Value-at-Risk and Extreme Returns*

Jon Danielsson
London School of Economics and
Institute of Economic Studies
at University of Iceland

Casper G. de Vries
Tinbergen Institute
and Erasmus University

November 26, 1997

Abstract

Accurate prediction of the frequency of extreme events is of primary importance in many financial applications such as Value-at-Risk (VaR) analysis. We propose a semi-parametric method for VaR evaluation. The largest risks are modelled parametrically, while smaller risks are captured by the non-parametric empirical distribution function. The semi–parametric method is compared with historical simulation and the J. P. Morgan RiskMetrics technique on a portfolio of stock returns. For predictions of low probability worst outcomes, RiskMetrics analysis underpredicts the VaR while historical simulation overpredicts the VaR. However, the estimates obtained from applying the semi–parametric method are more accurate in the VaR prediction. In addition, an option is used in the portfolio to lower downside risk. Finally, it is argued that current regulatory environment provides incentives to use the lowest quality VaR method available.

Keywords: Value-at-Risk, Extreme Value Theory, RiskMetrics, Historical Simulation, Tail Density Estimation, Financial Regulation

---

* Correspondence: *jond@hi.is*, http://www.hag.hi.is/˜jond/research. We wish to thank Frank X. Diebold, Philipp Hartmann, David Pyle, Ton Vorst, Eric Zivot, and workshop participants at the London School of Economics, University of Edinburgh, Cambridge University, Pompeu Fabra University, and Erasmus University for excellent comments. The University of Iceland research fund, the Fullbright program, and the Research Contribution of the Icelandic Banks supported this research. The first draft of this paper was written while the first author was visiting at the University of Pennsylvania, and we thank the university for its hospitality.

1 Introduction

A major concern for regulators and owners of financial institutions is catastrophic market risk and the adequacy of capital to meet such risk. Well publicized losses incurred by several institutions such as Orange County, Proctor and Gamble, and NatWest, through inappropriate derivatives pricing and management, as well as fraudulent cases such as Barings Bank, and Sumitomo, have brought risk management and regulation of financial institutions to the forefront of policy making and public discussion.

A primary tool for financial risk assessment is the Value-at-Risk (VaR) methodology where VaR is defined as an amount lost on a portfolio with a given small probability over a fixed number of days. The major challenge in implementing VaR analysis is the specification of the probability distribution of extreme returns used in the calculation of the VaR estimate.

By its very nature, VaR estimation is highly dependent on good predictions of uncommon events, or catastrophic risk, since the VaR is calculated from the lowest portfolio returns. As a result, any statistical method used for VaR estimation has to have the prediction of tail events as its primary goal. Statistical techniques and rules of thumb that have been proven useful in analysis and prediction of intra day and day-to-day risk, are not necessarily appropriate for VaR analysis. This is discussed in a VaR context by e.g. Duffie and Pan (1997) and Jorion (1997).

The development of techniques to evaluate and forecast the risk of uncommon events has moved at a rapid rate, and specialized methods for VaR prediction are now available. These methods fall into two main classes: parametric prediction of conditional volatilities, of which the J. P. Morgan RiskMetrics method is the best known, and non-parametric prediction of unconditional volatilities such as techniques based on historical simulation or stress testing methods.

In this paper we propose a new semi-parametric method for VaR estimation which is a mixture of these two approaches, where we combine non-parametric historical simulation with parametric estimation of the tails of the return distribution. These methods build upon recent research in extreme value theory, which enable us to accurately estimate the tails of a distribution. Hols and de Vries (1991) use extreme value analysis to measure risk in FOREX trading, and Longin (1997) applies extreme value analysis to risk in the U.S. stock market. The method used there for estimating the extreme value distribution selects the maxima in a series of non-overlapping sub–periods and therefore makes an inefficient use of the available sample information and demands large datasets. In contrast, Danielsson and de Vries (1997*a*) and Danielsson and de Vries (1997*b*) propose an efficient, semi-parametric method for estimating tails of the underlying distribution of financial returns, and this method is expanded here to the efficient estimation of ‘portfolio tails’.

In finance, it is natural to assume normality of returns in daily and multi-day conditional and unconditional volatility predictions, in applications such as derivatives pricing. As the volatility smile effect demonstrates, however, for infrequent events the normal model is less useful. Since returns are known to be fat tailed, the conditional normality assumption leads to a sizable underprediction of tail events. The popular RiskMetrics technique, in essence an IGARCH model, is based on conditional normal analysis with frequent parameter updates. The price one has to pay for the normality assumption and frequent parameter updating is that such model is not well suited for analyzing large risks. The normality assumption implies that one underestimates the chances of heavy losses, and the frequent updating implies a high variability in the estimates. For this reason, RiskMetrics focuses on the 5% quantile, or the probability of losses that occur once every 20 days. But these losses are so small that they can be handled by any financial institution. We argue below that RiskMetrics is ill suited for lower probability loses.

Furthermore, conditional parametric methods typically depend on conditional normality for the derivation of multi period VaR estimates. Relaxation of the normality assumption leads to difficulties due to the 'square-root-of-time' method, i.e. the practice of obtaining multi-period volatility predictions by multiplying the one day prediction by the square root of the length of the time horizon. Moreover, as Christoffersen and Diebold (1997) argue, conditional volatility predictions are not very useful for multi day predictions. We argue that the appropriate method for scaling up a single day VaR to a multi day VaR is a alpha-root rule, where alpha is the number of finite bounded moments, also known as the tail index. We implement the alpha-root method and compare it with the square-root rule.

By definition, extreme returns occur infrequently, and in addition do not appear to be related to a particular level of volatility, nor exhibit time dependence. Therefore, an unconditional approach is better suited for VaR estimation than conditional volatility forecasts, because it permits one to use all observations over a long span of time. One can either use the historical returns as a sampling distribution for future returns as in Historical simulation (HS) and stress testing, or use a form of kernel estimation to smooth the sampling distribution as in Butler and Schachter (1996)**.** The advantages of historical simulation have been well documented by e.g. Jackson, Maude and Perraudin (1997), Mahoney (1996), and Hendricks (1996). A disadvantage is that the low frequency and inaccuracy of tail returns leads to predictions which exhibit a very high variance, i.e. the variance of the highest order statistics is very high, and is in some cases even infinite. As a result, the highest realizations lead to poor estimates of the tails, which may invalidate HS as a method for stress testing. In addition, it is not possible to do out-of-sample prediction with HS, i.e. predict losses that occur less frequently than are covered by the HS sample period.

A primary use of value-at-risk is the determination of capital requirements for financial institutions, where capital is a function of the VaR. We argue that the current regulatory environment provides incentives for the financial institutions to choose the technique that produces the lowest VaR, typically the worst VaR method.

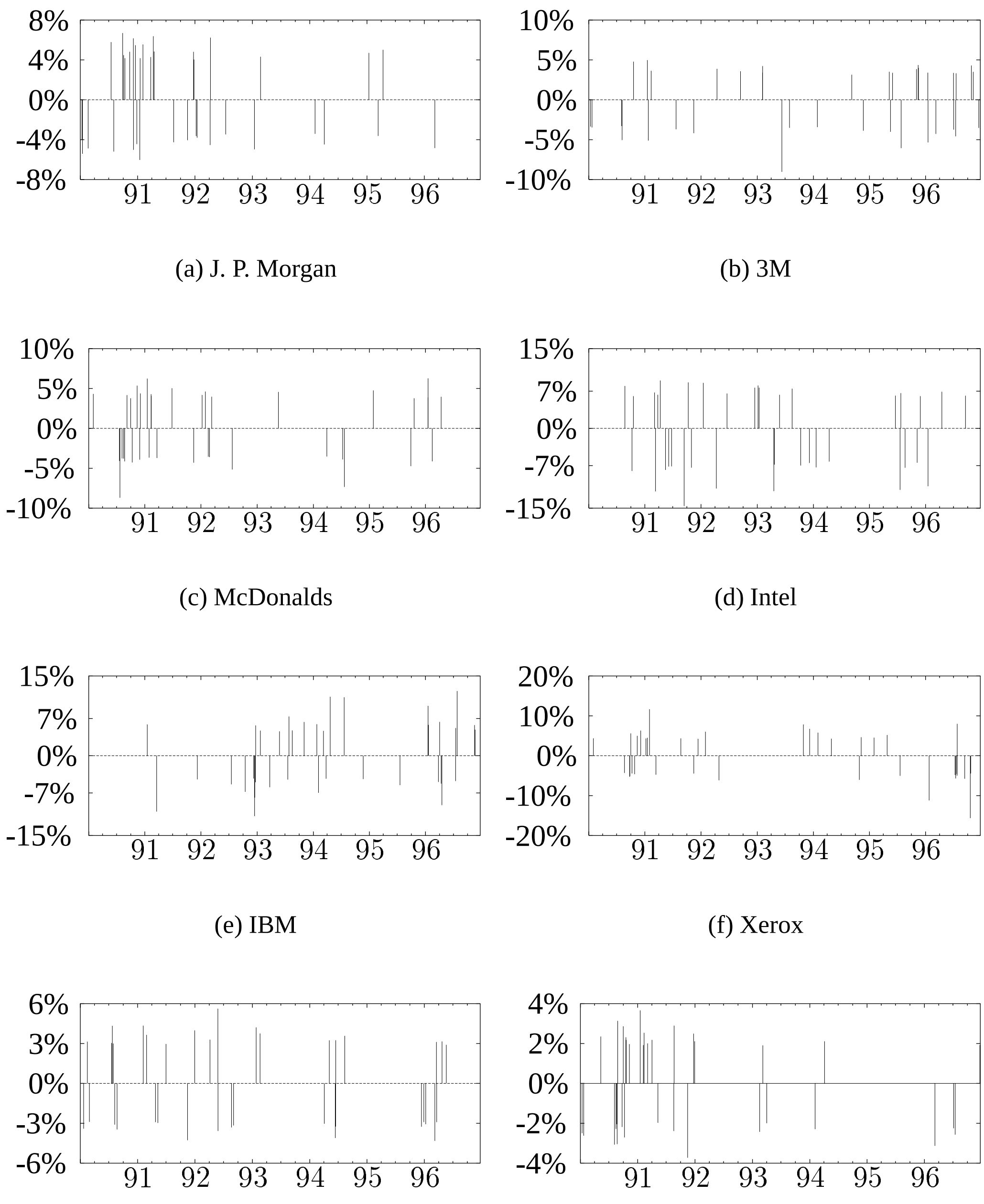
We evaluate various methods for VaR analysis, and compare the traditional methods with our tail distribution estimator using a portfolio of stocks. First we construct a number of random portfolios over several time periods, and compare the results of one step ahead VaR predictions. Second we investigate multi-day VaR analysis. Third we study the implications of adding an index option to the portfolio. Forth, the issues relating to the determination of capital are discussed. Finally, we discuss the practical implementations of these methods for real portfolio management, with special emphasis on the ease of implementation and computational issues.

