Statistical pitfalls in Solvency II Value-at-Risk models

Miriam Loois, MSc.

Supervisor: Prof. Dr. Roger Laeven

Student number: 6182402
Amsterdam Executive Master-programme in Actuarial Science
Faculty of Economics and Business/ Amsterdam Business School
University of Amsterdam

December 22, 2012

"Statistics is the grammar of science" - Karl Pearson

"Lies, damned lies, and statistics" - Mark Twain/Benjamin Disraeli.

Abstract

In the first part of this thesis, we compute the effect of misspecifications and misinterpretations of Solvency II Value-at-Risk models. We investigate the effect of backtesting using a rolling window without correcting for the use of the rolling window (a misinterpretation). We show that this leads to a significant increase in the probability of finding an historic extreme event in the data the model is calibrated on. If not corrected for this effect, backtesting with a rolling window can lead to false rejections of Value-at-Risk models. We illustrate this by analyzing the evaluation of the equity stress parameter for Dutch Pension Funds, which is said to be too low. We find that the rejection of the parameter in the report can be explained by the use of a rolling window. We propose a step by step approach to correctly backtest using a rolling window. To our knowledge, this is the first time the effect of this commonly made error is quantified. Next to that, we investigate the effect of a number of possible misspecifications of Value-at-Risk models. The normality assumption is at the heart of Solvency II. We compute how much the probability of finding an extreme event increases if normality is assumed, and effects like autocorrelation, clustered volatility, and heavy tailedness of distributions are neglected. Next to that, we have extended an existing analysis of the effect of scaling up the VaR of a Jump Diffusion process to a longer time horizon by the square root of time rule. By considering different time horizons, we identify three different regimes, and their implications on the use of the square root of time rule, of which two regimes do not occur in the current analysis.
For the misspecifications, we have computed possible theoretical effects, and investigated the practical relevance by computing the effects for realistic parameter values. For this purpose, historic interest rate data, and calibrations of Jump Diffusion models and GARCH models on historic equity (option) data have been used. For realistic parameter values, we find increases in the probability of finding an extreme event by a factor between 2 and 4. For backtesting with a rolling window, we even find an increase in the probability of finding an extreme event by a factor 7. Therefore, the found effects are relevant, and should be considered when calibrating VaR models.

In the second part of this thesis, statistical pitfalls of the one-year horizon in Solvency II are investigated. We compute long term probabilities of ruin and default for insurance companies all starting with an initial Solvency II ratio of 100%. We vary investment strategies, liability durations and cashflow patterns, and show that this highly influences levels of safety on time horizons beyond one year. In the light of the level playing field, we plead to consider longer time horizons in Solvency II beside the currently used one-year horizon, analogous to the regulations for Dutch Pension Funds. We show that for longer time horizons, probabilities of ruin and default are more sensitive for assumptions of mean reversion of risky assets, and therefore assumptions should be set with care.

Contents

1 Introduction

"Lies, damned lies, and statistics" - Mark Twain/Benjamin Disraeli.

Statistics is an interesting field of research, with ample applications in forensic science, biomedical science, psychology, economics, actuarial science, and many other fields. If used correctly, it can be a useful tool to determine whether a medicine works significantly better than a placebo, whether we can put someone to jail based on a match of his DNA or how the age of a driver influences the probability of making a car accident. The phrase above shows that not everybody shares this enthusiasm. It refers to the feeling that statistics can be misused to defend ones opinion on false grounds. It is possible to do so, because a statistical analysis often does not correspond with the intuition people have. This is illustrated by a famous example. A mother has two children, and we know one of them is a boy. If we ask a person about the probability the other child being a girl, many people will answer "one half". However, the correct answer is "2/3". It can be easy to mislead somebody (accidentally or on purpose). An example with serious consequences is the conviction of the Dutch nurse Lucia de Berk, based on erroneous statistics. Statistics is an important tool for insurance companies, and models based on statistics are used to make risk management decisions. In this thesis we will treat some statistical pitfalls that can be easily made, which from a risk management perspective can have serious consequences.

Solvency II is a framework of regulatory solvency requirements for European insurance companies. It will be more risk-based compared to the current Solvency I framework. Based on latest available information, it will be effective no earlier than 2016. Solvency II consists of three pillars, and contains qualitative, quantitative, as well as governance aspects. Pillar I sets quantitative requirements for the amount of capital an insurance company should hold. This amount is risk based. If a company takes more risk, for example by investing in risky assets, or by giving risky guarantees (from the perspective of the insurance company) to the policy holders, this is reflected in a higher required capital. Insurance companies can use the standard formula to compute this capital, but they can also develop an internal model. For every risk (interest rate, equity, longevity, expense, etc.), a model should be made to describe a "1-in-200 year" shock, and those risks should be aggregated to a Solvency Capital Requirement for the insurance undertaking as a whole. According to the directives of Solvency II [1], "Each of the risk modules ... shall be calibrated using a Value-at-Risk measure, with a 99.5 % confidence level, over a one-year period."

In this thesis, we describe some statistical pitfalls that may occur in Solvency II Value-at-Risk (VaR) models. This thesis contains two parts. In the first part we describe the statistical pitfalls that can occur when Value-at-Risk models are calibrated and scaled to a longer time horizon (for example, by the famous square root of time rule), and VaR models are backtested using a rolling win-

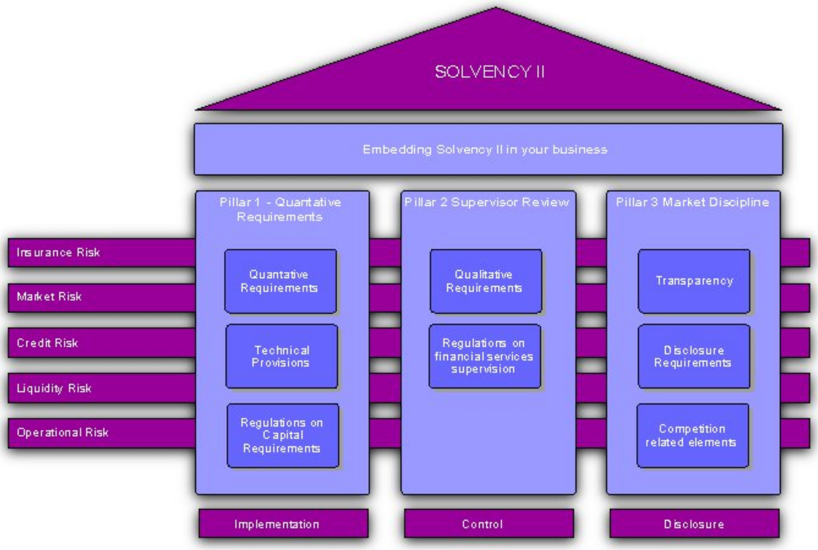


Figure 1: The three pillars of Solvency II. Source: www.ariscommunity.com

dow. The second part will show some pitfalls related to the one-year horizon in Solvency II. Solvency II solely focuses on a one-year horizon, where other regulatory frameworks also take longer time horizons into consideration.

For Solvency II internal models, it is obliged that the model is validated. The European Insurance and Occupational Pensions Authority (EIOPA) has written guidelines for the validation procedure. One of the requirements is that the model is backtested [2]:

“Many assumptions are set based on an analysis of historical data. There is therefore a presumption that past performance is a good indicator of future performance. Back-testing may be used to assess the validity of this underlying assumption”

If the model is correctly specified, one expects that the probability of finding an extreme event in history is 0.5% per year. If more events are found than expected, this can be a reason to question the validity of the model. Therefore it is important that the backtest is performed in a correct way. In practice, often more historic extreme events are found than we would expect based on the confidence level of the stress. For example, if a 1-in-200 year stress is calibrated, and in the last 40 years 2 events exceeded this stress, this corresponds to a p-value of 1.7%, and we can say the backtest has failed. We will describe different causes of finding more than expected extreme events. We will distinguish between two different categories: a misspecification of the model, and a

	Error	True data generating process	Assumed data generating process	Presentation effect error
Misspecification 1	Neglecting autocorrelation in residuals	MA(1)	Random Walk	Increase in probability finding an extreme event
Misspecification 2	Wrong distribution of errors	Student-t	Random Walk	Increase in probability finding an extreme event
Misspecification 3	Neglecting clustered volatility	GARCH(1,1)	Random Walk	Increase in probability finding an extreme event
Misspecification 4	Assuming square root of time rule Value-at-Risk	Jump Diffusion, yearly VaR	Jump Diffusion, scaled up monthly VaR	Increase in yearly VaR

Table 1: Overview of misspecifications treated in this thesis. The third column represents the true data generating process, and the fourth columns describes the assumption made when calibrating the VaR model. The error that is made by making this assumption is shown in the second column. The last column describes how the effect of the error is quantified. The effect can be quantified by computing the increase in the probability of finding an extreme event due to the error made, or by the difference between the true VaR and the VaR under the erroneous assumption.

misinterpretation of the model.

If the model is misspecified, this can lead to an underestimation or overestimation of the 1-in-200 stress over a one year horizon. We will investigate four misspecifications. For all four misspecifications, we will assume there is a true data generating process, and some erroneously assumed data generating process or statistical rule. We will create data based on the true process, and analyse it, based on the erroneous assumption. Then we will study the impact of the error made, in terms of an increase in the probability of finding an extreme event, or an under- or overestimation of the Value-at-Risk. The erroneous assumption will be linked to the normality assumption, since the normality assumption is at the heart of Solvency II. This is reflected in the aggregation of risks in the Standard Formula, where the total Solvency Capital Requirement is computed by computing correlated sums of individual risks. This aggregation method is almost exclusively correct if the underlying risk factors are normally distributed.

In practice, even if the true data generating process is assumed, wrong parameters can be calibrated due to a limited data set. We will leave this out of consideration, and, for simplicity, assume an infinite set of data is available. A summary of the misspecifications and the assumed data generating processes is given in Table 1.

We will consider four misspecifications of a data generating process describing a dataset. The effect of autocorrelation in the residuals of a data generating process will be investigated by considering an MA(1) model. If the errors (i.e.

the deviation between the realized and one-step-ahead predicted value of a time series) experience autocorrelation, a large error is more likely to be followed by another large error, and this has an effect if one predicts a number of time steps ahead. The practical relevance of this effect is shown by an example based on historic interest rate data.
To investigate the misspecification of a distribution, we will investigate the effect of errors that are Student-t distributed as opposed to a normal distribution. The Student-t distribution has heavy tails, which means that the probability mass in the tails of the distribution is relatively higher compared to a normal distribution. Furthermore, we will investigate the effect of neglecting clustered volatility in the residuals by means of a GARCH(1,1) model, where the standard deviation of an error depends on the standard deviation and realized error of the previous timestep. In the study of the Jump Diffusion effect, we will take the article of J. Danielsson and J.P. Zigrand [4] as a starting point. They have studied the impact of the square root of time rule, when a daily VaR is scaled up to a 10-day VaR. We will extend this analysis by comparing a monthly VaR with a yearly VaR, which is a more relevant timescale for insurance companies. The square root of time rule has been investigated in numerous articles. Diebold et. al. [3] warn for the use of the square root rule for scaling volatility from a one day horizon to multiple days. McNeil [5] and McNeil and Frey [6] have investigated a time series that follows a GARCH process, and find a power law for scaling the daily Value-at-Risk up to 50 days. In times of average volatility, they find a scaling factor of 0.6. The square root of time rule corresponds to a scaling factor of 0.5, and thus, using the square root rule underestimates risk. They disagree with Danielsson and de Vries [7], who find that for heavy tailed distributions, the square root rule overestimates the risk at longer horizons. The analysis of McNeil is based on Monte Carlo simulations, and Danielsson and de Vries use a theoretical assumption to compute the distribution of the tails of a sum of heavy tailed variables. Our results will show that for a Jump Diffusion process, the square root of time rule can lead both to an under and overestimation of risk, depending on the frequency of the jumps, relative to the time horizon considered.

If a model is misspecified, this can result in a wrong required capital of the insurance undertaking. If the required capital is underestimated, the company is less safe than the regulator, customer, and the company itself expect. This will lead to incorrect risk management policies. Therefore, it is important that one is aware of possible pitfalls.

We will consider one type of misinterpretation: backtesting with a rolling window. To properly backtest a Value-at-Risk model over a one year horizon, independent historic yearly stresses should be evaluated. In practice, the availability of historic data is often limited. Therefore, the model is often backtested using a rolling window. Instead of evaluating the stresses between year-end data, also stresses between January and January, February and February, until December to December are computed. This increases the probability of finding an extreme event, since more events are considered, and a shock that occured in February may have been recovered by the end of the year. If one does not correct for this effect, it can lead to a false rejection of the model. For example, it is possible that based on a rolling window, 3 extreme events in 40 years of data does not lead to a significant p-value (rejection of the model), but based on a fixed window, the model can be rejected. If extreme events are computed based on a rolling window, but are compared with p-values based on a fixed window, a model can be falsely rejected. Misspecifications and misinterpretations are different of nature. If a misspecification is found, the model should be adjusted. A misinterpretation can lead to a false rejection of the model. If a misinterpretation occurs, not the model itself should be adjusted, but the way it is evaluated should be adapted.

The second part of this thesis will show some pitfalls of the one-year horizon in Solvency II. One of the ideas behind the European Solvency II regulations is that it creates a level playing field. We will show that this is not the case, but the strictness depends on the distribution of the liabilities in time, and the investment strategy used. Although Solvency II is a huge improvement with respect to Solvency I since it is risk based, it can be improved by also looking at different time horizons.
We will distinct between within-horizon risk and end-of-horizon risk. Within-horizon risks reflects the risk that at a certain point in time, the market value of the assets is not sufficient to fund the market value of the liabilities. We will refer to this as a ruin. End-of-horizon risk, or default risk, refers to the risk that an insurance company does not have enough assets to make a payout, and it goes bankrupt. First, we will investigate an insurance company having one liability, at different points in time. The advantage of this simplified model is that we can compute probabilities of ruin and default analytically. We will also investigate different cashflows patterns, namely a constant cashflow over time, and a decreasing cashflows pattern. Ruin probabilities and default probabilities will be computed using simulations. We will add mean reversion to the model, and analyse the impact of the results on the level playing field of Solvency II.

The outline of this thesis is as follows. In Sections 2, 3, 4 and 5, we will quantify the effect of neglecting autocorrelation in residuals, assuming a wrong distribution of errors, neglecting clustered volatility, and erroneous scaling of VaR of a Jump Diffusion process respectively. Then, in section 6, we will study backtesting with a rolling horizon. We will compute the effect for AR(1) and Random Walk models, and illustrate the effect by evaluating different historic equity indices. In particular, we will look at the conclusions of an evaluation report of the equity stress parameter used by pension funds. We will propose a step by step approach for backesting with a rolling window in a correct manner.
In the second part of this thesis, we will start by introducing the terminology of time horizons and probability of ruin and default in section 7. Then, in section 8, we will analytically compute probabilities of ruin and default under different investment strategies and we will compare effects for different time horizons. In

section 9 we will study the effect of different cashflow models, and in section 10 we will add mean reversion to the model. We will show implications for the level playing field in section 11, and we will end with conclusions.

Part I
Statistical pitfalls in VaR models

2 Misspecification: Neglecting autocorrelation in residuals

In this section we treat a first possible misspecification of a VaR model, that of neglection of autocorrelation in the residuals (i.e. the deviation of the observed value from the one-step-ahead predicted value). We will start with a theoretical approach. We will quantify the effect of neglecting autocorrelation if data is generated by an MA(1) process, for different values of the autocorrelation parameter. Then we will consider the practical relevance of this issue by applying the model to interest rates.

For Solvency II purposes, we are interested in a stress that can occur over a one year horizon. Often in practice, models are calibrated on weekly, monthly or quarterly data. The stress found over a given time period is scaled up to a horizon of one year. An advantage of this method lies in the fact that the number of independent data periods increases. Often, a limited data history is available, and using for example monthly data instead of yearly data increases the number of data points with a factor 12. However, this method also has a potential hazard. If scaling is done incorrectly, this can lead to under- or overestimation of the stress over a one year horizon. One possible mistake that can be made is the neglection of autocorrelation in the (monthly) residuals. Positive autocorrelation in the errors will increase the stress, because a large error is more likely to be followed by another large error.

Here, we will quantify the effect of neglecting autocorrelation in the residuals. As pointed out in the Introduction, our approach will be to assume a true generating process, and an incorrect alternative data generating process. We will assume we have an infinite set of data available, and study the effect of analyzing the data under the incorrect assumptions. We will assume the true data generating process is an MA(1) process. We will study the impact of incorrectly assuming that the data follows a Random Walk. We assume the Random Walk is calibrated at time periods of $\frac{1}{k}$ year, and is scaled up to a yearly horizon to determine the yearly VaR. We study the MA(1) model because it is a simple model that contains the feature we are interested in, namely, autocorrelation in the residuals. Because it is a basic model, it is suitable to distill the effect we are interested in.

The specification of the MA(1) model is as follows:

$$\begin{aligned} y_t &= y_{t-1} + \epsilon_t + \rho\epsilon_{t-1} \\ \epsilon_t &\sim N(0, \sigma^2) \;\; i.i.d. \end{aligned} \tag{1}$$

The autocorrelation in the residuals is described by the parameter ρ. The value of the time series at time t depends on the error of the previous time step. If $\rho > 0$, an increase in y is more likely to be followed by another increase. If

$\rho < 0$, the effect of an increase is likely to be offsetted by a decrease in the next time step.
By iterating this equation, we can find a relation between the value of the time series at time t, and the value of the previous year $t-k$.

$$\begin{aligned} y_t &= y_{t-1} + \epsilon_t + \rho\epsilon_{t-1} \\ &= y_{t-2} + \epsilon_{t-1} + \rho\epsilon_{t-2} + \epsilon_t + \rho\epsilon_{t-1} \\ &= \ldots \\ &= y_{t-k} + \rho\epsilon_{t-k-1} + \sum_{n=1}^{k-1}(1+\rho)\epsilon_{t-n} + \epsilon_t \end{aligned} \tag{2}$$

Based on this relation, we can determine the one year standard deviation, taking into account the autocorrelation in the residuals.

$$\begin{aligned} E(y_t|y_{t-k}) &= y_{t-k} + \rho\epsilon_{t-k-1} \\ Var(y_t|y_{t-k}) &= \sum_{n=1}^{k-1}(1+\rho)^2\sigma^2 + \sigma^2 \\ &= \sigma^2(1+(k-1)(1+\rho)^2) \equiv \sigma^2_{ma}(k,\rho) \end{aligned} \tag{3}$$

Now assume we do not model the autocorrelation in the residuals. Instead, we will naively assume the data follows a Random Walk. Because the autocorrelation explains part of the variance in the data, the calibrated standard deviation of the error based on the calibration of the Random Walk will be higher than the standard deviation found by calibrating the MA(1) model. When we scale up the error to a horizon of one year, we will multiply the variance by the number of steps k. We can do this, because we assume the data follows a Random Walk. Therefore, all errors are independently and identically normally distributed, and the variance of the sum of k errors is equal to k times the variance of one error.

$$\begin{aligned} \tilde{y}_t &= \tilde{y}_{t-1} + \eta_t \\ \eta_t &= \epsilon_t + \rho\epsilon_{t-1} \\ Var(\eta_t) &= \sigma^2(1+\rho^2) \\ E(\tilde{y}_t|\tilde{y}_{t-k}) &= \tilde{y}_{t-k} \\ Var(\tilde{y}_t|\tilde{y}_{t-k}) &= k\sigma^2(1+\rho^2) \equiv \sigma^2_{rw}(k,\rho) \end{aligned} \tag{4}$$

If the security level is α (for example $\alpha = 0.005$), the calibrated stress value based on the Random Walk will be $\sigma_{rw}\Phi^{-1}(\alpha)$. The probability p that an extreme event will be found over a one year horizon, based on the true MA(1) process, equals:

$$\begin{aligned} \sigma_{ma}\Phi^{-1}(p) &= \sigma_{rw}\Phi^{-1}(\alpha) \\ p &= \Phi\left(\Phi^{-1}(\alpha)\frac{\sigma_{rw}}{\sigma_{ma}}\right) \end{aligned} \tag{5}$$

Under the assumption that the model is specified correctly, the probability to find an extreme event in n years of data is $1-(1-\alpha)^n$. Due to the incorrect assumption of the Random Walk, the probability however is $1-(1-p)^n$. This results in a relative factor R of:

$$R = \frac{1-(1-p)^n}{1-(1-\alpha)^n} \tag{6}$$

Example 1: Step by step calculation

Assume $k = 12$ (monthly data), $\sigma = 1$, $\rho = 0.3$ $n = 30$, and $\alpha = 0.005$.

