Value-at-Risk Analysis of Stock Returns Historical Simulation, Variance Techniques or Tail Index Estimation? *

Rob van den Goorbergh
Dept. of Economics
Tilburg University
P.O. Box 90153
5000 LE Tilburg
The Netherlands
e-mail: r.w.j.vdnGoorbergh@kub.nl

Peter Vlaar
Econometric Research and
Special Studies Dept.
De Nederlandsche Bank
P.O. Box 98
1000 AB Amsterdam
The Netherlands
e-mail: p.j.g.Vlaar@dnb.nl

March 1999

Abstract

In this paper various Value-at-Risk techniques are applied to the Dutch stock market index AEX and to the Dow Jones Industrial Average. The main conclusion are: (1) Changing volatility over time is the most important characteristic of stock returns when modelling value-at-risk; (2) For low confidence levels, the fat tails of the distribution can best be modeled by means of the t-distribution; (3) Tail index estimators are not successful, due to the fact that they can not cope with the volatility clustering phenomenon.

Keywords: Value-at-Risk, AEX, Dow Jones, Capital Requirements

J.E.L. Code: G28, G24

*This paper was written while the first author was affiliated with De Nederlandsche Bank. The authors thank Siem Jan Koopman for helpful comments and remarks.

1 Introduction

In the financial literature, three types of risk are distinguished, viz *business risk*, *strategic risk* and *financial risk*.[1] Business risk pertains to the risk firms face solely on account of their presence in some product market. This type of risk stems from uncertainty in such activities as technological innovations, product design and marketing. Strategic risk results from fundamental changes in the economic or political environment. A case in point is the expropriation of land and the nationalisation of businesses in communist countries in this century. This type of risk is typically very hard to quantify. And finally, there's financial risk, which is caused by movements in financial markets. For instance, changes in the prices of financial assets may affect the investment portfolio of a financial institution and bring about huge losses or gains. Although all three kinds of risk are important, we shall solely be concerned with financial risk for the remainder of the analysis.

Financial risk can be broken down further into various categories. There's *market risk* brought on by changes in the prices of financial assets and liabilities, *credit risk* caused by the unwillingness or inability of counterparties to fulfil their contractual obligations, *liquidity risk* resulting from insufficient market activity, *operational risk* due to inadequate systems, management failures or fraud, and *legal risk* that arises when a counterparty does not have the authority to engage in a transaction. Our focus will be entirely on the analysis of market, or price risk.

An extensive literature exists on the analysis of market risk. The availability of information from financial markets allows us to empirically examine this type of risk better than any other kind. In order to examine market risk, it has to be measured. There are different ways to do this. One commonly used measure of the price risk of an investment in some financial asset, is the standard deviation of the price of that asset. But if one is particularly interested in the maximum down-side risk one is exposed to, the so-called *Value-at-Risk*, *VaR* for short, might be a more suitable instrument. It was made popular by US investment bank J.P. Morgan, who incorporated it in their risk management model *RiskMetrics*™, to which we will come back later. Loosely speaking, the Value-at-Risk of a portfolio is the maximum loss that may be suffered on that portfolio in the course of some *holding period*, during which the composition of the portfolio remains unchanged. The length of this holding period is short-term, usually one day to a few weeks. So the Value-at-Risk of an investor's portfolio is the maximum amount of money he or she may lose in a short period of time.

A VaR always relates to some *confidence level*, typically in the range of 95 to 99.9 percent. So, strictly speaking, the VaR doesn't really pertain to the *maximum* loss that may be incurred, but it tells us the worst portfolio result that happens once every so many days.

Three aspects need to be kept in mind when judging the Value-at-Risk of a portfolio. In the

[1] See e.g. Jorion (1997).

first place, we need to know the initial value of the portfolio. For analytical purposes, the initial portfolio value is usually normalised to 100 currency units, but it could be any other amount, of course. A second ingredient is the holding period to which the VaR pertains. And finally, the confidence level is of importance. Evidently, the higher the confidence level, the larger the Value-at-Risk of the portfolio. By varying the confidence level, one is able to explore a whole *risk profile*, i.e. the entire distribution of results is revealed.

In recent years, many financial institutions have embraced Value-at-Risk as an easy to understand instrument to assess information about their portfolio positions. Apart from the conceptional attractiveness, its popularity is encouraged by the Basle Committee who permitted banks to calculate their capital requirements for market risks by means of VaR models. In order to calculate the VaR for this purpose, banks can choose between either historical simulation, variance-covariance techniques and Monte Carlo simulation. A priori, it is not clear which method provides the best results. In this paper, we will compare variance-covariance and historical simulation techniques applied to two stock market indices, the Dutch AEX and the Dow Jones Industrial Average. Several confidence levels will be evaluated, including the 99% level which is used to determine capital requirements.

The outline of the remaining sections will be as follows. In Section 2, the international regulations are outlined and the statistical evaluation of back-tests are given. Section 3 compares results for several VaR techniques applied to the AEX. Section 4 shows the results for the Dow Jones. Here more emphasis is given on the evaluation of failure rates through time. Finally, Section 5 concludes.

2 Evaluation of VaR

2.1 Regulations

Increasing financial instability has led to a call for tighter regulations. On 15 July 1988, the central banks of the G-10 countries[2] plus Luxembourg and Switzerland signed the *Basle accord*, an agreement which had the purpose to provide a level playing field for banks by imposing minimum capital requirements applying to all member countries. The Basle accord requires a bank's capital to be equal to at least 8 percent of the total risk-weighted assets of the bank.

The 1988 Basle accord has had to endure much criticism on several fronts. Besides its ad hoc character the two main points of criticism were that the accord did not recognise that credit risk may be reduced by diversifying across issuers, industries and geographical locations. Moreover, the

[2]The G-10 consists of Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, Sweden, the United Kingdom and the United States.

regulations were in terms of book values rather than in terms of market values, which insufficiently reflect the true risk.

Aware of the hiatus in the 1988 agreement, which focused chiefly on credit risk, the Basle Committee produced a number of proposals on market risk, which would be incorporated in the 1996 amendment to the capital accord. In April 1993, the so-called *Standard Model approach* was called into being. This is a building-block approach to determine a bank's capital charge which works as follows. First, capital requirements are computed for portfolios exposed to interest rate risk, exchange rate risk, equity risk and commodity risk. Then the bank's total capital charge is obtained simply by adding up the requirements corresponding to these four categories. The main shortcoming of this rather rigid rule is, of course, that it fails to take diversification across risks into account. The improvement on the previous regulations is that it is not just restricted to credit risk, but considers other important types of risk as well.

Two years later, in April 1995, the Basle Committee came forth with another set of proposals, which was nothing short of a regulatory innovation: the *Internal Model approach*. For the first time banks would be allowed to use their own risk management models to determine their VaR and with it their capital requirement. This capital requirement follows simply by multiplying the VaR by an add-on factor. This add-on factor, sometimes called the *hysteria* factor, may vary between three and four, depending on the accuracy of the bank's model in the past.[3] The hysteria factor is intended to provide additional protection against environments that are much less stable than historical data would lead to believe.

Of course, this liberal legislation required a number of preconditions. In the first place, the holding period and the confidence level for the VaR calculations needed to be set. The Committee decided to fix the holding period at 10 trading days. This 10-day period is motivated by the fact that in cases of severe market stress such as October 1987, markets can become very illiquid. Unprofitable parts of the portfolio can in such cases not be timely liquidated. Although banks should report a 10-day VaR, these VaR calculations should be based on daily return data, which is motivated by the fact that banks alter the composition of their portfolio rather on a daily than on a biweekly basis. Furthermore, the confidence level was set at 99 percent, corresponding to a loss exceeding the VaR once every hundred days, or two to three times a year.

In addition, the Internal Model approach required that a *back-test* be performed over a period of one year, or 250 trading days. A back-test is a means of verifying whether an internal model is adequate or not. Realised day-to-day returns on the bank's portfolio are compared to the VaR of the bank's portfolio. By counting the number of times the actual portfolio result was worse than

[3]Actually, the capital charge is set at the higher of the previous day's VaR, and the average VaR over the last 60 business days, times the add-on factor.

the VaR, the supervisor obtains an idea about how well the bank's internal model predicts its true market risk exposure. If this number corresponds to approximately 1 percent of the back-test trading days—the prescribed left tail probability—, the model will be fine, but if the number of violations is too high, a penalty is imposed resulting in an extra capital charge. This is reflected in an increase in the hysteria, or scaling factor. The following table presents the precise penalty directives.

Table 1: The relationship between the scaling factor and back-test results.

Zone	Number of violations	Scaling factor
	0	3.00
	1	3.00
Green Zone	2	3.00
	3	3.00
	4	3.00
	5	3.40
	6	3.50
Yellow Zone	7	3.65
	8	3.75
	9	3.85
Red Zone	more than 9	4.00

Table 1 shows that if the number of times the VaR is exceeded over a period of 250 trading days is not more than 4, we find ourselves in the Green Zone, and the multiplicative factor is 3, which means the capital requirement is three times the VaR of the bank's portfolio. This scaling factor is gradually increased as we pass through the Yellow Zone, until the Red Zone is entered where a maximum scaling factor of 4 is imposed. In that case the bank is likely to be obliged to revise its internal model for risk management.

In addition to the Internal Model approach directives, banks have to perform *stress tests* which may subject them to extra capital charges. These stress tests examine the effect of simulated large movements in important financial variables on the bank's portfolio. For instance, one could specify a scenario where the equity index changes by plus or minus 10 percent, or volatilities change by plus or minus 20 percent of their current values. The advantage of this stress testing, or *scenario analysis*, is that it may cover situations completely absent from the historical data, whereas VaR models are typically solely based on the historical data; stress testing forces management to consider events that they might otherwise ignore. The main drawback of scenario analysis is that it is completely subjective; stress tests do not specify the likelihood of worst-case situations. Still, scenario analysis can be an important means in laying bare the weaknesses in a portfolio.

The Internal Model approach presents banks with a trade-off between basic capital requirements and penalty costs. If a bank's internal model underestimates its market risk, the basis of the capital requirement, which is the VaR, is low, but the scaling factor may be set higher than 3 if the model is too bad. On the other hand, if the VaR estimate resulting from the internal model is relatively high, the profit opportunity set of the bank could be seriously restricted unless this is offset by a mild scaling factor. There has been some criticism on the Internal Model approach with respect to this trade-off. The discrepancy between the minimum and maximum scaling factor is thought to be too small. This gives banks an impetus to report too low a VaR, so that profit opportunities are kept on a high level, because the penalty for using a model that undervalues the VaR is relatively small. That is why there is a call for higher penalties, see e.g. Lucas (1997) and Vlaar (1998).