2 Properties of Extreme Returns

Value-at-Risk analysis is highly dependent on extreme returns or spikes. The empirical properties of the spikes, are not the same as the properties of the entire return process. A major result from empirical research of returns, is the almost zero autocorrelation and significant positive serial correlation in the volatility of returns. As a result volatilities can be relatively well predicted with a parametric model such as GARCH. If, however, one focuses only on spikes, the dependency is reduced.

Table 1 lists the number of trading days between the daily extremes for the SP-500 index along with the rank of the corresponding observation. Figure 1 shows the 1% highest and lowest returns on the daily SP-500 index in the 1990's along with the 7 stocks used below in testing the VaR estimation techniques. No clear pattern emerges for these return series. In some cases we see clustering, but typically the extreme events are randomly scattered. Furthermore, there does not appear to be strong correlation in the tail events. There were two days when 5 assets had tail events, no days with 4 tail events, 5 days with 3 events, 21 days with two events, 185 days with one event, and 1558 days with no tail events. For the SP-500, two of the upper tail observations are on adjacent days but none of the lower tailed observations, and in most cases there are a number of days between the extreme observations. One does not observe market crashes many days in a row. There are indications of some clustering of the tail events over time. However, the measurement of a spike on a given day, is not indicative of a high probability of a spike the following few days. The modelling of the dependence structure of spikes would therefore be different than in e.g. GARCH models. If the threshold level, indicating the beginning of the tails, rises as the sample size increases, the spikes eventually behave like a Poisson process. In other words, for certain dependent processes, like ARCH, volatility clustering vanishes at the level of the extreme

Figure 1: 1% largest and smallest daily returns on stocks in portfolio.



(a) J. P. Morgan

(b) 3M

(c) McDonalds

(d) Intel

(e) IBM

(f) Xerox

(g) Exxon

(h) SP-500

Table 1: Daily SP-500, 1990-96. Time Between Extreme Returns

Upper Tail			Lower Tail		
date	days	rank	date	days	rank
90-08-27	74	2	90-01-23	6	6
90-10-01	24	4	90-08-07	136	3
90-10-18	13	8	90-08-17	8	12
90-10-19	1	10	90-08-22	3	15
90-11-09	15	14	90-08-24	2	4
91-01-17	46	1	90-09-25	21	14
91-02-06	14	16	90-10-10	11	5
91-02-11	3	5	91-05-13	148	17
91-03-05	15	13	91-08-20	69	10
91-04-02	19	9	91-11-18	63	1
91-08-21	99	3	93-02-17	315	9
91-12-23	86	6	93-04-05	33	16
91-12-30	4	11	94-02-07	214	11
93-03-08	300	17	96-03-11	527	2
94-04-05	273	12	96-07-08	82	13
96-12-19	686	15	96-07-16	6	7

Note: The first observations, with rank 7 and 8 respectively for the upper and lower tail, are not shown since we do not have the number of days to a previous observation

realizations. This is demonstrated by de Haan, Resnick, Rootzen, and de Vries (1989). Therefore, for computing the VaR, which is necessarily concerned with the most extreme returns, the ARCH effect is of little importance. Hence it suffices to assume that the highest and lowest realizations are i.i.d. This is corroborated by the evidence from Christoffersen and Diebold (1997) that when the forecast horizon is several days, conditional prediction performs no better than using the unconditional distribution as predictive distribution. The reason is that most current history contains little information on the likelihood that a spike will occur, especially in the exponential weighting of recent history by RiskMetrics.

Another important issue is pointed out by Dimson and Marsh (1996) who analyze spikes in 20 years of the British FTSE-A All Share Index, where they define spikes as fluctuations of 5% or more. They find 6 daily spikes, however they also search for non-overlapping multi day spikes, and find 4 2-day spikes, 3 3-day, 3 4-day, 8 weekly, and up to 7 biweekly. Apparently, the number of spikes is insensitive to the time span over which the returns are defined. This is an example of the fractal property of the distribution of returns and the extremes in particular, and is highly relevant for spike forecasting when the time horizon is longer than one day.

2.1 Tail Estimation

Extreme value theory is the study of the tail of distributions. Several researchers have proposed empirical methods for estimation of tail thickness. The primary difficulty in estimating the tails is the determination of the start of the tails. Typically, these estimators use the highest/lowest realizations to estimate the parameter of tail thickness which is called the tail index. Hill (1975) proposed a moments based estimator for the tail index. The estimator is conditional on knowing how many extreme order statistics for a given sample size have to be taken into account. Hall (1990) suggested a bootstrap procedure for estimation of the start of the tail. His method is too restrictive to be of use for financial data, e.g., it is not applicable to the Student-t distribution, which has been used repeatedly to model asset returns. Recently Danielsson and de Vries (1997*a*) and Danielsson, de Haan, Peng and de Vries (1997) have proposed general estimation methods for the number of extreme order statistics that are in the tails. A brief formal summary of these results is presented in Appendix A.

It is known that only one limit law governs the tail behavior of data drawn from almost any fat tailed distribution[1]. The condition on the distribution $F(x)$ for it to be in the domain of attraction of the limit law is given by (7) in appendix A. Since financial returns are heavy tailed, this implies that for obtaining the tail behavior we only have to deal with this limit distribution. By taking an expansion of $F(x)$ at infinity and imposing mild regularity conditions one can show that for most heavy tailed distributions the second order expansion of the tails is:

$$F(x) \simeq 1 - ax^{-\alpha}\left[1 + bx^{-\beta}\right], \quad \alpha, \beta > 0 \tag{1}$$

for x large, while $a,\ b,\ \alpha,$ and β are parameters. In this expansion the key coefficient is α which is denoted as the tail index which indicates the thickness of the tails. The parameter a determines the scale; the other two parameters b and β are the second order equivalents to a and α. For example, for the Student-t or the non-normal stable densities, α equals the degrees of freedom or the characteristic exponent. For the ARCH process α equals the number of bounded moments of the unconditional distribution of the ARCH innovations.

Hill (1975) proposed a moments based estimator of the tail index which is estimated conditional on a threshold index M where all values $x_i > X_{M+1}$ are used in the estimation. The X_i indicate the decreasing order statistics, $X_1 \geq X_2 \geq ... \geq X_M \geq ... \geq X_n$,

1. Danielsson and de Vries (1997*a*) discuss this issue in details.

in a sample of returns $x.$ Danielsson and de Vries (1997*a*) discuss the following estimator for the tail probabilities, given estimates of α and the threshold:

$$\hat{F}(x) = p = \frac{M}{n}\left(\frac{X_{M+1}}{x}\right)^{\hat{\alpha}}, \quad x > X_{M+1} \tag{2}$$

where n is the number of observations, and p is the probability. This applies equally to the lower tails. By taking the inverse of $\hat{F}(x)$ we obtain an extreme quantile estimator:

$$\hat{x}_p = \hat{F}^{-1}(x) = X_{M+1}\left(\frac{M}{np}\right)^{\frac{1}{\hat{\alpha}}}. \tag{3}$$

Note that $\hat{F}(x)$ is always conditional on a given sample. In order to use the distribution $\hat{F}(x)$ we need to specify the parameters α and the random variables M and X_{M+1}, before we can obtain a quantile estimate for a probability. The empirical and estimated distribution functions of the SP-500 index are presented in Figure 3. Some practical issues of the tail estimation are discussed below.