The calibrated standard deviations based on the Random Walk and the MA(1) process will be:

$$\begin{aligned} \sigma_{rw} &= \sqrt{(1+0.3^2)\cdot 12} \approx 3.62 \\ \sigma_{ma} &= \sqrt{1+11\cdot(1+0.3)^2} \approx 4.43 \end{aligned} \tag{7}$$

The calibrated 1-in-200 stress based on the Random Walk is $\Phi^{-1}(0.005)\cdot 3.62 \approx -9.32$. The true 1-in-200 stress is larger and equal to $\Phi^{-1}(0.005)\cdot 4.43 \approx -11.40$.

The probability of finding an extreme event over a one year horizon is $\Phi(\Phi^{-1}(0.005)\frac{3.62}{4.43}) \approx 1.8\%$. The probability of finding at least one extreme event will be $1-(1-1.8\%)^{30} \approx 41\%$, while we expect to find an extreme event with a probability of $1-(1-0.5\%)^{30} \approx 14\%$.

If the correct MA(1) model had been calibrated, the probability to find at least one extreme event in 30 years of data would have been 14%. Because autocorrelation has been neglected, and a Random Walk model has been calibrated to determine the 1-in-200 stress, the stress found is too low, and the probability to find an event more extreme than the proposed stress increases from 14% to 41%. This is a relative increase of $R = \frac{41\%}{14\%} \approx 3$.

Results for different values of ρ, k and n

In the graph we see the factor R for different values of k and different time intervals n, assuming $\rho = 0.1, \rho = 0.3$ and $\rho = 0.5$. We see that R is inversely related to n. This can be explained by the fact that when n is large, the probability of finding an extreme event is larger, and can increase less, because a probability can not exceed one.
We also see that the effect increases with k. This is intuitively clear, since the error of neglecting autocorrelation accumulates with every time step.
We see that if $k = 1$, the factor is smaller than 1. This is caused by the fact that the stress based on the Random Walk is overestimated, since not all information

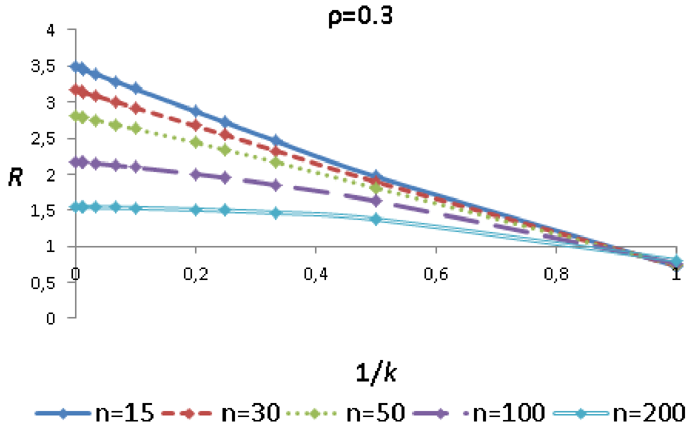


Figure 2: Relative increase R in probability of finding an extreme event if data generated by an MA(1) process, is analyzed as if it was generated by a Random Walk process. R can take values larger than 3. Different lengths of historic data are considered. Confidence level $\alpha = 0.005$ and $\rho = 0.3$. The relative effect increases if the model is calibrated on a shorter dataset. An increase in the frequency of the data k leads to an increase in the underestimation of risk.

is taken into account. If $k = 1$, errors are not scaled up, because the model is already based on yearly data, and therefore the effect of the autocorrelation does not play a role, only the overestimation of the standard deviation of the residuals.
If the size of the autocorrelation in the residuals increases, R increases. This is also clear, because a large error has higher impact on the next time step when the autocorrelation increases.

We see the effect of neglecting autocorrelation is substantial. Therefore, it is very important that autocorrelation of residuals is investigated when a model is based on fractional year data. To show the practical relevance, and study realistic parameter values, we will apply the model to interest rates.

Example 2: Practical relevance for interest rates

In this example, we will show the effect of autocorrelation in the residuals when interest rates are modeled. We will assume Euro swap rates to follow an MA(1) process (true data generating process), or Random Walk (erroneous assumption). Note that interest rates are often modeled by a Vasicek model, which in discrete time, corresponds to an AR(1) model, which contains mean reversion. Here, we will neglect mean reversion, but instead, we will include the moving average term in the time series. We use monthly Euro swap yields with a ma-

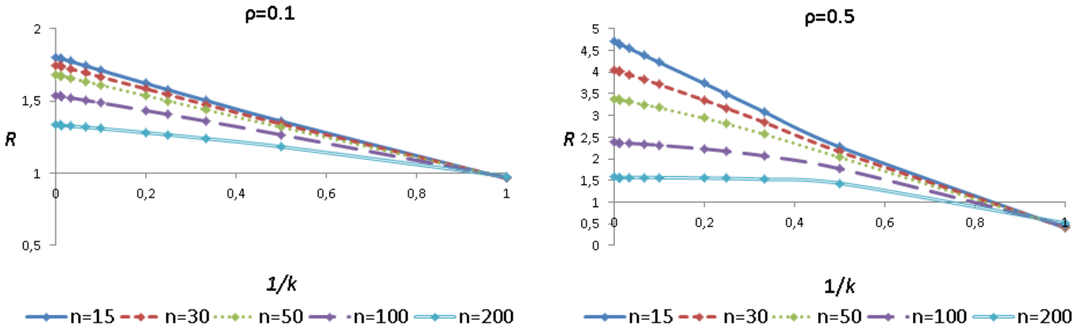


Figure 3: Same specifications as in Fig. 2, except $\rho = 0.1$ (left) and $\rho = 0.5$ (right) instead of $\rho = 0.3$. An increase in ρ leads to an increase in R.

turity of one year, from August 2002 to April 2012.

If we calibrate the Random Walk model, we find a monthly standard deviation of 0.208%, which results in a yearly 1-in-200 stress of 1.85%. If we calibrate the MA(1) model we find $\rho = 0.37$, and a monthly standard deviation of 0.191%. If we scale up this standard deviation, taking the autocorrelation coefficient into account, we find a yearly 1-in-200 stress of 2.29%, which is significantly higher than we one based on the Random Walk.

Neglecting the moving average term will lead to an underestimation of the yearly stress. This is illustrated in Figure 4, where we compare the 99.5% stresses of those two models with the observed changes in the swap rate between year ends. The most extreme decrease of interest rates took place between YE 2007 and YE 2008 (-2.15%). This event falls within the stress based on the MA(1) model, but does not fall within the stress based on the Random Walk.

Conclusions

In Example 2, we found a value of $\rho = 0.37$ for the autocorrelation parameter in monthly swap rates. Therefore, the effects shown in Figure 2 ($\rho = 0.3$) correspond to realistic parameter values. Neglecting autocorrelation in the residuals can therefore lead to an increase in the probability of finding an extreme event by a factor 3. This is a significant effect. Therefore, the possible pitfall of neglecting autocorrelation in residuals is of practical relevance.

Figure 4: Yearly changes in 1-year swap rates, compared to 1-in-200 year stresses based on a Random Walk model and an MA(1) model. Changes in monthly 1-year swap rates experience autocorrelation ($\rho = 0.37$). Neglecting this autocorrelation results in an underestimation of the yearly 1-in-200 stress. If autocorrelation is modelled explicitly, the model captures all historic stresses between year-ends. This example shows the practical relevance of the theoretically computed effects in Fig. 2 and Fig. 3.

3 Misspecification: Wrong distribution errors

In this section, we will investigate the effect of a misspecification of the distribution of errors. A wrong choice of the distribution of errors is a possible source of an under- or overestimation of a Value-at-Risk. Often, normality of errors is assumed. In practice, distributions of observed errors often have heavy tails, compared to a normal distribution. If this is neglected, and normality is assumed, risk can be underestimated. A well known example of a distribution with heavy tails is the Student-t distribution. If data is normally distributed, and the sample size of the data is small, the mean can be shown to be Student-t distributed. This distribution is therefore used to perform a t-test, to test if a parameter significantly differs from some predetermined value. Due to the heavy-tailed property of this distribution, the Student-t distribution is often used in practice to model heavy-tailed data. In this section, we will assume a yearly stress follows a Student-t distribution (the data generating process), but the stress is modeled by a normal distribution instead.

Assume a yearly stress follows a Student-t distribution with mean zero and ν degrees of freedom. The standard deviation of the stress is then equal to:

$$\sigma_\nu = \sqrt{\frac{\nu}{\nu - 2}} \tag{8}$$

The parameter ν determines the heaviness of the tails. A low value of ν corresponds to heavy tails. If ν approaches infinity, the distribution converges to a normal distribution.

Now, assume we model the stress by a normal distribution with the same standard deviation σ_ν, and mean zero. We consider the confidence level α. The stress we will propose, based on the incorrect assumption of a normal distribution, is $\sigma_\nu \Phi^{-1}(\alpha)$. The probability p of observing a stress event under the true Student-t distribution is:

$$\begin{aligned} T^{-1}(p,\nu) &= \sigma_\nu \Phi^{-1}(\alpha) \\ p &= T(\sigma_\nu \Phi^{-1}(\alpha),\nu) \end{aligned} \tag{9}$$

$T^{-1}(p,\nu)$ is the inverse cumulative Student-t distribution, the equivalent of Φ^{-1} for the normal distribution. The relative increase R of finding an extreme event in n years of data (assuming independence of the yearly errors) equals:

$$R = \frac{1-(1-p)^n}{1-(1-\alpha)^n} \tag{10}$$

Example

Assume $\nu = 5$, $n = 1$ and $\alpha = 0.005$. The standard deviation of the Student-t distribution equals:

$$\sigma_\nu = \sqrt{\frac{5}{5-2}} \approx 1.29. \tag{11}$$

The calibrated 1-in-200 stress, based on the normal distribution is
$\sigma_\nu \Phi^{-1}(0.005) \approx -3.32$.
The probability of finding an extreme event, based on the true Student-t distribution equals $T(1.29 \cdot \Phi^{-1}(0.005), 5) \approx 1.04\%$. This is a relative increase of $R = \frac{1.04\%}{0.5\%} \approx 2.09$.

In Figure 5, the probability density of the normal distribution and of the Student-t distribution with five degrees of freedom are plotted. We can see that the probability density of the Student-t distribution is higher in the tail. Therefore, the probability of finding an event larger than the stress (-3.32 in the example) is higher.

In Figure 6 we can see that the effect of erroneously assuming normality is higher for lower degrees of freedom of the Student-t distribution, and that the effect if higher for more extreme confidence intervals. A lower degree of freedom corresponds to fatter tails. In the limit of $\nu \to \infty$, the Student-t distribution converges to a normal distribution, and the effect vanishes.
If the confidence interval is set to 10%, assuming a normal distribution can lead to an overestimation of the stress. This is caused by the fact that the Student-t distribution has fatter tails, but the probability density is also more peaked. As we can see in Figure 5, there is a range where the density of the normal

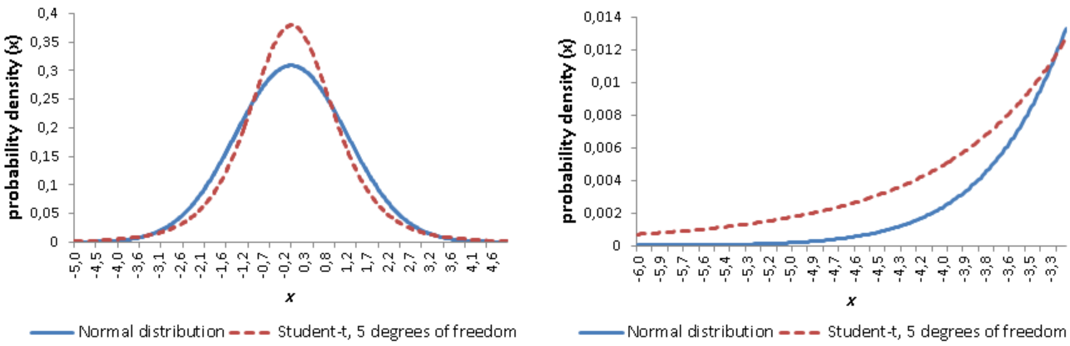


Figure 5: Probability density of a normal distribution and a Student-t distribution with 5 degrees of freedom, both with a standard deviation of 1.29. The tails of the Student-t distribution contain more probability mass compared to the normal distribution.

distribution is higher. In the case of $\alpha = 10\%$, this effect is more important than the higher mass of the Student-t distribution in the tails.

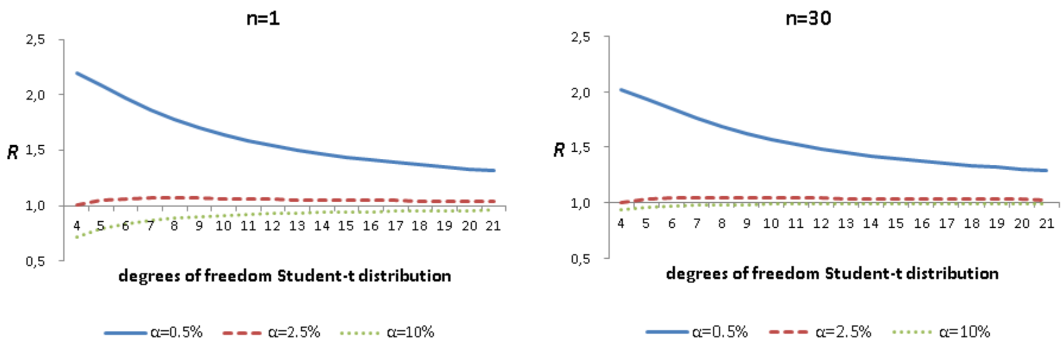


Figure 6: Relative increase R in probability of finding an extreme event if data is generated by a Student-t distribution, but analyzed as if the errors are normally distributed, for confidence intervals $0.5\%, 2.5\%$ and 10%. If the heavy tails are neglected, the probability of finding an extreme event can increases with a factor 2 for $\alpha = 0.5\%$. Since there is less probability mass in the middle of the Student-t distribution, a ratio smaller than one is found for $\alpha = 10\%$. Therefore, neglecting heavy tailedness of a distribution of errors can lead both to an underestimation and overestimation of stress. The case $n = 1$ (one Student-t distributed yearly error) and $n = 30$ (30 Student-t distributed yearly errors) are considered. The value of n has a limited effect on the results.

The above analysis is based on one Student-t distributed variable. It is also interesting to investigate the effect if monthly, or quarterly data is considered, and the error is scaled up to a yearly stress. To investigate this, we simulated a series of Student-t distributed variables. We computed the variance of the errors, and scaled up this variance by a factor k. For independent, identically distributed normally distributed variables, the square root of time can be applied to scale up the standard deviation. Then, we computed non-overlapping sums of k random Student-t variables. We computed the probability of finding a value more extreme than the 1-in-200 stress based on the scaled up variance. This results in a factor R. In the graph we can see the effect of different values of k, and different degrees of freedom of the Student-t distribution. The case $k = 1$ corresponds to the analysis above. We see that again, the effect decreases when the number of degrees of freedom increases, because a lower degree of freedom corresponds to more heavy tails. Further, we see that the effect decreases with increasing k. This is interesting. If Student-t distributed variables are summed, the heavy tails partly vanish, and the resulting distribution will become more close to a normal distribution.

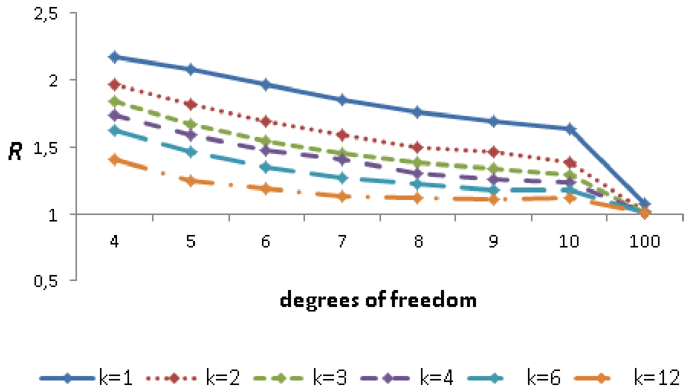


Figure 7: Relative increase R in probability of finding an extreme event ($\alpha =$ 0.5%) if data is generated by a Student-t distribution, but analyzed as if the errors are normally distributed. In contrast to Fig. 6, the data now consists of k Student-t distributed errors. Under the assumption of normality, the square root of time rule is used to scale up the variance. The effect of neglecting the heavy tails of the Student-t distributed errors decreases with increasing degrees of freedom, since the Student-t distribution converges to a normal distribution for infinite d.o.f. The effect decreases with increasing k.

Assuming a normal distribution, while the data is Student-t distributed, can lead to an underestimation of the stress. The probability of finding an extreme event can increase with a factor of more than 2. If a longer data series is backtested, for example 30 years, the effect is only significant if one is interested in severe stresses, for example the stress corresponding to a confidence level of 0.5%. For less severe stresses, like the 2.5% confidence level, the effect is limited. If the sum of Student-t variables is considered, the effect of the heavy tails decreases.

4 Misspecification: Neglecting clustered volatility of residuals

In this section, we investigate the effect of neglecting clustered volatility. First, we will theoretically derive the effect of neglecting clustered volatility if data is generated by GARCH(1,1) model, but is assumed to be generated by a Random Walk. Then, we will consider the practical relevance of this pitfall, by quantifying the effect for realistic parameter values.

In time series models, it is often assumed that the volatility of the errors is constant (homoscedasticity). In practice, this is often not the case. When a large shock has occurred, for example a large movement in a stock market index, it can be more likely that in the next time step, again a large movement will occur. The family of GARCH models is a class of models that incorporates this effect. In a GARCH model, the volatility of an error depends on the volatility and realized error of previous time steps. Periods of low volatility can be followed by periods with higher volatility. If one neglects this effect, stress can be underestimated. In this section we will quantify this effect for GARCH models. We will investigate a GARCH(1,1) model. In a GARCH(p,q) model, the volatility of an error depends on the error and volatility of the previous p and q time steps. We choose to assume the most basic GARCH model, in which the volatility is only related to the error and volatility of the previous time step. The specification of the GARCH(1,1) model is as follows:

$$\begin{aligned} y_t &= y_{t-1} + \epsilon_t \\ \epsilon_t &= \sigma_t z_t \\ z_t &\sim N(0,1) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned} \tag{12}$$

The long term expected volatility of y equals:

$$E(\sigma_t^2) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \tag{13}$$

Therefore, we will only consider combinations of parameters where $\alpha_1 + \beta_1 < 1$.

