Forecasting Extreme Financial Risk: A Critical Analysis of Practical Methods for the Japanese Market

Jón Daníelsson and Yuji Morimoto

The various tools for risk measurement and management, especially for value-at-risk (VaR), are compared, with special emphasis on Japanese market data. Traditional Generalized Autoregressive Conditional Heteroskedasticity (GARCH)-type methods are compared to extreme value theory (EVT). The distribution of extremes, asymmetry, clustering, and the dynamic structure of VaR all count as criteria for comparison of the various methods. We find that the GARCH class of models is not suitable for VaR forecasting for the sample data, due to both the inaccuracy and the high volatility of the VaR forecasts. In contrast, EVT forecasting of VaR resulted in much better VaR estimates, and more importantly, the EVT forecasts were considerably more stable, enhancing their practical applicability for Japanese market risk forecasts.

Key words: Risk; Regulation; Extreme value theory; Volatility; Value-at-risk

Jón Daníelsson: Financial Markets Group, London School of Economics
(E-mail: j.danielsson@lse.ac.uk)
Yuji Morimoto: Morgan Stanley Dean Witter (E-mail: yuji.morimoto@alum.mit.edu)

The paper was originally written while the first author was a visiting scholar at the Bank of Japan and the second author was working for the Bank of Japan. We thank the Bank of Japan for its hospitality. The paper was presented at a Bank of Japan conference, "New Approaches to Financial Risk Management," in January and February 2000, where it was very well received. Two anonymous referees gave it excellent reviews. Views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Japan or Morgan Stanley Dean Witter. The first author's papers can be downloaded from his homepage at cep.lse.ac.uk/~ond/research or www.risk.is.

I. Introduction

The financial industry, both private firms and regulators, has become increasingly aware of the impact of risk in tradable assets. There are many reasons for this: deregulation makes risk-taking activities available to banks; technology both fosters risk-taking and makes the measurement of risk more accurate; and increasing competition means banks need to engage in increasingly risky activities simply to stay competitive. As a result, market risk measurement and management, which, until recently, was an arcane part of the banking business, has been thrust into the forefront of issues facing bankers and regulators alike. In response to this, supervisory authorities require banks to practice risk management, and report risk measures to them. In addition to publicly mandated risk measuring, many banks choose to measure and manage risk internally. Banks' approach to risk management ranges from a reluctant minimum compliance to regulations to comprehensive internal risk management programs. Since market risk management is a recent phenomenon, innate banking conservatism, in many cases, hinders the adoption of modern market risk methods. These techniques have been, to a large extent, developed in the U.S.A., where banks have, by and large, the best risk management systems in the world, undoubtedly underpinning their preeminent role in global finance.

In the analysis of risk management, one has to distinguish between external and internal practices. Banks in all major financial centers are required to comply with the so-called 'Basel rules,' regulations proposed by the Basel Committee on Banking Supervision (1996), which consists of members from Central Banks and other supervisory authorities, such as the Japanese Financial Services Agency (FSA). The grist of these regulations is the use of internal models by banks, i.e. banks model risk internally and report the outcome to the regulators in the form of Value-at-Risk (VaR). VaR is the minimum amount of losses on a trading portfolio over a given period of time with a certain probability. VaR has to be reported daily. While the VaR measure has been rightly criticized by risk managers for being inadequate, it bridges the gap between the need to measure risk accurately and for non-technical parties to be able to understand the risk measure. As such, it is the minimum technical requirement from which other, more advanced, measures are derived. In addition, while VaR, as a risk measure, may be inadequate, at least it serves to force recalcitrant banks to practice minimum risk management. The concept of VaR does have several shortcomings, the main one being that it is only the minimum amount of losses, while expected losses are more intuitive. For example, if the VaR is ¥1 billion, then we do not know if maximum possible losses are ¥1.1 billion, or ¥10 billion. In response, other measures have been proposed, e.g. expected shortfall, defined as the expected loss, conditional on exceeding the threshold. For a discussion on the properties of risk models and their application to regulatory capital, see Daníelsson (2000a). Interestingly, the criticism that current VaR measures, being excessively volatile, result in excessive fluctuations in bank capital, is not valid in practice. While it is unquestionably true that most VaR measures are excessively volatile, this has little or no regulatory impact on financial institutions. The reason is that regulatory VaR is a loss that happens more than twice a year, and a financial institution that cannot handle that loss is faced with more serious problems than fluctuating regulatory capital. While regulatory capital is three or four times the 10-day 99 percent VaR level, capital for most banks is of an order of magnitude higher than that. This is both due to the fact that credit risk capital is much larger than market risk capital, and also that a bank that has the minimum market risk capital will be viewed as a very high risk by prospective clients. For example, in 1996, the average reported daily VaR by the JP Morgan Bank was US$36 million implying regulatory capital of US$340 million. This surprisingly small amount is only a small fraction of JP Morgan's capital in 1996.

Internal risk management is a different issue. Here, banks are able to employ a wide variety of techniques to deal with the myriad of risk management issues. For example, capital has to be allocated to various risky activities, fund managers and traders have to be monitored, etc. As a result, internal requirements for financial risk measurement are more complex and diverse than regulatory requirements. For example, risk measurement serves diverse managerial purposes, ranging from integrated risk management to allocation of position limits. When risk measurement methods are used to allocate position limits to individual traders, or set mandate letters for fund managers, high volatility of risk measures is a serious problem, because it is very hard to manage individual positions with highly volatile position limits.

Fundamentally, all statistical risk measuring techniques fall into one of three categories, or a hybrid thereof: fully parametric methods based on modelling the entire distribution of returns, usually with some form of conditional volatility; the nonparametric method of historical simulation; and parametric modelling of the tails of the return distribution. All these methods have pros and cons, generally ranging from easy and inaccurate to difficult and precise. No method is perfect, and usually the choice of a technique depends on the market in question, and the availability of well-trained financial engineers to implement the procedures.