2.2 Multi Period Extreme Analysis

The method for obtaining multi period predictions follows from the work of Feller (1971, VIII.8). Feller shows that the tail risk for fat tailed distributions is, to a first approximation, linearly additive. From (1) we know that for a single period return, $\Pr[X > x] \approx ax^{-\alpha}$ when x is large. Then for the T-period return we have

$$\Pr[X_1 + X_2 + ... + X_T > x] \approx Tax^{-\alpha}. \tag{4}$$

The implication for portfolio analysis of this result has been discussed in the specific case of non-normal stable distributions by Fama and Miller (1972, p. 270). In that case $\alpha < 2$ and the variance is infinite. Muller, Pictet and de Vries (1995) are the first to discuss the finite variance case when $\alpha > 2.$ It is well known that the self-additivity of normal distributions implies that the $T^{1/2}$ scaling factor for multi period VaR, i.e. the 'square-root-of-time rule' implemented in RiskMetrics. But for heavy-tailed distributions this factor is different for the largest risks. Heavy tailed distributions are self-additive in the tails, see e.g. (4). This implies a scaling factor $T^{1/\alpha}$ for VaR in a T-period analysis. With finite variance where $\alpha > 2$ and hence $T^{1/2} > T^{1/\alpha}$, i.e. the scaling factor for heavy tailed distributed returns is smaller than for normal distributed returns. In comparison with the normal model, there are two counterbalancing forces. If daily returns are fat tailed distributed, then there is a higher probability of extreme losses and this increases the one day possible loss vis-a-vis the normal model. This is a level effect. But there is also a slope effect. Due to the above result, the multiplication factor (slope) used to obtain the multi-day extreme is smaller for fat tailed distributed returns than for normal returns. For this reason, extreme predictions from the two models may cross if we consider different time horizons. This is demonstrated in Table 6.

2.3 Monte Carlo Evidence

In order to evaluate the performance of the estimated tail distribution in (2) Danielsson and de Vries (1997*a*) do extensive Monte Carlo experiments to evaluate the properties of the estimator. In Table 2 a small subset of the results is presented. We generate repeated samples of size 2000 from a Student-t distribution with 4 degrees of freedom and compare the average maxima, denoted here as the sample maxima by historical simulation (HS), from the samples with the average predicted value by $\hat{F}(x)$, denoted as extreme value (EV). The specific distribution was chosen since its tail behavior is similar to a typical return series. The Monte Carlo results are reported in Table 2.

Table 2: Predicted and Expected Maxima of Student-t(4)

In Sample Prediction, 2000 observations	Theoretical	Average Values
Sample Maxima by HS	8.610	$10.67\ (4.45)\ [4.90]$
Forecast Maximas by EV	8.610	$8.90\ (1.64)\ [1.66]$
Out of Sample Prediction		
Forecast Maximas by EV for Sample of Size 4000	10.306	$10.92\ (2.43)\ [2.50]$
Forecast Maximas by EV for Sample of Size 6000	11.438	$12.32\ (3.02)\ [3.14]$

Sample size = 2000, simulations 1000, bootstrap iterations = 2000. Standard errors in parenthesis, RMSE in brackets. HS denotes estimation by historical simulation and EV estimation by the tail estimator with method proposed by Danielsson and de Vries (1997*b*).

Out-of-sample predictions were obtained by using the estimated tail of the distribution to predict the value of the maxima of a sample of size 4000 and 6000, the true values are reported as well. We can see that the tail estimator performs quite well in predicting the maxima while the sample averages yield much lower quality results. Note that the variance of HS approach is much higher than the variance by EV method. Moreover, HS is necessarily silent on the out of sample sizes 4000 to 6000, where EV provides an accurate estimate. Obviously, if one used the normal to predict the maximas, the result would be grossly inaccurate, and would in fact predict values about one third of the theoretical values. See also Figures 3 and 4 in Section 4 below for a graphical illustration of this claim.

3 Value-at-Risk and Common Methods

The formal definition of Value-at-Risk (VaR) is easily given implicitly:

$$\Pr\left[\Delta P \Delta t \leq VaR\right] = \pi, \tag{5}$$

where $\Delta P \Delta t$ is a change in the market value of portfolio P over time horizon Δt with probability π. Equation (5) states that a loss equal to, or larger than the specific VaR occurs with probability π. Or conversely, (5) for a given probability π losses, equal to or larger than the VaR, happen. In this latter interpretation the VaR is written as a function of the probability π. Let $F\left(\Delta P \Delta t\right)$ be the probability distribution of $\Delta P \Delta t$, then

$$F^{-1}\left(\pi\right) = VaR; \tag{6}$$

where $F^{-1}\left(\cdot\right)$ denotes the inverse of $F\left(\cdot\right)$. The major problem in implementing VaR analysis is the specification of the probability distribution $F\left(\cdot\right)$ which is used in the calculation in (5).

Two methods are commonly used to evaluate VaR:

1. Historical Simulation (Non Parametric, Unconditional Volatility)
2. Parametric Methods (Fully Parametric, Conditional Volatility)

Both these methods are discussed in this section. The semi-parametric extreme value (EV) method falls in between these two methodologies.

3.1 Historical Simulation

A popular method for VaR assessment is historical simulation (HS). Instead of making distributional assumptions about returns, past returns are used to predict future returns.

The advantage of historical simulation is that few assumptions are required, and the method is easy to implement. The primary assumption is that the distribution of the returns in the portfolio is constant over the sample period. Historical simulation has been shown to compare well with other methods, see e.g. Mahoney (1996), however past extreme returns can be a poor predictor of extreme events, and as a result historical simulation should be used with care. The reason for this is easy to see. By its very nature HS has nothing to say about the probability outcomes which are worse than the sample minimum return. But HS also does not give very accurate probability estimates for the in sample extreme as is demonstrated below. Furthermore, the choice of sample size can have a large impact on the value predicted by historical simulation. In addition, the very simplicity of HS makes it difficult to conduct sensitivity experiments, where a VaR is evaluated under a number of scenarios.

A major problem with HS is the discreteness of extreme returns. In the interior, the empirical sampling distribution is very dense, with adjacent observations very close to each other. As a result the sampling distribution is very smooth in the interior. The closer one gets to the extremes, the longer the interval between adjacent returns becomes. This can be seen in Table 3 where the 7 largest and smallest returns on the stocks in the sample portfolio and SP-500 Index for 10 years are listed.

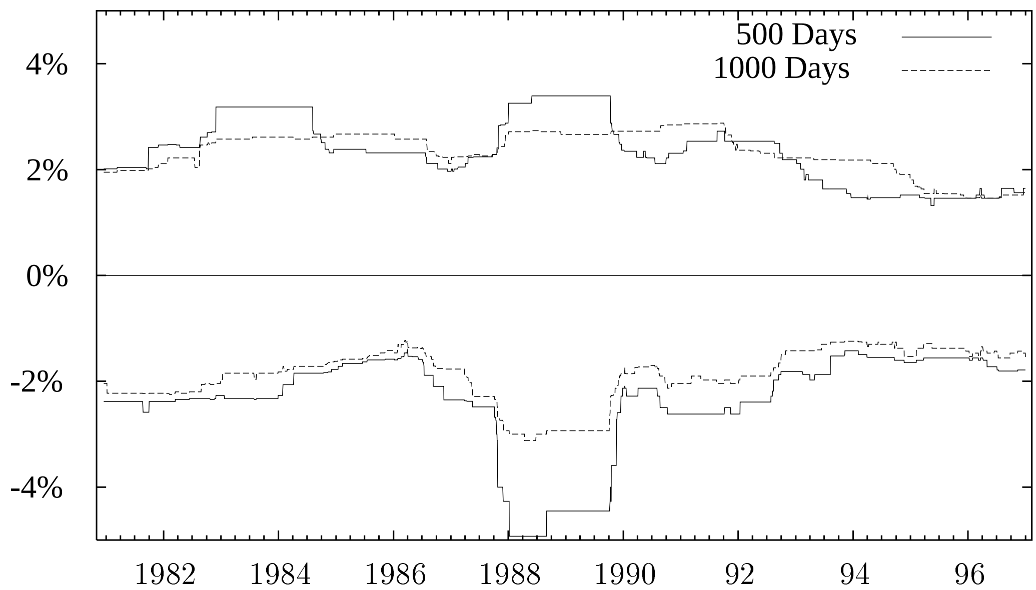
Table 3: Extreme daily returns 1987 - 1996

JPM	25%	12%	8.8%	6.7%	6.5%	6.4%	6.3%
	−41%	−6.7%	−6.3%	−6.1%	−6.0%	−5.8%	−5.7%
MMM	11%	7.1%	5.9%	5.7%	5.7%	5.0%	4.8%
	−30%	−10%	−10%	−9.0%	−6.2%	−6.1%	−5.6%
MCD	10%	7.9%	6.3%	6.2%	5.4%	5.0%	5.0%
	−18%	−10%	−8.7%	−8.5%	−8.3%	−7.3%	−6.9%
INTC	24%	11%	9.9%	9.0%	8.9%	8.6%	8.6%
	−21%	−21%	−16%	−15%	−14%	−12%	−12%
IBM	12%	11%	11%	10%	9.4%	7.4%	6.5%
	−26%	−11%	−11%	−9.3%	−7.9%	−7.5%	−7.1%
XRX	12%	8.0%	7.8%	7.5%	7.1%	6.8%	6.3%
	−22%	−16%	−11%	−8.4%	−7.5%	−6.9%	−6.2%
XON	17%	10%	6.0%	5.8%	5.8%	5.6%	5.4%
	−27%	−8.7%	−7.9%	−6.6%	−6.3%	−5.7%	−5.4%
SP-500	8.7%	5.1%	4.8%	3.7%	3.5%	3.4%	3.3%
	−23%	−8.6%	−7.0%	−6.3%	−5.3%	−4.5%	−4.3%

These extreme observations are typically the most important for VaR analysis, however since these values are clearly discrete, the VaR will also be discrete, and hence be either underpredicted or overpredicted. We see that this effect is somewhat more pronounced for the individual assets, than for the market portfolio SP-500, due to diversification. Furthermore, the variance of the extreme order statistics is very high, and in some cases infinite. As a result, VaR estimates that are dependent on the tails, will be measured discretely, with a high variance, making HS in many cases a poor predictor of the VaR. Results from a small Monte Carlo (MC) experiment demonstrating this are presented in Section 2.3.