The distribution of the errors of a Garch(1,1) model is leptokurtic (heavy tailed). Heavy tails can also cause underestimation of stress, as we have shown in Section 3. If data is Student-t distributed (which is heavy tailed), but the data is treated as if it is normally distributed, this can lead to an increase in the probability of finding an extreme event by a factor 2. Therefore, when analyzing the effect of neglecting clustered volatility, we will have to investigate which part of the effect is caused by the leptokurtic property of the errors, and which part results from the clustered volatility itself.

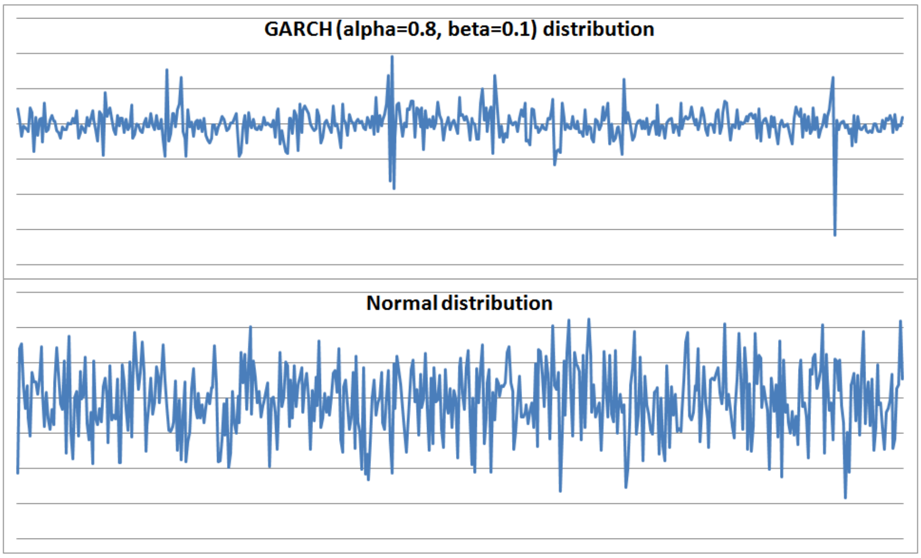


Figure 8: 500 Randomly generated numbers from a normal distribution and a GARCH distribution ($\alpha_1 = 0.8, \beta_1 = 0.1$).

We used the statistical program R to simulate a GARCH(1,1) process. The modeling process is graphically shown in Figure 9. We considered k steps per year, and simulated $N = 1.000.000$ years. For every year n, we simulated a set of k GARCH errors $\epsilon_{n,i}$. We computed the variance σ^2 of the errors, and scaled up the variance with a factor k to naively predict the yearly stress. We summed up k errors per year, and evaluated how often this value exceeded the 1-in-200 stress based on the scaled variance (assuming independent normally distributed errors). This results in an estimate for the probability of finding an event more extreme than the proposed stress:

$$p = \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}\Big\{ \sum_{j=1}^{k} \epsilon_{n,j} > \sigma\sqrt{k}\Phi^{-1}(1-\alpha) \Big\}. \tag{14}$$

Then, the ratio between the probability p, and the confidence level α is computed:

$$R = \frac{p}{\alpha}. \tag{15}$$

If $R > 1$, the yearly stress is underestimated. The code used can be found in Appendix C.

In Table 2, results are shown for $k = 12$ and different values of α_1 and β_1. The relative factor R is shown. For $\alpha_1 = 0$, there is no effect. If $\alpha_1 = 0$, the volatility of the error only depends on the volatility of the previous error, and does not depend on the realized error of the previous time step. Therefore, the

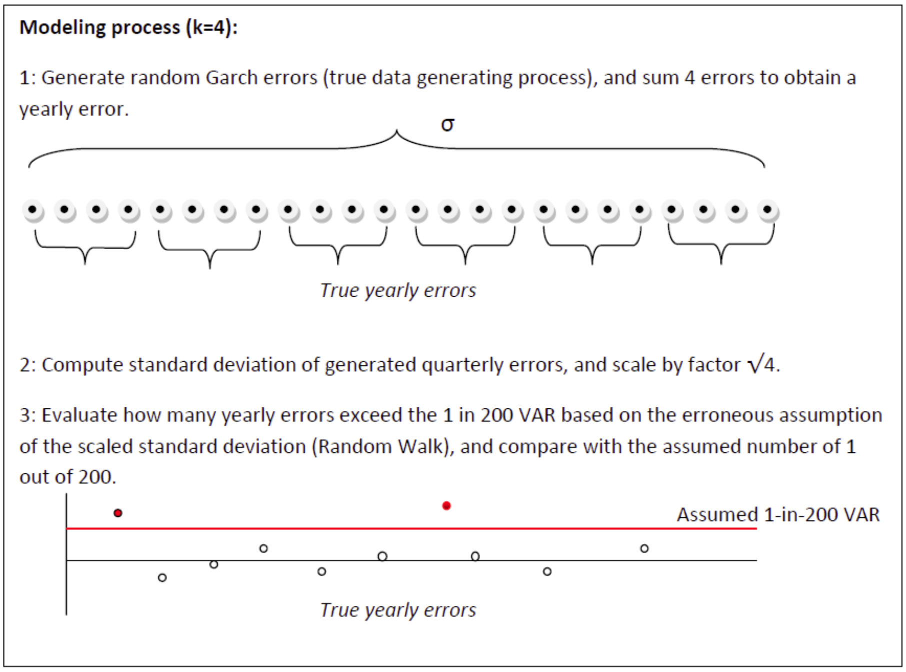


Figure 9: Modeling process performed in R.

process is deterministic, and will converge to a situation where the standard deviation is fixed and equal to $\sigma_t^2 = \frac{\alpha_0}{1-\beta_1}$. The GARCH process reduces to a Random Walk with independent normally distributed errors. If $\alpha_1 \neq 0$, both α_1 and β_1 have an increasing effect on the probability of finding an extreme event. If α_1 increases further, R decreases. The relative factor R is graphically shown in Figure 10, for $k = 12$ (monthly data), and $k = 4$ (quarterly data). The maximum effect we find is of order two. The effect for $k = 4$ is slightly larger than the effect for $k = 12$.

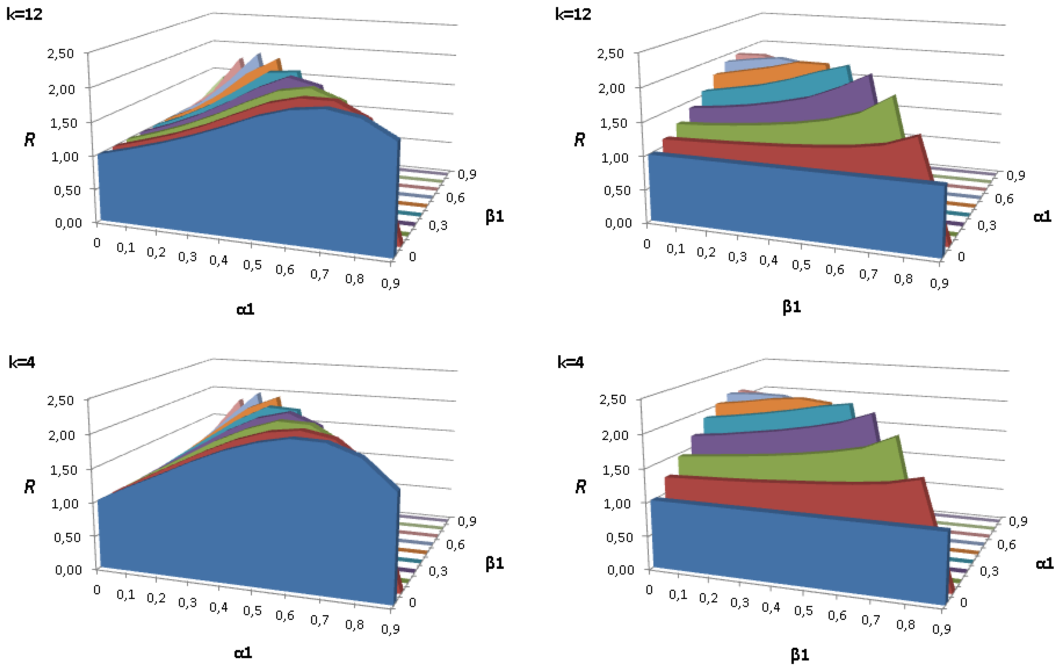


Figure 10: Effect of different specified GARCH(1,1) models on the probability of finding a 1-in-200 event over a one year horizon. The relative increase in probability R is shown. Parameter β_1 has an increasing effect on the probability of finding an extreme effect, a higher value of α_1 has an increasing effect, but the effect decreases if α_1 rises further. The effect can partly be explained by the leptokurtic property of the GARCH-errors. The graph shows the effect for monthly errors ($k = 12$, upper graphs), and quarterly errors ($k = 4$, lower graphs) seen from two different angles. Effects for $k = 4$ are slightly larger.

The effect of the heavy tails of the distribution

As we have seen in Section 3, an underestimation of stress can also be caused by heavy tailedness of the distribution of errors. To estimate which part of the underestimation is caused by the leptokurtic property of the GARCH errors, we assume the errors are Student-t distributed, and use the results of Section 3.

If $\alpha_1 = 0.4$, $\beta_1 = 0.1$ and $k = 12$, we find a relative factor of 2. The kurtosis of the simulated GARCH errors is equal to 68. If we assume that the errors are Student-t distributed, this corresponds to a number of degrees of freedom slightly above 4. For k=12, and 4 degrees of freedom, the effect of the heavy tails of the Student-t distribution corresponds to a factor 1.41. (see Section 3)

	α_1										
β_1		**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
	0	1.00	1.12	1.25	1.39	1.57	1.76	1.89	1.94	1.86	1.60
	0.1	1.00	1.13	1.27	1.44	1.65	1.84	1.96	1.96	1.72	
	0.2	1.00	1.14	1.30	1.50	1.73	1.93	2.01	1.84		
	0.3	1.00	1.16	1.35	1.59	1.85	2.05	1.94			
	0.4	1.00	1.18	1.41	1.70	2.00	2.04				
	0.5	1.00	1.22	1.50	1.88	2.13					
	0.6	1.00	1.26	1.65	2.11						
	0.7	1.00	1.34	1.93							
	0.8	1.00	1.51								
	0.9	1.00									

Table 2: Underlying numbers of Figure 10, for $k = 12$ (upper part of the figure).

If $\alpha_1 = 0.6$, $\beta_1 = 0$ and $k = 12$, we find a relative factor of 1.89. The kurtosis of the simulated GARCH errors is equal to 12.5. If we assume that the errors are Student-t distributed, this corresponds to 4.5 degrees of freedom. For k=12, and 4.5 degrees of freedom, the effect of the heavy tails of the Student-t distribution corresponds to a factor of 1.33.

We see that the heavy tails of the distribution of the errors explain approximately 40% of the increased probability of finding a 1-in-200 event. Since the maximum factor R is of order two, the effect of clustered volatility is significant, but clearly lower than the effect of neglecting autocorrelation in the residuals we found in Section 2.

Realistic parameter values

On page 129 of the Handbook of Financial Time Series [8], a GARCH(1,1) model has been fitted on monthly and daily Microsoft and S&P data. The parameters found are all around 0.1 for α and around 0.9 for β. Based on our results, for these parameter values, scaling up the monthly errors by the square root of time can lead to an increase in the probability of finding an extreme event by a factor around 1.5. As we have shown, part of this effect is caused by the leptokurtic property of the errors. For these realistic parameter values, the effect of neglecting volatility is significant. However, the effect is small compared to the effect of neglecting autocorrelation in the residuals we found in Section 2. There, we found a factor 3 for realistic parameter values.

5 Misspecification: The use of the square root of time rule for a Jump Diffusion process

The geometric Brownian motion is a well known and commonly used stochastic process. The famous Black-Scholes-Merton model assumes stock returns follow a Brownian motion, and uses this assumption to price for example equity options. Given a certain period in time, the stock return between time 0 and time t is normally distributed with a standard deviation proportional to the square root of t. Therefore, the distribution of the returns is symmetric. In practice, one often observes that the distribution of returns is not symmetric, and once in a while, an equity index crashes. To describe this non-symmetric and heavy tailed behaviour, Merton [9] introduced a model that adds an extra term to the geometric Brownian motion. He assumes that once in a while a jump occurs. Jumps are Poisson distributed, and if a jump occurs, the index decreases by a certain percentage.

In this section, we will analyze the impact of scaling up the Value-at-Risk of a time variable following a Jump Diffusion process by the square root of time rule. Work has been performed on this subject by Danielsson e.a [4]. This article focuses on a time scale inspired by the Basel regulations for banks. The effect is investigated of scaling up a 1-day VaR to a 10-day VaR by the square root of time rule. We will take this article as a starting point, and first we will reproduce their results for this time scale. Then, we will investigate the impact if a monthly VaR is scaled up to a yearly VaR, since this timescale is more relevant for Solvency II. We will see that the change to a longer time scale significantly changes the results. We will identify three different ranges of jump frequencies, all with their own characteristic behavior. At the end of this section, we will consider realistic parameter values.

In previous sections, we computed ratios between probabilities of finding an extreme event. In this section we will compute ratios of Value-at-Risk, to be able to make a comparison with the results in the article of Danielsson e.a.

The Jump Diffusion model has been studied intensively in many papers. Kou [10] has extended the original model of Merton, such that it contains both upward and downward jumps, to model option prices. We use the specification of the Jump Diffusion model as presented in [4]. There, the Jump Diffusion process is specificated in its most basic form. The size of a jump, if it occurs, is fixed by the parameter δ, without a random term added to it:

$$dy_t = \left(\mu + \frac{1}{2}\sigma^2\right)y_t dt + \sigma y_t dW_t - (1-\delta)y_t dq_t. \tag{16}$$

Parameter μ represents a drift term, σ equals the standard deviation of the underlying Brownian motion W_t, and $(1-\delta)$ is the fraction of y that is "lost" if a crash occurs. The Poisson process dq_t has value 1 with probability λdt and is zero otherwise. If $\delta = 1$ or $\lambda = 0$, the process reduces to a geometric Brownian motion. The value of y_T given y_t equals:

$$y_T = y_t e^{\mu(T-t)+\sigma(W_T-W_t)}\delta^{q_T-q_t}. \tag{17}$$

Defining $x = \log(y)$, and given a random realization of the Brownian Motion ($R_{normal}(0, \sigma^2\Delta t)$) and the Poisson process ($R_{poisson}(\Delta t, \lambda)$), the stress over a time period Δt equals:

$$x(t) - x(t - \Delta t) = \mu\Delta t + \sigma R_{normal}(0, \sigma^2\Delta t) + R_{poisson}(\Delta t, \lambda)\log(\delta). \tag{18}$$

Using this equation, we can simulate daily (monthly) returns in groups of 10 days (12 months). Adding up those groups, we can compute an approximation of the 1 day (1 month) and 10 day (12 month) Value-at-Risk at a certain confidence level α. We will compare the 10 day (12 month) VaR with the 1 day (1 month) VaR scaled up with the square root of time. Then, we can see wether the square root of time rule underestimates or overestimates the VaR. We compute the ratio between the VaR's, analogous to the article of Danielsson:

$$ratio = \frac{VaR_\alpha(k\eta)}{\sqrt{k}VaR_\alpha(\eta)}, \tag{19}$$

where η equals 1 day (1 month), and $k = 10$ ($k = 12$). Simulations and computations are performed in the statistical program R. The code used can be found in Appendix C.

The results of the simulations can be found in Figure 11 and Table 3 and Table 4. The left graph of Figure 11 is based on the parameters of the article of Danielsson. We find similar results, the effects are small. The square root of time rule can lead to a small underestimation of the risk. The effect increases when the frequency of the jumps increases. The frequency of the jumps is low compared to the time scale: the average time between two jumps is typically in the order of tens of years. The time scale considered is 1 to 10 days. The probability that a jump will occur within those days is low compared to the confidence interval of $\alpha = 0.01$.

If we consider the second graph in Figure 11, and Table 3, we see different behavior. We have considered average times between jumps in the order of tens of years, and also 200 and 400 years, to explore limiting behavior. The confidence interval is set to $\alpha = 0.005$, consistent with the Solvency II framework. We scaled up the monthly simulated VaR by a factor $\sqrt{12}$, and compared it with the simulated yearly VaR. We see that for high jump frequencies, the square root of time rule overestimates the VaR, for intermediate frequencies it is underestimated, and then, for lower frequencies, the effect fades away. Furthermore, for this Solvency II time scale, effects are more extreme than for the Basel inspired time scale. We will consider the three regimes and their behavior separately.

High jump frequencies

If the jump frequency is high, for example $1/\lambda = 10$ years, the yearly VaR is overestimated. This can be explained as follows.
If δ is smaller than one, the size of the jump is typically much higher than the size of a deviation caused by the Brownian motion. The probability a jump will occur within 1 month is approximately 1/120, and falls within the 1-in-200 event considered. If we scale up this stress by using the square root of time rule, this results in a yearly stress of $(1 - \exp(\log(d)\sqrt{12}))$.

The probability that 1 jump occurs within 1 year, and the probability that 2 jumps occur within 1 year, both fall within the confidence interval considered. Therefore, if d is small and dominates the 1-in-200 stress, the yearly stress is approximately $(1 - d^2)$. Therefore, the ratio between the VaR's (defined in equation 19) is approximately:

$$ratio \approx \frac{(1 - d^2)}{1 - \exp(\log(d)\sqrt{12})}. \tag{20}$$

For small values of d, this results in a number close to one. Both on a monthly basis and on a yearly basis, the stress will be approximately 100%, and therefore they are almost equal. For intermediate regions of d, this will lead to a number smaller than one, and therefore, the yearly stress is overestimated. As d approaches 1, the jump size is small compared to the variation caused by the ordinary geometric Brownian motion. Therefore, the process is dominated by the ordinary geometric Brownian motion, for which the square root of time rule is appropriate, and the ratio between the VaR's converges to 1.

Intermediate jump frequencies

For intermediate frequencies, for example $1/\lambda = 30$ years, the probability of a jump falls outside the monthly VaR, but inside the yearly VaR. Therefore, the monthly VaR is dominated by the geometric Brownian motion. When this stress is scaled up by the square root of time rule, it results in a yearly stress of $1 - \exp(-\sigma\sqrt{12}\Phi^{-1}(0.995))$. For small d, the yearly VaR is dominated by the jump, and is approximately $1 - d$. Therefore, for small d the ratio will be approximately:

$$ratio \approx \frac{1 - d}{1 - \exp(-\sigma\sqrt{(}12)\Phi^{-1}(0.995))}. \tag{21}$$

This results in a ratio larger than 1, and therefore the VaR is underestimated by the square root of time rule.

Low jump frequencies

If the frequency of the jumps is very low, for example $1/\lambda = 400$ years, the

Table 3: Scaling up a monthly VaR to a yearly VaR: Results for $\alpha = 0.005$, $\sigma = 0.15/\sqrt{12}$, $\mu = 0$, and $k = 12$. Underlying numbers of Fig. 11.