Below, we compare some of these procedures, with a special focus on extreme value theory (EVT) and the issue of dependence. In Chapter II, we discuss the general properties of financial returns, and how they relate to risk management, along with a brief review of common risk methods. Then, in Chapter III, we present an extensive discussion of EVT from a risk perspective, examining the pros and cons of that method. After that, Chapter IV contains a discussion on extensive empirical results. Mathematical derivations are contained in Appendix.

II. Distribution of Returns and Risk Forecasting

In order to predict risk, one needs to model the dynamic distribution of prices. However, even though financial practitioners usually prefer to work with the concepts of profit and loss (P/L), it is not well-suited for risk management, with returns being a preferred measure. There are two equivalent ways to calculate returns.

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad \text{and} \tag{1}$$

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right). \tag{2}$$

The compound returns in (2) are generally preferred for risk analysis, both due to their connection with common views of the distribution of process and returns, as well as the link with derivatives pricing, which generally depends on (2). In empirical analysis of the empirical properties of returns on liquid traded assets, three properties, or stylized facts, emerge which are important from a risk perspective:

1. non-normality of returns,
2. volatility clustering, and
3. asymmetry in return distributions.

In addition, the volatility of risk forecasts is problematic in practical risk applications. We address each issue in turn.

A. Distribution of Returns

1. Non-normality and fat tails

The fact that returns are not normally distributed is recognized both by risk managers and supervisory authorities:

> ".....as you well know, the biggest problem we now have with the whole evolution of risk is the fat-tail problem, which is really creating very large conceptual difficulties. Because, as we all know, the assumption of normality enables us to drop off the huge amount of complexity in our equations... Because once you start putting in non-normality assumptions, which is unfortunately what characterizes the real world, then these issues become extremely difficult."
> Alan Greenspan (1997)

The non-normality property implies the following for the relative relationship between the return distribution and a normal distribution with the same mean and variance:

1. the center of the return distribution is higher,
2. the sides of the return distribution are lower, and
3. the tails of the return distribution are higher.

This implies that the market is either too quiet or too turbulent relative to the normal. Of these, the last property is most relevant for risk. The fat-tailed property results in large losses and gains being more frequent than predicted by the normal.

An assumption of normality for the lower tail is increasingly inaccurate, the farther into the tail one considers the difference. For example, if one uses the normal distribution to forecast the probability of the 1987 crash by using the preceding years' data and the normal, one roughly estimates that a crash of the '87 magnitude occurs once in the history of Earth. Most financial analysis, until recently, has been based on the assumption of normality of returns. The reason for that is the mathematical tractability of the normal, since non-normal distribution is very difficult to work with. While normality may be a relatively innocuous assumption in many applications, in risk management it is disastrous, and non-normality needs to be addressed.

2. Volatility clustering

The second stylized fact is that returns go through periods when volatility is high and periods when it is low. This implies that, when one knows that the world is in a low volatility state, a reasonable forecast is for low financial losses tomorrow. Most risk models will take this into account, usually with a form of the GARCH model (see Bollerslev [1986]). The reason for the prevalence of the GARCH model is that it incorporates the two main stylized facts about financial returns, volatility clustering and unconditional non-normality. The popular RiskMetrics method is a restricted GARCH model where $\omega=0$ and $\alpha+\beta=1$. The most common form of the GARCH model is the GARCH(1,1):

$$r_t \mid r_{t-1} \sim N\left(\mu, \sigma_t^2\right) \tag{3}$$

$$\sigma_t^2 = \omega + \alpha\sigma_{t-1}^2 + \beta r_{t-1}^2,\ \omega,\ \alpha,\ \beta > 0,\ \alpha + \beta < 1. \tag{4}$$

One of the appealing properties of this model is that, even if the conditional distribution is normal, the unconditional distribution is not normal. This model will perform reasonably well in forecasting common volatility, e.g. in the interior 90 percent of the return distribution. However, as the normal GARCH model implies that the thickest tails of the conditional forecast distribution are only as fat as the normal with highest volatility forecast, they are still normal. This, coupled with the fact that the parameters of the GARCH model are estimated using all data in the sample equally weighted, and that, say, only 1 percent of events are in the tails, results in the estimation being driven by common events. Furthermore, because of the functional form of the normal distribution, the tail observations have lower weight than the center observations, and hence contribute too little to the estimation. This is the same reason why the Kolmogorov-Smirnoff test is not robust in testing for non-normality in fat-tailed data. Note that the JP Morgan RiskMetrics technique (J.P. Morgan [1995]) is based on a restricted form of the GARCH model.

For this reason, a GARCH model with Student-*t* innovations is sometimes used for risk.

$$\frac{r_t \mid r_{t-1}}{\sigma_t} \sim t_{(v)} \tag{5}$$

where σ_t is as in (4), and the degrees of freedom parameter, ν, is estimated along with the other parameters. Since the Student-*t* is fat-tailed (see Chapter III), it has much better properties for risk forecasting. There are three reasons why the Student-*t* is not ideal for risk forecasting:

1. its tails generally have a different shape than the return distribution,
2. it is symmetric, and
3. it does not have a natural multivariate representation.

As a result, while the GARCH Student-*t* will, in most cases, provide much better risk forecasting than a normal GARCH, its limitations imply that the model should be used with care.

3. Extreme clustering

While it is a stylized fact that returns exhibit dependence, it is less established whether extremes are clustered. We suggest the following definition:

Definition 1 (Extreme clustering) *Data are said to have extreme clusters if the arrival time between extreme events is not i.i.d.*

Furthermore, data with extreme clustering may have the following property:

Hypothesis 1 *Dependence in time between quantiles decreases with the lower probability of the quantiles.*

The truthfulness of the hypothesis depends on the underlying distribution of returns. Hence, the validity of the hypothesis will, of course, crucially affect the choice of optimal risk forecast technique. For example, Hypothesis 1 is true for the popular GARCH model. As we argue below, this may imply that, for very low probability quantiles, one should use an unconditional risk forecast technique, even if the true data generating process is GARCH. In addition, the rate at which dependence decreases will affect the modelling choice. If the rate is rapid, then unconditional risk forecast methods are preferred, but if the rate is slow, long memory models should be used. Since the GARCH model, and most related models, fall into the rapid decline class, unconditional models are preferred to GARCH models.