2.2 Statistical evaluation

In the sections to come a number of techniques to assess the VaR of a portfolio will be discussed. All of these techniques will be applied to an imaginary investment portfolio, and evaluated by a back-testing procedure. To this end, the sample of portfolio returns is split up into an *estimation sample* and an *evaluation sample* for each technique. The estimation sample is used to estimate the model in question and to predict the VaR of the portfolio, whereupon a statistically sound back-test is conducted by means of the evaluation sample in order to assess the adequacy of the model.

The back-test will be performed in line with the Basle Internal Model approach regulations; we count the number of days in the evaluation sample the portfolio result was worse than is to be expected on the basis of our VaR estimate, and divide this number by the evaluation sample size. The resulting figure is called the *failure rate*. Subsequently, this failure rate is compared to the left tail probability p that was used to determine the VaR estimates. If these tally, we succeed in predicting our portfolio's Value-at-Risk accurately; if they differ substantially, the model has to be rejected.

A *likelihood ratio test* developed by Kupiec (1995) will be used to find out whether a Value-at-Risk model is to be rejected or not. Let N be the number of times the portfolio loss is worse than the true Value-at-Risk in a sample of size T. Then the number of VaR violations has a binomial distribution, $N \sim \mathcal{B}(T, p)$. Ideally, the failure rate, N/T, should be equal to the left tail probability, p. Thus, the relevant null and alternative hypotheses are

$$H_0 : N/T = p$$
$$H_a : N/T \neq p,$$

and the appropriate likelihood ratio statistic is

$$LR = 2\left[\log\left(\left(\frac{N}{T}\right)^{N}\left(1-\frac{N}{T}\right)^{T-N}\right) - \log\left(p^{N}\left(1-p\right)^{T-N}\right)\right]. \qquad (1)$$

This likelihood ratio is asymptotically χ_1^2 distributed under the null that p is the true probability the VaR is exceeded. With a certain confidence level, say $100 \cdot (1-\alpha)$ percent, we can construct nonrejection regions that indicate whether a model is to be rejected or not. In accordance with convention, α is set at 0.05. For a number of left tail probabilities and evaluation sample sizes the nonrejection regions are tabulated below. Clearly, the smaller the left tail probability, the more difficult it gets to confirm deviations, especially when the evaluation sample size is small.

Table 2: Nonrejection regions according to Kupiec's back-test.

left tail	evaluation sample size			
probability	250	500	750	1000
5.00%	$7 \leq N \leq 19$	$17 \leq N \leq 35$	$27 \leq N \leq 49$	$38 \leq N \leq 64$
1.00%	$1 \leq N \leq 6$	$2 \leq N \leq 9$	$3 \leq N \leq 13$	$5 \leq N \leq 16$
0.50%	$0 \leq N \leq 4$	$1 \leq N \leq 6$	$1 \leq N \leq 8$	$2 \leq N \leq 9$
0.10%	$0 \leq N \leq 1$	$0 \leq N \leq 2$	$0 \leq N \leq 3$	$0 \leq N \leq 3$
0.01%	$0 \leq N \leq 0$	$0 \leq N \leq 0$	$0 \leq N \leq 1$	$0 \leq N \leq 1$

The size of the test is 5%.

One may wonder whether we can relate Kupiec's back-test to the Basle accord regulations for the Internal Model approach. Note that the Kupiec back-testing procedure is based on a *two-sided* statistical test; both high and low failure rates lead to a possible rejection of a model. In particular, models that are too conservative are turned down, whereas the Basle accord directives only penalise models that underestimate the Value-at-Risk. For banks, however, it is not only important to know whether their model underpredicts the Value-at-Risk, but it is also important to know whether the model is too conservative because that would unnecessarily jeopardise their profit opportunities. Hence, for banks, the two-sided back-test is fit for the job of model evaluation. For bank supervision purposes, only too low VaR calculations are important, and therefore a one-sided test, for instance a quantile of the binomial distribution, is more appropriate.

For an evaluation sample size of 250 trading days, and a left tail probability of 1 percent, as prescribed by the Basle Committee, the critical value of the one-sided test equals 6 at a confidence level of 95 percent. That is to say, a VaR model should be rejected if it renders VaR predictions that are violated more than 5 times in a back-testing period of 250 trading days. The Basle penalty scheme is slightly more stringent as it starts imposing an extra capital charge for VaR violations over 4.[4] The subsequent gradual increase of the multiplication factor has no clear statistical

[4] The probability of finding 5 or more violations of the VaR in 250 days is about 10.8%.

background. It merely reflects the loss function of the Basle Committee.

3 Value-at-Risk techniques applied to the AEX

In this section, various VaR techniques will be applied to the the Dutch stock market index AEX. First, the data are analysed in order to get an idea of the stylised facts of stock market returns. Then, Variance techniques are discussed, beginning with static variances and followed by conditional heteroskedasticity techniques. Third, historical simulation and tail index estimation are investigated. Finally, the results for the AEX are summarised.

Throughout the analysis, a holding period of one day will be used. Various values for the left tail probability level will be considered, ranging from the very conservative level of 0.01 percent to the less cautious 5 percent. The various VaR models will be estimated using the data preceding the last thousand days of the sample, which will be used to evaluate them by means of Kupiec's back-test. So all models are evaluated 'out-of-sample' as opposed to 'in-sample'.

3.1 Data

There is an abundance of possibilities when it comes to the composition of a portfolio. One may include stocks, bonds, stock options, bond options, futures contracts, barrier options, caps, floors and many more financial instruments. Financial institutions typically hold portfolios comprising a multitude of assets. In this paper, however, we shall concentrate on the stock market. For an analysis of Value-at-Risk of portfolios invested in the bond market we refer to Vlaar (1998).

The analysis of portfolio risk can become very complicated indeed if one tries to examine large (stock) portfolios straight away. Moreover, information on the frequently changing composition of real life portfolios that would be needed to construct a realistic portfolio, is hard to come by. That is why we shall restrict ourselves, as far as the empirical evaluation of VaR techniques is concerned, to the simplest non-trivial portfolio possible, namely a portfolio consisting of a single stock index, starting with the AEX.

The AEX is an equity basket consisting of a 25 Dutch stocks in different weights. Throughout the analysis, it will be used as a representative stock. We have a sizeable time series of 4039 daily data at our disposal running from 3 January 1983 to 17 December 1998. During that time span, a period of 16 years, the index rose from 100 to 1113, or about 16 percent a year. Some descriptive statistics of the data are shown in Table 3.

The table shows that the average daily return is about 0 percent, or at least negligibly small compared to the sample standard deviation. This is why the mean is often set at zero when modelling daily portfolio returns, which reduces the uncertainty and imprecision of the estimates; see e.g. Figlewski (1994). Also, a model's parsimony is considered a virtue.

Table 3: Summary statistics for the AEX: 1983–1998.

number of observations	4038
mean	0.060%
median	0.079%
maximum	11.164%
minimum	-12.805%
standard deviation	1.187%
skewness	-0.523
kurtosis	14.534

From an inspection of the correlogram of the daily return series we conclude that there is hardly any evidence of autocorrelation in the data, at least for the first eighth lags. See Table 4.

Table 4: Autocorrelation in AEX returns: 1983–1998.

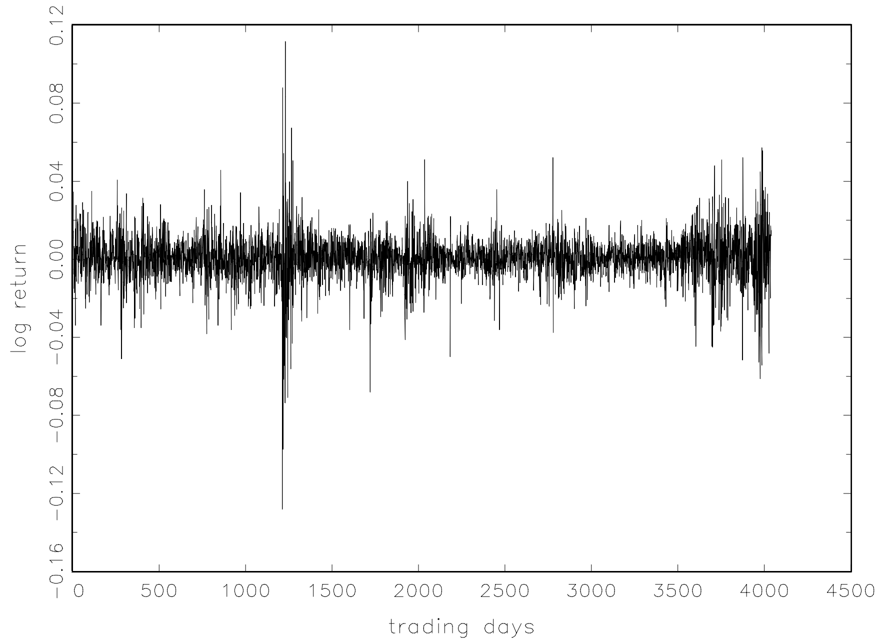
lag	ACF	PACF	Box-Ljung	prob
1	-0.006	-0.006	0.166	0.684
2	-0.016	-0.016	1.162	0.559
3	-0.002	-0.002	1.178	0.758
4	0.017	0.017	2.315	0.678
5	0.024	0.025	4.729	0.450
6	-0.008	-0.007	4.999	0.544
7	0.005	0.005	5.086	0.650
8	0.038	0.038	10.900	0.207

A plot of the log returns displays the volatility clustering phenomenon: large and small swings tend to cluster, see Figure 1. As is intuitively immediately clear, this will turn out to be important for the measurement of risk; the Value-at-Risk will be higher in tumultuous times than when the financial markets are in smooth waters. Furthermore, the maximum and minimum statistics are quite large in absolute value, indicating the presence of extreme returns. This is supported by the occasional extreme spikes in Figure 1 and the high sample kurtosis, which is indicative of the fatness of the tails of the distribution. Clearly, the most striking example of extreme instability in our data set was the stock market crash of October 1987 (round about trading day number 1200 in Figure 1).

3.2 Variance methods

In this subsection, a multitude of VaR techniques will be treated that are all based on some assumption concerning the distribution of portfolio returns. In particular, these so-called *variance methods* all relate the Value-at-Risk of a portfolio directly to the variance or standard deviation of the portfolio returns. Intuitively, the larger the variance of a portfolio return, the more likely the occurrence of large swings in the portfolio value, and the larger the Value-at-Risk.

Figure 1: Log return series of the Dutch stock index, AEX.



By way of introduction we will discuss the at first sight rather naïve *static methods*, that disregard the volatility clustering phenomenon mentioned in the previous chapter. After that, we shall look at a methodology that does consider volatility clustering, namely the *Generalised Autoregressive Conditional Heteroskedasticity*, or *GARCH* model Bollerslev (1986).