		δ					
$1/\lambda$		**0.01**	**0.2**	**0.4**	**0.6**	**0.8**	**1**
	10	0.99	0.87	0.77	0.70	0.76	1.00
	12	0.99	0.86	0.74	0.68	0.78	1.00
	15	0.99	0.85	0.73	0.67	0.84	1.00
	17	2.33	1.98	1.61	1.23	0.92	1.00
	20	2.64	2.23	1.80	1.37	0.97	1.00
	30	2.86	2.39	1.90	1.42	1.01	1.00
	50	2.97	2.45	1.92	1.38	1.01	1.00
	200	1.40	1.40	1.34	1.10	1.00	1.00
	400	1.07	1.07	1.07	1.04	1.00	1.00
	∞	1.00	1.00	1.00	1.00	1.00	1.00

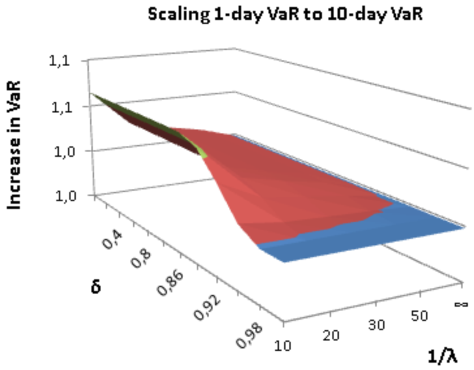
Table 4: Scaling up a 1-day VaR to a 10-day VaR: Results for $\alpha = 0.01$, $\sigma = 0.15/\sqrt{250}$, $\mu = 0$, and $k = 10$. Underlying numbers of Fig. 11.

		δ											
$1/\lambda$		**0.01**	**0.2**	**0.4**	**0.6**	**0.8**	**0.83**	**0.86**	**0.89**	**0.92**	**0.95**	**0.98**	**1.0**
	10	1.07	1.07	1.07	1.07	1.07	1.06	1.06	1.06	1.04	1.01	1.00	1.00
	20	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.02	1.00	1.00	1.00
	30	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.01	1.00	1.00	1.00
	50	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.00	1.00	1.00
	∞	1.00	1.010	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

probability of a jump falls both outside the 1-in-200 event on a monthly and on a yearly basis. This is comparable with the situation investigated by [4], where a 1-day VaR is scaled up to a 10-day VaR, and $1/\lambda$ is in the order of tens of years. Therefore, the influence of the jumps is small, and only small underestimations of the yearly VaR are found.

Realistic parameter values

The question rises which parameter values are realistic. Andersen and Andreasen [11] have calibrated a Jump Diffusion model on options on the S&P 500 index, and found a jump frequency of approximately once per 11 years. The parameter δ can be deducted from other calibrated parameters and is approximately 0.45. He et. al. [12] have performed a similar analysis and found similar results. Scott [13] has calibrated a Jump Diffusion model on the S&P 500 index itself, and found a jump frequency of approximately once per 15 years. The parameter δ also has to be deducted from other parameters, and is approximately 0.85. This value is higher than the value calibrated on option prices,



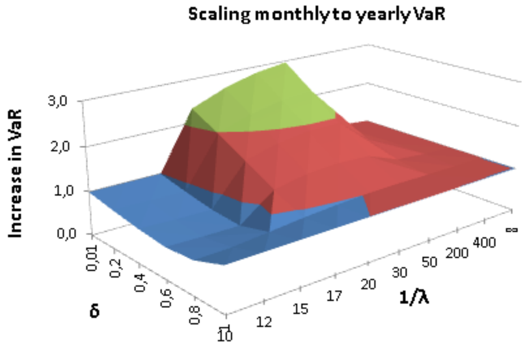


Figure 11: The ratio between the true VaR of a Jump Diffusion process, and the VaR obtained by scaling up the VaR of a shorter time horizon by the square root of time rule. If the ratio is larger than one, the VaR is underestimated by the square root of time rule. Parameter α represents the confidence level, k the number of steps of scaling, and σ equals the standard deviation of the underlying Brownian Motion of the Jump Diffusion process. Parameter μ represents the drift of the Brownian motion, δ the loss if a jump occurs, and $1/\lambda$ the frequency of the jumps. Up: Results for $\alpha = 0.01$, $\sigma = 0.15/\sqrt{250}$, $\mu = 0$, and $k = 10$. Scaling up the VaR from 1 to 10 days can lead to an underestimation of the VaR, but effects are small. Down: Results for $\alpha = 0.005$, $\sigma = 0.15/\sqrt{12}$, $\mu = 0$, and $k = 12$. Scaling up the monthly VaR to a yearly VaR can both lead to an under- and overestimation of the VaR. Differences are caused by the relative size of the time scale considered compared to the frequency of the jumps. We can distinguish three regimes, if the VaR is scaled from timescale t_1 to a longer timescale t_2. The jumps can significantly influences both t_1 and t_2 (high jump frequencies), only t_2 (intermediate jump frequencies), or none of the two (low jump frequencies). In the upper graph, we only see the regime of low jump frequencies. In the lower graph, we see all three regimes.

and therefore corresponds to a lower average jump size.

A jump frequency of $1/\lambda = 11$ years and $\delta = 0.45$ leads to an overestimation of the yearly VaR by a approximately 33% (i.e. a ratio of approximately 0.75). If $1/\lambda = 15$ years, it is close to the border between the regime of high frequencies and the regime of intermediate frequencies. Therefore, realistic parameter values can result both in an underestimation and an overestimation of the stress. The size of the effect can be substantial. Therefore, the practical relevance of this phenomena is high. If a Jump Diffusion model is used to determine a one year stress for equity indices, this model should be calibrated on yearly data instead of weekly or monthly data, to prevent a modeling error that results in an incorrect stress.

6 Misinterpretation: Backtesting with a rolling window

6.1 Introduction

In previous sections, we have investigated the effect of misspecifications of VaR models. In this section, we treat a misinterpretation, namely backtesting with a rolling window.

If a model is backtested, historical stress events are compared with a proposed stress. For example, if a decline of the market value of equity of 40% is proposed as a 1-in-200 stress event over a one year horizon, one can backtest this stress by computing historic declines of equity indices such as the MSCI World index. If non-overlapping yearly returns are evaluated, on average, only once every 200 year such a stress may occur. If significantly more events have occurred, the stress parameter does not correspond to a 99.5 % Value-at-Risk, and a higher stress should be used.

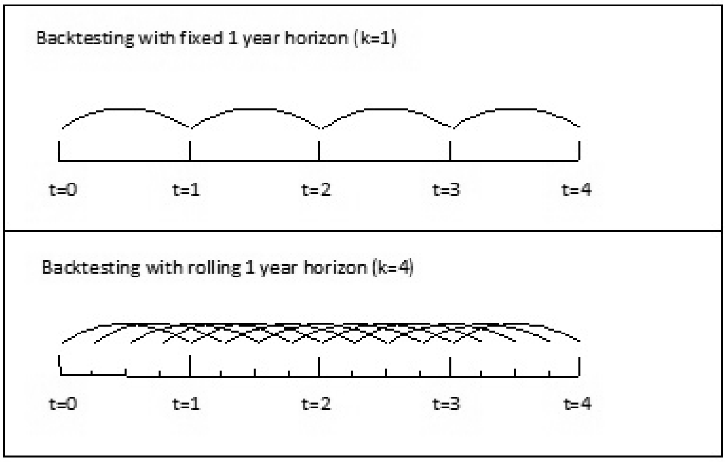


Figure 12: Up: Backtesting with a fixed window. Only non-overlapping yearly interval are considered. Down: Backtesting with a rolling quarterly window. The yearly events considered overlap each other.

In order to backtest a model by this rule of thumb (a 1-in-200 stress may occur on average only once per 200 year), a long history of observations must be available, and we have to compare non-overlapping yearly events with a proposed stress parameter. For example, if we want to backtest an equity stress parameter, we can choose to use returns between year ends, which are all non-overlapping. Since we are interested in a 99.5% stress, we need a multiple of 200 years of data, for example 2000 years, or even 20000 years of data. In practice, only a limited amount of data is available. For equity markets, around 200 years of data may be available, and other markets, like the market for inflation linked swaps, are much younger. Since a limited amount of data is available, often not only non-overlapping historic event are evaluated, but also overlapping yearly events, by using a rolling window. For example, the stress between January 1993 and January 1994, and the stress between March 1993 and March 1994 are evaluated. A rolling window is used in stead of a fixed window. The difference between those two approaches is graphically shown in Figure 12. Using a rolling window increases the probability to find an event more extreme than some proposed stress. It is possible an event occurs between March and March, which we would not have noticed if we had only considered events between year ends. Therefore, the rule of thumb does not apply for backtests based on a rolling window. One should account for the fact that the probability of finding an extreme event increases by the use of the rolling window.

If data sets are very long, it is intuitively clear that more extreme events will be found if yearly stresses for all months are evaluated, instead of only end-of-year stresses. If a data set is sufficiently large, the number of extreme events found will approximately increase by a factor 12. For shorter data sets, this rule does not apply, and intuition often fails. Often, intuition tells people that in a limited data set, for example 30 years, no 1-in-200 event may occur, even if a rolling window is used. If nevertheless an extreme event has occurred, they conclude the proposed stress parameter must be too low. They apply the rule of thumb without correcting for the use of a rolling window. However, it may be well possible that if they had taken into account the effect of the rolling window, the conclusion would be that it is not unlikely to find an extreme event in 30 years of data. Next to that, a rolling window is often used for shorter data sets, to increase the number of observations. Therefore, the effect of the rolling window is particularly relevant for data sets which are short compared to the time horizon and significance level of the stress parameter. If a data set of 30 years is used to evaluate a 1-in-200 year stress, it is likely a rolling window will be used. If a data set of 30 years is used to evaluate a 1-in-10 day event, it is more likely non-overlapping periods are used.

In this section, we quantify the possible impact of neglecting the use of a rolling window for two basic time series models, the Random Walk and the AR(1) model. We will also propose a method to correct for the effect of the rolling window, and will provide a step by step manual for correctly backtesting with a rolling window. We will use this method to backtest equity stress parameters for the MSCI World and S&P index. We will illustrate the practical relevance of this phenomena by analyzing an example of a backtest for an equity stress parameter for Dutch pension funds. In an official evaluation report, the rule of thumb (a 1-in-200 stress may occur on average only once per 200 year) is applied, without correcting for the use of a rolling window.

6.2 The model

To investigate the possible effect of backtesting with a rolling window without correcting for it, we will investigate an AR(1) model. Since this is a basic model, it enables us to focus on the effect of the rolling window in its most basic form. We assume data generated by the AR(1) process covers n years, and each year consists of k steps. For example, if monthly data is used k will be equal to twelve.

The AR(1) model has the following specification, which is iterated k times (k time steps corresponds to a period of one year):

$$\begin{aligned} y_t &= a_k \cdot y_{t-1} + \epsilon_t \sim N(0, \sigma_k^2) \\ &= a_k^k \cdot y_{t-k} + \sum_{j=0}^{k-1} a^j \cdot \epsilon_{t-j} \end{aligned} \tag{22}$$

To be able to compare results for different values of k, we choose a_k such that the conditional expectation of y over a one year horizon is independent of k.

$$\begin{aligned} E(y_t|y_{t-k}) &= a_k^k \cdot y_{t-k} \\ a_k &= a^{1/k} \end{aligned} \tag{23}$$

We choose the standard deviation of the errors σ_k such that the variance of y over a one year horizon is equal to 1.

$$\begin{aligned} Var(y_t|y_{t-k}) &= \sigma_k^2 \cdot \frac{1 - a_k^{2k}}{1 - a_k^2} = \sigma_k^2 \cdot \frac{1 - a^2}{1 - a^{2/k}} = 1 \\ \sigma_k^2 &= \frac{1 - a^{2/k}}{1 - a^2} \end{aligned} \tag{24}$$

In the case of a random walk, $a = 1$, and the variance of the errors is equal to $\frac{1}{k}$. In next sections, we will use this model to compute the probability of finding at least one extreme event if a rolling window is used. First, we will treat a special case of the AR(1) model, the Random Walk, and then we will treat the AR(1) process itself.

6.3 Random Walk

In this section, we will compute the probability of finding at least one extreme event if a rolling window is used. This probability will depend on the length of the data set n (in years), the frequency of the data k and the significance level of the stress α.

In the case of a Random Walk, the model specified in Section 6.2 reduces to:

$$y_t = y_{t-1} + \epsilon_t \sim N\Big(0, \frac{1}{k}\Big). \tag{25}$$

The variance on a yearly basis is equal to 1. We now compute the distribution of observed yearly stresses S_n:

$$S_t = y_n - E(y_t | y_{t-k}) = \epsilon_t + \epsilon_{t-1} + ... + \epsilon_{t-k+1} \tag{26}$$

Since all errors are independent and identically distributed with variance $\frac{1}{k}$, the expectation of S is equal to zero, and the variance of S is equal to 1, like we demanded. However, since we consider a rolling horizon, the stresses itself are not independent.

$$S_{t+1} = S_t - \epsilon_{t-k} + \epsilon_{t+1} \tag{27}$$

The covariance structure between two stresses can be computed as follows:

$$\begin{aligned} Cov(S_t, S_{t-x}) &= E(S_t \cdot S_{t-x}) \\ &= E((\epsilon_t + ... + \epsilon_{t-k+1}) \cdot (\epsilon_{t-x} + ... + \epsilon_{t-x-k+1})) \\ &= \begin{cases} \frac{|k-x|}{k} & |x| < k \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{28}$$

If x is larger than, or equal to k, the stresses are based on non-overlapping time periods, and therefore, the covariance between the stresses is equal to zero. The variance of the stresses is equal to one ($x = 0$). The covariance is largest for $x = 1$, because this corresponds to the largest possible overlap in time periods. If x increases, the overlapping between the periods in time the stresses are based on will decrease, and therefore the covariance will also decrease. Using this expression, the covariance matrix of a series S_t equals:

$$\Sigma = \frac{1}{k} \begin{pmatrix} k & k-1 & ... & 2 & 1 & 0 & 0 & 0 \\ k-1 & k & k-1 & ... & 2 & 1 & 0 & 0 \\ k-2 & k-1 & k & & & 2 & 1 & 0 \\ ... & & & k & & & 2 & 1 \\ 1 & & & & k & & & \\ 0 & 1 & & & & k & & \\ 0 & 0 & 1 & & & & k & \\ 0 & 0 & 0 & 1 & & & & k \end{pmatrix} \tag{29}$$

Now we can compute the probability of finding at least one stress higher than some threshold. Since the errors are normally distributed with a standard deviation of one, the VaR over a one year horizon simple corresponds to the inverse cumulative normal distribution evaluated at confidence level α. We consider the Solvency II treshhold tr ($\alpha = 0.005$):

$$tr = \Phi^{-1}(0.005) \approx -2.57. \tag{30}$$

If we consider n years of data and k timesteps in a year, there are $(n-1)k$ observed stresses. We define $P(n, k)$ as the probability of finding at least one stress more extreme then the threshhold. This probability can be written as the integral over a multivariate normal distribution with the covariance matrix specified above.

$$
\begin{aligned}
P(n,k) &= P(\exists n \in \{k+1,...,y\cdot k\}|S_n < tr) \\
&= 1 - P(\nexists n \in \{k+1,...,y\cdot k\}|S_n < tr) \\
&= 1 - \int_{x_1=tr}^{\infty} ... \int_{x_{(y-1)k}=tr}^{\infty} N(0,\Sigma)dx_1..dx_{(y-1)k} \\
N(0,\Sigma) &= \frac{1}{\sqrt{2\pi^{(y-1)k}|\Sigma|}} e^{-\frac{1}{2}(x_1..x_{(y-1)k})'\Sigma^{-1}(x_1..x_{(y-1)k})} \qquad (31)
\end{aligned}
$$

In the special case of $k = 1$ (backtesting with a fixed window), all stresses are independent, and the covariance matrix reduces to the identity matrix. In this case, the probability of finding an extreme event can be computed easily:

$$P(n, k = 1) = 1 - 0.995^{y-1}. \qquad (32)$$

In Figure 13, the results are shown for different values of k and n. These results are based on simulations in Scilab, and verified in R by computing the integral over the multivariate normal distribution (equation 31).
The figure shows both the absolute probability of finding an extreme event, and the relative probability, compared to the case $k = 1$.

$$R(n,k) = \frac{P(n,k)}{P(n,k=1)} \qquad (33)$$

We see that if a rolling window is used, the probability of finding a 1-in-200 event can increase with a factor of more than 4.5. For example, as we can see in Figure 13, if 30 years of data is used, the probability of finding a 1-in-200 event based on a fixed horizon ($k = 1$, $n = 30$) is equal to 0.14. This probability increases to 0.48 if a rolling window on a monthly basis ($k = 12$, $n = 30$) is used. As expected, the effect increases with increasing k.
The relative effect of the rolling horizon increases if n decreases. This if intuitively clear. If a longer data set is used, the probability of finding an extreme event is higher, and can increase less. For example, if 100 years of data is used, the probability of finding a 1-in-200 event is already 0.4, and cannot increase as much as when $n = 30$. As we have mentioned before, the effect of the rolling horizon is particularly important if the length of the data set is short compared to the time horizon and significance level of the stress parameter (in this case, 200 years).

6.4 AR(1)

For general values of a, the AR(1) model is specified as follows:

$$y_t = a^{1/k} y_{t-1} + \epsilon_t \sim N(0, \frac{1-a^{2/k}}{1-a^2}) \qquad (34)$$

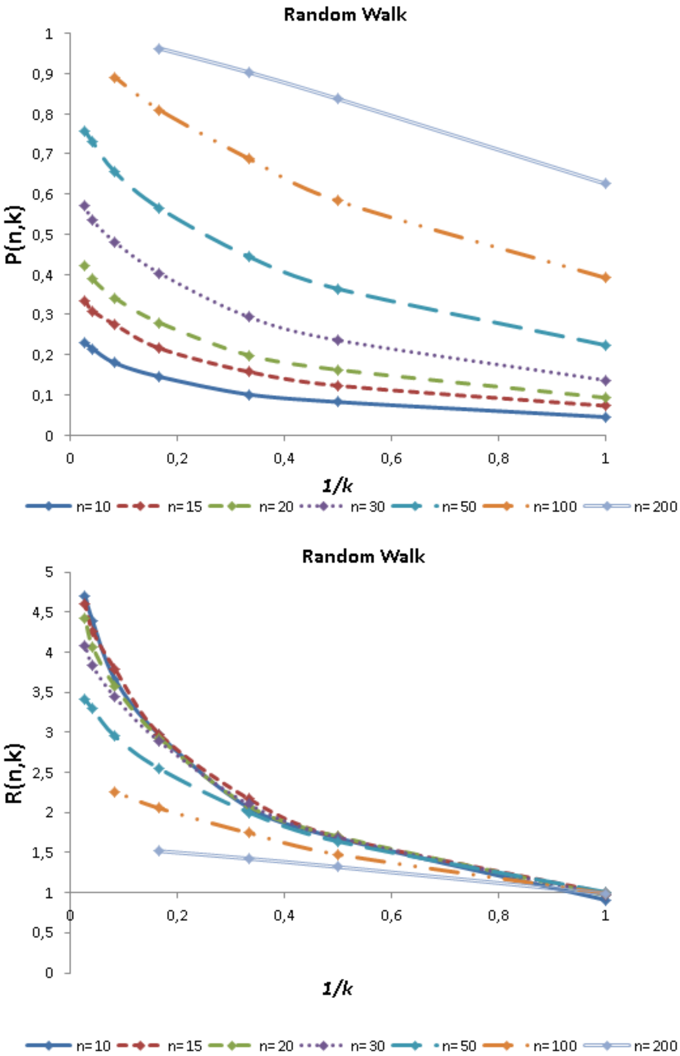


Figure 13: Up: Probability of finding an event more extreme than the proposed VaR at $\alpha = 0.005$, when a rolling window is used instead of a fixed window. Down: Relative increase in probability R. Using a rolling window increases the probability of finding an extreme event. The size of the effect increases as the length of the data (n years) considered decreases. Increasing the number of data points k used in a year also increases the effect.