We propose a test for extreme dependence in order to aid the modelling decision. Specifically, we decide on a threshold, and count the number of consecutive non-exceedances between two exceedances. If the observations exceed the threshold independently, the counted numbers are geometrically distributed. If the sample counted numbers are far from geometric distribution, it is reasonable evidence of extreme clustering, and can be tested by the χ^2 goodness of fit test.

Specifically, if r_t is the return, then the statistic ϕ_n, defined as

$$\phi_t = \begin{cases} 0 & \text{if } r_t > \lambda \\ 1 & \text{if } r_t \le \lambda \end{cases} \quad \lambda >> 0, \tag{6}$$

becomes increasingly more i.i.d. as λ increases.

The data used are the TOPIX index, oil price index WTI, SP-500 index, JPY/USD exchange rate, and Tokyo Stock Exchange Second section index TSE2. In addition, various subsets of the data were used. See Table 1 for summary statistics. We tested the independence of extremes for returns, as well as the residuals of normal and Student-*t* GARCH forecasts. The thresholds chosen were 5 percent, 2.5 percent, 1 percent and 0.5 percent for each tail. These results are presented in Tables 2-4.

Table 2 shows the dependence in the returns. As expected, at the lower probability levels (5 percent, 2.5 percent), there is significant dependence, but at the 0.5 percent level, the data are mostly independent. Interestingly, at the 1 percent level, for the lower tail there is dependence, but not for the upper tail. This is consistent with the fact that the upper tail is usually thinner than the lower tail. The longest datasets have dependence at the 0.5 percent level. These results indicate that dependence decay is slow, suggesting that long memory models may be appropriate for risk forecasting.

Table 3 shows the results from testing for dependence of residuals in normal GARCH residuals, and Table 4 does the same for a Student-*t* GARCH model. As expected, there is much less dependence in residuals. However, for some of the datasets, the GARCH models fail to eliminate dependence. Even worse, if the extreme dependence is not of the GARCH type, the residuals may exhibit spurious extreme dependence.

Table 1 Summary Statistics

Data	From	To	Obs.	Mean	S.D.	Skew.	Kurt.	AC(1)
TOPIX	08/01/49	07/30/99	14,179	0.032%	0.88%	–0.44	19.30	0.16
TOPIX	08/01/49	07/31/59	3,007	0.052%	0.90%	–0.12	12.41	0.25
TOPIX	08/01/59	07/31/69	3,006	0.021%	0.72%	–0.36	6.46	0.16
TOPIX	08/01/69	07/31/79	2,896	0.039%	0.69%	–1.44	17.28	0.22
TOPIX	08/01/79	07/31/89	2,803	0.063%	0.49%	–2.41	70.62	0.10
TOPIX	08/01/89	07/30/99	2,467	–0.023%	1.25%	0.30	7.81	0.11
WTI	06/01/83	07/30/99	4,021	–0.010%	2.45%	–1.38	28.87	0.01
SP500	08/01/49	07/30/99	12,632	0.035%	0.85%	–1.78	50.89	0.10
SP500	08/01/49	07/31/59	2,518	0.055%	0.71%	–0.83	10.27	0.10
SP500	08/03/59	07/31/69	2,515	0.017%	0.63%	–0.50	12.82	0.16
SP500	08/01/69	07/31/79	2,543	0.005%	0.86%	0.29	5.45	0.24
SP500	08/01/79	07/31/89	2,529	0.048%	1.09%	–3.75	83.59	0.06
SP500	08/01/89	07/30/99	2,527	0.053%	0.88%	–0.50	9.31	0.01
JPY/USD	08/01/79	07/30/99	5,093	–0.012%	0.71%	–0.81	10.65	0.02
JPY/USD	08/01/79	07/31/89	2,489	–0.018%	0.66%	–0.39	5.48	0.04
JPY/USD	08/01/89	07/30/99	2,604	–0.007%	0.76%	–1.06	13.20	0.00
TSE2	08/01/69	07/30/99	8,166	0.031%	0.70%	–0.79	13.10	0.43
TSE2	08/01/69	07/31/79	2,896	0.049%	0.63%	–1.02	13.96	0.47
TSE2	08/01/79	07/31/89	2,803	0.051%	0.54%	–1.76	30.21	0.32
TSE2	08/01/89	07/30/99	2,467	–0.012%	0.91%	–0.31	7.29	0.45

Table 2 Extreme Dependence in Returns

Data	Lower tail 5%	2.5%	1%	0.5%	Upper tail 5%	2.5%	1%	0.5%
TOPIX 49-99	***	***	***	***	***	***	***	***
TOPIX 49-59	***	***	***		***	***	***	***
TOPIX 59-69	***	***	***		***	***		
TOPIX 69-79	***	***	***		***	***		
TOPIX 79-89	***	***	***	***	***	***	**	***
TOPIX 89-99	***	***	***		***	**		
WTI 83-99	***	***	***	***	***	***	***	***
SP500 49-99	***	***	***	***	***	***	***	***
SP500 49-59		**						
SP500 59-69	***	***	***		***	**	**	
SP500 69-79	***	***	***	***	***	***	***	***
SP500 79-89	***	***	***		***	***	***	***
SP500 89-99	***	***	***		***	***		
JPY/USD 79-99	***	***	***		***	***		
JPY/USD 79-89	***	***	***	***	***	***	**	***
JPY/USD 89-99	***	***	***		***	**		
TSE2 69-99	***	***	***	***	***	***	***	***
TSE2 69-79	***	***	***	***	***	***	***	***
TSE2 79-89		**						
TSE2 89-99	***	***	***		***	**	**	

Notes: ***indicates significance at the 1% level, **at the 2.5% level and *at the 5% level.