In Figure (2) we plot the 99th percentile of the S&P for the past 500 and 1000 days, i.e. the 5th and 10th largest and smallest observations for the past 500 and 1000 days respectively. It is clear from the figure that the window length in assessing the proba-

Figure 2: 1% largest and smallest returns on SP-500 over 500 and 1000 Day Windows



bility of spikes is very important, and this creates a serious problem. Note how rapidly the percentile changes when new data enter and exit the window. In VaR prediction with HS, the inclusion or exclusion of one or two days at the beginning of the sample can cause large swings in the VaR estimate, while no guidelines exist for assessing which estimate is the better.

Butler and Schachter (1996) propose a variation of HS by use of a kernel smoother to estimate the distribution of returns, which is in essence an estimation of the distribution of returns. This type of methodology has both advantages and drawbacks. The advantage is that a properly constructed kernel distribution provides a smooth sampling distribution. Hence sensitivity experiments can be readily constructed, and valuable insight can be gained about the return process. Furthermore such distribution may not be as sensitive to the sample length as HS is. Note that these advantages are dependent on a properly constructed kernel distribution. In kernel estimation, the specific choice of a kernel and window length is extremely important. Almost all kernels are estimated with the entire data set, with interior observations dominating the kernel estimation. While even the most careful kernel estimation will provide good estimates for the interior, there is no reason to believe that the kernel will describe the tails adequately. Tail bumpiness is a common problem in kernel estimation, however, if tail bumpiness is observed in kernel estimation of returns it is simply an artifact of the specific methods used. Returns are in general unimodal. Note especially that financial data are thick tailed with high excess kurtosis. Therefore, a Gaussian kernel, which assumes that the estimated distribution has the same shape as the normal, is unsuitable for financial data.

3.2 Parametric Forecasting

In parametric forecasting, the predicted future volatility of an asset is an explicit function of past returns, and the estimated model parameters. The most common models are the unconditional normal with frequently updated variance estimate, or explicit models for conditional heteroscedasticity like the GARCH model, with normal innovations. The popular RiskMetrics approach which uses the frequently updated normal model is asymptotically equivalent to an IGARCH model. This implies a counterfactual hypothesis of an unconditional infinite variance. However since in most cases only short horizon conditional forecasts are made, this does not affect the results significantly. GARCH models with normal innovations have proved valuable in forecasting common volatilities, however they perform poorly in predicting extreme observations, or spikes, in returns. The normality assumption is primarily a matter of convenience, and a GARCH model with non-normal innovations can easily be estimated, with the most common specification being the Student-t. The advantage of Student-t innovations is that they are thick tailed and hence will in general provide better predictive densities; note that the Student-t contains Gaussian errors as a special case. The disadvantages of non-normal innovations are several, e.g. multivariate versions of such models are typically hard to estimate and recursive forecasts of future volatilities are difficult for most distributions, since they are typically not self additive.

There are several reasons for the failure of RiskMetrics to adequately capture the tail probabilities. For example the normal likelihood function weighs values close to zero higher than large values so the contribution of the large values to the likelihood function is relatively small. Since most observations are in the interior, they dominate the estimation, especially since tail events are maybe 1-2% of the observations. While a GARCH model with normal innovations preforms poorly, it does not imply that parametric forecasting will in general provide biased VaR estimates, however such a model would have to be constructed with the tails as the primary focus. See Jackson, Maude and Perraudin (1997) for discussion on this issue.

There is yet another problem with the way RiskMetrics implements the GARCH methodology. Instead of going by the GARCH scheme for predicting future volatilities, RiskMetrics ignores GARCH and simply uses the square-root-of-time method which is only appropriate under an i.i.d. normal assumption. If the predicted next day volatility is $\hat{\sigma}^2_{t+1}$, then the predicted T day ahead volatility is $T\hat{\sigma}^2_{t+1}$ in the RiskMetrics analysis. This implies that for the next T days, returns are essentially assumed to be normally distributed with variance $T\hat{\sigma}^2_{t+1}$. The underlying assumption is that returns are i.i.d., in which case there would be no reason to estimate a conditional volatility model.

Table 4: Observed Extreme Returns of the daily SP-500, 1990-1996, and the Probability of that Return as Predicted by the Normal GARCH, Student-t GARCH model, the Extreme Value Estimation Method, and the Empirical Distribution.

Observed	Probabilities			
Return	Normal	Student-t	EV Estimator	Empirical
-3.72%	0.0000	0.0002	0.0007	0.0006
-3.13%	0.0000	0.0010	0.0015	0.0011
-3.07%	0.0002	0.0021	0.0016	0.0017
-3.04%	0.0032	0.0071	0.0016	0.0023
-2.71%	0.0098	0.0146	0.0026	0.0028
-2.62%	0.0015	0.0073	0.0029	0.0034
3.66%	0.0000	0.0011	0.0004	0.0006
3.13%	0.0060	0.0096	0.0009	0.0011
2.89%	0.0002	0.0022	0.0013	0.0017
2.86%	0.0069	0.0117	0.0014	0.0023
2.53%	0.0059	0.0109	0.0025	0.0028
2.50%	0.0007	0.0038	0.0026	0.0034

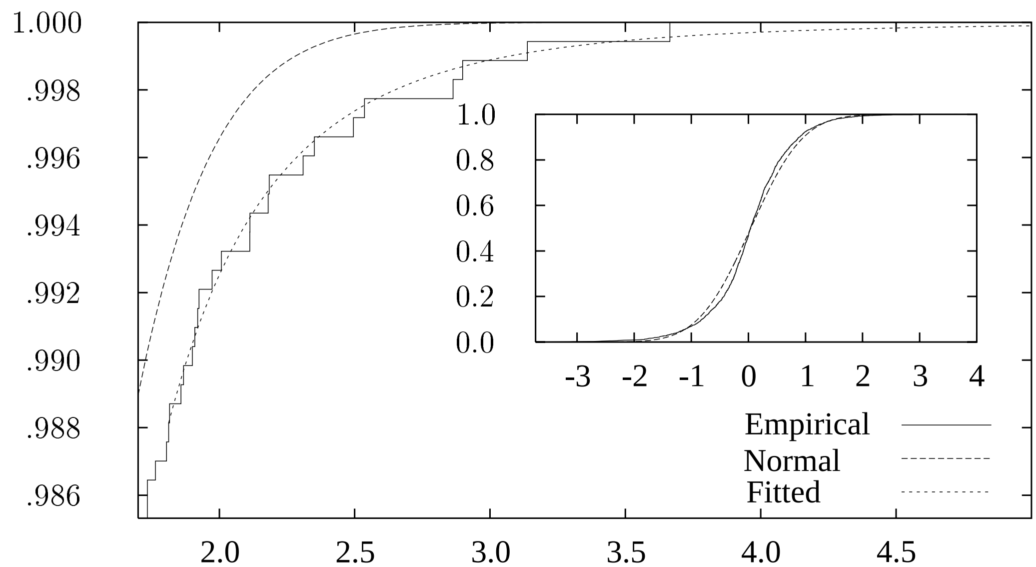
Note that this problem can be bypassed by using T day data to obtain T day ahead predictions as suggested in the RiskMetrics manual.

In Table 4 we show the six highest and lowest returns on the daily SP-500 index from 1990 to 1996, or 1771 observations. We used the normal GARCH and Student-t GARCH models to predict the conditional volatility, and show in the table the probability of an outcome equal to or more extreme than the observed return, conditional on the predicted volatility for each observation. In addition we show the probability as predicted by the extreme value estimator, and values of the empirical distribution function. We see from the table that the normal GARCH model performs very poorly in predicting tail events, while the Student-t GARCH model gives somewhat better results. Both methods are plagued by high variability and inaccurate probability estimates, while the extreme value estimator provides much better estimates.

