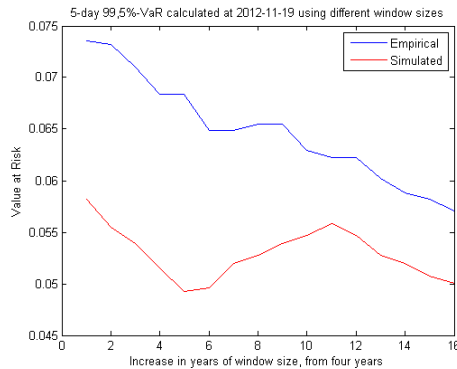


Figure 11: 5-day 99.5%- VaR calculated at 2012-11-19 with different window sizes. The first point represents the model using data ranging from 2007-11-19 until 2012-11-19, the second point uses data from 2006-11-19 until 2012-11-19 and so on.

5.2.1 Different probabilities given to VaR

We also found it interesting to try the model on different thresholds of Value at Risk. We wanted to try it at some of the most common levels of VaR, i.e. 99% and 95%. By using the same date as origin (2012-11-19) we will get comparable data and plots, which we will present below. When we use these lower thresholds, we automatically see the data overstepping the threshold more often and it will basically work as if we would have more data. Another good thing for the estimations is that when using lower VaR levels, the extreme points which our distribution fitting does not capture is given less importance.

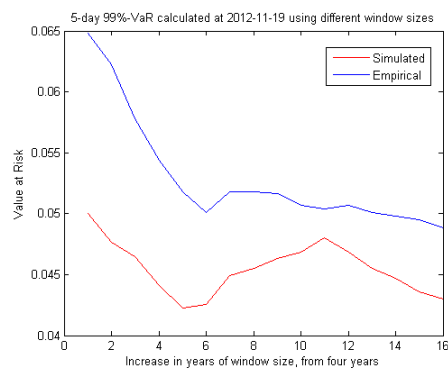


Figure 12: 5-day 99%-VaR calculated at 2012-11-19 with different window sizes.

For the 99% plot, we can see that it behaves very much like the 99,5% plot, indicating similar levels of performance.

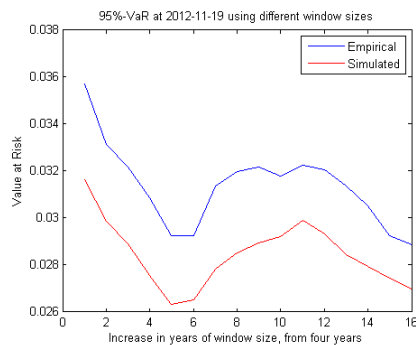


Figure 13: 5-day 95%-VaR calculated at 2012-11-19 with different window sizes.

For the 95% plot however, we see the model being very adaptive to changes in the data. We still have an underestimation, but it seems close and to be constant which indicates that our model is good at capturing changes in the data. Although, it *should* be much easier to fit a model to lower VaR levels,

since it gives us more out of bounds-data to work with.

5.3 Simulated data and testing

After noticing that we would not be able to fully model the market changes of our portfolio, we still wanted to see if our model could produce good VaR estimates if the data would follow some multivariate t-distribution exactly. To do this, we just simulated multivariate t-variables with the exactly the same marginals and parameters as the ones extracted from the data.

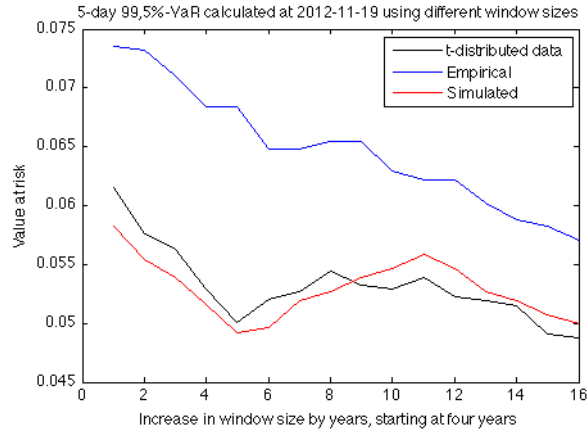


Figure 14: Plot including data created from a simulation

As we can see, we get as good results as can be expected when we use data having the exact distribution as proposed. To verify this statistically, we performed a Kupiecs proportion of failures, or POF-test, as well. It is an unconditional coverage test, which simply compares the number of actual VaR exceedances with the sought for $VaR\%$. It looks like this:

$$POF = 2 \log \left(\left(\frac{1 - \hat{\alpha}}{1 - \alpha} \right)^{T - I(\alpha)} \left(\frac{\hat{\alpha}}{\alpha} \right)^{I(\alpha)} \right)$$

$$\hat{\alpha} = \frac{1}{T} I(\alpha)$$

$$I(\alpha) = \sum_{t=1}^T I_t(\alpha)$$

The results of this vary a lot, since the data is simulated stochastically. The test gives an answer which is $\chi^2(1)$ distributed, and it can be interpreted as follows: $1 - \chi^2(POF, 1)$ is the probability this could happen, given that the model in fact gives a true estimation of VaR . Even for relatively few simulations, we have mostly used $N = 10$ years, we seldom obtain a probability lower than $\approx 60\%$. This is to be considered good, but as we know the exact distribution of our simulated data, something would be very bad if it did not.

Just to have seen it, we performed a Kupiecs test for the real data as well, building the model using the first half of the data, and then testing it on the second half. As expected, it yielded *very* low probabilities of being correct. The empirical data seems to yield about 100% more exceedances than our model expects.

5.4 Comparing VaR with ES

VaR has endured criticism because it does not say anything about the shape of the loss quantile, it only tells at what level it starts. Just to see for ourselves how the tail loss looks like we provide a plot including ES for the 99,5% level as well.

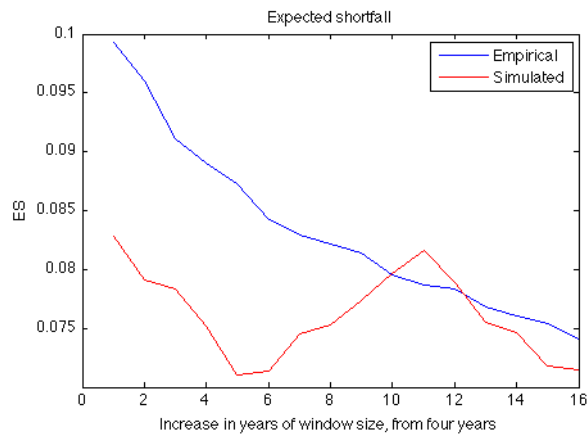


Figure 15: 99,5% Expected shortfall

As we can see, the relationship between the different ES-curves differs to the relationship between the corresponding VaR -curves. This mainly has to do with the last crisis being so influential on the ES no matter how far back we go in the data. The model misses to capture these extremes, naturally making it look like the VaR curve shifted upwards. This tells us that many of the losses larger than VaR , are in fact significantly larger.

6 Discussion

6.1 Overall succession

The object of this thesis was to implement a model from which we would be able to calculate 5-day $VaR_{99,5\%}$ for a large portfolio. Unfortunately, we cannot say that our model accomplishes this, primarily due to the difficulties that lies in fitting a distribution to the returns of the portfolio. There are almost always, and almost for all the different securities, more than a few extreme outliers in the data. This gives us trouble, since not even the distribution which we still think fits the data the best can incorporate this. However, had it been easier

to fit the distributions to the data, we feel that our model would perform very well. The easiest way to see this is shown in the last subsection.

6.2 The difference given by window sizes

If we follow the result found in Figure 11 we see a concurring result in reference to the empirical result. Even though our simulated result does not fit the precise empirical result, it still has similar trend. In the case of $VaR_{99,5\%}$ the difference between the empirical and simulated starts out at $\sim 1,7\%$ of the portfolio's value, where the model use five years data (about 1300 data points). When the model use 16-20 years of data it gives basically the same difference between the empirical and simulated 5-day VaR , $0,6\% - 0,7\%$.