Table 3 Extreme Dependence in Normal GARCH Residuals

Data	Lower tail 5%	2.5%	1%	0.5%	Upper tail 5%	2.5%	1%	0.5%
TOPIX 49-99	**				**			**
TOPIX 49-59					***	**		
TOPIX 59-69	***							
TOPIX 69-79								
TOPIX 79-89								
TOPIX 89-99							**	
WTI 83-99								
SP500 49-99	***							
SP500 49-59								
SP500 59-69					*	*		
SP500 69-79	**				**			**
SP500 79-89					***	**		
SP500 89-99	***							
JPY/USD 79-99								
JPY/USD 79-89								
JPY/USD 89-99							**	
TSE2 69-99								
TSE2 69-79	***							
TSE2 79-89								
TSE2 89-99					*	*		

Notes: ***indicates significance at the 1% level, **at the 2.5% level and *at the 5% level.

Table 4 Extreme Dependence in Student-*t* GARCH Residuals

	Lower tail				Upper tail			
Data	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%
TOPIX 49-99	**				**	**		**
TOPIX 49-59					***	***		
TOPIX 59-69	***							
TOPIX 69-79								
TOPIX 79-89								
TOPIX 89-99							**	
WTI 83-99								
SP500 49-99	***							
SP500 49-59								
SP500 59-69					**	*		
SP500 69-79	**				**			**
SP500 79-89								
SP500 89-99							**	
JPY/USD 79-99								
JPY/USD 79-89	**							
JPY/USD 89-99								
TSE2 69-99	***	**			***			
TSE2 69-79	**							
TSE2 79-89	**	**			**			
TSE2 89-99	*				**			

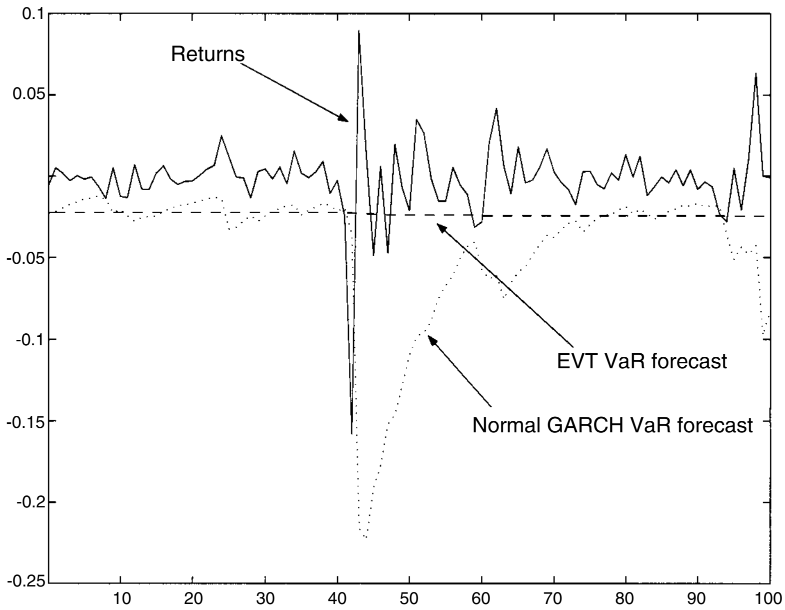
Notes: ***indicates significance at the 1% level, **at the 2.5% level and *at the 5% level.

4. Violation clustering

While the concept of extreme clustering applies to the actual data, it is also of interest to consider clustering in violations of the VaR. We propose the following definition:

Definition 2 (Violation clustering) *When the time between VaR violations is not i.i.d.*
Violation clustering depends on the data and VaR forecast method. In practice, it is probably impossible to create a VaR model that does not result in some type of violation clustering, either globally or, at least, locally. A form of local violation clustering can be seen in Figure 1, where we forecast VaR for the TOPIX index by both normal GARCH and EVT methods (see below for a more detailed discussion of the VaR forecasts). We see that the GARCH method lags behind the returns, but overcompensates by producing an extremely large VaR forecast. The EVT method is hardly affected by the event. The overly sharp reaction in the GARCH model, combined with a rapid decrease, suggests that the long memory models would be much preferable to the GARCH model, and in the absence of a long memory model, the EVT method is preferred.

Figure 1 Risk Overshooting in TOPIX VaR Forecast



5. Asymmetry

One feature of conditional volatility models, such as the GARCH model, is the implicit assumption of symmetry of the return distribution. As discussed below, this is not correct. Usually, one of the tails is fatter than the other. For example, for equities, the lower tail is commonly thicker than the upper tail. In general, if the market trend is upwards, the upper tail is thinner than the lower tail, i.e. the market moves in small steps in the direction of the trend, and in large jumps away from the trend. In the 1990s, most major stock indices have a relatively thicker lower tail, except the NIKKEI. Since the GARCH model, and its immediate extended forms, are symmetric, due to conditional normality or Student-*t* distributional assumptions, they tend to underpredict losses relative to gains, and hence, for an unsuspecting risk manager, will provide false comfort. While, in theory, it is straightforward to extend the GARCH model to take into account asymmetry, this necessarily complicates the model.