4 Extreme Value Theory and VaR

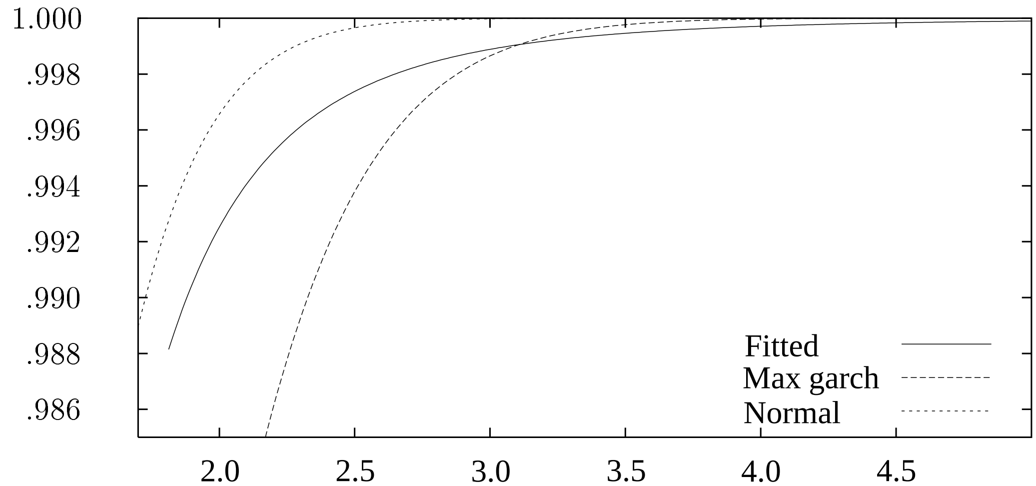
Accurate prediction of extreme realizations is of central importance to VaR analysis. VaR estimates are calculated from the lower extreme of a portfolio forecast distribution; therefore, accurate estimation of the lower tail of portfolio returns is of primary importance in any VaR application. Most available tools, such as GARCH, are however designed to predict common volatilities, and therefore have poor tail properties. Even historical simulation (HS) has less than desirable sampling properties out in the tails. Therefore, a hybrid technique that combines sampling from the empirical distribution for common observations with sampling from a fitted tail distribution has the potential to perform better than either HS or fully parametric methods by themselves.

Figure 3: Distribution of SP-500 returns 1990-1996 with fitted upper tail.



In Figure 3 the empirical distribution of the SP-500 index is plotted along with the fitted tail estimator distribution $F(x)$ and the estimated normal distribution. We see the problems with HS in the tails from Figure 3, e.g. discreteness of observations and the inability to provide out-of-sample low probability predictions. The normal distribution clearly underestimates the probability of the highest returns. On the other hand, the fitted distribution is a smooth function through the empirical distribution, both in and out of sample. For comparison, in figure 4 we plot the fitted distribution along with the normal distribution estimated from the sample mean and variances, and the distribution obtained from the normal GARCH(1,1) process if one conditions on the maximum observed past volatility. This means that the normal distribution, with the variance of the largest of the one day GARCH volatility predictions, is plotted. This gives the normal GARCH the maximum benefit of the doubt. Since this conditional distribution

Figure 4: Distribution of SP-500 returns 1990-1996 and Highest GARCH Prediction.



is still normal, it underestimates the extreme tails. There are several advantages in using the estimated tail distribution in VaR estimation. For example:

- In HS, the presence of an event like the '87 crash in the sample, will cause a large VaR estimate. However, since a '87 magnitude crash only occurs rarely, say once every 60 years, the presence of such an event in the sample will produce downward biased VaR estimates. And hence imposes too conservative capital provisions. By sampling from the tail distribution, the probability of a '87 type event will be much smaller, leading to better VaR estimates.
- The empirical distribution is sampled discretely out in the tails, with the variance of the extreme order statistics being very high. This implies that a VaR that relies on tail realizations will exhibit the same properties, with the resulting estimates being highly variable. A Monte Carlo example of this is given in Table 2.
- By sampling from the tail of the distribution, one can easily obtain the lowest return that occurs with a given probability, say 0.1%, greatly facilitating sensitivity experiments. This is typically not possible with HS by itself.
- The probability theory of tail observations, or extreme value theory, is well known, and the tail estimator therefore rests on firm statistical foundations. In contrast, most traditional kernel estimators have bad properties in the tails.

4.1 Estimated Tails and Historical Simulation

We propose combining the HS for the interior with the fitted distribution from (1) along the lines of Danielsson and de Vries (1997*a*). Recall from above that the fitted distribution, $\hat{F}(x)$, is conditional on one of the highest order statistics X_{M+1}. Therefore we can view X_{M+1} as the start of the tail, and use $\hat{F}(x)$ as the sampling distribution for extreme returns. Below this threshold X_{M+1} we can see the empirical distribution for interior returns. This can be implemented in the following algorithm, where $X_{M^{\text{upper}}+1}$ and $X_{M^{\text{lower}}-1}$ are the thresholds for the upper and lower tail respectively, and T is the window size.

Draw x^s from $\{x_t\}_{t=1}^T$ with replacement
if $x^s < X_{M^{\text{lower}}-1}$ **then**
  draw x^s from $\hat{F}(x)$ for the lower tail
else
  if $x^s > X_{M^{\text{upper}}+1}$ **then**
    draw x^s from $\hat{F}(x)$ for the uper tail
  else
    keep x^s
  end if
end if

Note that this guarantees that the combined density integrates out to one. We can then view x as one draw from the combined empirical and extreme value distributions, and we denote the method as the combined extreme value estimator and historical simulation method.

4.2 Tails of Portfolios

In general, multiple assets are used to construct a portfolio. We can implement simulations of portfolio returns with one of two methods, post fitting or pre-sampling. Results from implementing both methods are presented in Table 5 and discussed below. Note that while we would not necessarily expect correlation in the tails of stock returns, tail correlation is often expected in exchange rates, e.g. in the EMS, large movements happen often at the same time period for several countries.

4.2.1 Post Fitting

In post fitting, one proceeds along the lines of the combined extreme value and historical simulation procedures and applies the current portfolio weights to the historical prices to obtain a vector of simulated portfolio returns. This is exactly as in historical simulation. Subsequently, the tails of the simulated returns are fitted, and any probability-VaR combination can be read from the fitted tails. This procedure has several advantages. No restrictive assumptions are needed, the method can be applied to the largest of portfolios, and does not require significant additional computation time over HS. The primary disadvantage is that it carries with it the assumption of constant correlation across returns, while in many cases one observes systematic changes in correlation over time. However, in the results below this does not seem to cause any significant problems.

4.2.2 Presampling

In the presampling method, each asset is sampled independently from the hybrid extreme value estimator and empirical distribution, and subsequently scaled to obtain properly correlated returns. Then the value of the portfolio is calculated. The scaling is achieved as follows. Let Σ_t be the covariance matrix of the sample, and $L_t L_t' = \Sigma_t$ be the Cholesky transformation. The number of assets in the portfolio is K and the number of simulations is N. We then draw a KN matrix of simulated returns, denoted as $\tilde{X}_n$. Let the covariance matrix of $\tilde{X}_n$ be denoted by Ω_n, with the Cholesky transformation $M_n M_n' = \Omega_n$. Scale $\tilde{X}_n$ to an identity covariance matrix by $M_n^{-1}\tilde{X}_n$, which can then be scaled to the sample covariance by L_t. The matrix of simulated returns X is:

$$X_n = L_t M_n^{-1} \tilde{X}_n.$$

If $w = \{w_i\}_{i=1}^K$ is the vector of portfolio weights, the simulated return vector R is:

$$R_n = \sum_{i=1}^{K} w_i X_{t,n,i} \quad n = 1, N.$$

By sorting the simulated portfolio returns R, one can read off the tail probabilities for the VaR, in the same manner as in HS. By using this method, it is possible to use a different covariance matrix for sub samples than for the whole sample. This may be desirable when the covariance matrix of returns changes over time where it may yield better results to replace the covariance matrix Ω with the covariance matrix of the last part of the sample.

5 Estimation

To test the performance of our VaR procedure, we selected 6 US stocks randomly as the basis for portfolio analysis in addition to the JP Morgan bank stock price. The stocks in the tables are referred by their ticker tape symbols. The window length for HS and the combined extreme value–empirical distribution procedure was set at 6 years or 1500 trading days. Note this is much larger than the regulatory window length of one year. The reason for this long period is that for accurate estimation of events that happen once every 100 days, as in the 1% VaR, one year is not enough for accurate estimation. In general, one should try to use as a large sample as is possible. Using a smaller sample than 1500 trading days in the performance testing was not shown to improve the results. Performance testing starts at Jan. 15. 1993, and the beginning of the sample is 1500 days before that on Feb. 12, 1987. It is a stylized fact in empirical studies of financial returns, that returns exhibit several common properties, regardless of the underlying asset. This extends to the tails of returns. In Table 12 we present summary statistics on a wide range of financial returns for the period 1987-1996, and is clear that the tails all have similar properties. Summary statistics for each stock return are listed in Table 8 for the entire sample period, and in Table 9 for the 1990-1996 testing period. The corresponding correlation matrixes are presented in Tables 10 and 11. The sample correlations drop in the 1990's. Given this change in correlation, we tested changing correlations in the pre-fitting method, but it did not have much impact for our data, and therefore we do not report those results here.