Even though the empirical result is by no means the "correct solution" to our simulation, it can be considered to be a quite good indication. This is why we consider the empirical result to be a good guideline of the actual expected result.

Empirically, the data will on average overstep the $99,5\%$ threshold only every two hundred trading day, which is only a little more often than once a year. Naturally, we would optimally want to be able to use very much data because of this. To get a statistically significant result we would need more data than what we have, and that is why we consider the empirical result as a indication rather than the pursued result for our model's simulation.

One feature of the plot 5.2 is the simulated result's incline. In the smallest window, i.e. 5-10 years' windows, we find the greatest absolute incline. With the next interval, 10-16 years, the incline is lower and in the last window, 16-20 years, it is the lowest. This is because with more data, the result generally smooths out as the extreme periods get somewhat diluted. With enough data, and a cyclic movement pattern, the result will move towards an average. A smaller window may be preferred to a larger window because of this, since the most relevant data probably gets a bigger effect on the result. The downside with a small data window is though that the extreme tail events will be much harder to capture, as well as the fact that less data gives less statistical certainty. Weighting these two factors against each other is critical to any result. According to the performance of our model, we still think using a window of >10 years is to prefer whenever possible.

6.3 Different thresholds of VaR

If we compare the Figures 11, 12 and 13, we see something happening to the empirical curve. It is moving closer and closer to the simulated data, as the threshold is lowered. With the threshold at 95% the empirical and simulated data have the same shapes and changes, seen in 5.2.1, but with a $\sim 0,4\%$ distance.

As we said previously, the empirical result cannot be considered to be the answer sheet of our simulation, it can only be considered to be a quite good indication. The plots, seen in Figure 11, 12 and 13, tell us that the simulated result comes closer and closer to the empirical result the lower the threshold is, and thus,

we can come to the conclusion that our model underestimates the risk of the portfolio.

One common feature of the three plots is the simulated result's incline. It has the same trends and change of trends as $VaR_{99,5\%}$. With the smallest windows, i.e. 5-10 years, we find the greatest absolute incline. In the next interval, 11-15 years, the incline is lower and in the last window, 16-20 years, it is the lowest. This also enhances our belief in our model.

6.4 Summary

After starting our with some background information, we proceeding by explaining some different risk measures used in the thesis. After this, we explained our choice of distribution and the model used to solve the problem. We followed this up by showing how we handled the input data. Finally, we ran the model for several different cases and backtested the results, both with respect to empirical and simulated data.

6.5 Conclusion

We started out with the hope of being able to construct a model which effectively could simulate the real world events we were interested in. This proved a task too complex though, and when looking back on it in retrospect, it feels like a task *very* hard to solve in the limited time frame of this thesis.

For all statistical tasks where there is limited data available, assumptions and approximations are needed. Perhaps we settled on a few major of these too early, resulting in us having a model a little too complex a little too early, which was not 100% verified to work out. We struggled a lot to understand, adjust and implement this GHS-model. This left us short of time to be able to fully investigate other options later on in the project. However, the model we have implemented works very well for what it is designed for, i.e. modelling the development of multivariate t-distributed risk factors. We can make very fast simulations and still be able to estimate the true VaR to a high certainty. The problem lies in that there is seemingly very hard to approximate real market data with the t-distribution, at least for our case where we are interested in 5-day losses. This makes our model good in theory and perhaps for analytical purposes, but sadly we cannot recommend it being used in practice.