3.2.1 Static Models

Normality

The vast majority of VaR models based on variance techniques assume the normal distribution. As it is widely believed that high frequency financial return data have fatter tails than can be explained by the normal distribution, this artefact seems odd. However, the normal distribution entails some very convenient characteristics, that do not carry over to other distributions. First of all, the parameters of the normal distribution are usually easier to estimate as there is often an analytical solution for them. Especially for large models, this is an advantage. Second, and more important, is the additivity of the normal distributions. The sum of two normally distributed variables is also normally distributed. This characteristic is especially important for the calculation of multi-day VaRs based on one-day VaRs — a feature of the Basle guidelines. Assuming independence of normally distributed returns and a mean return of zero, it can easily be shown that:

$$VaR^{(T)} \approx \sqrt{T}\; VaR^{(1)}. \tag{2}$$

Formula (2) is known as the *square root of time rule*. It is quite useful when one needs to determine the T-periods-ahead risk of a portfolio on the basis of one-period-ahead return data. Owing to its simplicity, the square root of time rule is often applied in other VaR models, even though in most models this is inappropriate, either because of the fact that such models do not assume a normal or other sum-stable distribution[5], or because of the fact that the return data is not assumed to be independent.

Another interesting consequence of the additivity characteristic is the fact that the marginal distribution of a multivariate normal distribution is also normally distributed. Consequently the risk of a large multivariate linear portfolio can easily be expressed in the risks of the individual components.

Assuming independent and identically normally distributed (log) returns:

$$r_t \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$$

the Value-at-Risk boils down to

$$VaR = -W_0 \left(e^{\mu + \sigma \Phi^{-1}(p)} - 1 \right), \tag{3}$$

where W_0 is the initial value of the portfolio, and $\Phi(\cdot)$ represents the cumulative distribution function of the standard normal probability distribution. The parameters μ and σ can be efficiently estimated by means of (log) likelihood maximisation which for this simple model boils down to least squares estimation. Applying the normal distribution to our AEX portfolio, using an evaluation sample consisting of the last 1000 days and an estimation sample consisting of the preceding data, we end up with the following parameter estimates. The maximum mean log likelihood is 3.0371008.

parameter	estimate	standard error
μ	0.0470%	0.02106%
σ	1.1608%	0.00035%

Let us apply this normal fit to calculate the VaR. The table clearly shows the poor performance of the normal model. For the smaller left tail probabilities, the failure rates are quite high. This means that the mass in the tails of the distribution is underestimated, which is consistent with the assertion that stock returns are heavy tailed. Hence, the actual VaR is higher in these cases. Kupiec back-tests reject the normal model for left tail probabilities of 1 percent or smaller. Only

[5] The class of sum-stable distributions is characterised by the fact that sums of random variables from a sum-stable distribution again follow that sum-stable distribution.

for a left tail probability of 5 percent is the normal model not rejected.

left tail probability	VaR estimate	standard error	failure rate (times 1000)
5.00 %	1.85	0.032	53
1.00 %	2.62	0.039	26†
0.50 %	2.90	0.042	22†
0.10 %	3.48	0.049	12†
0.01 %	4.18	0.057	8†

Throughout the analysis a dagger (†) indicates the failure rate differs significantly from the corresponding left tail probability according to Kupiec's (two-sided) back-test at a 95 percent confidence level.

The Student-t Distribution

Now that we've concluded the normal model underestimates the probability mass out in the tails, it seems natural to try a distribution that has 'fatter' tails, i.e. a distribution that generates more realisations in the tails than is to be expected on the basis of a normal distribution. The Student-t distribution can deal with this phenomenon called *leptokurtosis*. The Student-t probability distribution is characterised by three parameters: the location parameter, μ, the scale, $\gamma > 0$, and the degrees of freedom parameter, $\nu > 0$. A random variable X that is Student-t distributed has a mean equal to μ, provided $\nu > 1$, and a variance equal to $\nu\gamma^2/(\nu-2)$, if $\nu > 2$.[6] For $\nu \to \infty$, the t-distribution reduces to the normal distribution with mean μ and variance γ^2. The smaller ν gets, the fatter the tails are.

Let us assume the log portfolio returns r_t follow a t-distribution. Then the log likelihood function equals

$$\begin{aligned}\mathcal{L} &= T\left[\log\Gamma\left(\frac{\nu+1}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log\pi\nu - \log\gamma\right] - \\ &\quad \frac{\nu+1}{2}\sum_{t=1}^{T}\log\left(1+\left(\frac{r_t-\mu}{\gamma\sqrt{\nu}}\right)^2\right) \qquad (4)\end{aligned}$$

where $\Gamma(\cdot)$ is the *Gamma function*.[7] Unfortunately, in contrast to the normal distribution, no analytical expressions are available for the maximum likelihood estimates of μ, γ and ν. Likelihood optimisation has to be done numerically. Applying the Student-t fit to our AEX portfolio, one

[6] In fact, all moments up to (but excluding) the ν-th moment exist.

[7] The Gamma function is defined as $\Gamma(\alpha) = \int_0^\infty e^{-x}x^{\alpha-1}dx$. In particular, $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

gets the following parameter estimates. The maximum mean log likelihood is 3.1807817.

parameter	estimate	s.e.
μ	0.0763%	0.0161%
γ	0.7329%	0.0168%
σ	1.1549%	0.0434%
ν	3.3489	0.2159

The Value-at-Risk can be expressed as follows:

$$VaR = -W_0 \left(e^{\mu + \gamma F_\nu^{-1}(p)} - 1 \right), \tag{5}$$

where $F_\nu(\cdot)$ is the cumulative distribution function of a standardised t distributed random variable. The VaR estimates are:

left tail probability	VaR estimate	failure rate (times 1000)
5.00 %	1.57	69†
1.00 %	2.95	21†
0.50 %	3.72	10†
0.10 %	6.16	0
0.01 %	12.10	0

As the table indicates, most failure rates are not very close to the corresponding left tail probabilities, certainly for the less conservative levels. Kupiec back-tests lead to rejection of the model for the 5, 1 and 0.5 percent level. At the more cautious levels, the failure rates do not differ significantly from their left tail probabilities. Here, the t-distribution clearly outperforms the normal fit. The probability mass far out in the tails is far more adequately captured by this heavy tailed distribution. However, the power of Kupiec's back-test is rather limited at these levels.

In the static normal model we were able to express the T-periods-ahead Value-at-Risk in terms of the one-period-ahead Value-at-Risk. Although the square root of time rule is not valid in the case of a t-distribution, it will be applied anyway, as there is no analytical alternative. In order to get a proper 10-day ahead VaR, one would need to perform Monte Carlo simulation.

Mixtures of Distributions

As the t-distribution does not capture the lower confidence levels, whereas the normal distribution does not capture the high ones, an alternative might be to consider mixtures of distributions. Most of the time, stock return deviations are relatively moderate. But every once in a while large swings occur and the volatility seems to jump to a higher level. One might say that a large proportion of the data is generated by a distribution with a relatively low variance, and that the remainder follows a distribution with a relatively high variance. This view is inspired by the literature on target zone exchange rates where one clearly discerns different volatility levels associated with different realignment probabilities, see e.g. Vlaar and Palm (1993).

We shall look at the simplest mixture distribution, namely a time-independent mixture of two normal distributions, the static Bernoulli-normal jump model, where we assume that stock returns follow a normal distribution with variance σ^2 with probability $1 - \lambda$, and a normal distribution with a higher variance $\tau^2 = \sigma^2 + \delta^2 > \sigma^2$ with probability λ. Thus the volatility level is in times of instability augmented by δ^2. Empirically, the stock return mean does not seem to be affected by volatility jumps, so we will model it by a constant parameter μ. The log likelihood function reads

$$\mathcal{L} = \sum_{t=1}^{T} \log \left\{ \frac{1-\lambda}{\sigma} \varphi \left(\frac{r_t - \mu}{\sigma} \right) + \frac{\lambda}{\tau} \varphi \left(\frac{r_t - \mu}{\tau} \right) \right\}, \tag{6}$$

where $\varphi(\cdot)$ is the density function of the standard normal distribution. Likelihood maximisation has to be done numerically as there are no analytical expressions for the ML estimators. The maximum mean log likelihood applied to the AEX amounts to 3.1733492.

parameter	estimate	s.e.
μ	0.0798%	0.0167%
σ	0.8151%	0.0184%
δ	2.7903%	0.2026%
λ	0.0879	0.0134

Note that the jump probability estimate is about 8 percent. This means that 92 percent of the time we find ourselves in the low variance distribution, and 8 percent of the time the high variance distribution prevails. As for the evaluation of the model, far out in the tail the performance is quite good. For higher confidence levels the performance is disappointing however. For the left tail probability of 5 percent the failure rate is even higher than for the Student-t.

left tail probability	VaR estimate	failure rate (times 1000)
5.00 %	1.49	75†
1.00 %	3.37	14
0.50 %	4.42	5
0.10 %	6.33	0
0.01 %	8.42	0

3.2.2 The GARCH Model

An important omission of the static models is that they do not take the volatility clustering into account. By far the most popular model to model this phenomenon is the so called Generalised Autoregressive Conditional Heteroskedasticity, or GARCH, model introduced by Bollerslev (1986). It is an extension of the Autoregressive Conditional Heteroskedasticity, or ARCH, model by Engle (1982). In the GARCH model we start by defining an innovation η_{t+1}, i.e., some random variable with mean zero conditional on time t information, $\mathcal{I}_t$. This time t information is a set including the innovation at time t, $\eta_t \in \mathcal{I}_t$, and all previous innovations, but any other variable available at time t as well. In finance theory, η_{t+1} might be the innovation in a portfolio return.

In order to capture serial correlation of volatility, or volatility clustering, the GARCH model assumes that the conditional variance of the innovations, $\sigma_t^2 \equiv \text{Var}[\eta_{t+1}|\mathcal{I}_t]$, depends on the latest past squared innovations—as is the assumption in the less general ARCH model—, possibly augmented by the previous conditional variances. In its most general form, the model is called GARCH(p, q), and it can be written as

$$\sigma_t^2 = \omega + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} \alpha_i \eta_{t-i+1}^2. \tag{7}$$

As (7) shows, p lags are included in the conditional variance, and q lags are included in the squared innovations. In this section, we shall regard these innovations as deviations from some constant mean portfolio return:

$$r_{t+1} = \mu + \eta_{t+1}, \tag{8}$$

so that σ_t^2 is also the conditional variance of the portfolio returns. We can write the innovation η_{t+1} as $\sigma_t \varepsilon_{t+1}$, where ε_{t+1} is assumed to follow some probability distribution with zero mean and unit variance, such as the standard normal distribution.