Again, by construction, the variance over a horizon of one year is equal to one. We will follow the same procedure as we did for the Random Walk. The observed stress S_t depends on an observed stress at some other time, if they are based on overlapping time periods.

$$S_t = y_t - E(y_t|y_{t-k}) = \epsilon_t + a^{1/k}\epsilon_{t-1} + ... + a^{(k-1)/k)}\epsilon_{t-k+1} \tag{35}$$

The covariance structure between two stresses is equal to:

$$\begin{aligned} Cov(S_t, S_{t-x}) &= E(S_t \cdot S_{t-x}) \\ &= \sum_{j=0}^{k-x-1} a^{(x+2j)/k} \cdot \frac{1-a^{2/k}}{1-a^2}, |x| < k \\ &= \begin{cases} a^{x/k} \cdot \frac{1-a^{2(k-x)/k}}{1-a^2} & |x| < k \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{36}$$

If x is larger or equal to k, the two stresses are based on independent periods in time, and the covariance between them is zero.
In Figure 14 we can see the results for the AR(1) model based on R calculations of integral of equation 31. In this figure, y is constant and equal to 30. We have calculated the effect for different values of the mean reversion parameter a and of k. The effect of the rolling window increases as k increases, as we expected. We see that the effect of the rolling horizon increases as a decreases. This is intuitively clear. A low value of a corresponds to a high mean reversion speed. Therefore, the correlation between the stresses within a year is smaller. The probability of finding an extreme event during the year, but not at the end of the year is higher, since the stress diminished by mean reversion. If $a = 0$ and $k = 24$, the probability of finding an extreme event increases by a factor 7.

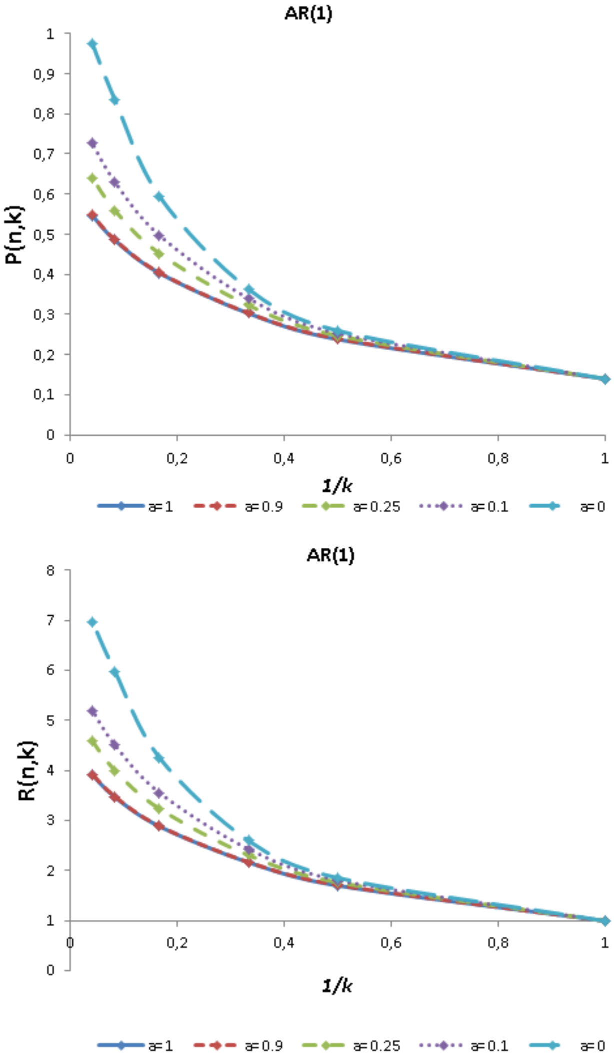


Figure 14: Up: Probability of finding and event more extreme than the proposed VaR at $\alpha = 0.005$, when a rolling window is used in stead of a fixed window. Down: Relative increase in probability R. Using a rolling window increases the probability of finding an extreme event. The size of the effect increases as parameter a decreases. Increasing the number of data point k used in a year also increases the effect. Parameter n (length of the data in years) is constant and equal to 30.

6.5 How to correctly backtest using a rolling window: A step by step approach

In previous sections, we have quantified the error that can be made by backtesting an AR(1) model using a rolling window. In this section, we will explain step

by step how a backtest using a rolling window should be executed correctly. In Figure 15, those steps are summarized.
If a model is backtested using a rolling window, the probability to find an extreme event increases. To judge the outcome of a backtest, one has to correct for this increased probability. We propose a method to account for this increased probability. We propose to compute the distribution of the number of extreme events in a given data set, taking into account the use of the rolling window, and test whether the value of the stress parameter can be rejected by setting up a Null-hypothesis and an alternative hypothesis. We will illustrate the steps by assuming we want to backtest a 1-in-200 stress parameter for equity risk over a one-year horizon. However, this method can also be applied to other risk factors and time series.

Step 1
First, we determine which data set we will use to backtest the equity stress parameter. We determine both the length of the data (y years), and the frequency the data, k. For example, we can use a data set of equity returns of the last 50 years. The data may be available on a quarterly basis, and we can choose to use all available data. It may also be possible that daily data is available. Then we can choose the frequency we want to consider. For example, we can choose to only use end-of-month data, or quarterly data. In the first case, $k = 12$, in the second case, $k = 4$. Note that backtesting using a rolling window is particularly of interest if the length of the data set is limited. If a large data set is available, for example 20000 independent observations, it is not necessary to use a rolling window. Then, k can be set to one, and the rule of thumb applies (a 1-in-200 stress may occur on average only once per 200 year). However, for financial time series, often a limited (relevant) historical data series is available, and backtesting using a rolling window is desirable. We will assume $y = 50$ and $k = 12$.

Step 2
In the second step, we determine the Null-hypothesis. We have to make an assumption about both the value of the stress parameter, and about the data generating process of the equity returns. For example, we can assume equity returns follow a Random Walk, and the 1-in-200 stress corresponds to a decrease of 40% of the equity index over one year. The alternative hypothesis is the hypothesis that the stress parameter is too low, and should be higher than 40%. In practice, modelling the stress parameter and backtesting is often an integrated process. Then, the Null-hypothesis of the underlying model will be equal to the model the stress parameter is based on. However, it is also possible a stress parameter is based on expert judgement. Then, it will be necessary to make an assumption about the underlying data generating process.
In this step, we also specify which confidence interval we will use to accept or reject the value of the stress parameter. For example, we may reject the value of the stress parameter if the p-value (to be computed in the coming steps), is smaller than 5%.

How to properly backtest a one-year VaR model by using a rolling window

Step 1:

Select a data set to be backtested, and determine the length of the data set *y* (in years), and the frequency of the data *k* (*k* observations per year).

Step 2:

Set a Null-hypothesis for a stress parameter, a one-year VaR for confidence level α based on an assumtpion of a data generating process . Determine a significance level β to reject or accept the Null-hypothesis.

Step 3:

Determine the probability distribution of the number of yearly extreme events found in *y* years of data with frequency *k*, by simulation under the Null-hypothesis.

Step 4:

Compute how often the observed yearly stresses in the data set exceed the stress of the Null-hypothesis. Denote this number *x*.

Step 5:

Compute the corresponding *p*-value of this event by adding all the probabilities from the simulated probability distribution of *x* or more extreme events.

Step 6:

Compare this *p*-value with the predetermined security level β, and reject or accept the size of the stress parameter.

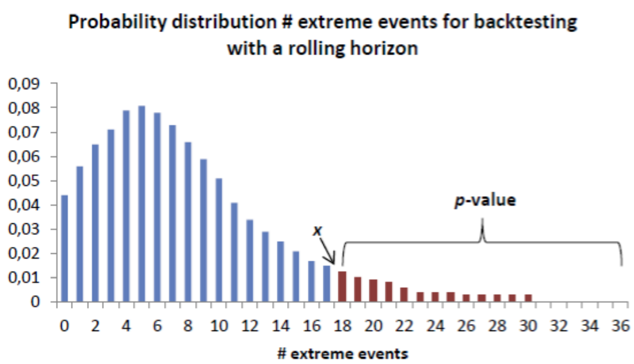


Figure 15: A step by step approach to correctly backtest using a rolling window.

Step 3
In the third step, we will determine the distribution of the number of extreme events in a data set of length y and frequency k under the Null-hypothesis. Often it will be necessary to perform simulations to find this distribution. In the case of our Null-hypothesis (a Random Walk, and a stress parameter of 40%), we simulate a large number of scenarios (for example 100.000). One scenario consists of $k \cdot y$ point (in our case $50 \cdot 12$). We simulate 100.000 Random Walks with a yearly standard deviation of $\frac{40\%}{\Phi^{-1}(0.995)}$, such that the 1-in-200 stress corresponds to -40%. For each of those scenarios, we compute $k \cdot (y-1)$ realized yearly stresses. Then we evaluate the number of realized yearly stresses that exceed -40%. For each scenario, we find a number of extreme events. Based on those numbers, we estimate the probability distribution of the number of events to be found that exceed the proposed stress parameter of 40%, in 50 year of data, which is available on a monthly basis. We denote the probability of finding n events $p(n)$.

Step 4
In the fourth step, we determine how often the stress parameter is exceeded in our real data set. In our example, we compute $12 \cdot (50-1)$ realized yearly returns. It may be possible we find 18 yearly returns worse than -40%. We call this number x.

Step 5
Then, we compute the p-value of finding 12 or more yearly returns worse than -40%. We use the probability distribution determined in step 3 to do so. We sum all probabilities corresponding to $x \geq 12$.

$$p_{value} = \sum_{n=x}^{n=\infty} p(n) \tag{37}$$

Step 6
In the last step we evaluate the backtest by comparing the p-value found in step 5 with the security level we have chosen in step 1. For example, if the p-value is 3%, and we set the confidence level to 5%, we conclude the stress parameter is too low.

The limiting case: $k = 1$***, and yearly stresses are independent:*** If we do not use a rolling window, but use yearly non-overlapping realizations, and we assume the yearly stresses are independent, the distribution of the number of extreme events can be computed analytically. We have y years of data, $y-1$ realized errors, and we set $k = 1$. Since the yearly stresses are independent, the probability of finding exactly n events more extreme than the stress parameter under the Null-hypothesis, is Binomially distributed.

$$p(n) = \binom{y-1}{n} \alpha^n \cdot (1-\alpha)^{y-1-n} \tag{38}$$

The probability to find at least x events is equal to

$$\begin{aligned} p_{value}(x) &= \sum_{n=x}^{n=\infty} p(n) \\ &= 1 - \sum_{n=0}^{n=x-1} p(n). \end{aligned} \tag{39}$$

In particular, the probability to find at least one event is equal to $1-(1-\alpha)^{y-1}$, where α is the level of security (for Solvency II, $\alpha = 0.005$).

6.6 Testing the model: S&P and MSCI world equity index

We have seen that the probability of finding an extreme event increases significantly when a rolling horizon is used to backtest a model. In this section, we will investigate whether this effect can explain the number of extreme events in the S&P and MSCI world equity index. In particular we will investigate a statement from an evaluation report of the FTK (Financieel Toetsings Kader, rules for Dutch Pension Funds to determine their required capital), about the size of the equity stress parameter of the FTK.
We will use the backtest method as outlined Section 6.5. We will assume equity returns follow a Random Walk with drift.

6.6.1 Evaluation of the FTK

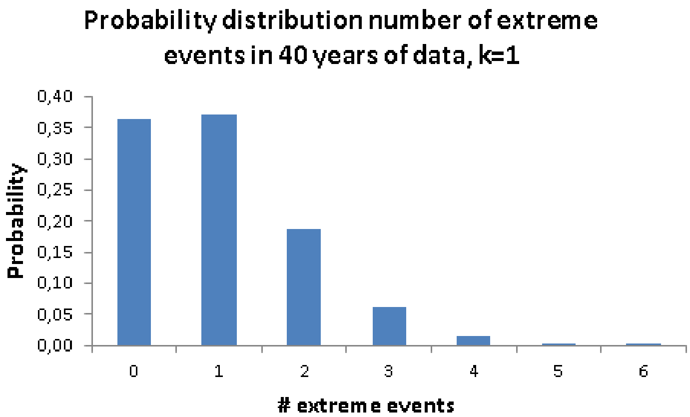
The parameter for an equity shock in the FTK corresponds to a negative equity return of 25%. Note that Pension Funds in the Netherlands should hold a required capital corresponding to a 97.5 % Value-at-Risk measure, where the security level of Solvency II is 99.5%. In the evaluation report, this equity stress parameter is backtested against the MSCI World index. Data since 1970 is used. The evaluation report states the following [14]:

"Sinds 1970 heeft deze index 4 keer een tuimeling meegemaakt die groter is dan de 25 procent daling waar het FTK op is gestoeld. Twee van deze gebeurtenissen vonden plaats in de afgelopen 7 jaar. De frequentie van dergelijke omvangrijke schokken is dus hoger dan op basis van de 97,5 procent zekerheidsmaat en een (log)normale verdeling mag worden verwacht."

The report states that since 1970, 4 stresses higher than 25% have taken place, of which two in the past 7 years. It concludes that this is higher than can be expected based on the confidence level of 97.5% and a (log)normal distribution. Therefore, the equity stress parameter is rejected based on a backtest.
We will investigate this statement by performing the backtesting procedure of Section 6.5.
We will assume equity returns follow a Random Walk with drift μ. The Null-hypothesis assumes the 97.5% VaR corresponds to an equity return of -25%.

$$H_0 : \mu + \Phi^{-1}(0.025) \cdot \sigma = -25\% \tag{40}$$

We will compute the p-value corresponding to the number of extreme events that have happened, to see whether we can reject the Null-hypothesis. We reject the backtest if the p-value is smaller than 5%.
We use monthly data of the MSCI world index, downloaded from www.msci.com. We will both consider $k = 1$ and $k = 12$, backtesting with a fixed horizon, and with a rolling monthly horizon. To compute the p-value, we performed simulations to estimate the distribution of the number of stresses that exceeds the threshold under the Null-hypothesis. The distribution of the number of extreme events in 40 years of data under the Null-hypothesis is plotted in Figure 16. In



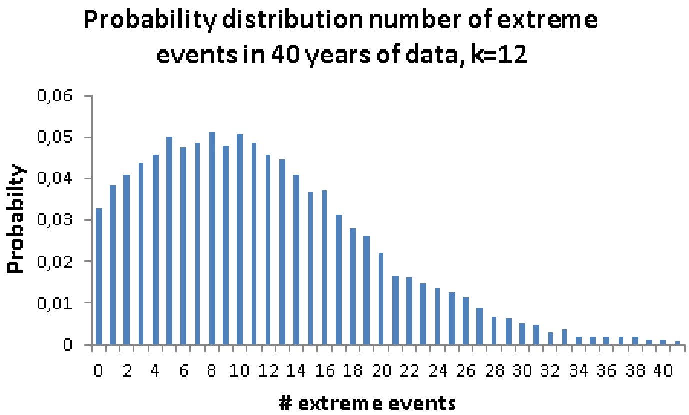


Figure 16: Probability distribution of the number of extreme events found in 40 years of data, based on the proposed procedure in Section 6.5. Up: Values for $k = 1$ (fixed window) have been computed analytically. Down: Values for $k = 12$ (rolling window, monthly evaluation) are based on simulations.

the case of a fixed horizon, we find two equity decreases larger than 25%. (-

	Historic events worse than -25%	p-value based on simulations
k=1	2	0.264
k=12	21	0.140

Table 5: Extreme events (corresponding to a decrease of more than 25%) found in the MSCI World index. P-values are computed by the method described in Section 6.5. P-values are higher than 5%, and therefore the Null-hypothesis, an equity stress of 25%, is not rejected. This is in contradiction with the findings of the evaluation report of the FTK, which is based on backtesting with a rolling window without correcting for this.

	Historic events worse than 2.5% percentile	p-value based on simulations
k=1	3	0.665
k=12	43	0.442

Table 6: Extreme events found in the S&P index, $\alpha = 2.5\%$.

27.8% YE 1973 to YE 1974, and -42.1%, YE 2007 to YE 2008). When a rolling window is used, 21 events are found. This corresponds to p-values of 0.264 and 0.140. Both p-values are larger than 5%. Therefore, we do not reject the Null-hypothesis.

The conclusion of the evaluation report seems partly unfounded. It is well possible the equity stress parameter is too low. Part of the reasoning of the report comes from the fact that the size of the violations are very high (-42.1% in 2008). We have only tested the number of violations, not their sizes. However, the argumentation that the equity stress parameter is too low because four violations have taken place is not correct. If a fixed window, from year end to year end is used, only two violations are found. If a rolling window is used, more violations are found, which is caused by the use of this rolling window. The rule of thumb is applied, without correcting for the use of the rolling horizon.

6.6.2 S&P index

We can perform the same analysis on the S&P index. This index is available as of 1871. If we estimate the parameters of a Random Walk with drift for the yearly equity returns, we find $\mu = 5.8\%$ and $\sigma = 18.9\%$. According to this distribution, the 0.5% point of the distribution corresponds to a decline of the equity index of 42.8%. The 2.5% point of the distribution corresponds to a decline of -31.2%.

In Table 7, a typical example of the effect of the rolling horizon can be seen. The data consists of 138 years of returns. According to the rule of thumb, maximal one 1-in-200 event may occur in this historic data. In Figure 17, the yearly returns are plotted, and the 0.5% and 2.5% VaR based on the fitted model are shown. If we consider a yearly horizon, only 1 extreme event has happened.

	Historic events worse than 0.5% percentile	p-value based on simulations
k=1	1	0.512
k=12	12	0.252

Table 7: Extreme events found in the S&P index, $\alpha = 0.5\%$.

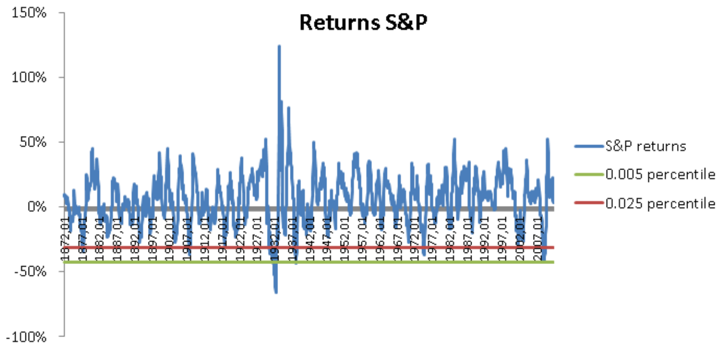


Figure 17: Historic S&P returns

However, based on a rolling window, 12 extreme events have happened (clustered around the begin of the '30's and the end of the '30's).

Naively one could think the model fails the backtest, since at two periods in time violations have taken place, in less than 200 years of data. However, the p-value equals 25.2%, and the Null-hypothesis cannot be rejected. The number of violations can be explained by the fact that a rolling horizon is used. The 2.5% stress parameter also passes the backtest as can be seen in Table 6.