5.1 VaR Prediction

5.1.1 Interpretation of Results

Results are reported in Table 5. The VaR return estimates for each method are compared with the realized returns each day. The number of violations of the VaR estimates were counted, and the ratio of violations to the length of the testing period was compared with the critical value. This is done for several critical values. This is perhaps the simplest possible testing procedure. Several authors, most recently Dave and Stahl (1997), propose much more elaborate testing procedures, e.g. the likelihood based method of Davé and Stahl which is used to test a single portfolio. However by using a large number of random portfolios one obtains accurate measurements of the performance of the various methods, without resorting to specific distributional assumptions, such as the normality assumption of Davé and Stahl. In addition, while the green, yellow, and red zone classification method proposed by BIS, may seem attractive for the comparison, it is less informative than the ratio method used here.

Table 5: Estimation Results: Average Number of Realized Portfolios that were Larger than VaR Predictions

Confidence Level	5%	2.5%	1%	0.5%	0.25%	0.1%	0.05%	0.025%	0.01%	0.005%
Expected Number of days with Exceedances	50	25	10	5	2.5	1	0.5	0.25	0.1	0.05
Expected Frequency										
in Days	20	40	100	200						
in Years					1.5	3.8	7.7	15	38	77
RiskMetrics	52.45	30.26	16.28	10.65	7.29	4.85	3.55	2.72	2.00	1.58
	(7.39)	(4.41)	(3.13)	(2.73)	(2.27)	(2.06)	(1.81)	(1.66)	(1.45)	(1.29)
Historical Simulation	43.24	20.50	7.66	3.69	1.90	0.95	0.75	0.75	0.75	0.75
	(10.75)	(7.22)	(3.90)	(2.39)	(1.57)	(1.03)	(0.89)	(0.89)	(0.89)	(0.89)
Extreme Value Presampling	44.02	22.35	9.32	4.82	2.54	1.21	0.68	0.37	0.09	0.09
	(11.62)	(7.66)	(4.26)	(2.56)	(1.71)	(1.27)	(0.98)	(0.71)	(0.31)	(0.31)
Extreme Value Post Fitting	43.14	20.84	8.19	4.23	2.35	1.06	0.59	0.33	0.12	0.06
	(11.10)	(7.35)	(3.86)	(2.55)	(1.72)	(1.13)	(0.82)	(0.62)	(0.35)	(0.23)

Daily observations in testing = 1000 over period 930115 to 961230. Window size in HS and EV = 1500, initial staring date for window 870210. Random portfolios = 500. Simulation size in presampling tail estimator = 10000. Standard errors in parenthesis. Probabilities expressed in percentages with sum=100%. EV results in the table are based on the prodecdure proposed by Danielsson and de Vries (1997*b*)

The test sample length was 1000 trading days, and the window size in HS and EV was 1,500. For the 1% risk level, we expect a single violation of the VaR every 100 days, or 10 times over the entire testing period. This risk level is given in the fourth column from the left in Table 5. At this risk level RiskMetrics yields too many violations, i.e. 16.3, on average, while the other methods give too few violations, or from 7.6 for HS to 9.3 for the presampling EV method, on average. If the number of violations is higher than the expected value, it indicates that the tails are underpredicted, thinner or lower than expected, and conversely too few violations indicate that the estimated tail is thicker than expected. In addition to the tail percentages, we show the implied number of days, i.e. how frequently one would expect a tail event to occur. If the number of days is large, we transform the days into years, assuming 260 trading days per year.

5.1.2 Comparison of Methods

For the 5th percentile, RiskMetrics performs best. The reason for this is that at the 5% level we are sufficiently inside the sample so that the conditional prediction performs better than unconditional prediction. However, as we move to the tails, RiskMetrics consistently underpredicts the tail, with ever larger biases as we move farther into the tails. For example, at the 0.1% level RiskMetrics predicts 5 violations, while the expected number is one. Therefore RiskMetrics will underpredict the true number of losses at a given risk level. Historical simulation has in a way the opposite problem, in that it consistently overpredicts the tails. Note that for HS we can not obtain estimates for lower probabilities than one over the sample size, or in our case probabilities lower than once every 1500 days. Hence the lowest prediction, 0.75, is repeated in the last four columns in the table. Obviously for smaller sample sizes HS is not able to predict the VaR for even relatively high probabilities. Both EV estimators have good performance, especially out in the tails. The presampling version of the EV estimator can not provide estimates for the lowest probability. The simulation size was 10,000 and this limits the lowest probability at 1/10,000. The post fitting version has no such problems. It is interesting to note that the EV estimators do a very good job at tracking the expected value of exceedances. Even at the lowest probability, the expected value is 0.05 while the post fitting EV method predicts 0.06.

5.1.3 Implication for Capital Requirements

A major reason for the implementation of VaR methods is the determination of capital requirements (CR). Financial regulators determine the CR according to the formula

$$\text{CR} = 3 * \text{VaR} + \text{constant}$$

Individual financial institutions estimate the VaR, from which the CR are calculated. If the banks underestimate the VaR they get penalized by an increase in the multiplicative factor or the additive constant. The multiplicative constant may be increased to 4. If, however, the financial institution over estimates the VaR, it presumably gets penalized by shareholders. Hence accurate estimation of the VaR is important. The scaling factor 3 appears to be somewhat arbitrary, and has come under criticism from financial institutions for being too high. Stahl (1997) argues that the factor is justified by applying Chebyshev's inequality to the ratio of the true and model VaR distributions. In this worst case scenario, Stahl calculates 2.7 as an appropriate scaling factor at the 5% level, 4.3 at the 1% level, and increasing with lower probabilities. But according to Table 5, this factor is much too high or conservative. By comparing the RiskMetrics and the EV results at the 5% level, we see that they are very close to the expected number of violations, and in that case a multiplicative constant close to one would be appropriate. At the 0.1% level, RiskMetrics has five times the expected number of violations and in that case a large multiplicative constant may be appropriate, but the EV method gives results close to the expected value, suggesting that the constant should be close to one if EV is used for VaR. While a high scaling factor may be justified in the normal case, by using the estimate of the tails, as we do with the EV method, the multiplicative factor can be much lower. Note that HS, implies too high capital requirements in our case, while RiskMetrics implies too low CR. The extreme value estimator method appears to provide accurate tail estimates, and hence the most accurate way to set capital requirements.

Danielsson, Hartmann and de Vries (1997) raise an issue regarding implications for incentive compatibility. The banks want to keep capital requirements as low as possible, and are faced with a sliding multiplicative factor in the range from three to four. Given that using a simple normal model implies considerably smaller capital requirements than the more accurate historical simulation or extreme tail methods, or even RiskMetrics, and that the penalty for under predicting the VaR is relatively small, i.e. the possible increase from 3 to 4, it is in the banks best interest to use the VaR method which provides the lowest VaR predictions. This will, in general be close to the worst VaR method available. This may explain the current prevalence among banks of using a moving average normal model for VaR prediction. It is like using a protective sunblock, because one has to, but choosing the one with lowest protection factor because its cheapest, with the result that one still gets burned.

5.2 Multi Day Prediction

While most financial firms use one day VaR analysis for internal risk assessment, regulators require VaR estimates for 10 day returns. There are two ways to implement a multi day VaR. If the time horizon is denoted by T, one can either look at past non-overlapping T day returns, and use these in the same fashion as the one day VaR analysis, or extrapolate the one day VaR returns to the T day VaR. The latter method

Table 6: 10 Day VaR Prediction on December 30, 1996 in Millions of US Dollars for a $100 Milion Trading Portfolio

Risk Level	5%	1.0%	0.5%	0.10%	0.05%	0.005%
EV						
One day	$0.9	$1.5	$1.7	$2.5	$3.0	$5.1
10 day	$1.6	$2.5	$3.0	$4.3	$5.1	$8.9
RM						
one day	$1.0	$1.4	$1.6	$1.9	$2.0	$2.3
10 day	$3.2	$4.5	$4.9	$5.9	$6.3	$7.5

has the advantage that the sample size remains as it is. Possibly for this reason, RiskMetrics implements the latter method by the so called 'square-root-of-time' rule which implies that returns are normal with no serial correlation. However, for fat tailed data, a $T^{1/\alpha}$ is appropriate. See section 2.1 for further discussion on this issue.

It is not possible to backtest the $T = 10$ day VaR estimates because we have to compare the VaR predictions with non-overlapping T day returns. This implies that the sample available for testing is T times smaller than the one day sample. Since we are looking at uncommon events, we need to backtest over a large number of observations. In our experience, 1000 days is a minimum test length. Therefore, for 10 day VaR we would need 10,000 days in the test sample.

In order to demonstrate the multi day VaR methods, we use the one day VaR at the last day of our sample, December 30, 1996 to obtain 10 day VaRs. This is the VaR prediction on the last day of the results in Table 5, the number of random portfolios was 500. In Table 6 we present the one day and 10 day VaR predictions from RiskMetrics type and extreme value post fitting methods. The numbers in the table reflect losses in millions of dollars on a portfolio of 100 million dollars. We see in Table 6 the same result as in Table 5, i.e. RiskMetrics underpredicts the amount of losses vis-a-vis EV at the 0.05% and 0.005% probabilities, while for the 10 day predictions RiskMetrics over predicts the loss, relative to EV, even for very low risk levels. Recall that EV uses the multiplicative factor $T^{1/\alpha}$ while RiskMetrics uses $T^{1/2}$. The average α was 4.6, with the average scaling factor of 1.7 which is much smaller than $T^{1/2} = 3.7$. As a result, at the 0.05% level RiskMetrics predicts a 10 day VaR of $6.3m while EV only predicts $5.1m, on average.