A great many empirical studies have proved it unnecessary to include more than one lag in the conditional variance, and one lag in the squared innovations. This is why our point of departure will be the GARCH(1,1) model:

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\eta_t^2, \tag{9}$$

with $\omega > 0$, $\beta \geq 0$ and $\alpha \geq 0$ to ensure positive variances. Formula (9) nicely demonstrates the essence of the volatility clustering feature in the GARCH model. If the market was volatile in the current period, next period's variance will be high, which is intensified or offset in accordance with the magnitude of the return deviation this period. If, on the other hand, today's volatility was relatively low, tomorrow's volatility will be low as well, unless today's portfolio return deviates from its mean considerably. Naturally, the impact of these effects hinges on the parameter values. Note that for $\alpha + \beta < 1$, the conditional variance exhibits *mean reversion*, i.e., after a shock it will eventually return to its unconditional mean $\omega/(1 - \alpha - \beta)$. If $\alpha + \beta = 1$ this is not the case, and we have *persistence*.

Normality

In order to estimate these parameters by means of likelihood maximisation, one has to make assumptions about the probability distribution of the portfolio return innovations η_{t+1}. First, we shall consider Gaussian innovations,

$$\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, 1), \tag{10}$$

so that

$$\eta_{t+1}|\mathcal{I}_t \sim \mathcal{N}(0, \sigma_t^2), \tag{11}$$

leading to a conditional log likelihood of η_{t+1} equal to

$$\ell_t(\eta_{t+1}) = -\log\sqrt{2\pi} - \frac{1}{2}\log\sigma_t^2 - \frac{\eta_{t+1}^2}{2\sigma_t^2}. \tag{12}$$

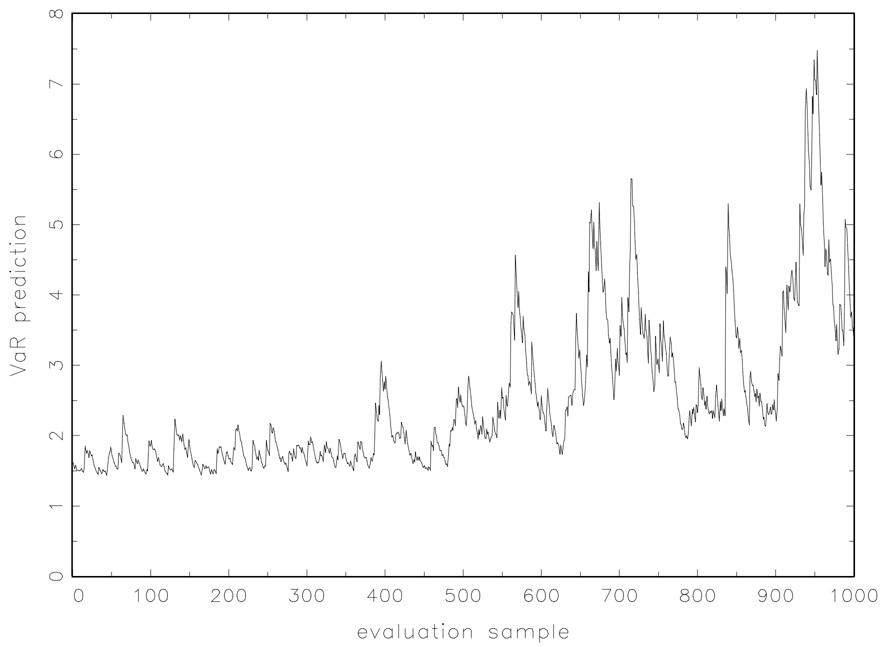
The log likelihood of the whole series $\eta_1, \eta_2, \ldots, \eta_T$ is

$$\mathcal{L} = \sum_{t=1}^{T} \ell_t(\eta_{t+1}). \tag{13}$$

Let us apply the above to our AEX dataset. The mean log likelihood attained its maximum at 3.1874773. The rest of the results are tabulated below. The evaluation of the VaR is shown in Figure 2. These VaR predictions correspond to a left tail probability of 1 percent. The VaR fluctuates a lot over time with a maximum value of almost five times the minimum one.

parameter	estimate	s.e.
μ	0.0727%	0.0167%
ω	0.0005%	0.0001%
α	0.1063	0.0138
β	0.8587	0.0175

Figure 2: GARCH-normal VaR predictions at the 1 percent level.



left tail probability	average VaR estimate	failure rate (times 1000)
5.00 %	1.76	54
1.00 %	2.51	17†
0.50 %	2.78	10†
0.10 %	3.34	3
0.01 %	4.02	1

The failure rates are not very satisfactory apart from the 54 in 1000 violations at the 5 percent level. Still, the failure rates look a lot better than the static normal approach in the previous section, especially for the more extreme events. We cannot reject the GARCH(1,1) model with normal innovations on the basis of a Kupiec back-test for the 0.1 and 0.01 percent left tail probabilities. This can be explained by the fact that conditional heteroskedasticity causes fat-tailedness. It can be shown that if the *conditional* distribution is normal and conditional heteroskedasticity is present, the *unconditional* distribution is still fat-tailed. However, in view of the diminishing power of the back-test out in the tail, one may utter the conjecture that, far out in the tail, the normal GARCH model still underpredicts the probability mass.

RiskMetrics™

As an alternative to the GARCH(1,1) model, or as a special case of this model really, US investment bank J.P. Morgan introduced *RiskMetrics*™, a VaR assessment method that basically restricts both μ and ω to 0, and α to $1-\beta$ in formula (9).[8] Furthermore, the parameter β, called the *decay factor* and renamed λ, is set at 0.94 for daily data.[9] This makes estimation elementary, since there are no parameters to estimate left. The portfolio return variance conditional on time t information is just:

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1-\lambda) r_t^2, \tag{14}$$

or,

$$\sigma_t^2 = \lambda^t \sigma_0^2 + (1-\lambda) \sum_{k=0}^{t-1} \lambda^k r_{t-k}^2, \tag{15}$$

where σ_0^2 is some initial variance level. The effect of this initial variance dies out as time passes since $\lambda < 1$. Hence, in practice, it may be fixed quite arbitrarily to the sample variance of the whole series, for instance.

From the formulae above, it is clear that the conditional variances are modelled using an *exponentially weighted moving average*: the forecast for time t is a weighted average of the previous forecast, using weight λ, and of the latest squared innovation, using weight $1-\lambda$. See formula (14). The RiskMetrics™ approach essentially boils down to keeping track of the return data and using these along with the decay factor to update the conditional volatility estimates.

Note that the imposition of the restriction that α and β should sum to unity implies *persistence* in the conditional variance, i.e., a shock moving the conditional variance to a higher level does not die out over time but 'lasts forever'. If there are no shocks offsetting this volatility increase, the conditional variance will remain high, and won't display mean-reverting behaviour. Hence, today's volatility affects forecasts of volatility into the future.

[8] See J.P. Morgan Bank (1996).

[9] The determination of this 'optimal' decay factor is based on minimisation of the mean squared variance forecast error with respect to λ for a number of asset return series. The decay factor of 0.94 is a weighted average of individual optimal decay factors.

The results of applying the RiskMetrics$^{\text{TM}}$ method to our AEX portfolio are depicted in the table hereunder.

left tail probability	average VaR estimate	failure rate (times 1000)
5.00 %	1.78	56
1.00 %	2.51	21†
0.50 %	2.77	13†
0.10 %	3.31	6†
0.01 %	3.97	1

The back-test results are rather disappointing: most failure rates exceed the respective left tail probabilities considerably. Not surprisingly, Kupiec tests reject the model for all left tail probabilities except for the 5 percent level, at which the RiskMetrics$^{\text{TM}}$ approach performs exceptionally well, and the extreme 0.01 percent level where model verification is problematic. Particularly at the 1 percent and 0.5 percent levels does RiskMetrics$^{\text{TM}}$ do worse than the GARCH(1,1) with normal innovations, probably due to the fact that all parameters are fixed.

The merit of the RiskMetrics$^{\text{TM}}$ approach becomes clear once we consider larger portfolios; other methods generally have to see that they manage a rapidly increasing number of parameters as portfolios become more extensive, whereas J.P. Morgan's approach deals with this issue by fixing the parameters.

Student-t Innovations

To remedy the problems in the tails one may try other assumptions regarding the distribution of the innovations. By analogy with the previous section we will look at the heavier tailed Student-t distribution. The conditional log likelihood of η_{t+1} reads

$$\begin{aligned} \ell_t(\eta_{t+1}) = \; & \log\Gamma\left(\frac{\nu+1}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log\pi(\nu-2) - \log\sigma_t - \\ & \frac{\nu+1}{2}\;\log\left(1+\left(\frac{\eta_{t+1}}{\sigma_t\sqrt{\nu-2}}\right)^2\right). \end{aligned} \tag{16}$$

Note that our assumption of a unit—and hence finite—error variance restricts the degrees of freedom parameter ν to values above 2. Applying the GARCH(1,1) model with Student-t errors to our AEX portfolio with our usual estimation and evaluation periods, we find the following parameter estimates at a maximum mean log likelihood of 3.2275759.

parameter	estimate	s.e.
μ	0.0796%	0.0152%
ω	0.0003%	0.0001%
α	0.0780	0.0131
β	0.8952	0.0173
ν	5.7728	0.5615

The average VaR estimates are presented in the table that follows.

left tail probability	average VaR estimate	failure rate (times 1000)
5.00 %	1.65	58
1.00 %	2.72	10
0.50 %	3.22	4
0.10 %	4.56	1
0.01 %	7.04	0

The GARCH(1,1) with t distributed innovations performs quite well for all confidence levels. For none of the left tail probability levels are we able to reject the model on the basis of a Kupiec back-test. On the whole its performance is superior to the GARCH(1,1) with normal innovations. If we compare the dynamic model of this section to the static Student-t of the previous section, we can conclude that the GARCH(1,1) model with t errors does also better than its static pendant. Ergo, it is the best model so far.

Bernoulli-Normal Mixture

The second fat-tailed distribution considered is the Bernoulli-normal mixture. As there are two normal distributions in the Bernoulli-normal jump process, the way GARCH effects are modelled is not unique. We will assume that the variance of the first normal distribution follows a GARCH process, whereas the second equals the first plus the jump variance.

$$r_{t+1}|\mathcal{I}_t \sim (1-\lambda)\mathcal{N}(\mu, \sigma_t^2) + \lambda\mathcal{N}(\mu, \tau_t^2), \tag{17}$$

where $\tau_t^2 = \sigma_t^2 + \delta^2$ with δ^2 the jump variance, and

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\eta_t^2, \tag{18}$$

with ω, β and $\alpha > 0$, and $\eta_t = r_t - \mu$.

Numerical likelihood optimisation results in the following. The mean maximum likelihood is 3.2334819.

parameter	estimate	s.e.
μ	0.0763%	0.0153%
λ	0.0369	0.0096
δ	0.0255	0.0031
ω	$1.8706 \cdot 10^{-6}$	$5.0650 \cdot 10^{-7}$
α	0.0602	0.0090
β	0.9033	0.0144

left tail probability	VaR estimate	failure rate (times 1000)
5.00 %	1.61	62
1.00 %	2.53	16
0.50 %	3.27	10†
0.10 %	5.17	1
0.01 %	7.33	0

The performance is somewhat disappointing. The Kupiec back-test rejects the model at 0.5 percent. The Student-t GARCH model clearly outperforms the Bernoulli normal mixture at all probabilities except the lowest two for which they score the same. Compared with the static mixture, both at 1 and 0.5 percent, the failure rates are worse.