6.7 Conclusions on backtesting with a rolling window

Using a rolling window significantly increases the probability of finding an event that is more extreme then some predetermined value. If a model is backtested, a fixed window has to be used, or one has to correct for this phenomena. Using a rolling window without correcting for this increased probability can lead to rejecting a model on false grounds. The effect of the rolling window increases with decreasing length of historic data that is used. We propose a method for backtesting taking the effect of the rolling window into account. The fact that a rolling window has been used partly explains why the FTK parameter for equity stress has been rejected. If we backtest this parameter by taking into account the effect of the rolling window, the parameter passes the backtest.

Part II

Statistical pitfalls of the one-year time horizon in Solvency II

7 Time horizons, probabilities of ruin and probabilities of default

Solvency II focuses on a time horizon of one year. An insurance company meets the Solvency II requirements if the own funds (market value of the assets minus market value of the liabilities) of the company are high enough to bear a 1-in-200 year loss that may occur, exactly in one year from now. Solvency II is risk based and meant to create a level playing field for all insurance companies in Europe. In the Netherlands, the regulatory rules for Pension Funds focuses on different time horizons. Comparable with insurance undertakings, Pension Funds have to show that their own funds are high enough to bear a 1 in 40 year loss over a one year horizon. Next to that, they have to perform a continuity analysis, which focuses on the long term (15 years). We will investigate pitfalls that may occur by only focussing on this one year horizon in Solvency II. In particular, we will examine the level playing field, by considering different kinds of insurance undertakings, all with an initial Solveny II ratio of 100%. Based on the same starting point, we will examine how the different insurance undertakings perform at longer time horizons, compared to each other.

The effect of time horizons on optimal portfolio choice has been subject of intensive research. The behavior of asset classes and the relation between different asset classes can depend on the time horizon considered, for example caused by mean reversion behavior of returns. Therefore, the optimal asset mix can depend on the time horizon considered. If a certain risky asset class shows mean reverting behavior, the risk over a longer horizon is relatively low compared to the expected excess return. This is sometimes referred to as time diversification of risk, or the term structure of risk [15], [16].

The vast majority of the research on time diversification of risk focuses at the end point of the time horizon. Often, what happens within this horizon is neglected. However, this within-horizon behavior can be of particular importance for insurance companies. Liabilities need to be covered by sufficient assets at every point in time (or, as often as official figures are reported). Next to the risk of default at maturity, within-horizon risk arises: The risk of underfunding at a certain point in time, although it is well possible the obligations can be met at maturity. As M. Kritzman and D. Richm [19] conclude, end of horizon risk diminishes with time, but within-horizon risk rises as the investment horizon expands. They conclude this is a new challenge for research on time diversification of risk. Due to regulatory requirements, it is necessary that insurance undertakings also take within-horizon risk into consideration when making an asset allocation decision. P. Devolder [17] has studied both the end-of-horizon and within-horizon risk for long term guarantees, and concludes the one-year horizon in Solvency II is probably not the right horizon for this kind of product.

An insurance undertaking can have different kinds of liabilities. Often, liabilities depend on stochastic variables like mortality rates, and can contain optionalities. In the remainder of this thesis, we will assume an insurance undertaking has a set of deterministic cashflows it has to pay at different times in future. The customer pays an amount of money, and in return, it will receive fixed amounts at fixed points in future. Of course, the customer wants to be as sure as possible that the insurance undertaking will have enough assets to pay the liabilities. Therefore, the market value of the assets and liabilities will be monitored on a frequent basis. If the market value of the assets is lower than the market value of the liabilities at a certain point in time, we will refer to this as a ruin. However, it is well possible that the insurance undertaking recovers from this situation, and at the time the cashflow has to be paid, it has enough assets to do so. It is possible for an insurance company to recover from a situation of underfunding, because the liabilities are valued at a market consistent, risk neutral basis. In reality, the insurance undertaking can make more profit than the risk free interest rate, for example by investing in equity. If at a certain point in time a cashflow needs to be paid, but the undertaking does not have enough assets to do so, we will refer to this as a default. In next sections, we will consider both default and ruin probabilities, and we will also consider the ruin probability after exactly one year, since this probability should be lower than 0.5%, as demanded by the Solvency II regulations. A ruin corresponds to within-horizon risk, a default corresponds with end-of-horizon risk.

8 Modeling one cashflow

In this section, we will compute probabilities of default and ruin for insurances undertakings of which the liabilities solely consist of one guarantee at maturity. The advantage of using this simplified example is the possibility to compute probabilities of ruin and probabilities of default analytically. The model we use is based on [17]. We will compute probabilities of ruin and default for different maturities N, and we will consider different investment strategies. We will compare insurance companies which all exactly meet the Solvency requirements at time zero. Their probabilities of default and ruin in time will differ due to different investment strategies, and different maturities of the guarantee.

We consider a nominal guarantee at time N. The value of the liability at time N is equal to $L(N) = e^{rN}$, where r is the risk free interest rate. The market consistent value of the liability at time 0 is therefore equal to 1. Assets will partly be invested in a riskless asset and partly in a risky asset. The return of the riskless asset is equal to the risk free rate r. The value of the risky assets follows a geometric Brownian Motion, with mean δ and standard deviation σ. A proportion β $(0 < \beta \leq 1)$ is invested in the risky asset, and assets may or may not be continuously rebalanced. Let $w(t)$ be a standard Brownian motion. If the assets are rebalanced, the development of the value of the assets in time is:

$$A(t) = A(0)e^{(\beta\delta+(1-\beta)r-\beta^2\sigma^2/2)t+\beta\sigma w(t)} \tag{41}$$

If the assets are not rebalanced, the value of the assets in time equals:

$$A(t) = A(0)(1-\beta)\cdot e^{rt} + A(0)\beta\cdot e^{(\delta-\sigma^2/2)t+\sigma w(t)}. \tag{42}$$

We will investigate both within-horizon and end-of-horizon risk, and therefore we are interested in the following probabilities:

- The probability of default, P_d. This is the probability that at time N, the assets are not sufficient to meet the obligations. This refers to end-of-horizon risk.
- The probability of ruin, P_r. This is the probability that at some time before N, the assets are lower than the liabilities. This refers to within-horizon risk.
- The naive probability of default that is implied by Solvency II, P_n. This probability is based on the presumption that companies that meet the Solvency requirements only go bankrupt once every 200 years, and probabilities of default are independent and equal to $\frac{1}{200}$.

In formulas, these probabilities can be written as:

$$
\begin{aligned}
P_d &= P(A(N) < L(N)) \\
P_r &= 1 - P(A(t) \geq L(t), \forall t \in [0, N]) \\
P_n &= 1 - \alpha^N
\end{aligned} \tag{43}
$$

Here, α is the level of security over a 1 year horizon, 0.995 for Solvency II.

In the following subsections we will define different asset allocation strategies. We will compute the required Solvency II ratio, and compute the probability of ruin and default under those different strategies, and for different time scales. Then, we will analyze the results.

8.1 Strategies

To cover the liabilities at time t, assets $A(t)$ will be available. The market value of the liabilities at $t = 0$ is equal to 1. The assets at time $t = 0$ are equal to $A(0) = 1 + SC$. The capital SC serves as a buffer for shocks in the value of the risky asset. We will consider different rebalance strategies, and different investment strategies of the SC. If the assets are rebalanced, at every point in time the percentage β of the risky asset is the same. If the risky assets increase more in value than the risk free assets, risky assets will be sold and riskless assets will be bought back. If at a certain point in time, β is less than its value at start, the opposite will happen, and extra risky assets will be bought. This is a procyclical strategy. If we do not rebalance the assets, and we do nothing, only at time $t = 0$, β is fixed, but β will vary over time, depending on the performance of the risky asset compared to the riskless asset. Next to that, we will distinguish between two different investment strategies for the SC. The SC can be invested conform the other assets (a percentage β in the risky asset), or the entire SC can be invested in the risk free asset.

We will consider four different strategies:

- Invest SC in current asset mix, rebalance assets (reb, cam)
- Invest SC in risk free asset, rebalance assets (reb, rf)
- Invest SC in current asset mix, do nothing (dn, cam)
- Invest SC in risk free asset, do nothing (dn, rf)

Now, we will compute the N-year Solvency Capital (i.e. the SC that is necessary such that the probability that the value of the assets at time N is larger or equal to the value of the liabilities equals $1 - \alpha$), and the probabilities of ruin and default for the first strategy. The assets are rebalanced, and the SC is invested in the current asset mix. To compute the probability of ruin, we will use the law of the minimum of a geometric Brownian motion, see Appendix A. The derivations for the other three strategies are similar, and shown in Appendix B. For the first strategy, the assets evaluate as follows:

$$A(t) = e^{((\beta\delta+(1-\beta)r-\beta^2\sigma^2/2)t+\beta\sigma w(t))}(1+SC) \tag{44}$$

First, we will compute the N year Solvency Capital. Note that the SC required for Solvency II corresponds to the case $N = 1$.

$$\begin{aligned}
P(A(N) < L(N)) &= 1-\alpha \\
P(e^{((\beta\delta+(1-\beta)r-\beta^2\sigma^2/2)N+\beta\sigma w(N))}(1+SC) < e^{rN}) &= 1-\alpha \\
P(((\beta\delta+(1-\beta)r-\beta^2\sigma^2/2)N+\beta\sigma w(N)) < \log(\frac{e^{rN}}{1+SC}) &= 1-\alpha \\
P(w(1) < \frac{-\log(1+SC)-(\beta(\delta-r)+\beta^2\sigma^2/2)N}{\beta\sigma\sqrt{N}}) &= 1-\alpha \\
\log(1+SC) = -\beta\sigma\Phi^{-1}(1-\alpha)\sqrt{N} &- (\beta(\delta-r)+\beta^2\sigma^2/2)N \qquad (45)
\end{aligned}$$

In the last step, we used the fact that the standard Brownian motion at time one, $w(1)$, is normally distributed with mean zero and standard deviation one. Rewriting the last expression, we obtain the following expression for the N-year Solvency Capital:

$$SC_{reb,cam}(N) = e^{-\beta\sigma\Phi^{-1}(1-\alpha)\sqrt{N}-(\beta(\delta-r)-\beta^2\sigma^2/2)N} - 1 \tag{46}$$

Analogous, we can compute the probability of default at maturity:

$$\begin{aligned}
P_{d,reb,cam} &= P(e^{((\beta\delta+(1-\beta)r-\beta^2\sigma^2/2)N+\beta\sigma w(N))}(1+SC) < e^r) \\
&= P(((\beta\delta+(1-\beta)r-\beta^2\sigma^2/2)N+\beta\sigma w(N)) < \log(\frac{e^r}{1+SC}) \\
&= P(w(1) < \frac{-\log(1+SC)-(\beta(\delta-r)+\beta^2\sigma^2/2)N}{\beta\sigma\sqrt{N}}) \qquad (47)
\end{aligned}$$

We obtain the following expression for the probability of default at maturity:

$$P_{d,reb,cam} = \Phi\Big(\frac{-\log(1+SC)-(\beta(\delta-r)-\beta^2\sigma^2/2)N}{\beta\sigma\sqrt{N}}\Big) \tag{48}$$

If $SC_{reb,cam}$ based on equation 46 is filled in into equation 48, the desired value of 0.005 is obtained.

Now we will compute the ruin probability, given a Solvency Capital SC is invested in the current asset mix.

$$
\begin{aligned}
P_r &= P(min_{0\le t\le N}\frac{A(t)}{L(t} < 1) \\
&= P(min_{0\le t\le N}e^{(\beta\delta+(1-\beta)r-\beta^2\sigma^2/2)t+\beta\sigma w(t)}\frac{(1+SC)}{e^{rt}} < 1) \\
&= P(min_{0\le t\le N}e^{(\beta(\delta-r)-\beta^2\sigma^2/2)t+\beta\sigma w(t)} < \frac{1}{1+SC})
\end{aligned}
\tag{49}
$$

Using the law of the minimum of a geometric Brownian motion in Appendix A, we obtain:

$$P_{r,reb,cam} = \Phi\Big(\frac{-\log(1+SC)-(\beta(\delta-r)-\beta^2\sigma^2/2)N}{\beta\sigma\sqrt{N}}\Big) \tag{50}$$

$$+\frac{1}{1+SC}^{\frac{2\beta(\delta-r)}{\beta^2\sigma^2}-1}\cdot\Phi\Big(\frac{-\log(1+SC)+(\beta(\delta-r)-\beta^2\sigma^2/2)N}{\beta\sigma\sqrt{N}}\Big). \tag{51}$$

The first part of this expression corresponds to the probability of default at maturity (end-of-horizon risk, see equation 48), the second part corresponds to the probability of a within-horizon ruin, given there would have been no default at maturity.

8.2 Analysis

We will analyse the results of the four different strategies, and different time horizons. We will use parameter values $r = 4\%$, $\delta = 7\%$, $\sigma = 16\%$ and $\beta = 20\%$.

8.2.1 Probabilities of default and ruin

In Figure 18, the probabilities of ruin and default, and the naive default probability for the second investment strategy are plotted. Assets are continuously rebalanced, and the value of the assets at time zero is exactly sufficient to cover a 1-in-200 shock over a one year horizon. Therefore, the probability of default for a one-year guarantee is exactly 0.5%. The probability of ruin for a one-year guarantee is approximately a factor two higher. Therefore, in 1% of the cases, at some point in time the market value of the assets is insufficient to cover the market value of the liabilities. Only half of the time the insurance undertaking is unable to pay out the liability at maturity ($t = 1$). In the other half of the cases, the insurance undertaking recovers from the situation of underfunding.

If we also consider guarantees at other times, we see that the probability of ruin increases substantially with time, but after a certain time, default probabilities start decreasing. This is consistent with the findings in [19], where it is concluded that end-of-horizon risk diminishes with time, but within-horizon risk rises as the investment horizon expands. If a company meets the solvency

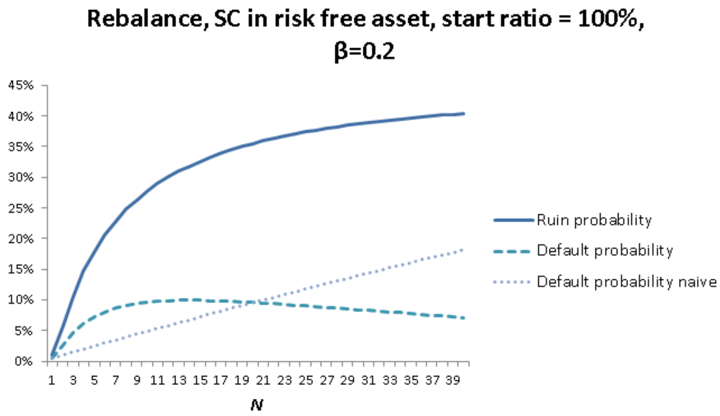


Figure 18: Probabilities of ruin and default. One liability at maturity is considered, between one and 40 years from now. The ruin probability keeps increasing with time (within-horizon risk), the default probability first increases, but decreases after a certain point. The gap between the ruin probability and default probability increases with the time horizon.

requiremets, the probability of default after one year is exactly 0.5%, but after 40 years, the probability of ruin increases to 40%. We see that for maturities shorter than twenty years, the probability that a company defaults at some time (if measured on a continuous basis) is much larger than people expect ("Naive probability of default").

The large gap between the probability of default and probability of ruin for large N implies that if measured on a continuous basis, a large number of companies will go bankrupt that would have been able to meet the obligations at maturity. The question is whether this is desirable, and if this is in contradiction with the level playing field of Solvency II. Different insurance companies can all start with the same solvency ratio of 100%. One expects that those companies are equally safe. However, the probabilities of ruin and default substantially differ, depending on the time the guarantee has to be paid.

8.2.2 N-year Solvency Capital

In Figure 19, the Solvency Capital is plotted, such that the probability of default at time N is equal to 0.5%. All four investment strategies are considered. We see that the investment strategy largely influences the Solvency Capital for longer maturities. For the chosen parameter values, in general, a rebalancing strategy requires more capital. For very long guarantees, the Solvency Capital of the rebalancing strategies starts decreasing. This is not the case if assets are not rebalanced.

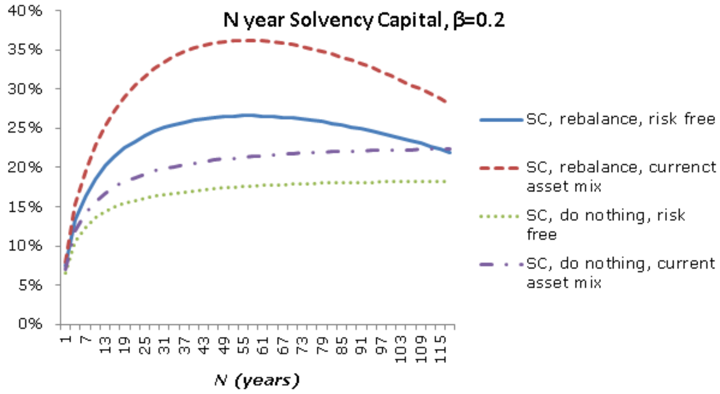


Figure 19: Necessary starting capital such that the probability of a default at maturity is equal to 0.5% (one nominal guarantee at maturity N). Four different strategies, (rebalancing yes or no, investing the SC in the current asset mix, or risk free asset mix), are considered. The investment strategy can have significant impact on the required N-year Solvency Capital.

8.2.3 Horizon effects in strategies

In Figure 20, probabilities of ruin and default are compared for the four different investment strategies. If the available capital at $t = 0$ is exactly equal to the required capital, the rebalance strategy leads to lower probabilities of default and ruin. However, the initial required SC is also higher. Therefore, a solvency ratio of 100% is "worth more" if it is based on a rebalance strategy. After exactly one year, probabilities of ruin are the same, but for longer horizons, the probability of ruin and default will be lower if assets are rebalanced.

If we start with a fixed SC of 0.1, we find horizon effects in investment strategies. For a short time horizon, probabilities of ruin and default are larger if assets are rebalanced, which is consistent with the higher required SC for the rebalance strategy. However, for longer time horizons, we see interesting behavior. After approximately 12 years, the probability of ruin and default of the rebalance strategy is lower compared to the case where assets are not rebalanced. Therefore, given an available capital at time 0, the optimal strategy (assuming we want to minimize the probability of ruin and default) depends on the relevant time horizon for the insurance company.