6.6 Possible improvements

We have many ideas on how to improve what we have done. First of all, since the biggest trouble has been the distribution fitting, that is also what we have spent most time on. From what we know, the multivariate t-distribution is amongst the most advanced distributions used together with importance sampling today. It is also said to be amongst the best in representing market data. Since this did not work all the way for us, there must be something possibly better. With the disclaimer of the fact that we have not had the opportunity to investigate much in the fields of the distributions from the multivariate extreme value theory, we feel like they probably could take care of the extreme events associated

with really high VaR levels. We have not found anything on combining this with the importance sampling needed in order to achieve statistically better results though, which we think demonstrates how difficult and complex it must be.

An other possible improvement, which we originally thought we would be able to implement, was to incorporate volatility clustering or the like of it. Eventually, we fell short of time in the end so we had to leave it by. If we would have been able to solve this in a good way, it could perhaps have helped us capture the most extreme values a little better.

We have used a rolling 5 day window in order to calculate the 5 day losses from the data. This is not optimal due to the bias introduced by using the same pay off more than once. However, because of the limitations in the amount of data, we still felt this was the best approach. In a perfect world, we would have had much more relevant data and would have been able to use distinct windows instead of rolling.

A fourth, minor, idea was to be more careful with the data handling, and particularly around the more common holidays such as Christmas and Easter. As we ended up having trouble with the distribution fitting in the end, this was eventually not of any concern.

7 References

References

- [1] Acerbi C., Nardio C., Sirtori C., *Expected Shortfall as a Tool for Financial Risk Management*. Available at: <http://arxiv.org/pdf/cond-mat/0102304v1.pdf>
- [2] Artzner P., Delbaen F., Eber J.M., Heath D., *Coherent measures of risk*. Mathematical Finance, vol 9:203-228, July 1999
- [3] Dodd O., Gakhovitch A., *The holiday effect in Central and Eastern European financial markets*, Investment Management and Financial Innovations, vol 8, Issue 4, 2011
- [4] Dowd K., *Measuring market risk*. John Wiley & Sons Ltd, 2002
- [5] Embrechts P., Frey R., McNeil A. J., *Quantitative Risk Management*. Princeton University Press, 2005
- [6] Fuh C.D., Hu I. , Hsu Y.H., Wang R.H., *Efficient Simulation of Value at Risk with Heavy-Tailed Risk Factors*. Operations Research, Nov-Dec 2011.
- [7] Glasserman P., *Monte Carlo methods in financial engineering*. Springer-Verlag, 2003
- [8] Glasserman P., Heidelberger P., Shahabuddin P., *Portfolio Value-at-Risk with Heavy-Tailed Risk Factors*. Mathematical Finance, July 2002
- [9] Jimenez-Martin J.A., McAleer M., Perez-Amaral T., *The Ten Commandments for Managing Value-at-Risk Under the Basel II Accord*, Department of Quantitative Economics, Complutense University of Madrid, March 2009
- [10] King M., *Speech by Mervyn King, Governor of the Bank of England*. Leeds, 21 October 2008. Available at: <http://www.bankofengland.co.uk/publications/Documents/speeches/2008/speech362.pdf>
- [11] Nelsen, R., *An Introduction to Copulas*, Springer Lecture Notes in Statistics no 139, 2006
- [12] Sköld M., *Computer Intensive Statistical Methods*, Lund University, 2006
- [13] Wissén P., Wissén U., *Finanskrisen förklarad*. Fores+Ivrig, 2011
- [14] *Bloomberg's Market Data Site*. Updated daily. Information gathered November 26th, 2012, <http://www.bloomberg.com/quote/MXWO:IND>

A Appendix

A.1 Translation of the Bloomberg tickers

| Bloomberg ticker | Commonly used name |
|------------------|--|
| KOSPI Index | Korea Stock Exchange Index |
| NKY Index | Nikkei 225 Index Fund |
| OMX Index | OMX Stockholm 30 Index |
| UKX Index | FTSE 100 Index (London Stock Exchange) |
| DAX Index | Deutsche Borse AG German Stock Index |
| SWG2PR Index | Swedish government bonds, maturity 3-5 yrs |
| SWG3PR Index | Swedish government bonds, maturity 5-7 yrs |