3.3 Historical simulation

3.3.1 Plain

A different approach for VaR assessment is called Historical Simulation (HS). This technique is nonparametric and does not require any distributional assumptions. This is because HS uses essentially only the empirical distribution of the portfolio returns. It works as follows.

The sample of returns is split up into a number of equally long (overlapping) subsamples. The length of a subsample, or *window*, is called the *window size*. If the sample size is T, and the window size is n, we can construct $T - n + 1$ subsamples, so that two subsequent subsamples have all but one datum in common. Next, we pick the p-th percentile of each subsample, say R_t^p. This leaves us with an estimate of the portfolio's VaR for each subsample:

$$\widehat{VaR}_{t+1|t} = -W_0 R_t^p. \tag{19}$$

Thus, in order to obtain an estimate of next day's VaR at time t, we use the portfolio return at time t and the $n-1$ preceding returns. Hence, Historical Simulation is just taking sample percentiles over a moving sample.

Suppose we want to use HS to predict a portfolio's Value-at-Risk at a confidence level of 1 percent and the window size is chosen to be 250 trading days. Then the 1 percent sample percentile is some amount between the second worst portfolio loss and the third worst portfolio loss. We decide to determine the Value-at-Risk through interpolation; in this case the VaR lies halfway between the second worst portfolio loss and the third worst portfolio loss.

Note that HS puts the same weight on all observations in the chosen window, including old data points, which may be an undesirable feature. The measure of risk may change abruptly once an old observation is dropped from the window. The choice of the window size n is open to debate. In case of a short window size, VaR estimates will be very sensitive to accidental outcomes from the recent past. A long window size, on the other hand, has the disadvantage that past data are included which might no longer be relevant to the current situation.

Let us employ the HS approach for our AEX portfolio. To this end we set the evaluation period equal to the last thousand trading days, and predict the VaR for different window sizes and left tail probabilities. In the tables below we report the average of the VaR estimates across time for these window sizes and left tail probabilities, as well as the accompanying failure rates. In the last two columns the results can be found for the maximum window size given the evaluation period of a thousand trading days. These results are best compared with the results for the variance techniques, especially the static ones.

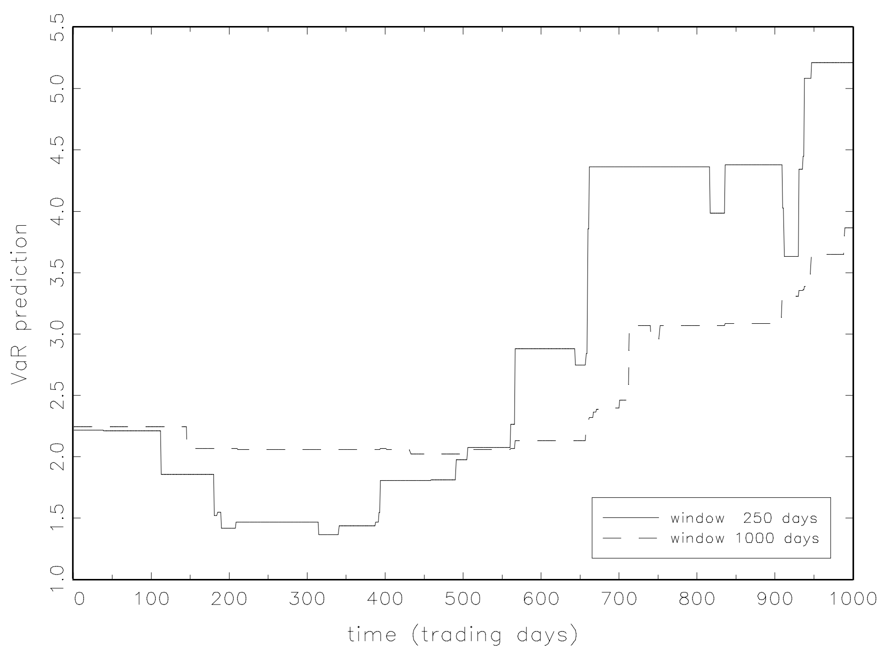
window size	250		500	
left tail probability	average VaR estimate	failure rate (times 1000)	average VaR estimate	failure rate (times 1000)
5.00 %	1.64	67†	1.44	75†
1.00 %	2.80	15	2.53	22†
0.50 %	3.19	9	3.08	12†
0.10 %	—	—	—	—

window size	1000		$T - 1000 = 3038$	
left tail probability	average VaR estimate	failure rate (times 1000)	average VaR estimate	failure rate (times 1000)
5.00 %	1.28	94†	1.56	69†
1.00 %	2.44	22†	3.17	16
0.50 %	2.84	12†	4.18	7
0.10 %	4.33	5†	7.25	0

The failure rates are not always close to their corresponding left tail probabilities, and usually exceed them. Apparently, the Value-at-Risk estimates for the last thousand days in the sample are too low for most window sizes and left tail probabilities. The market risk must have been smaller in the preceding period. Broadly speaking, in this sample, the HS approach seems to accurately predict the VaR for the more conservative left tail probabilities but only for a window size of 250 days and for the maximum window size.

Note that in case of our 0.1 percent left tail probability, HS makes no sense for windows smaller than 1000 days, because for those windows it is not possible to determine the 0.1 sample percentile. In general, it is not possible to make VaR predictions using HS for left tail probability levels smaller than the reciprocal of the window size. Hence, the HS approach is useless for predicting extreme risks if one is not willing to employ a window size of substantial length.

Figure 3: AEX Value-at-Risk predictions using HS.



Depending on the choice for the window size, the VaR predictions may be viscose or volatile: large windows keep data points for a long time which can lead to sustained periods of constant VaR predictions, while small windows eject data points relatively fast resulting in swift changes in the VaR predictions. See Figure 3 for an illustration of this. (The VaR predictions in Figure 3 are based on a left tail probability level of 1 percent and window sizes of 250 and 1000 days.) For a detailed evaluation of Value-at-Risk models using HS, see Hendricks (1996).

3.3.2 Tail index estimators

The fact that no distribution has to be estimated is an important advantage of historical simulation. However, the method also involves some drawbacks. First, as mentioned before, it is not possible to calculate VaRs for confidence intervals that are smaller then one over the window size. Second, the empirical distribution function is a step function. Especially far out in the tails of the distribution this discrete approximation of the true distribution function can cause biased results. As risk management is especially concerned with extreme observations, these objections might be serious. Both drawbacks can be remedied by fitting a smooth function through the tail of the distribution. This is essentially what tail index estimation does. The form of this function follows from *Extreme Value Theory.*[10] Extreme Value Theory concerns the asymptotic behaviour of extreme order statistics, such as the maximum and the minimum. This is what makes Extreme Value Theory of particular importance to Value-at-Risk analysis: we want to study the behaviour of extremely low returns that cause large losses. It can be shown that for every fat-tailed distribution, the distribution of the tail converges to the tail of a *Pareto* distribution:[11]

$$F_\alpha(x) = 1 - s^\alpha x^{-\alpha},\ x > s,$$

where s can be regarded as a threshold level above which the Pareto distribution holds.[12]

Hill's Estimator

Under the assumption that this threshold level s is known, the maximum likelihood estimator of the reciprocal of the tail index, $\gamma \equiv 1/\alpha$, is easily obtained:

$$\hat{\gamma}_H \equiv \widehat{1/\alpha}_{ML} = \frac{1}{n}\sum_{i=1}^{n} \log \frac{X_i}{s} \tag{20}$$

This estimator is known as the Hill estimator, introduced in Hill (1975). In practice, the threshold level s needs to be determined. In case the random sample $X_1, X_2, \ldots, X_n$ is known to be Pareto

[10] See e.g. Bain and Engelhardt (1987), Chapter 8, Embrechts, Klüppelberg, and Mikosch (1997), and Feller (1971), Section VIII.8.

[11] See De Haan and Stadtmüller (1996).

[12] In the analysis of Value-at-Risk, x represents the portfolio's *loss.*

distributed, we might estimate s by $X_{1:n}$, the minimum order statistic, but the fact is, that we are not dealing with a Pareto distribution, but with a distribution that has a right tail that looks—at least to a first approximation—like the tail of a Pareto distribution. As a consequence, there has to be some level, say s, below which the Pareto law applies. The idea is to estimate equation 20 with only the observations that are greater than s, with s replaced by the minimum observation that is tolerable. The problem is of course to determine the optimal number (m) of extreme observations to include.[13]

Once the tail index has been estimated, large quantile estimates can be obtained as follows. Consider two tail probabilities p and q, $p < q$. Let x_p and x_q denote the corresponding quantiles. Hence, $p \approx a x_p^{-\alpha}$ and $q \approx a x_q^{-\alpha}$. Combining these two expressions yields $x_p \approx x_q(q/p)^{1/\alpha}$. Set $x_q = X_{n-m+1:n}$, the $n-m+1$-th ascending order statistic, which is in fact the threshold level. Then $q \approx m/n$ by the empirical distribution function. The quantile x_p may then be estimated by

$$\hat{x}_p = X_{n-m+1:n} \left(\frac{m}{np} \right)^{\widehat{1/\alpha}}. \tag{21}$$

Let us apply the Hill tail index estimation approach to our fictitious investment in the AEX. In this example, the optimal cut off level $\hat{m}_n$ turns out to be 47. Given the 3038 observations in the estimation sample, this means that the Pareto approximation is only optimal for left tail probabilities up to 1.5%. The tail index estimate equals 2.81, implying finite first and second moments but unbounded higher moments for the distribution of AEX returns.

left tail probability	VaR estimate	failure rate (times 1000)
5.00 %	1.79	55
1.00 %	3.15	17†
0.50 %	4.01	9
0.10 %	7.00	0
0.01 %	15.20	0

Only at the 1 percent left tail probability level does the Kupiec back-test reject the model. At the 0.5 percent level, the model is not rejected, but only just. Far out in the tail, there is no reason to assume the model over- or undervalues the Value-at-Risk. Do bear in mind, however, that there is almost no power at these extreme levels. Note that the corresponding Value-at-Risk estimates are considerably higher than the ones we saw in the previous chapter. One might very well argue

[13] We apply the approach of Daníelsson and De Vries (1998). They choose the threshold level such that the estimator's asymptotic mean squared error is minimized.

that the extreme portfolio risk resulting from the tail index estimation method is more accurate than the extreme risk predicted by the more conventional techniques because of the statistical foundations behind tail shape of the distribution, provided we are willing to make the assumption of fat tails.