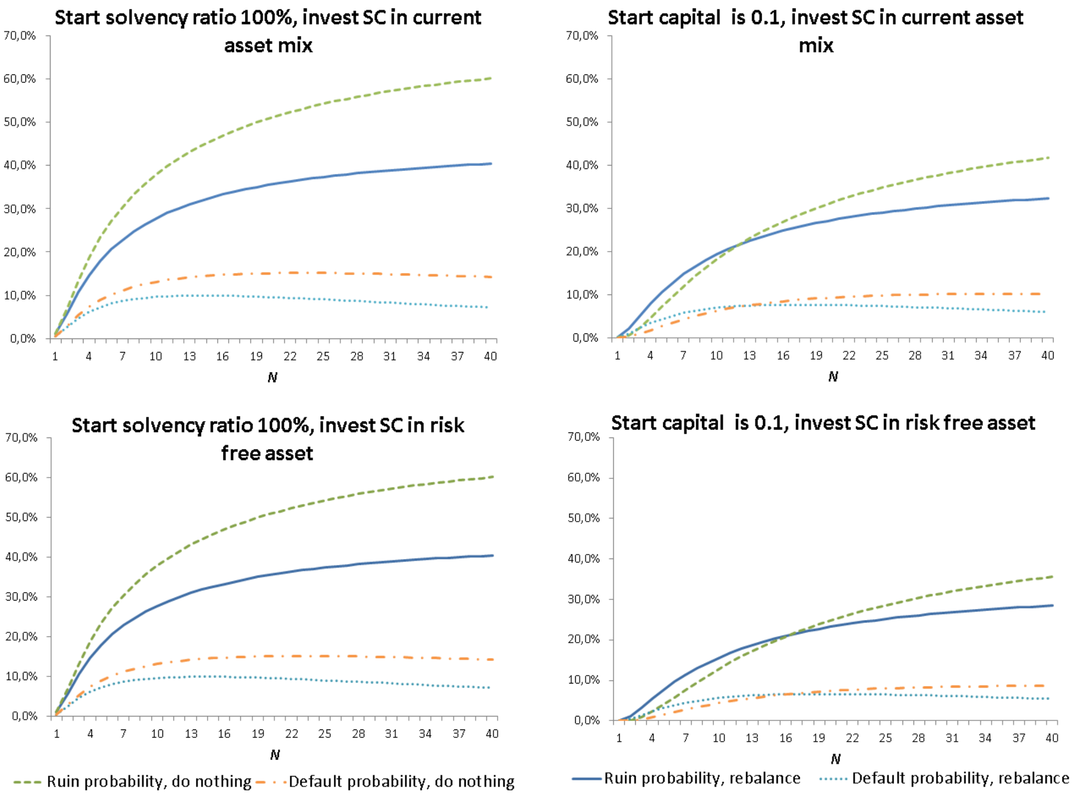


Figure 20: Probabilities of ruin and default under four different investment strategies. If the Solvency II ratio is 100% at start, a rebalancing strategy leads to lower probabilities of ruin and default for longer time horizons, compared to a strategy where assets are not rebalanced. If the value of the assets is 10% higher than the value of the liabilities at start, the optimal investment strategy to minimize probabilities of default or ruin depends on the time horizon considered.

9 Modeling different cashflow patterns

In the previous section, only one cashflow at maturity was considered. In reality, an insurance undertaking can have many cashflows at many times. We will consider two extra cashflows patterns, a constant cashflow over time, and cashflows linearly decreasing to zero over time. We will investigate the effect on the probabilities of ruin and default. It is very hard, if possible, to compute this analytically. Therefore, we have used simulations, in Scilab. We assumed again that the portfolio consists of a risk free asset and a risky asset. At every time step, a return of the riskless asset, the risky asset, and the liabilities is determined. The risky asset is assumed to follow a geometric Brownian mo-

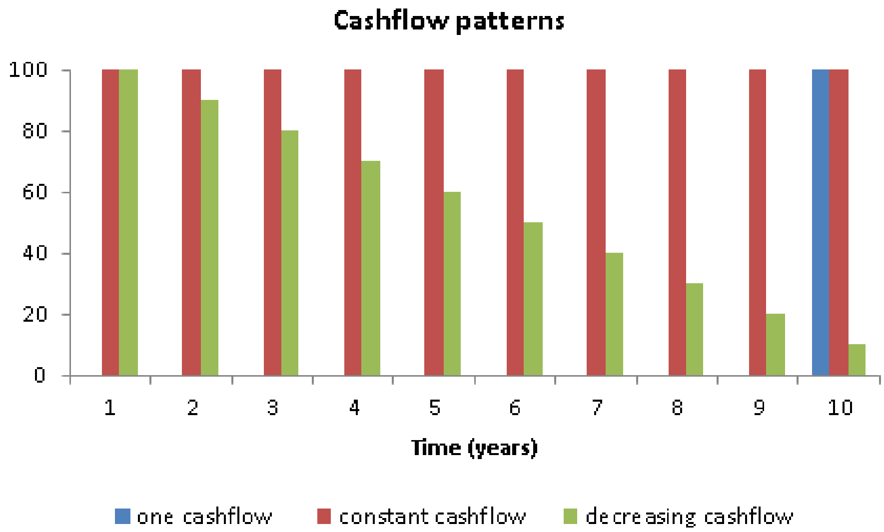


Figure 21: Three different cashflow patterns considered.

tion, and at every time step, the assets are rebalanced. If there is a payout, it is subtracted from the value of the assets. We simulated 12 points per year ($k = 12$). If at a certain time, the value of the assets becomes negative (due to a payout), there is a default. If at a certain time, the value of the assets is less than the value of the liabilities, there is a ruin. The code used can be found in Appendix D.

We computed the probability of ruin and default for cashflows of 10 and 30 years ($y = 10$, $y = 30$). In the case of thirty years, the first cashflows pattern corresponds to one cashflow at time 30. The second pattern correspond to a constant cashflow at all points in time (30 years times 12 simulated points per year). The last pattern corresponds to cashflows linearly decreasing to zero between time zero and 30 years. First, we performed simulations to find the start ratio (value of assets divided by liabilities at time 0), such that the probability of a ruin after one year equals 0.5%. Therefore, the start ratio always corresponds to a Solvency II ratio of 100%. Note that consequently the size of the cashflows does not influence the results.

The results of the simulations are graphically shown in Figure 22. We see that the ruin probability and the default probability are highest for the first cashflow pattern, with one cashflow at maturity, followed by the constant cashflow pattern. The probabilities of default and ruin are lowest if the cashflows are decreasing. Simulations for 10 and 30 years give similar results. If one cashflow

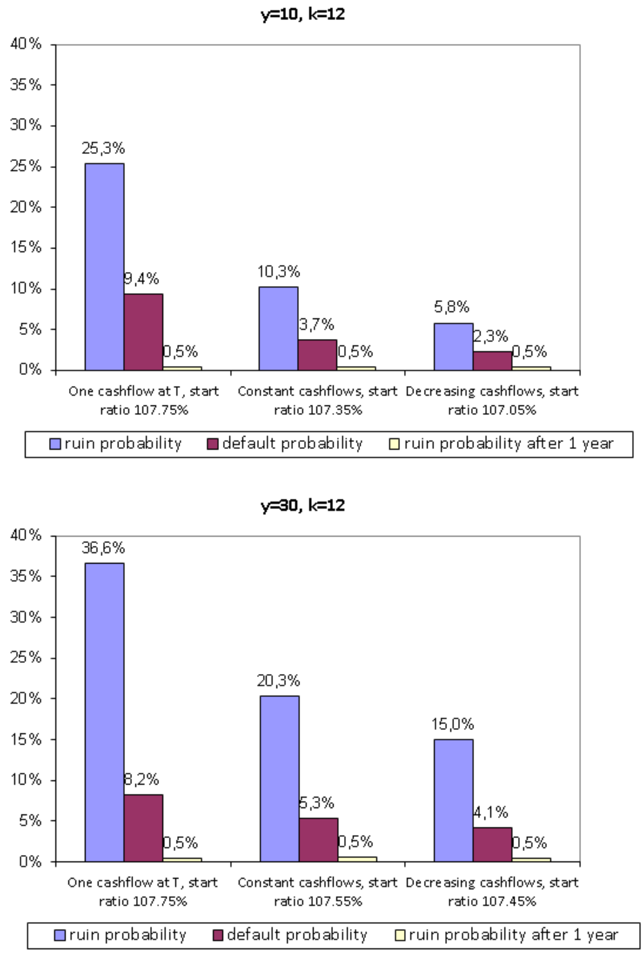


Figure 22: Probabilities of ruin, probability of ruin after exactly one year and probability of default. The value of the assets at $t = 0$ is chosen such that the ruin probability at $t = 1$ is 0.5%. Three different cashflows are considered. Cashflows which end in 10 years and in 30 years are considered. The asset portfolio consists of a risky and risk free asset, and assets are rebalanced on a monthly basis. The cashflows pattern has substantial influence on the probabilities of ruin and default. One cashflow at maturity results in the highest probability of default, followed by the constant cashflows pattern. The decreasing cashflow pattern leads to the lowest probability of default and ruin. Therefore, insurance companies which both have a Solvency II ratio of 100% are not equally safe over a longer time horizon. This can have implications for the level playing field of Solvency II.

has to be paid in 30 years from now, the probability of ruin is higher, and the probability of default is lower, compared to the 10 year case. This is consistent with the analytical results of Section 8.

At first sight, it may be surprising that one cashflow at maturity results in the highest probability of default. The duration of the liabilities for this cashflow pattern is longer than the duration of the other two patterns. One could expect that since default probabilities start decreasing for longer durations, this pattern would result in the lowest probability of default. Probably the lower probability of default of the other two patterns is caused by the fact that at time zero, more assets are available to cover the other cashflows. If the risky asset performs well or expected in the first years, an excess return is made over the entire provision. However, after the first years the provision of the liabilities has decreased, due to the first payouts that have been made. Therefore, the own funds of the insurance undertaking, relative to the value of the liabilities, are higher. This will have a decreasing effect on the probability of ruin. Therefore, the cashflow pattern which contains a higher concentration of liabilities in the first years (the decreasing cashflows pattern), results in the lowest probability of default.

In the previous section, we have shown that insurance companies which all start with a Solvency II ratio of 100%, have different probabilities of ruin and default at longer time horizons. The differences are caused by different investment strategies and different maturities of the guarantee. Here we have shown that also the cashflow pattern influences probabilities of ruin and default. Therefore, it is necessary to take those factors into account in Solvency II, and one should not focus solely on the time horizon of one year.

10 Adding mean reversion to the model

In the previous sections, we have assumed the value of the risky asset follows a geometric Brownian motion. Therefore, the standard deviation of the return is proportional to the square root of time. In this section, we will investigate the effect of adding mean reversion to the model. There has been much debate on the question whether risky assets show mean reverting behavior. Here, we will not question whether risky assets show mean reversion behavior or not. We are interested in the sensitivity of the results for the assumption made about the mean reverting behavior of the risky asset.
We will assume the stochastic part of the risky asset returns follows an Ornstein-Uhlenbeck process instead of a Brownian motion. The specification of the Ornstein-Uhlenbeck process is as follows:

$$dX_t = -aX_t dt + \sigma dW_t. \tag{52}$$

W_t follows a standardized Brownian motion. Parameter a represents the mean reversion speed. If $a = 0$, there is no mean reversion, and the model reduces to an ordinary Brownian motion. The value of X at time t depends on the value of X at time zero as follows [22]:

$$X_t = X_0 e^{-at} + \epsilon_t \sim N\Big(0, \frac{\sigma^2}{2a}(1 - e^{-2at})\Big) \tag{53}$$

A deviation of X is corrected and pulled towards the mean of zero by the term e^{-at}. The size of the standard deviation converges to $\frac{\sigma}{\sqrt{2a}}$ as $t \to \infty$. For a Brownian motion, the standard deviation goes to infinity as t goes to infinity.

Because the distribution of the risky asset returns at a certain time is known, we can compute the probability of default after one year and at maturity. It is possible to compute the ruin probability, by using results from investigations of the first passage times of an Ornstein Uhlenbeck process [22], but we will leave this for future research.

For the geometric Brownian motion, the assets develop as follows:

$$A(t) = e^{((\beta\delta + (1-\beta)r - \beta^2\sigma^2/2)t + \beta\sigma w(t))}(1 + SC) \tag{54}$$

Now, assuming mean reversion, and $\tilde{\sigma} = \frac{\sigma}{\sqrt{2a}}\sqrt{(1 - e^{-2at})}$ the development of the value of the assets in time equals:

$$A(t) = e^{((\beta\delta + (1-\beta)r)t - \beta^2\tilde{\sigma}^2/2 + \beta\tilde{\sigma} w(1))}(1 + SC) \tag{55}$$

Using this, we can compute the probability of ruin and default, analogous to section 8.1.

$$SC = e^{-\beta\tilde{\sigma}\Phi^{-1}(1-\alpha) - \beta(\delta - rN) - \beta^2\tilde{\sigma}^2/2} - 1 \tag{56}$$

$$P_d = \Phi\Big(\frac{-\log(1 + SC) - (\beta(\delta - rN) - \beta^2\tilde{\sigma}^2/2)}{\beta\tilde{\sigma}}\Big) \tag{57}$$

The analytical results for the case of one guarantee at maturity are shown in Figure 24. The results of simulations in Scilab, analogous to section 9, are shown in Figure 23. The results are consistent with the expectations. If the mean reversion speed increases, the probability of default at maturity significantly decreases, especially for longer time horizons. The standard deviation of the returns of the risky asset decreases due to the mean reversion. Therefore, on the long run, investing in a risky asset is profitable, and will lead to excess return. It is therefore more likely that in the long run, the insurance undertaking is able to pay its obligations. We have used relatively small mean reversion speeds. If the speed equals 0.05, approximately 5% of the deviation of the return with respect to the mean is corrected over one year. Even if this relatively low speed is used, the effects are enormous.

Adding mean reversion also affects the relative size of the probabilities of default of different cashflows pattern. If we compare Figure 23 and Figure 22, we see that probabilities of default are lower if mean reversion is added, like we expect. In the absence of mean reversion, the decreasing cashflows pattern leads to the lowest probability of default. If mean reversion is added, one cashflow at maturity leads to the lowest probability of default.

The assumption for the mean reversion speed substantially influences probabilities of ruin and default, especially for longer time horizons. Therefore, if longer time horizons are considered, one should be aware of this sensitivity, and extra care should be taken when setting the assumptions.

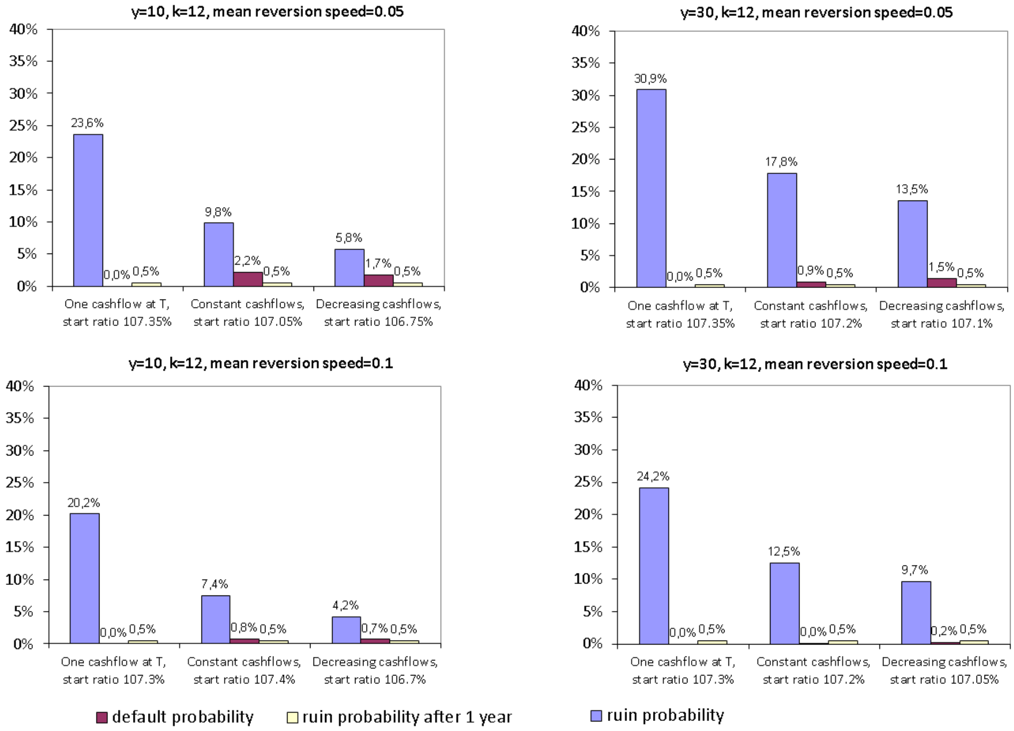


Figure 23: Default and ruin probabilities if mean reversion is added to the model. If we compare the results with Fig 22, we see that adding mean reversion substantially decreases the default probability, for all cashflows patterns. In the absence of mean reversion, a decreasing cashflows pattern leads to the lowest probability of default. If mean reversion is added, this role is taken over by the pattern which consists of one cashflow at maturity. If a longer time horizon is considered (i.e. longer than the one-year horizon in Solvency II), one should be aware that results are more sensitive to the assumptions chosen.

11 Implication for the level playing field

One of the purposes of Solvency II is to create a level playing field for all insurance companies in Europe. Having this in mind, one expects that two insurance companies, both with a solvency ratio of 100%, are equally safe. As we have seen, the probability that the undertaking can meet its obligations to the customer can differ substantial depending on the investment strategy, the time the payouts have to be made, and the cashflow pattern. Therefore, the question rises if Solvency II should focus entirely on a one-year horizon, or whether also other time horizons need to be taken into consideration. In this perspective, we can learn from discussions in the field of Pension Funds in the Netherlands.

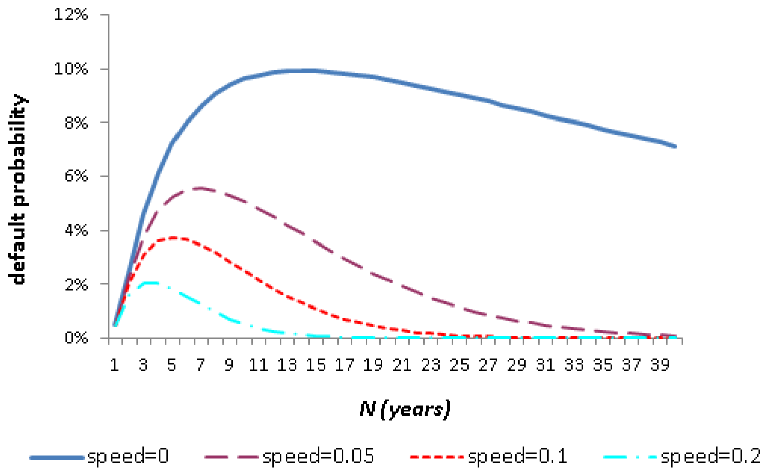


Figure 24: Analytical results of the influence of the mean reversion speed assumption on the default probability at maturity. One nominal guarantee at maturity is considered. Especially for longer guarantees, mean reversion substantially decreases the default probability.

In the Netherlands, the regulatory framework for Pension Funds, the FTK, consists of three parts:

- A solvency test over a one year horizon
- A continuity test over a longer horizon (15 years)
- A minimum funding ratio

As F. de Jong and A. Pelsser sketch in [23], the continuity test used to be seen as the most important of the three. However, currently the focus is on the one year solvency test. The asset mix of the pension fund is often adapted to reach one year nominal security. This leads to tensions between the long term objectives of the pension fund, and the short term regulatory demands. Therefore, they plead for a more prominent role of the continuity test, like it was intended in the FTK.
Pension funds and insurance companies cannot be compared one-to-one, since pension funds have the right to cut the rights of the participants in the fund if a severe market shock has occurred, and has the objective to compensate the participants for inflation. Insurance companies often have nominal obligations, or predetermined (inflation) guarantees. However, there is a parallel between the two. For both, the time horizon considered influences the conclusions drawn from a solvency study. On a one year horizon, regulation can

create a level playing field for funding security ("dekkingszekerheid"). However, the long term security of payouts ("uitkeringszekerheid" can differ substantial. Therefore, we plead for an extension of Solvency II, taking more time horizons into consideration.