The EGARCH model by Nelson (1991) is sometimes used for volatility forecasting.

$$y_t = \sqrt{h_t} z_t, \quad z_t \sim N(0,1)$$

$$\log(h_t) = \omega \left(1 + \sum_{i=1}^{q} \alpha_i L^i \right) \left(1 - \sum_{i=1}^{p} \beta_i L^i \right)^{-1} \left\{ \theta z_{t-1} + \gamma \left[|z_{t-1}| - E|z_{t-1}| \right] \right\},$$

where θ and γ are the parameters of asymmetry, and L^i is the lag operator. The advantage of this model is, primarily, its relation to stochastic volatility models. Hence, it provides a link to continuous time finance. For risk applications, the primary advantage is the explicit asymmetry, where risk forecasts depend on the direction of lagged marked movements. However, typically, when this model is estimated, the asymmetry parameters are non-significant, suggesting that this model is misspecified. As a result, in risk applications, this model does not have any special advantage over the GARCH model, and since it is more complex than the GARCH, it cannot be recommended for risk forecasting.

The best way to forecast risk would be to use skewed conditional distributions. This, however, is not commonly done, due to the difficulty of finding an appropriate skewed distribution and estimating its parameters.

6. VaR volatility

One feature of conditional volatility models, such as the GARCH, is that the volatility of risk forecasts is very high. Table 5 shows some sample statistics of predicted VaR numbers for the first two quarters of 1999 for the TOPIX index. We see that, in a portfolio of ¥1 billion, a normal GARCH model predicts VaR ranging from ¥18 million to ¥41 million. Since market risk capital is 3 × VaR, this would imply very large fluctuations in capital. If financial institutions actually set capital at this level, it would imply very high financing costs, and most likely result in the institution keeping a higher capital level than necessary. The reason why this is largely irrelevant, from a regulatory point of view, is that the regulatory VaR is typically of an order of magnitude higher than required. One area where this problem is not academic is in internal risk management, where VaR may be used, e.g. to set limits for traders. Since frequently fluctuating VaR limits would result in high variability in the size of positions, this would be considered unacceptable in most cases. As a result, in practice, most banks would use various techniques to dampen VaR volatility. In many cases, the covariance matrix is updated at infrequent intervals, e.g. once every three months. This, of course, usually results in large jumps in the limits (which are derived from the covariance matrix) four times a year, which in itself creates problems. Alternatively, a dampening function on the covariance matrix could be used, e.g. a moving average. A better way would be to use a long memory model for volatility. In this case, the variance is neither stationary nor non-stationary, implying a fractal dimension. The advantage is that these models provide persistence in shocks that falls between the extremely short horizon of the GARCH-type model and the infinite persistence of $I(1)$ models, like RiskMetrics. Long memory models are, however, notoriously difficult to work with, and have some way to go until they can be employed in regular risk management. Another common method is to dispense with conditional volatility-based methods altogether for the setting of limits, and perhaps use historical simulation or extreme value theory. See Daníelsson (2000b) for more discussion on risk volatility.

Table 5 VaR Volatility in the TOPIX Index (Percentage Returns)

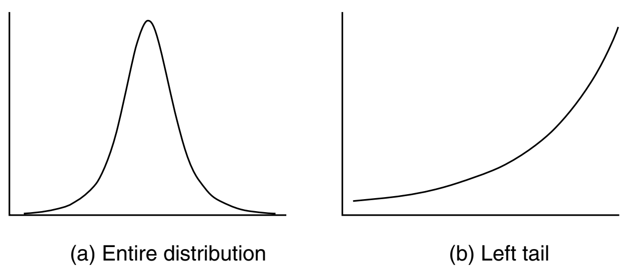
Name	From	To	Mean	S.E.	Min.	Max.	Violations
GARCH normal	1954/07/29	1999/07/30	–1.71	1.04	–28.28	–0.59	1.65%
GARCH-*t*	1954/07/29	1999/07/30	–1.86	1.09	–22.30	–0.65	1.25%
EVT	1954/07/29	1999/07/30	–2.20	0.70	–3.94	–1.03	1.17%
GARCH normal	1996/01/04	1999/07/30	–2.64	0.91	–7.56	–1.43	1.48%
GARCH-*t*	1996/01/04	1999/07/30	–2.90	0.92	–7.13	–1.75	0.91%
EVT	1996/01/04	1999/07/30	–3.25	0.21	–3.94	–2.90	0.91%
Returns	1999/01/04	1999/03/31	0.26	1.18	–2.41	3.70	
GARCH normal	1999/01/04	1999/03/31	–2.59	0.64	–4.12	–1.80	
GARCH-*t*	1999/01/04	1999/03/31	–2.79	0.61	–4.21	–2.04	
EVT	1999/01/04	1999/03/31	–3.18	0.03	–3.31	–3.09	
Returns	1999/04/01	1999/06/30	0.18	1.00	–2.01	3.24	
GARCH normal	1999/04/01	1999/06/30	–2.57	0.37	–3.57	–2.02	
GARCH-*t*	1999/04/01	1999/06/30	–2.82	0.40	–3.84	–2.23	
EVT	1999/04/01	1999/06/30	–3.17	0.03	–3.19	–3.09	

Note: These are summary statistics of VaR predictions.

III. Extreme Value Theory

In common statistical methods, such as GARCH, all observations are used in the estimation of a forecast model for risk, even if only one in 100 events is of interest. This, obviously, is not an efficient way to forecast risk. The basic idea behind extreme value theory (EVT) is that, in applications where one only cares about large movements in some random variable, it may not be optimal to model the entire distribution of the event with all available data. Instead, it may be better only to model the tails with tail events. For example, an engineer designing a dam is only concerned with the dam being high enough for the highest waters. Regular days when the waters are average simply do not matter. Extreme value theory is a theory of the behavior of large, or extreme, movements in a random variable, where extreme observations are used to model the tails of a random variable (see Figure 2). EVT has been widely used in diverse fields, such as engineering, physics, chemistry, and insurance. However, it is only recently that it has been applied to financial risk modelling.