An interesting feature is that if we want to know the VaR over a longer holding period on the basis of daily data, say T days ahead, we can apply the so-called *α-root of time rule*:[14]

$$VaR_p^{(T)} \approx \sqrt[\alpha]{T}\ VaR_p^{(1)}. \tag{22}$$

This approximation follows from extreme value theory.

Compare this to the square root of time rule mentioned earlier. Suppose that the tail index is larger than two, $\alpha > 2$. (This means that the variance is finite.) The reciprocal of the tail index is smaller than a half in this case. Ergo, if we are willing to believe we are dealing with a heavy tailed distribution with finite variance, the square root of time rule overrates the T-periods-ahead Value-at-Risk. For instance, if we consider the AEX Value-at-Risk estimate at the 1 percent left tail probability level, the 10-days-ahead Value-at-Risk will be about $\sqrt[2.81]{10}\ 3.15 = 7.15$ currency units which is less than the 9.96 currency units that would result had we used the square root of time rule.

Least squares tail Estimator

A different approach to the estimation of the tail-index was suggested by Van den Goorbergh (1999). As the basic assumption of fat tails implies the tails decline by a power, like the Pareto distribution, we have that below a certain threshold level

$$F(x) \approx ax^{-\alpha}, \tag{23}$$

where a, $\alpha > 0$. After taking logarithms we find

$$\log F(x) \approx \log a - \alpha \log x. \tag{24}$$

The idea of the new approach is to estimate a and α in the above equation by means of least squares, where x is replaced by the extreme portfolio returns and $F(x)$ by the empirical distribution function of the portfolio returns. The optimal cut-off point is determined by estimating the equation for a range of possible values, and choosing the one for which the R^2 is the highest.

The Value-at-Risk for a portfolio with initial value W_0 can subsequently be estimated as:

$$VaR \approx -W_0 \left[\exp\left(-(\hat{a}/p)^{1/\hat{\alpha}}\right) - 1\right]. \tag{25}$$

[14] See e.g. Daníelsson and De Vries (1998).

According to this method, the optimal number of extremes to include for our investment in the AEX equals 151, which implies a proper approximation up until a probability of 5%. The tail index is estimated at 2.42 which implies that the mean and variance of the stock returns are finite, but all higher moments are unbounded. Moreover, the square root of time rule for converting the one-period-ahead Value-at-Risk into a T-periods-ahead Value-at-Risk overvalues the T-periods-ahead Value-at-Risk. The scaling constant estimate equals $2.61 \cdot 10^{-6}$.

left tail probability	VaR estimate	failure rate (times 1000)
5.00 %	1.68	60
1.00 %	3.24	17†
0.50 %	4.29	8
0.10 %	8.18	0
0.01 %	19.85	0

The performance of the least squares tail estimation approach is not very convincing for the higher left tail probability levels—the Kupiec back-test rejects the model at the important 1 percent level. It is quite similar to the performance of the Hill tail index estimation technique. Again, the VaR predictions are substantially higher than in the previous chapter. However, due to the possibility to apply the α-root of time rule instead of the square root of time rule, the resulting capital charge might very well be lower.

3.4 Review of the AEX Results

In this section we have encountered a multitude of models to assess the Value-at-Risk of a portfolio consisting of a single stock index. An empirical evaluation of these models was done by means of a fictitious investment in the Dutch stock index AEX. For each VaR model an evaluation period of a thousand trading days was used so that their performances can be compared. Here we repeat the results for all VaR techniques that we have seen.

Table 5 recapitulates the (average) AEX Value-at-Risk predictions for various left tail probabilities along with the corresponding failure rates multiplied by 1000. Recall that a dagger (†) marks failure rates significantly different from their left tail probability on the basis of Kupiec's back-test at a 95 percent confidence level. A dash (—) indicates no VaR predictions can be made due to an insufficiently large window size.

The various VaR techniques are classified according to their distinguishing characteristics. First, the results for the three static variance VaR methods are tabulated, followed by the GARCH related VaR techniques including J.P. Morgan's RiskMetrics$^{\mathrm{TM}}$. Finally, we have the nonpara-

Table 5: Mean VaR and failure rates for the AEX.

left tail probability	5.00 %	1.00 %	0.50 %	0.10 %	0.01 %
ideal failure rate (×1000)	50	10	5	1	0.1
static	1.85	2.62	2.90	3.48	4.18
normal	53	26†	22†	12†	8†
static	1.57	2.95	3.72	6.16	12.10
Student-t	69†	21†	10†	0	0
static mixture	1.49	3.37	4.42	6.33	8.42
of normals	75†	14	5	0	0
GARCH	1.76	2.51	2.78	3.34	4.02
normal innovations	54	17†	10†	3	1
J.P. Morgan's	1.78	2.51	2.77	3.31	3.97
RiskMetrics^{TM}	56	21†	13†	6†	1
GARCH	1.65	2.72	3.22	4.56	7.04
Student-t innovations	58	10	4	1	0
mixture of normals	1.61	2.53	3.27	5.17	7.33
with GARCH	62	16	10†	1	0
Historical Simulation	1.56	3.17	4.18	7.25	—
(maximum window size)	69†	16	7	0	—
tail index estimation	1.79	3.15	4.01	7.00	15.20
(Hill estimator)	55	17†	9	0	0
tail index estimation	1.68	3.24	4.29	8.18	19.85
(least squares principle)	60	17†	8	0	0

metric method of historical simulation and the related tail index estimation VaR techniques. For the historical simulation method, the results for the maximum window size are reported in order to make a more or less fair comparison with the other VaR techniques that all use an estimation sample of that size.

Once again, the superiority of the GARCH(1,1) model with Student-t innovations is demonstrated: at none of the left tail probabilities examined is this model rejected. Moreover, at almost all probability levels it outperforms the other models in that its failure rates are closest to the ideal failure rates in most cases. Do bear in mind, however, that this doesn't mean one should use the Student-t GARCH model in each and every application; the model happens to perform well in our specific portfolio setting. Had we for instance concentrated on the bond market instead of the stock market, we might have reached different conclusions. Moreover, we cannot extrapolate our findings to the multivariate case of a portfolio consisting of more than one stock. Therefore, prudence is in order.

In Figure 4, the failure rates for all left tail probabilities up to 5% are shown. It is remarkable that the actual failure rates are almost always higher than expected for all methods and all left tail probabilities. For none of the methods is the actual failure rate within the 95% confidence band for all left tail probabilities, although the one for GARCH-t is very close. Regarding the GARCH models, the good performance of the normal distribution for confidence levels above 1%

stands out. RiskMetrics^{TM} performs very good at the 5% level, but not for confidence levels lower than 3%. Indeed, the most extreme losses are not well predicted by this method.

4 An Application to the Dow Jones

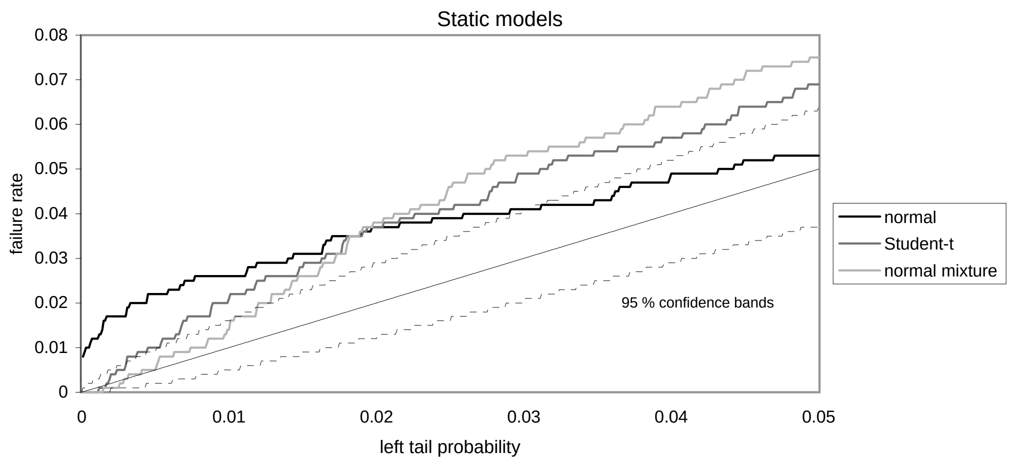
4.1 Introduction

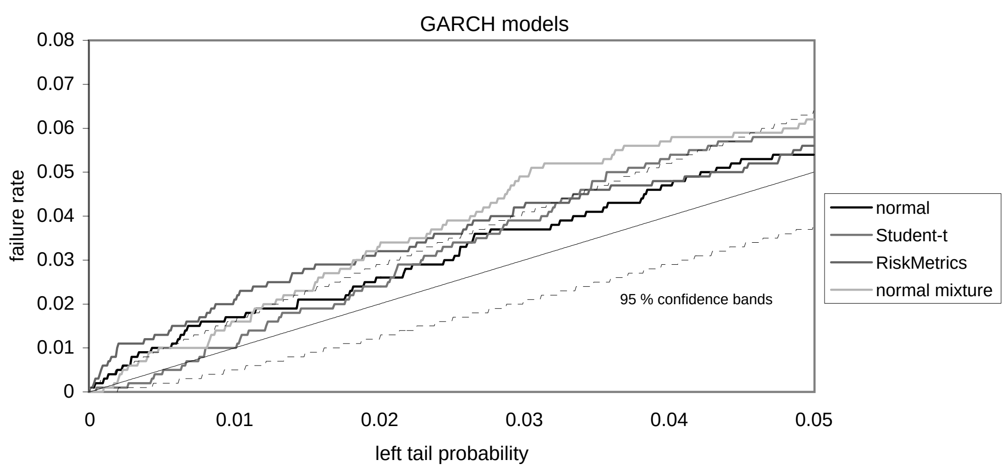
In the previous section, Historical Simulation, variance methods and tail index estimation were applied to an imaginary investment in the Dutch AEX stock index in order to elucidate the various VaR techniques presented. For each technique, a period of about 3000 daily data was used to estimate the Value-at-Risk, and a back-test was performed using the subsequent period of 1000 trading days. Of course, in practice, banks that use variance or tail methods are not likely to keep parameters fixed for a very long time. They hope to reduce the effect of parameters changing in time as much as possible by regularly re-estimating their VaR models. In this chapter we will follow such a strategy of updating parameter estimates by re-estimating VaR models each year.