If the Solvency II framework is extended to longer time horizons, a new challenge will rise. As we have seen, the long term results can be very sensitive to assumptions, like mean reversion behavior of risky assets. In sections 9 and 10, the required Solvency Capital over a one year horizon is typically around 7% of the provision, regardless of the mean reversion speed and cashflow pattern. For longer horizons, default probabilities are very sensitive to the mean reversion speed. When analyzing the results, one should be aware of the impact assumptions can have.

12 Conclusions

In the first part of this thesis, we investigated misspecifications and misinterpretations of Value-at-Risk models. If VaR models are backtested using a rolling window without correcting for this (a misinterpretation), the probability of finding an extreme event can typically increase by a factor 4 if data follows a Random Walk, and even by a factor 7 when data follows an AR(1) process. This can lead to a false rejection of the model. Therefore, this effect should be taken into consideration when conclusions are drawn from a backtest with a rolling window. In the evaluation report of the FTK, the validity of the equity stress parameter for Dutch pension funds is backtested using a rolling window. In the report, it is concluded that the stress parameter is too low, given the shocks that have occurred in the past years. We have shown that the rejection of this parameter can largely be explained by the use of a rolling window.
We investigated four misspecifications. If autocorrelation in errors is neglected, but normality is assumed, the probability of finding an extreme event can increase substantial. This effect increases if the measure for autocorrelation increases. Based on historic interest rate data, we find that for relevant parameter values, the probability can increase by a factor 3. If errors are Student-t distributed, or clustered volatility is neglected, we find theoretical factors around 2. Based on historic equity data, we have shown that for relevant parameter values, the effect of neglecting clustered volatility can result in an increase by a factor around 1.5. Scaling up the monthly Value-at-Risk of a Jump Diffusion process to a yearly VaR by using the square root of time rule, can theoretically both lead to an under- and overestimation of the yearly VaR, depending on the frequency of the jumps. Based on calibrations of Jump Diffusion models on historic equity (option) data, we find that for realistic parameter values, both underestimations and overestimations of the VaR can be found.

In the second part of this thesis, we have studied insurance companies all starting with a Solvency II ratio of 100%. We have shown that within-horizon risk increases with time, but end-of-horizon risk can decrease after a certain point. The gap between the two, representing companies that ruin at a certain point in time, but would have been able to recover from this situation and fulfil all their obligations, increases with time. Next to that, probabilities of default and ruin over longer time horizons depend on the investment policy, the cashflow pattern of the liabilities, and the duration of the liabilities. Therefore, we conclude that Solvency II should not focus solely on the one-year horizon, but should also take longer time horizons into account. Probabilities of default decrease substantially if risky assets are assumed to be mean reverting, in particular for long dated liabilities. If longer time horizons are taken into account, the sensitivity of the results for the assumptions made will increase. Therefore, if longer time horizons are considered, assumptions like mean reversion should be chosen with care.

References

[1] DIRECTIVE 2009/138/EC, 2009

[2] CEIOPS-DOC-33/09

[3] F. Diebold, A. Hickman, A. Inoue, and T. Schuermann, Converting 1-day volatility to h-day volatility: Scaling by $\sqrt{h}$ is worse than you think, Discussion paper series, no. 97-34, Wharton, 1997

[4] J. Danielsson and J.P. Zigrand, On time-scaling of risk and the square-root-of-time rule, Journal of Banking & Finance 30, pp. 2701-2713, 2006

[5] A.J. McNeil, Extreme value theory for risk managers, RISK Books, pp. 93-113, 1999

[6] A.J. McNeil and R. Frey, Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, Journal of Empirical Finance, 7, pp. 271-300, 2000

[7] J. Danielsson and C. de Vries, Value-at-Risk and Extreme Returns, Annales d'economie et de statistique, No 60, pp. 236-269, 2000

[8] T.G. Andersen, R.A. Davis, J-P Kreiss and T. Mikosch, Handbook of Financial Time Series, Springer, May 2009

[9] R. Merton, Option pricing when underlying stock returns are discontinuous, Journal of Financial Economics, 3, pp. 125-144, 1976

[10] S.G. Kou, A Jump-Diffusion model for option pricing, Management Science, Vol. 48 No. 8, pp. 1086-1101, 2002

[11] L. Andersen and J. Andreasen, Jump-Diffusion Processes: Volatility Smile Fitting and Numerical Methods for Option Pricing, Review of Derivatives Research, 4, pp. 231-262, 2000

[12] C. He, J. S. Kennedy , T. Coleman , P. A. Forsyth , Y. Li and K. Vetzal, Calibration and hedging under jump diffusion, Review of Derivatives Research, Vol. 9(1), pp. 1-35, 2006

[13] L.O. Scott, Pricing stock options in a Jump-Diffusion model with stochastic volatility and interest rates: applications of fourier inversion methods, Mathematical Finance, Vol. 7 No. 4, pp. 413-424, 1997

[14] Rapport Evaluatie Financieel Toetsingskader, www.rijksoverheid.nl, April 2010

[15] J.Y. Campbell and L.M. Viceira, The Term Structure of the Risk-Return Tradeoff, NBER Working Paper 11119

[16] R.P.M.M. Hoevenaars, R. D.J. Molenaar, P.C. Schotman and T.B.M. Steenkamp, Strategic asset allocation with liabilities: Beyond stocks and bonds, Journal of Economic Dynamics & Control, 32 pp. 2939-2970, 2008

[17] P. Devolder, Solvency requirements for long term guarantee: risk measure versus probability of ruin, European Actuarial Journal, Vol. 1, 2, pp. 199-214, 2011

[18] S.R. Thorley, The Time-Diversification Controversy, Financial Analysts Journal, Vol.51, No.3, pp. 68-76, 1995

[19] M. Kritzman and D. Richm, The Mismeasurement of Risk, Financial Analysts Journal, Vol. 58, No.3, pp. 91-99, 2002

[20] M. Barth, A comparison of risk-based capital standards under the expected policyholder deficit and the probability of ruin approach, The Journal of Risk and Insurance, Vol. 67, No.3, pp. 387-413, 2000

[21] S. Yen and S. Lin, Re-examining the contributon of Intra-Horizon risk measures, online paper

[22] C. Yi, On the first passage time distribution of an Ornstein-Uhlenbeck process, Journal of Quantitative Finance, Vo.10, 9, 2010

[23] F. de Jong and A. Pelsser, Herziening Financieel Toetsingskader, nea paper 33, Netspar economische adviezen, 2010

A Minimum of a geometric Brownian motion

Analogous to P. Devolder [17], we will use the law of the minimum of a geometric Brownian motion to compute probabilities of ruin.

Let $w(t)$ be a standard Brownian motion.

$$
\begin{aligned}
S(t) &= e^{(\mu-\sigma^2/2)t+\sigma w(t)} \\
z &= min_{0\leq s\leq t} S(s) \\
0 &< L \leq 1 \\
P(z \leq L) &= \Phi(d_1) + L^{\frac{2\mu}{\sigma^2}-1} \cdot \Phi(d_2) \\
d_{1,2} &= \frac{\log(L) \mp (\mu - \sigma^2/2)t}{\sigma\sqrt{t}}
\end{aligned}
\tag{58}
$$

B The N-year SC, probabilities of ruin and default for different investment strategies

Investing the SC in the risk free asset, rebalance assets

$$A(t) = e^{((\beta\delta+(1-\beta)r-\beta^2\sigma^2/2)+\beta\sigma w(t))} + SC \cdot e^{rt} \tag{59}$$

$$SC_{reb,rf}(N) = 1 - e^{\beta\sigma\Phi^{-1}(1-\alpha)\sqrt{N}+(\beta(\delta-r)-\beta^2\sigma^2/2)N} \tag{60}$$

$$\begin{aligned} P_{r,reb,rf} &= \Phi\Big(\frac{\log(1-SC)-(\beta(\delta-r)-\beta^2\sigma^2/2)N}{\beta\sigma\sqrt{N}}\Big) \\ +(1-SC)^{\frac{2\beta(\delta-r)}{\beta^2\sigma^2}-1} &\cdot \Phi\Big(\frac{\log(1-SC)+(\beta(\delta-r)-\beta^2\sigma^2/2)N}{\beta\sigma\sqrt{N}}\Big) \end{aligned} \tag{61}$$

$$P_{d,reb,rf} = \Phi\Big(\frac{\log(1-SC)-(\beta(\delta-r)-\beta^2\sigma^2/2)N}{\beta\sigma\sqrt{N}}\Big) \tag{62}$$

Investing the SC in the current asset mix, do nothing

$$A(t) = (1-\beta)(1+SC)\cdot e^{rt} + \beta(1+SC)\cdot e^{(\delta-1/2\sigma^2)t+\sigma w(t)} \tag{63}$$

$$SC_{dn,cam}(N) = \frac{e^{\sigma\Phi^{-1}(1-\alpha)\sqrt{N}+((\delta-r)-\sigma^2/2)N}-1}{1-1/\beta-e^{\sigma\Phi^{-1}(1-\alpha)\sqrt{N}+((\delta-r)-\sigma^2/2)N}} \tag{64}$$

$$\begin{aligned} P_{r,dn,cam} &= \Phi\Big(\frac{\log(\frac{1-SC/\beta+SC}{1+SC})-((\delta-r)-\sigma^2/2)N}{\sigma\sqrt{N}}\Big) \\ +(\frac{1-SC/\beta+SC}{1+SC})^{\frac{2(\delta-r)}{\sigma^2}-1} &\cdot \Phi\Big(\frac{\log(\frac{1-SC/\beta+SC}{1+SC})+((\delta-r)-\sigma^2/2)N}{\sigma\sqrt{N}}\Big) \end{aligned} \tag{65}$$

$$P_{d,dn,cam} = \Phi\Big(\frac{\log(\frac{1-SC/\beta+SC}{1+SC})-((\delta-r)-\sigma^2/2)N}{\sigma\sqrt{N}}\Big) \tag{66}$$

Investing the SC in the risk free asset, do nothing

$$A(t) = (1-\beta+SC)\cdot e^{rt} + \beta\cdot e^{(\delta-1/2\sigma^2)t+\sigma w(t)} \tag{67}$$

$$SC_{dn,rf}(N) = \beta(1-e^{\sigma\Phi^{-1}(1-\alpha)\sqrt{N}+((\delta-r)-\sigma^2/2)N} \tag{68}$$

$$P_{r,dn,rf} = \Phi\Big(\frac{\log(1 - SC/\beta) - ((\delta - r) - \sigma^2/2)N}{\sigma\sqrt{N}}\Big)$$
$$+(1 - SC/\beta)^{\frac{2(\delta-r)}{\sigma^2}-1} \cdot \Phi\Big(\frac{\log(1 - SC/\beta) + ((\delta - r) - \sigma^2/2)N}{\sigma\sqrt{N}}\Big) \tag{69}$$

$$P_{d,dn,rf} = \Phi\Big(\frac{\log(1 - SC/\beta) - ((\delta - r) - \sigma^2/2)N}{\sigma\sqrt{N}}\Big) \tag{70}$$

C R Code

Code used in section 4

```
alpha← c(rep(0,10),rep(0.1,9), rep(0.2,8),rep(0.3,7),rep(0.4,6),
rep(0.5,5),rep(0.6,4),rep(0.7,3),rep(0.8,2),0.9)
beta←c(c(0:9)/10,c(0:8)/10,c(0:7)/10,c(0:6)/10,c(0:5)/10,
c(0:4)/10,c(0:3)/10,c(0:2)/10,c(0:1)/10,0)
p←rep(0,55)
num←1000000
k←12
for (i in 1:55){
a←alpha[i]
b←beta[i]
set.seed(1)
garch.sim=garch.sim(alpha=c(0.1,a),beta=b,n=k*num)
sd←sd(garch.sim)
data←matrix(garch.sim,nrow=k,ncol=num)
data←t(data)
A←rowSums(data)
B←A[A>qnorm(0.995)*sd*sqrt(k)]
n←length(B)
prob←n/num
p[i]←prob
}
```

Code used in section 3

```
nu←c(rep(4,6),rep(5,6),rep(6,6),rep(7,6),rep(8,6),rep(9,6),rep(10,6),rep(100,6))
k←c(rep(c(1,2,3,4,6,12),8))
p←rep(0,48)
num←1000000
for (i in 1:48){
a←nu[i]
b←k[i]
set.seed(1)
rand←rt(b*num,a)
sd←sd(rand)
data←matrix(rand,nrow=b,ncol=num)
data←t(data)
A←rowSums(data)
B←A[A>qnorm(0.995)*sd*sqrt(b)]
n←length(B)
prob←n/num
p[i]←prob
```

```
}
```

Code used in section 5

```
lambda ←c(rep(0,6),rep(1/(10*12),6),rep(1/(20*12),6),rep(1/(30*12),6),rep(1/(50*12),6))
delta ←rep(c(0.01,0.2,0.4,0.6,0.8,1),5)
sigma ←0.15/sqrt(12)
mu ←0
p ←rep(0,30)
p2 ←rep(0,30)
sdwaarde ←rep(0,30)
quantwaarde ←rep(0,30)
quantwaardeopgeschaald ←rep(0,30)
num ←1000000
k ←12
for (i in 1:30){
l ←lambda[i]
d ←delta[i]
set.seed(1)

    errornormal ←rnorm(num*k)
errorpoisson ←rpois(num*k,l)
reeks ←mu+sigma*errornormal+log(d)*errorpoisson
sd ←sd(reeks)
sdwaarde[i] ←sd
quant ←quantile(reeks,0.005)
quantwaarde[i] ←quant
data ←matrix(reeks,nrow=k,ncol=num)
data ←t(data)
A ←rowSums(data)
quant2 ←quantile(A,0.005)
quantwaardeopgeschaald[i] ←quant2
B ←A[A<(-qnorm(0.995)*sd*sqrt(k))]
C ←A[A<quant*sqrt(k)]
n ←length(B)
prob ←n/num
p[i] ←prob
m ←length(C)
prob2 ←m/num
p2[i] ←prob2
}
ratio ← (1-exp(quantwaardeopgeschaald))/(sqrt(k)*(1-exp(quantwaarde)))
```

D Scilab Code

Code computation probabilities of ruin and default:

```
stacksize("max")
clear

//number of simulations
N=5000
//number of years
y=30
//steps per year
k=12
// x = percentage equity. (1-x) in risk free bonds. Sigma standard deviation
equity.
x=0.2
deltat=1/12
//start interest rate
rstart=0.04

T=k*y

//read scenarios from file
rente=fscanfMat("interest_scenario.txt")
equity=fscanfMat("equity_scenario.txt")

//choose startratio such that one year default probability of 0.5%
start_ratio_vec=[1.08,1.08,1.08]
liability_vec=[1,2,3]

yeardefaultvec=zeros(size(start_ratio_vec,"c"),1)
ruinvec=zeros(size(start_ratio_vec,"c"),1)
enddefaultvec=zeros(size(start_ratio_vec,"c"),1)

for v=1:size(start_ratio_vec,"c")

// set the liabilities to one of the three specifications

liabcode=liability_vec(v)
liab=zeros(T,1)
if liabcode==1 then
liab(T,1)=100;
elseif liabcode==2 then
for i= 1:T
liab(i)=100;
end
```

```
elseif liabcode==3 then
for i= 1:T
liab(i)=100*(T-i+1);
end
end

     start_ratio=start_ratio_vec(v)
L=zeros((k*y+1),N)
E=zeros((k*y+1),N)
B=zeros((k*y+1),N)
A=zeros((k*y+1),N)
R=zeros((k*y+1),N)

discount_t0=zeros(T,1)
for i=1:T
discount_t0(i)=(1/(1+rstart))^(i/k)
end

duration=zeros(T,1)

L0=discount_t0'*liab
A0=L0*start_ratio

L(1,:)=L0
A(1,:)=A0
E(1,:)=x*A0
B(1,:)=(1-x)*A0
R(1,:)=(A(1,1)-L(1,1))/L(1,1)

for t=2:(T)
liab_new=liab(t:size(liab,"r"))
timevec=zeros(size(liab_new,"r") ,1)
for j=1:size(liab_new,"r")
timevec(j)=j/k
end
payout=liab(t-1,1)
for n=1:N
discount=zeros(size(liab_new,"r"),1)
for j=1:size(liab_new,"r")
discount(j)=(1/(1+rente(t,n)))^(j/k)
end
E(t,n)=E(t-1,n)*equity(t,n)/equity(t-1,n)
L(t,n)=discount'*liab_new
duration(t-1)=(discount'*(liab_new.*timevec))/(discount'*liab_new)
B(t,n)=B(t-1,n)*((1+rente(t-1))^(1/k)+duration(t-1)*(rente(t-1,n)-rente(t,n)))
A(t,n)=E(t,n)+B(t,n)-payout
```

```
E(t,n)=A(t,n)*x
B(t,n)=A(t,n)*(1-x)
R(t,n)=(A(t,n)-L(t,n))/L(t,n)
end
end

for t=(T+1):(T+1)
duration(T)=1/k
payout=liab(T,1)
for n=1:N
E(t,n)=E(t-1,n)*equity(t,n)/equity(t-1,n)
B(t,n)=B(t-1,n)*(1+rente(t-1))^(1/k)+duration(T)*(rente(t-1,n)-rente(t,n)))
A(t,n)=E(t,n)+B(t,n)-payout
E(t,n)=A(t,n)*x
B(t,n)=A(t,n)*(1-x)
end
end

//default at maturity
default=zeros(N,1)
for n=1:N
if(size(find(A(:,n)<0),"c")>0) then default(n)=1;
else default(n)=0;
end
end

//probability of ruin
ruin=zeros(N,1)
//probability of default after one year
sol=zeros(N,1)
RU=A-L
for n=1:N
if(size(find(RU(:,n)<0),"c")>0) then ruin(n)=1;
else ruin(n)=0;
end
if(size(find(RU(k,n)<0),"c")>0) then sol(n)=1;
else sol(n)=0;
end
end

p_default=sum(default)/size(default,"r")
p_ruin=sum(ruin)/size(ruin,"r")
p_default1year=sum(sol)/size(sol,"r")

yeardefaultvec(v)=p_default1year
ruinvec(v)=p_ruin
```

```
enddefaultvec(v)=p_default
end
```

Interest rate and equity scenarios

```
stacksize(”max”)
clear
seed=1
//number of simulations
N=5000
//number of years
y=50
k=12
speed=0.1
// x = percentage equity. (1-x) in risk free bonds. Sigma standard deviation
equity.
x=0.2
sigma_eq=0.16
sigma_eq_k=sigma_eq/sqrt(k)
//sigma_int=0.1
sigma_int=0
sigma_int_k=sigma_int/sqrt(k)
rho=0

delta=0.07
deltat=1/12
rstart=0.04
hulpspeed=exp(-speed*deltat)

T=k*y

rente=zeros((T+1),N)
equity=zeros((T+1),N)
ou=zeros((T+1),N)

rand1=rand(T,N,”normal”)
rand2help=rand(T,N,”normal”)
rand2=rho*rand1+sqrt((1-rho)^2)*rand2help

for j=1:N
rente(1,j)=rstart
equity(1,j)=1
ou(1,j)=0
for i=1:T
rente(i+1,j)=rente(i,j)*exp(rand1(i,j)*sigma_int_k-0.5*sigma_int_k^2)
ou(i+1,j)=ou(i,j)*hulpspeed + rand2(i,j)*sigma_eq*sqrt((1-exp(-2*speed*deltat))/(2*speed))
```

```
equity(i+1,j)=equity(i,j)*exp(delta/k-0.5*sigma_eq^2*(1-exp(-2*speed*deltat))/(2*speed))
*exp(ou(i+1,j)-ou(i,j))

end
end
```