Figure 2 The Tail



A. Theoretical Background

Classification of tails of distribution is often arbitrary. For example, a high kurtosis is, perhaps, the most frequently used indication of fat tails. This, however, is an incorrect use of kurtosis, which measures the overall shape of the distribution. One way to demonstrate this is by Monte Carlo experiments. We generated repeated samples of size 2000 from a known fat-tailed distribution, the Student-t (3). For each sample, we excluded the largest and smallest 40 values. Our random sample was, hence, truncated from above and below, and clearly thin-tailed. However, the average excess kurtosis was 7.1, falsely indicating fat tails. In addition, various models are frequently labeled as fat-tailed, e.g. the conditionally normal stochastic volatility (SV) model. It can be shown that, according to the criteria below, the SV model is thin-tailed. Note, in contrast, that the GARCH model is fat-tailed, since the return process feeds back to the volatility process.

Formally, random variables fall into one of three tails shapes, fat, normal, and thin, depending on the various properties of the distribution.

1. The tails are thin, i.e. the tails are truncated. An example of this is mortality.
2. The tails are normal. In this case, the tails have an exponential shape. The most common member is the normal distribution.
3. The tails are fat. The tails follow a power law.

It is a stylized fact that financial returns are fat. Hence, we only need to consider the third case. An extremely important result in EVT is that the upper tail[1] of any fat-tailed random variable (x) has the following property:

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

where α is known as the tail index, and $F(\cdot)$ is the asymptotic distribution function. The reason why this is important is that, regardless of the underlying distribution of x, the tails have the same general shape, where only one parameter is relevant, i.e. α. If the data are generated by a heavy-tailed distribution, then the distribution has, to a first order approximation, a Pareto-type tail:

$$P\{X > x\} \approx ax^{-\alpha}, a > 0, \ \alpha > 0, \tag{8}$$

as $x \to \infty$. Daníelsson and de Vries (1997b) and Daníelsson and de Vries (1997a) demonstrate that the distribution of the lower tail is

$$F(x) \cong \frac{m}{n}\left(\frac{x}{X_{m+1}}\right)^{-\alpha}, \tag{9}$$

1. This applies trivially to the lower tails as well.

where n is the number of observations, m is the number of observations in the tail, and X_{m+1} is the decreasing order statistics, $X_1 \geq X_2 \geq \ldots \geq X_m \geq \ldots \geq X_n$. The tail index α is the key parameter, in that it governs how rapidly the tails go down.

- If $\alpha = \infty$, the tails decline exponentially, and the distribution is usually normal.
- If $\alpha < \infty$, the tails are fat, and in that case,
  - α is the number of finite moments,
  - for the student-t distributions, α is the degrees of freedom, and
  - for the stable distributions, α is the characteristic exponent.

B. Applications of EVT

Many applications exist for EVT in financial institutions. By estimating the tails of return distributions, a financial institution is able to obtain better forecasts of out-of-sample events, as well as better measurements of extreme in-sample quantiles. The primary application considered here is risk management, but in addition, EVT has important implications for measurement of diversification effects under heavy tails, which implies that standard portfolio theory, which is based on risk-return profiles in a mean variance framework, is not strictly correct. Also, derivatives pricing, especially for exotic options, is strongly affected by the presence of heavy tails. This is as predicted by EVT. However, risk management remains the primary use of EVT within the financial industry.

When EVT is used for VaR forecasting, the parameters of the lower tail are estimated, i.e. α and m: (9) forms the basis of the VaR forecast. Conditional on the estimates of α and m, we obtain X_{m+1}. These three variables, in addition to the sample size n, enable us to calculate any probability–quantile combination (P, Q) where $P < m/n$. Specifically, if we need to calculate $\text{VaR}^{p\%}$, we ascertain whether $p \geq m/n$. If that is the case, we use the p^{th} sample quantile from the data to get $q(p)$, i.e. historical simulation is used to provide the VaR forecast. If $p < m/n$, (9) is used to obtain the corresponding quantile $q(p)$. Knowing the quantile forecast $q(p)$, calculating the VaR is straightforward.

C. Estimation

If the tails follow the Pareto distribution is (8) exactly, i.e.

$$F(x) = 1 - ax^{-\alpha}, \qquad (10)$$

estimation of the tail indexed α is straightforward. Hill (1975) shows that a maximum likelihood estimator of the tail index is

$$\frac{1}{\hat{\alpha}} = \frac{1}{m}\sum_{i=1}^{m}\log\frac{X_i}{X_{m+1}}. \qquad (11)$$

The estimator in (11) is known as the Hill estimator, m is the number of observations in the tail, and X_{m+1} is the quantile (threshold) where the tail begins. If the Pareto approximation in (10) is only asymptotically correct, the Hill estimator is a consistent method of moments estimator. For an appropriately chosen m, it is well known that this estimator currently has the best empirical properties for the tail index of financial returns. However, the choice of m is not trivial. Choosing m is the same as determining where the tail begins. Choosing m arbitrarily, e.g. as 1 percent of the sample size, is not recommended. Hall (1990) proposes a subsample bootstrap method for the determination of m. His method relies on strong assumptions of the tail shape, and Daníelsson and de Vries (1997a) propose an automatic double subsample bootstrap procedure for determination of the optimal threshold, m^*.

1. The optimal threshold

The choice of the optimal threshold m^* is very important, since the estimates of α and, hence, risk, will vary highly by changing m. The optimal choice of m^*.depends on the following results.

1. The inverse estimate of the tail index $1/\hat{\alpha}$ is asymptotically normally distributed.
2. Therefore, we can construct the asymptotic mean square error (AMSE) of $1/\hat{\alpha}$.
3. Since, in general, $1/\hat{\alpha}$ is biased and is subject to estimation variance, we need to take the following into account.

   (a) As both the variance and bias are affected by the choice of m, it is optimal to choose m where the bias and variance vanish at the same rate.