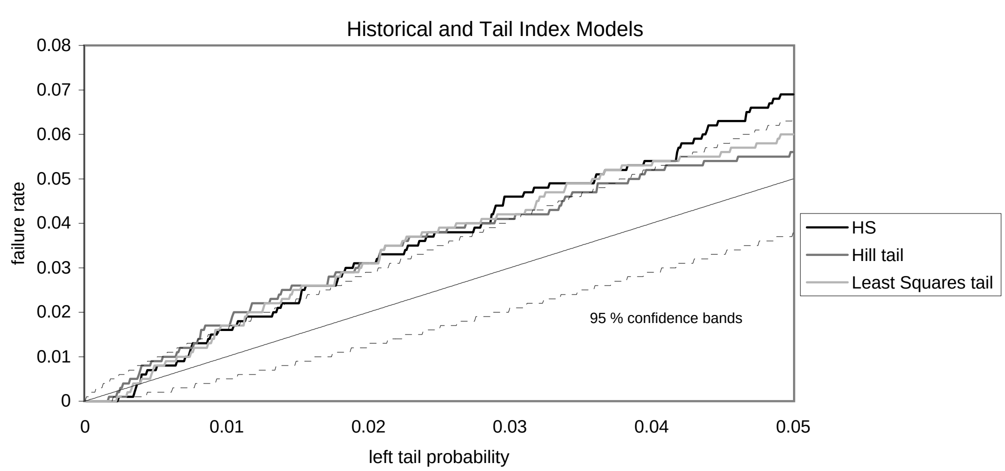
This time we will apply our various VaR models to a fictitious investment in the most well-known American stock index, the Dow Jones Industrial Average. The main reason for doing so is the profuse availability of daily data which is a crucial requirement for evaluating Value-at-Risk models; the Dow Jones Industrial Average goes back to the nineteenth century. We shall concentrate on the five most recent postwar decades to see which VaR techniques can accurately predict the Value-at-Risk of an investment in the Dow Jones index. Moreover, we want to see whether this Value-at-Risk has changed over time, and, if it has, to what extent.

In econometric forecasting a balance always has to be found regarding the number of past observations included in the forecast. Including many past observations has the advantage of reducing the random error in the estimates, but including too many entails the danger of an overemphasis of irrelevant, old data points. The very nature of Value-at-Risk analysis—and large quantile estimation in general—, obliges one to use a substantial amount of data in any case, especially with regard to Historical Simulation and tail index estimation, putting this balance into a different perspective than for other fields of study which do not require large data sets per se. In view of this, we estimate the various VaR models presented in the previous chapters for each year from 1960 onwards using the previous 10 years as estimation sample. On each occasion, the VaR model is evaluated by means of a back-test using the portfolio returns of that year. Subsequently, failure rates for the entire sample are computed by adding up the VaR violations for each of the usual left tail probabilities, and relating these to the total number of trading days in the sample, thereby increasing the power of discernment of Kupiec's back-test in particular for the more cautious confidence levels.

Figure 4: Performance of VaR methods on the AEX for various left tail probabilities.







4.2 Estimation results

The estimation results for the entire sample are presented in the tabulated form, whereas the failure rates at the 1% confidence level for the individual years are presented in graphical form (Figure 5). For each confidence level, Table 6 reports first the average failure rate. If this failure rate was significantly different from the expected failure rate, it is printed in bold face. Next, the standard deviation of the yearly failure rates is given, to give an idea of the dispersion of the rate through time. Third, the number of years in which the failure rate was significantly higher than expected (one-sided test) is noted. Next, the *weighted sum of squared violation errors*, $WSSVE$ is given, as a measure of overall performance. For each year the difference between the actual and the expected number of VaR violations is squared, and multiplied by the fraction of trading days of that year in relation to the total number of trading days in the entire sample. In order to get the $WSSVE$, these numbers are added up across time. These four numbers give an idea of the accuracy of the VaR. Next, the average VaR is shown to give an idea of the efficiency of the calculations. For the 1% level, the average capital requirement is also given. It is computed by multiplying the estimated VaR by the hysteria factor that results from the previous year's back test result, and either the square root of 10 —for the variance methods and for historical simulation— or the α-root of 10 —for the tail estimators. The lowest number in each row is printed in italics.

The VaR techniques can be roughly classified into two categories: the conditional methods and the unconditional methods. Unconditional methods are those that yield a constant VaR prediction for each trading day in the year under examination. The static variance normals model, as well as the tail index estimation methods belong to this class. In contrast, conditional methods predict a different VaR for each trading day in the year under examination depending on the previous days' portfolio returns. Both Historical Simulation and the GARCH related VaR techniques meet this qualification. The fact that conditional VaR methods incorporate new information each day should make their predictive performance superior to the unconditional VaR methods if the distribution of portfolio returns changes through time. Indeed, Table 6 clearly demonstrate the superiority of the GARCH related VAR techniques when it comes to the accuracy of the VaR. These methods handle periods with heavy fluctuations much better than the other methods. Do note that historical simulation also performs considerably worse than the GARCH models. Apparently HS hinges too much on the assumption that—within the window—the distribution of portfolio returns is time-independent and moving this window each day cannot correct this sufficiently. GARCH related methods are able to deal with this changing distribution much better by attaching decaying weights to the past observations, so that past portfolio returns become less and less important as time passes. The conditionality of the historical simulation method might be improved by shortening the window. Indeed for a ten year window, HS is almost an unconditional method.

Table 6: Summary results for the Dow Jones VaR

		normal	Student-t	Mixture	Garch-normal	RiskMetrics	Garch-t	Garch-mix	historical	Hill tail index	LS tail index
failure rate	5%	0.0503	**0.0612**	**0.0613**	0.0501	*0.0484*	0.0539	**0.0547**	**0.0558**	**0.0618**	0.0523
st.dev.		0.0503	0.0524	0.0520	0.0221	*0.0161*	0.0215	0.0217	0.0474	0.0528	0.0501
#years		7	12	11	5	*2*	5	6	10	13	9
WSSVE		157.69	179.34	176.84	30.46	*16.24*	29.80	30.74	142.56	183.30	156.80
VaR		1.40	1.27	*1.26*	1.34	1.36	1.30	1.28	1.30	1.26	1.37
failure rate	1%	**0.0184**	**0.0136**	**0.0133**	**0.0131**	**0.0153**	*0.0102*	**0.0139**	**0.0135**	**0.0178**	**0.0139**
st.dev.		0.0292	0.0238	0.0223	0.0080	0.0082	*0.0062*	0.0089	0.0225	0.0294	0.0248
#years		8	6	8	5	7	*1*	7	7	9	7
WSSVE		57.86	36.16	31.75	4.65	5.91	*2.36*	5.87	32.36	58.12	39.57
VaR		1.99	2.16	2.20	1.91	1.92	2.03	*1.89*	2.11	2.04	2.20
Capital charge		20.05	21.58	22.02	19.04	19.51	19.89	18.94	21.07	*13.38*	14.82
failure rate	0.5%	**0.0136**	**0.0068**	0.0060	**0.0075**	**0.0103**	*0.0048*	0.0057	0.0063	**0.0099**	**0.0076**
st.dev.		0.0240	0.0140	0.0136	0.0053	0.0062	*0.0041*	0.0046	0.0117	0.0210	0.0168
#years		10	5	6	6	11	*1*	2	7	7	7
WSSVE		40.83	12.51	11.74	2.19	4.17	*1.04*	1.36	8.64	29.07	18.24
VaR		2.20	2.61	2.90	*2.12*	2.12	2.35	2.21	2.62	2.53	2.74
failure rate	0.1%	**0.0070**	**0.0018**	**0.0019**	**0.0035**	**0.0049**	0.0016	*0.0014*	0.0016	**0.0023**	**0.0019**
st.dev.		0.0145	0.0048	0.0057	0.0042	0.0045	*0.0022*	0.0026	0.0037	0.0059	0.0053
#years		9	4	4	8	15	*1*	2	3	4	4
WSSVE		15.58	1.54	2.12	1.48	2.22	*0.32*	0.45	0.89	2.32	1.81
VaR		2.64	3.90	4.42	2.54	*2.54*	3.17	3.97	4.49	4.17	4.64
failure rate	0.01%	**0.0038**	*0.0003*	**0.0006**	**0.0019**	**0.0020**	*0.0003*	**0.0004**		*0.0003*	*0.0003*
st.dev.		0.0072	0.0014	0.0021	0.0027	0.0027	*0.0011*	0.0012		0.0014	0.0014
#years		13	*2*	4	15	16	3	4		*2*	*2*
WSSVE		4.16	0.13	0.30	0.67	0.69	*0.07*	0.10		0.13	0.13
VaR		3.17	6.64	6.01	3.06	*3.05*	4.58	6.43		8.59	10.19

Note: For each confidence level the following statistics are given: (1) the mean failure rate, (2) the standard deviation of the yearly failure rates, (3) the number of years, out of 39, with a failure rate significantly higher than the prescribed confidence level according to a 5% significance level, (4) the weighted sum of squared violation errors, and (5) the mean VaR. For the 1% confidence level the mean resulting capital requirement is also computed. Mean failure rates that are significantly different from the prescribed confidence level according to the Kupiec back-test are in bold face. Lowest figures in a row are in italics.

Figure 5: Yearly VaR violations for the Dow Jones at the 1 percent level.