5.3 Options

The inclusion nonlinear derivatives like options in the portfolio does not cause much extra difficulty. In general, one might has to price the option, by means of risk neutral probabilities, in the process of calculating the VaR. However, the risk neutral measure is not observed, at least not directly. This is a generic problem for any VaR method, and for this reason RiskMetrics proceeds under the assumption of risk neutrality, and the assumption followed here as well. The extreme value method can be used to generate the data for the underlying asset, and these simulated data can be used to price the option under risk neutrality. A structured Monte Carlo method is easily implemented by the post fitting method.

For simulation of returns on an European option, the path of returns on the underlying is simulated from the current day until expiration, sampling each day return from the combined empirical and estimated distributions, as described above, with the mean subtracted, and summing up the one day returns to obtain a simulated return for the entire period, y_i. If P^F is the future spot price of the asset, then a simulated future price of the underlying is $P^F \exp[y_i]$, and the simulated payoff follows directly. By repeating this N times we get a vector of simulated options payoffs, which is discounted back with the rescaled three month t-bill rate, the vector is averaged, and the price of the option is subtracted. We then update the current futures price by one day through an element from the historical return distribution of the underlying, and repeat the simulation. This is done for each realization in the historical sample. Together this gives us the value of the option, and a vector of option prices quoted tomorrow. Finally we calculate the one day option returns and can treat these returns as any other asset in the portfolio

We used the same data as in the VaR exercise above, and added a European put option on the SP-500 index to the portfolio. The VaR was evaluated with values on September 4, 1997, the future price of the index was 943 and the strike price was 950. We used random portfolio weights, where the option received a weight of 4.9%, and evaluated the VaR on the portfolio with and without the VaR.

The results are in Table 7, where we can see that the option results in lower VaR estimates than if it is left out. Interestingly, the difference in monetary value are the

6 Practical Issues

There are several practical issues in implementing the extreme value method, e.g. the length of the data set, the estimation of the tail shape, and the calculation of the VaR for individual portfolios.

Table 7: Effect of inclusion of Option in Portfolio

Confidence level	VaR with option	VaR without option	Difference
95%	$\$895,501$	$\$1,381,519$	$\$486,019$
99%	$\$1,474,056$	$\$2,453,564$	$\$979,508$
99.5%	$\$1,823,735$	$\$2,754,562$	$\$930,827$
99.9%	$\$3,195,847$	$\$3,856,747$	$\$669,900$
99.99%	$\$7,130,721$	$\$6,277,714$	$-\$853,007$

For any application where we are concerned with extreme outcomes, or events that happen perhaps once every 100 days or less, as is typical in VaR analysis, the data set has to include a sufficient number of extreme events in order to obtain an accurate prediction of VaR. For example, if we are concerned with a 1% VaR, or the worst outcome every 100 days, a window length of one year, or 250 days is not very sensible. In effect the degrees of freedom are around two, and the VaR estimates will be highly inaccurate. This is recognized by the Basle Committee which emphasizes stress testing over multiple tumultuous periods such as the 1987 Crash and the 1993 ERM crisis. In this paper we use a window length of 1,500 days, or about 7 years, and feel that a much shorter sample is not practical. This is reflected when we apply our extreme value procedure to a short sample in Monte Carlo experiments. When the sample is small, say 500 days or two years, the estimate of the tail index is rather inaccurate. There is no way around this issue, historical simulation and parametric methods will have the same small sample problems. In general the sample should be as large as possible. The primary reason to prefer a relatively small sample size is if the correlation structure in the sample is changing over time. However, in that case one can use the presampling version of the tail estimator, and use a covariance matrix that is only estimated with the most recent realizations in the sample. In general one would expect lower correlation in extremes among stocks than e.g. exchange rates; and we were not able demonstrate any benefit for our sample by using a frequently updated covariance matrix. However, we would expect that to happen for a sample that includes exchange rates that belong to managed exchange rate systems like the EMS.

It is not difficult to implement the tails estimation procedure. Using the historical sample to construct the simulated portfolio is in general not computer intensive for even very large portfolios, and in most cases can be done in a spreadsheet like Excel. The subsequent estimation of the tails may take a few seconds at most using an add-in module with a dynamic link library (dll) to fit the tails. So the additional computational complexity compared with historical simulation is a few seconds.

7 Conclusion

Many financial applications are dependent on accurate estimation of downside risk, such as optimal hedging, insurance, pricing of far out of the money options, and the application in this paper, Value-at-Risk (VaR). Several methods have been proposed for VaR estimation. Some are based on using conditional volatilities, such as the GARCH based RiskMetrics method. Others rely on the unconditional historical distribution of returns, such as historical simulation. We propose the use of the extreme value method as a semi–parametric method for estimation of tail probabilities. We show that conditional parametric methods, such as GARCH with normal innovations, as implemented in RiskMetrics, underpredict the VaR for a sample of U.S. stock returns at the 1% risk level, or below. Historical simulation performs better in predicting the VaR, but suffers from a high variance and discrete sampling far out in the tails. Moreover, HS is unable to address losses which are outside the sample. The performance of the extreme value estimator method performs better than both RiskMetrics and historical simulation far out in the tails.

The reason for the improved performance of the EV method is that it combines some of the advantages of both the non-parametric HS approach and the fully parametric RiskMetrics method. By only modelling the tails parametrically, we can also evaluate the risk on observed losses. In addition, because we know that financial return data are heavy tailed distributed, one can rely on a limit expansion for the tail behavior that is shared by all heavy tailed distributions. The importance of the central limit law for extremes is similar to the importance of the central limit law, i.e. one does not have to choose a particular parametric distribution. Furthermore, this limit law shares with the normal distribution the additivity property, albeit only for the tails. This enables us to develop a straightforward rule for obtaining multi period VaR from the single period VaR, much like the normal based square root of time rule. At a future date, we plan to investigate the cross section implication of this rule, which may enable us to deal in a single manner with very widely diversified trading portfolios. We demonstrated that adding non-linear derivatives to the portfolio can be implemented quite easily by using a structured Monte Carlo procedure. We also observed that the present incentives are detrimental to implementing these improved VaR techniques. The current Basle directives rather encourage the opposite, and we would hope that, prudence nonwithstanding, positive incentives will be forthcoming to enhance future improvements in the VaR methodology and implementation thereof in practice.

A Extreme Value Theory and Tail Estimators

This appendix gives an overview of the statistical methods that are used in obtaining the estimated extreme tail distribution. The following is a brief summary of results in Danielsson and de Vries (1997*a*) which also provide all the proofs; the method has been applied by Danielsson and de Vries (1997*b*) .

Let x be the return on a risky financial asset where the distribution of x is heavy tailed. Suppose the distribution function $F(x)$ varies regularly at infinity with tail index α:

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-\alpha}, \quad \alpha > 0, \quad x > 0. \tag{7}$$

This implies that the unconditional distribution of the returns is heavy tailed and that unconditional moments which are larger than α are unbounded. The assumption of regular variation at infinity as specified in (7) is essentially the only assumption that is needed for analysis of tail behavior of the returns x. Regular variation at infinity is a necessary and sufficient condition for the distribution of the maximum or minimum to be in the domain of attraction of the limit law (extreme value distribution) for heavy tailed distributed random variables.

A parametric form for the tail shape of $F(x)$ can be obtained by taking a second order expansion of $F(x)$ as $x \to \infty$. The only non-trivial possibility under mild assumptions is

$$F(x) = \; 1 - ax^{-\alpha}\left[1 + bx^{-\beta} + o\left(x^{-\beta}\right)\right], \quad \beta > 0 \qquad \text{as } x \to 0 \tag{8}$$

The tail index can be estimated by the Hill estimator (Hill (1975)), where M is the random number of exceedances over a high threshold observation X_{M+1}.

$$\frac{1}{\alpha} = \frac{1}{M}\sum_{i=1}^{M} \log \frac{X_i}{X_{M+1}}, \tag{9}$$

The asymptotic normality, variance, and bias, are known for this estimator. It can be shown that a unique AMSE minimizing threshold level exists which is a function of the parameters and number of observations. This value can be estimated by the bootstrap estimator of Danielsson and de Vries (1997*a*). In this paper we employ the simpler procedure presented in Danielsson and de Vries (1997*b*).

It is possible to use (8) and (9) to obtain estimators for out of sample quantile and probability (P, Q) combinations given that the data exhibit fat tailed distributed innovations. The properties of the quantile and tail probability estimators below follow directly from the properties of $\widehat{1/\alpha}$. In addition, the out of sample (P, Q) estimates are related in the same fashion as the in sample (P, Q) estimates.