   (b) Therefore, the level where the AMSE is minimized gives us the optimal threshold level, m^*, i.e. we have

$$m^* = \min\left\{ \text{AMSE}\left[\left(\frac{1}{\hat{\alpha}} - \frac{1}{\alpha} \right)^2 \right] \right\} .$$

We obtain the AMSE by a bootstrap procedure. However, resampling of size T does not eliminate variation in the AMSE, and a subsample bootstrap is needed. Hall (1990) proposes a subsample procedure, where he assumes an initial m level to obtain an initial α, as a proxy for $1/\alpha$ in the AMSE, then obtains $1/\hat{\alpha}$ estimates for each subsample bootstrap realization, and chooses m^*_{sub} for the subsample where the average is minimized. This is then scaled up to the full sample m^* using $\hat{\alpha}_{sub}$ and the assumption of $\hat{\beta}_{sub} = \hat{\alpha}_{sub}$ where β is the parameter of the second order expansion of the limit law in (7). This however has two drawbacks:

1. a need to assume an initial α, and
2. assumption of a value for the second order parameter β (Hall argues that $\beta = \alpha$ is a good assumption, but for Student-t, $\beta = 2$).

Daníelsson and de Vries (1997a) propose an automatic procedure to determine m^*, where they use a double subsample bootstrap to eliminate reliance on the initial α and assumption of $\hat{\beta}$. The difference statistic between the Hill estimator and the Daníelsson-de Vries estimator converges at the same rate and has a known theoretical benchmark which equals zero in the limit. The square of this difference statistic produces a viable estimate of the MSE[$1/\alpha$] that can be minimized with respect to the choice of threshold. Furthermore, in order to attain the desired convergence in probability, instead of convergence of distribution that can be obtained, at most, by a full sample bootstrap, they show that one needs to create resamples of a smaller size than the original sample with a subsample bootstrap technique. The reason is that two samples with different sampling properties are needed for the estimation of the second order parameter β, which is used in scaling the subsample optimal threshold up to the full sample threshold, where the Hill estimator is used to obtain the full sample α estimate.

D. Issues in Application of EVT

While, sometimes, EVT is presented as a panacea for risk management, this is not correct. There are several issues which limit the applicability of EVT to the financial sector.

1. Sample size

While there is nothing intrinsic which limits the sample size in EVT applications, in practice, there are constraints. There are primarily two types of constraints which limit the sample size from below.

1. We need to observe some events which constitute extremes, and
2. most estimation methods for the threshold, m, depend on subsample bootstrap or even double subsample bootstrap, which implies that the first constraint has to be observed in all bootstrap subsamples.

If the Hill estimator is used for the determination of m, one requires, at the absolute minimum, that the number of observations in the tail is $m+1$. However, m is estimated with a subsample procedure, where the size of the subsample should be as small as is feasible. Practical experience and theoretical results indicate that the size of the subsample should be, perhaps, 10 percent of the sample size. Therefore, if the sample size is 1000, the subsample size is 100, and on average, one will observe 50 positive values, from which to estimate the subsample values. However, in repeated bootstraps, the particular bootstrap with the smallest number of positive observations will provide the upper constraint on feasible positive values. In practice, this implies that 1,000 observations is the absolute minimum, and 1,500 observations is preferable. Having more than 6000 observations does not seem to make much of a difference. The risk manager is, therefore, in a classical Catch22 situation; statistical demands may be inconsistent with the fact that the world is changing. Actually, this is a reflection of a greater problem. In many markets, especially emerging markets, there is simply not enough data to do sensible analysis, whether by EVT or any other method. Perhaps for this reason, most published work on risk management is focused on modelling risk in highly liquid, long running data series, such as SP-500.

2. EVT is usually univariate

While multidimensional EVT is actively being developed, it suffers from the same curse of dimensionality problems as many other techniques, such as GARCH. As a result, at the time of writing, such methods have very limited applicability in risk management. There are, however, signs that this is changing, especially in the very interesting work of Longin (1999), where he measures how the covariance changes as one moves into the tails.

3. Dependence

EVT, as presented here, assumes that the data is i.i.d. This, however, is not a theoretic limitation. Resnick ans Stărică (1996) have shown that the Hill estimator is consistent under certain types of dependence, such as GARCH.

E. Extreme Value Theory Forecasting

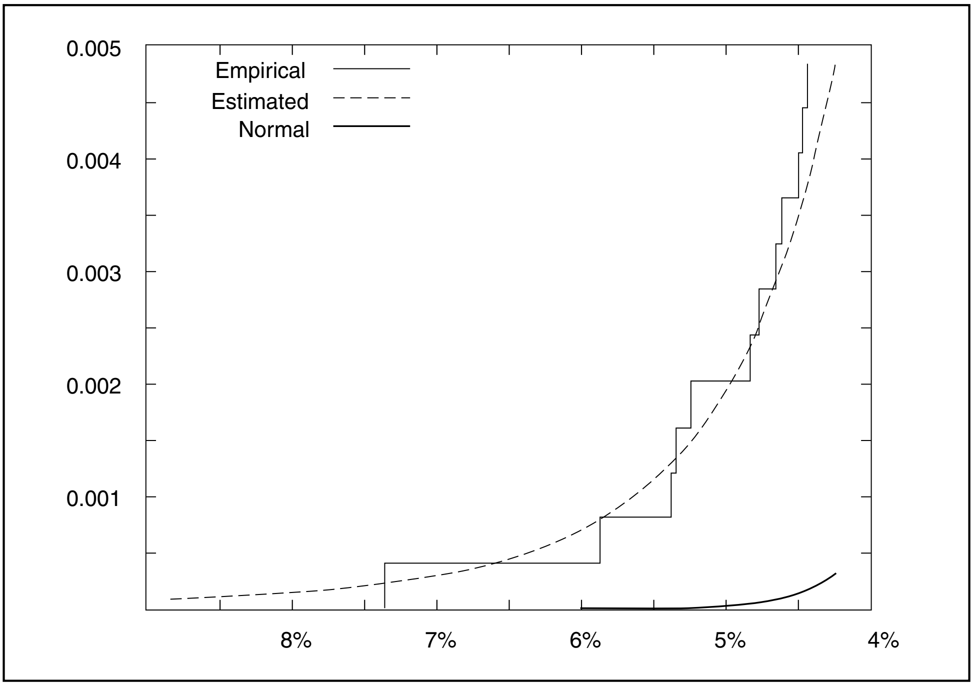
Results using EVT are presented in Table 6, and plots of the lower tail of the TOPIX in the 1990s are presented in Figure 3. EVT was used to predict the largest expected decrease and increase in each series over a 20-year interval. Of the datasets, oil prices are the most volatile, and a one-day increase of 29 percent is expected once every 20 years, with a one-day drop of 19 percent once in the same period. We note that the JPY/USD exchange rate is getting riskier over time, with a one-day increase of 9 percent. Curiously, the Second Section stocks are subject to less extreme risk than the TOPIX stocks. The main U.S.index, SP-500, is less risky than the major Japanese index, TOPIX, in the 1990s, and the TOPIX is especially risky on the downside.