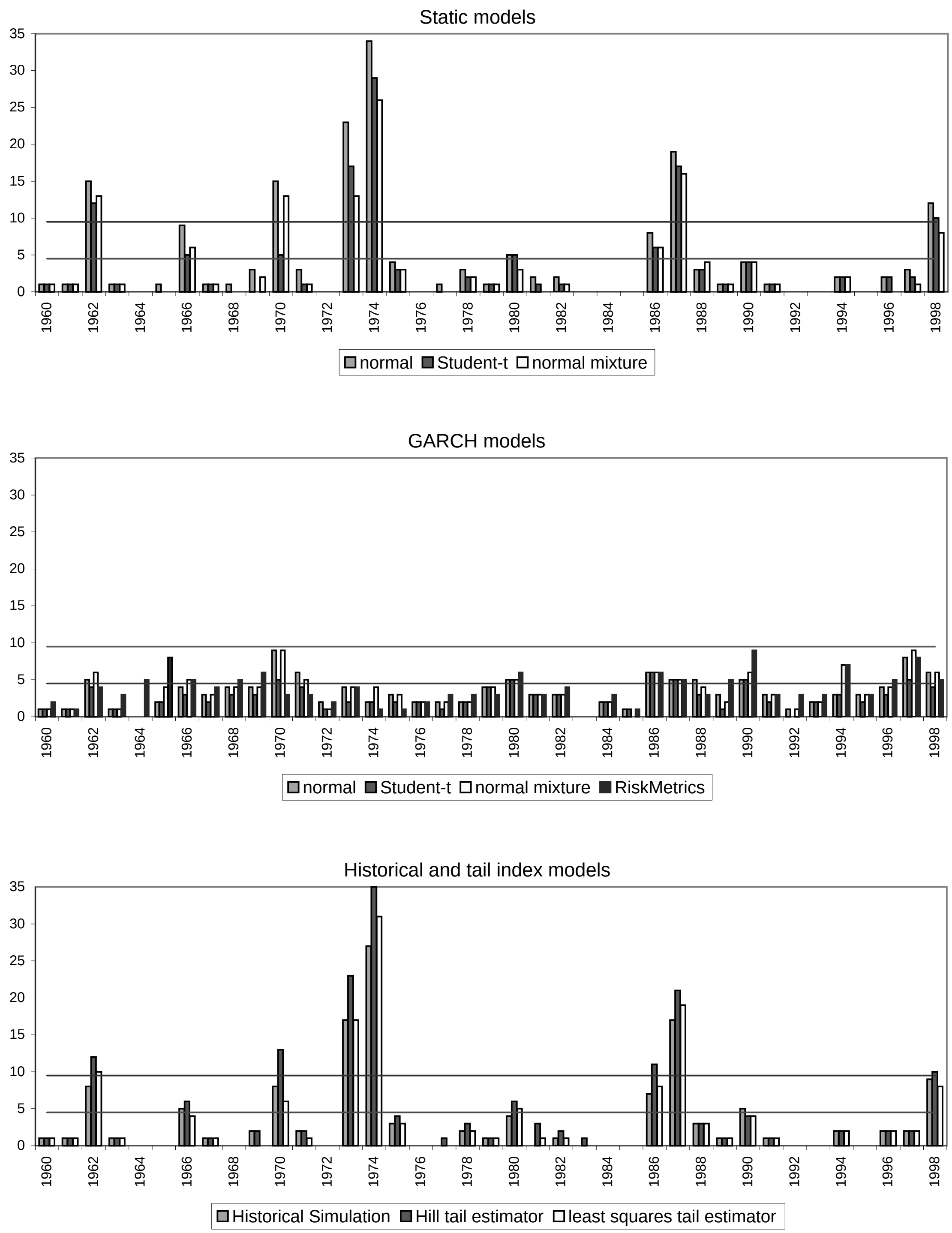


Figure 5 display the number of VaR violations for each year for the various VaR techniques at the 1 percent left tail probability level—the probability prescribed by the Basle regulations. Furthermore, two horizontal lines indicate the boundaries of the different zones of the Internal Model approach; if a bar rises above the lower horizontal line, the Green Zone of the penalty scheme described in Section 2.1 is left and the Yellow Zone is entered, and if the upper horizontal line is exceeded, we find ourselves in the Red Zone. Recall that in the Green Zone, the VaR scaling factor is three, in the Yellow Zone it is gradually increased, and in the Red Zone it reaches its maximum, four. In the Red Zone a bank is likely to be obliged to revise its internal model altogether.

These graphs clearly demonstrate the inability of the unconditional methods to cope with critical years such as 1962, 1973, 1974, 1987 and 1998, during which the Value-at-Risk is severely underestimated. On top of that, a substantial amount of years yield no VaR violations at all when ideally one would expect 2 or 3 violations. Due to the fact that the VaR is not adjusted during the evaluation year, the Value-at-Risk for what turns out to be a very risky year is undervalued if it is preceded by a relatively tranquil period, since VaR prediction are made on the basis of this relatively tranquil period, and no adjustments are made in the course of the evaluation year, even though new information that becomes available each trading day omens increased market volatility. Likewise, Value-at-Risk predictions based on periods with heavy fluctuations are too high if the year they pertain to is not very volatile. Things are quite different for the GARCH related approaches. Not even once is the Red Zone entered, and the number of Yellow Zone VaR violations is limited, in particular for the Student-t distribution. Note also that the RiskMetrics^{TM} approach does rather well given the fact that all its parameters are fixed.

The weighted sum of squared violation errors reaffirm the marked discrepancy between VaR techniques that use GARCH and those that do not, also for other left tail probabilities. The GARCH(1,1) model with Student-t innovations has the smallest WSSVE, hence, according to this criterion, it is the best model we have examined. Apparently, this model describes the down-side risk of investing in the Dow Jones Industrial Average best, as was the case for the AEX portfolio, owing to its ability to handle volatility clustering by means of a GARCH(1,1) specification for the variance, and the underlying heavy tailed Student-t distribution that takes care of the remaining leptokurtosis caused by extreme returns.

As for the tail index estimation techniques, that appeared to be tailor-made for estimating large quantiles the results are very disappointing. This is due to the fact that the method assumes an *iid* sample of random variables. Daníelsson and De Vries stress that this is by no means a very restrictive assumption. They claim that under various forms of time-dependence, the results still hold. In fact, De Haan, Resnick, Rootzen, and De Vries (1989) show that the stationary distribution of the ARCH process varies regularly at infinity, and Resnick and Starica (1998) argue that the Hill estimator is consistent under the assumption of ARCH. Consequently, the tail index estimation methods we have seen in this chapter should be able to predict the Value-at-Risk even though there is statistical evidence that the distribution of stock returns changes through time (volatility clustering). Here we see that these asymptotic results have not much practical importance. The results for the important 1 percent level clearly show that the 1 percent extremes are still correlated over time. The graphs show that, broadly speaking, the Hill estimator seems to perform even worse than the least squares method at this level. Even at the most cautious left tail probabilities, such as 0.1 and 0.01 percent, where one would expect the tail index techniques that focus on the prediction of extreme events to excel, the results are not very convincing. One might as well use the static Student-t distribution, that gives similar failure rates but at lower VaR costs. The best results however, are reached with the conditional GARCH-t model. Apparently, volatility clustering can not be disregarded, even at the lowest probability levels. Of course, at these levels, model accuracy can hardly be tested.

Given the poor performance of the tail index estimators, it is all the more striking that these methods result in the lowest capital requirement (if a maximum stress factor of 4 is maintained). This adverse result is primarily due to the α-root of time rule. If the square root of time rule would also be applied to these estimates, the capital requirement would become 20.59 (Hill) respectively 21.97 (LS). The dispersion of capital charges across methods is not very stimulating. The GARCH Student-t is clearly optimal at the 1% level, but GARCH-normal gives on average the lowest charge. Based on these results for linear stock portfolios, more differentiation in the penalty rate seems desirable.

5 Conclusions

In this paper the relatively novel risk management concept of Value-at-Risk has been examined. Many practitioners have embraced Value-at-Risk as an easy to understand measure of the downside risk on an investment portfolio. Value-at-Risk has not only found its way to the internal risk management of banks and other financial institutions, but we have seen that it has also been firmly rooted in the regulations that supervisors have imposed on them. And although these regulations have been subject to some criticism—the Basle Committee has rather arbitrarily set certain parameters, notably the (range of the) multiplication factor—, it is generally felt that they constitute a vast improvement on the former rigid legislation.

The study has been concerned with the Value-at-Risk analysis of the stock market. A wide variety of Value-at-Risk models has been presented and empirically evaluated by applying them to a fictitious investment in the Dutch stock market index AEX, mainly for illustrative purposes. Subsequently, a more rigorous approach was taken by applying all of the presented Value-at-

Risk techniques to another stock market index, the Dow Jones Industrial Average. The generous availability of historical daily return data on this index allowed us to more realistically imitate the behaviour of banks, namely by re-estimating and re-evaluating the Value-at-Risk models each year. The main conclusions are:

1. By far the most important characteristic of stock returns for modelling Value-at-Risk is volatility clustering. This can effectively be modelled by means of GARCH. Even at the lowest left tail probabilities (up to 0.01%), modelling GARCH effectively reduces average failure rates and the fluctuation of failure rates over time, whereas at the same time the average VaR is lower.

2. For left tail probabilities of 1% or lower, the assumed conditional distribution for the stock returns needs to be fat-tailed. The Student-t distribution seems to perform better in this respect then the Bernoulli-normal mixture. At the 5% level, the normal distribution performs best.

3. Tail index techniques are not successful, due to the fact that they do not cope with the volatility clustering phenomenon. At the 1% level the assumption of a constant VaR throughout a year resulted in up to 35 violations of the VaR in one year (1974). Even at the 0.1% level, the average number of VaR violations over the last 39 years was significantly higher than to be expected.

References

Bain, L.J. and M. Engelhardt, 1987, *Introduction to Probability and Mathematical Statistics.* Duxbury Press, Boston.

Basle Committee on Banking Supervision, 1995, *An Internal Model-Based Approach to Market Risk Capital Requirements*, Basle: Basle Committee on Banking Supervision (available at http://www.bis.org).

Basle Committee on Banking Supervision, 1996, *Supplement to the Capital Accord to Incorporate Market Risks*, Basle: Basle Committee on Banking Supervision (available at http://www.bis.org).

Bollerslev, T., 1986, Generalized Autogreressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307–327.

Daníelsson, J. and C.G. de Vries, 1998, Beyond the Sample: Extreme Quantile and Probability Estimation, Tinbergen Institute Discussion Paper TI 98-016/2.

Embrechts, P., C. Klüppelberg, and T. Mikosch, 1997, *Modelling Extremal Events.* Springer, Berlin.

Engle, R., 1982, Autogregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation, *Econometrica*, 50, 987–1008.

Feller, W., 1971, *An Introduction to Probability and its Applications.* Wiley, New York.

Figlewski, S., 1994, Forecasting Volatility Using Historical Data, New York University, Working Paper 13.

Van den Goorbergh, R.W.J., 1999, Value-at-risk analysis and least squares tail index estimation, Research Memorandum WO&E, 578, De Nederlandsche Bank (available at http://www.dnb.nl).

De Haan, L., S.I. Resnick, H. Rootzen, and C.G. de Vries, 1989, Extremal behaviour of solutions to a stochastic difference equation with applications to arch processes, *Stochastic Processes and their Applications*, 32, 471–487.

De Haan, L. and U. Stadtmüller, 1996, Generalized regular variation of second order, *Journal of the Australian Mathematical Society, series A*, 61, 381–395.

Hendricks, D., 1996, Evaluation of Value-at-Risk Models Using Historical Data, Federal Reserve Bank of New York Economic Policy Review April: 39–69.

Hill, B.M., 1975, A Simple General Approach to Inference about the Tail of a Distribution, *Annals of Statistics*, 35, 1163–1173.

Jorion, P., 1997, *Value at Risk.* Irvine, Chicago.

J.P. Morgan Bank, 1996, *RiskMetrics—Technical Document* (4 ed.), New York: J.P. Morgan Bank.

Kupiec, P., 1995, Techniques for Verifying the Accuracy of Risk Measurement Models, *Journal of Derivatives*, 2, 73–84.

Lucas, A., 1997, An Evaluation of the Basle Guidelines for Backtesting Internal Risk Management Models of Banks, VU Research Memorandum, Free University Amsterdam.

Resnick, S. and C. Starica, 1998, Tail estimation for dependent data, *Annals of Applied Probability*, forthcoming.

Vlaar, P.J.G. and F.C. Palm, 1993, The message in weekly exchange rates in the european monetary system: mean reversion, conditional heteroskedasticity and jumps, *Journal of Business & Economic Statistics*, 11, 351–60.

Vlaar, P.J.G., 1998, Value at Risk models for Dutch bond portfolios, *Journal of Banking and Finance*, forthcoming (available at http://www.dnb.nl).