To derive the out of sample (P, Q) estimator consider two excess probabilities p and t with $p < 1/n < t$, where n is the sample size. Corresponding to p and t are the large quantiles, x_p and x_t, where for x_i we have $1 - F(x_i) = i$, $i = t, p$. Using the expansion of $F(x)$ in (8) with $\beta > 0$ we can show that by ignoring the higher order terms in the expansion, and replacing t by M/n and x_t by the $(M+1)$-th descending order statistic one obtains the estimator

$$\hat{x}_p = X_{(M+1)} \left(\frac{m}{np} \right)^{\frac{1}{\hat{\alpha}}}. \tag{10}$$

It can be shown that the quantile estimator $\hat{x}_p$ is asymptotically normally distributed. A reverse estimator can be developed as well by a similar manipulation of (8).

$$\hat{p} = \frac{M}{n} \left(\frac{x_t}{x_p} \right)^{\hat{\alpha}}. \tag{11}$$

The excess probability estimator $\hat{p}$ is also asymptotically normal distributed.

B Figures

Table 8: Summary Statistics. Jan. 27 1984 to Dec. 31, 1996.

	JPM	MMM	MCD	INTC	IBM	XRX	XON
Mean	0.05	0.04	0.07	0.09	0.01	0.04	0.05
S.D.	1.75	1.41	1.55	2.67	1.62	1.62	1.39
Kurtosis	100.28	68.07	8.36	5.88	25.71	16.44	49.23
Skewness	−2.70	−3.17	−0.58	−0.36	−1.08	−1.06	−1.74
Minimum	−40.56	−30.10	−18.25	−21.40	−26.09	−22.03	−26.69
Maximum	24.63	10.92	10.05	23.48	12.18	11.67	16.48

JPM = J. P. Morgan; MMM = 3M; MCD = McDonalds; INTC=Intel; IBM=IBM;XRX=Xerox; XON = Exxon. Source DATASTREAM.

Table 9: Summary Statistics. Jan. 2 1990 to Dec. 31, 1996.

	JPM	MMM	MCD	INTC	IBM	XRX	XON
Mean	0.05	0.04	0.05	0.15	0.03	0.06	0.04
S.D.	1.45	1.19	1.48	2.34	1.72	1.60	1.12
Kurtosis	1.83	3.78	1.51	2.86	6.67	9.46	1.10
Skewness	0.28	−0.32	0.05	−0.36	0.25	−0.35	0.11
Minimum	−6.03	−9.03	−8.70	−14.60	−11.36	−15.63	−4.32
Maximum	6.70	4.98	6.27	9.01	12.18	11.67	5.62

Table 10: Correlation Matrix. Jan. 27 1984 to Dec. 31, 1996.

	JPM	MMM	MCD	INTC	IBM	XRX	XON
JPM	1.00						
MMM	0.49	1.00					
MCD	0.42	0.44	1.00				
INTC	0.30	0.36	0.29	1.00			
IBM	0.38	0.42	0.34	0.40	1.00		
XRX	0.35	0.39	0.34	0.32	0.35	1.00	
XON	0.44	0.48	0.37	0.24	0.35	0.30	1.00

Table 11: Correlation Matrix. Jan. 2 1990 to Dec. 31, 1996.

	JPM	MMM	MCD	INTC	IBM	XRX	XON
JPM	1.00						
MMM	0.28	1.00					
MCD	0.28	0.28	1.00				
INTC	0.24	0.21	0.21	1.00			
IBM	0.18	0.19	0.19	0.32	1.00		
XRX	0.23	0.23	0.22	0.21	0.19	1.00	
XON	0.20	0.25	0.21	0.12	0.10	0.12	1.00

Table 12: 10 Years, 2600 Daily Returns, 1987 - 1996. With Predicted Maximum Daily Drop in one Year (250 days)

							upper tail				lower tail			
Stock Index	mean	var	kurtosis	max	min	skew	α	M	$X_{(M+1)}$	max	α	M	$X_{(M-1)}$	max
Hang Seng	0.06	2.7	144.5	8.9	−40.5	−6.5	3.5	32	3.5	3.6%	2.2	49	−3.2	−3.3%
Straights Times	0.03	1.5	64.4	11.5	−23.4	−3.7	3.1	36	2.5	2.6%	2.2	59	−2.3	−2.3%
Word	0.03	0.6	25.7	7.9	−10.0	−1.4	3.5	37	1.6	1.7%	3.1	44	−1.6	−1.7%
DAX	0.03	1.4	12.2	7.3	−13.7	−1.1	2.9	52	2.3	2.3%	2.6	43	−2.7	−2.8%
FT All Share	0.03	0.7	25.9	5.7	−12.1	−2.0	2.9	58	1.5	1.5%	3.1	86	−1.3	−1.3%
SP-500	0.04	1.0	115.8	8.7	−22.8	−5.1	3.8	26	2.3	2.4%	2.5	51	−1.9	−1.9%
Miscellaneous Assets														
Gold Bulion	0.00	0.5	7.6	3.6	−7.2	−1.0	4.8	16	2.2	2.2%	3.0	33	−1.9	−1.9%
US Bonds	0.00	0.9	73.5	17.8	−10.7	1.7	2.4	86	1.3	1.4%	2.5	79	−1.4	−1.5%
US Stocks														
JPM	0.03	3.3	106.9	24.6	−40.6	−3.1	3.5	31	4.2	4.3%	3.1	48	−3.2	−3.3%
MMM	0.04	2.1	72.5	10.9	−30.1	−3.6	4.5	29	3.4	3.5%	2.4	52	−2.8	−2.9%
MCD	0.06	2.5	6.6	10.0	−18.3	−0.7	5.2	22	4.1	4.2%	3.0	45	−3.3	−3.4%
INTC	0.13	6.8	5.1	23.5	−21.4	−0.5	4.7	29	6.3	6.6%	2.8	37	−6.2	−6.4%
IBM	0.01	2.9	23.5	12.2	−26.1	−1.2	3.2	28	4.3	4.5%	2.9	38	−3.8	−3.9%
XRX	0.03	2.6	16.9	11.7	−22.0	−1.2	3.6	29	4.0	4.1%	2.7	50	−3.3	−3.4%
XON	0.04	2.0	56.7	16.5	−26.7	−2.0	3.5	34	3.1	3.2%	2.7	70	−2.3	−2.4%

References

Butler, J. S. and Barry Schachter. 1996. “Improving Value-at-Risk estimates by combining kernel estimation with historical simulation.”. mimeo, Vanderbilt University and Comptroller of the Currency.

Christoffersen, Peter F. and Francis X. Diebold. 1997. “Are volatility dynamics relevant for risk management.”. mimeo, Department of Economics, University of Pennsylvania.

Danielsson, Jon and Casper G de Vries. 1997*a*. “Beyond the sample: Extreme quantile and probability estimation.”. mimeo, Tinbergen Institute Rotterdam.

Danielsson, Jon and Casper G. de Vries. 1997*b*. “Tail index and quantile estimation with very high frequency data.” *Journal of Empirical Finance* 4:241–257.

Danielsson, Jon, L. de Haan, L. Peng and C.G. de Vries. 1997. “Using a bootstrap method to choose the sample fraction in tail index estimation.”. Tinbergen Institute Rotterdam, TI 97-016/4.

Danielsson, Jon, Philipp Hartmann and Casper G. de Vries. 1997. “The cost of conservativeness: Extreme returns, Value-at-Risk, and the Basle ’Multiplication Factor’.”. mimeo, London School of Economics.

Dave, Rakhal D. and Gerhard Stahl. 1997. “On the accuracy of VaR estimates based on the varianvce-covariance approach.”. mimeo, Olsen & Associates.

de Haan, L., S. I. Resnick, H. Rootzen, and C. G. de Vries. 1989. “Extremal behavior of solutions to a stochastic difference equation with applications to ARCH processes.” *Stochastic Process and their Applications* 32:213–224.

Dimson, Elroy and Paul Marsh. 1996. “Stress tests of capital requirements.”. mimeo, London Business School.

Duffie, Darrell and Jun Pan. 1997. “An overview of Value-at-Risk.” *The Journal of Derivatives* Spring 1997:7–49.

Fama, E. T. and M. H. Miller. 1972. *The theory of finance.* Dryden Press.

Feller, W. 1971. *An introduction to probability theory and its applications.* Vol ii. 2nd. ed. ed. New York: Wiley.

Hall, P. 1990. “Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems.” *Journal of Multivariate Analysis* 32:177–203.

Hendricks, Darryll. 1996. “Evaluation of Value-at-Risk models using histrorical 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.

Hols, M. C. and Casper G. de Vries. 1991. “The limiting distribution of extremal exchange rate returns.” *J. Appl. Econometrics* 6:287–302.

Jackson, Patricia, David J. Maude and William Perraudin. 1997. “Bank capital and Value-at-Risk.” *Journal of Derivatives* .

Jorion, Philippe. 1997. *Value-at-Risk.* Irvin.

J.P. Morgan. 1995. *RiskMetrics-technical manual.* Third ed.

Longin, F. M. 1997. "From Value-at-Risk to stress testing: the extreme value approach.". CERESSEC working paper 97-004.

Mahoney, James M. 1996. "Forecast biases in Value-at-Risk estimations: Evidence from foreign exchange and global equity portfolios.". mimeo, Federal Reserve Bank of New York.

Muller, Michel. M. Dacorogna Ulrich. A., Olivier. V. Pictet and Casper G. de Vries. 1995. "The distribution of extremal foreign exchange rate returns in extremely large data sets.". Tinbergen Institute Working Paper, TI 95-70.

Stahl, Gerhard. 1997. "Three cheers." *Risk* 10:67–69.