Table 6 EVT Tail Estimates (Percentage Returns)

Data	From	To	Obs.	Mean	S.D.	Skew.	Kurt.	AC(1)
TOPIX	08/01/49	07/30/99	3.83	4.02	37	30	7.34	–7.37
TOPIX	08/01/49	07/31/59	3.27	4.70	32	17	6.27	–8.26
TOPIX	08/01/59	07/31/69	4.74	3.15	20	36	6.85	–4.29
TOPIX	08/01/69	07/31/79	3.44	2.68	28	41	8.89	–5.34
TOPIX	08/01/79	07/31/89	2.92	2.88	48	29	8.26	–7.63
TOPIX	08/01/89	07/30/99	4.00	5.57	22	12	7.51	–9.73
WTI	06/01/83	07/30/99	4.25	2.82	14	32	29.5	–18.9
SP500	08/01/49	07/30/99	3.89	3.25	61	163	7.18	–6.01
SP500	08/01/49	07/31/59	4.59	2.85	33	32	7.57	–4.02
SP500	08/03/59	07/31/69	2.97	3.59	76	25	5.07	–5.78
SP500	08/01/69	07/31/79	3.46	4.81	31	23	4.79	–7.13
SP500	08/01/79	07/31/89	3.97	2.80	33	56	10.4	–6.83
SP500	08/01/89	07/30/99	3.91	3.33	29	28	7.51	–6.15
JPY/USD	08/01/79	07/30/99	4.60	3.61	25	34	6.29	–4.08
JPY/USD	08/01/79	07/31/89	6.89	6.61	14	11	3.80	–3.02
JPY/USD	08/01/89	07/30/99	3.98	2.66	29	36	9.46	–4.91
TSE2	08/01/69	07/30/99	5.20	3.13	20	34	7.56	–4.35
TSE2	08/01/69	07/31/79	4.69	2.41	23	51	8.50	–3.86
TSE2	08/01/79	07/31/89	5.95	3.37	16	32	4.53	–2.81
TSE2	08/01/89	07/30/99	4.55	7.97	17	11	5.71	–5.67

Notes: α_u and α_l are the tail index for the upper and lower tails, respectively. m_u and m_l are the number of order statistics in the upper and lower tails, respectively. Max. 20 and Min. 20 are forecasts of the expected largest one-day increase and decrease in the data over 20 years in percentages.

Figure 3 Lower Tails of the CDF for TOPIX 1990-1999



IV. Empirical Analysis

The empirical analysis is based on a variety of datasets from several periods. For most parts, our empirical protocol was strictly non-data-snooping, i.e. we did not look at the data before applying our procedures.

A. VaR Predictions

The most common model for VaR forecasting is the GARCH class of models. We use the GARCH model with normal and Student-t innovations to forecast VaR. The estimation window was 1000 observations, and the model is reestimated each day. We moved the 1000-day window to the end of the sample. Each day, we forecast the next day's VaR and count the number of violations. The results from this exercise, which are presented in Table 7, will not come as a surprise to readers of VaR literature, since many authors have tested this and reached the same conclusions. The normal GARCH model has the worst performance, followed by the Student-t GARCH, with EVT the best method. These results are especially interesting in light of the results on VaR volatility below.

B. Asymmetry

Both the normal and Student-t GARCH models are based on the assumption of a symmetric conditional distribution, with no asymmetric responses of returns on the risk forecasts. As a result, these models assume that the unconditional distribution of

Table 7 VaR Violation Ratios

		Lower tail				Upper tail			
Model	Data	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%
Normal	TOPIX	4.50%	3.07%	1.98%	1.47%	4.10%	2.71%	1.59%	1.21%
	SP500	5.13%	3.40%	1.96%	1.41%	4.81%	2.93%	1.66%	1.19%
	JPY/USD	5.65%	3.84%	2.50%	1.89%	5.15%	3.23%	2.09%	1.53%
	WTI	5.08%	3.05%	1.98%	1.39%	4.80%	2.86%	1.90%	1.51%
	TSE2	4.76%	3.05%	2.04%	1.53%	5.06%	3.29%	1.86%	1.35%
GARCH	TOPIX	5.14%	3.01%	1.65%	1.25%	4.05%	2.19%	1.05%	0.63%
	SP500	5.47%	3.33%	1.81%	1.12%	4.30%	2.31%	0.99%	0.59%
	JPY/USD	5.57%	3.42%	2.12%	1.64%	4.90%	2.84%	1.50%	1.00%
	WTI	5.04%	2.98%	1.71%	1.27%	4.40%	2.66%	1.43%	1.15%
	TSE2	5.25%	3.26%	1.82%	1.26%	4.34%	2.54%	1.22%	0.72%
GARCH-*t*	TOPIX	5.80%	3.02%	1.25%	0.76%	4.44%	2.11%	0.71%	0.28%
	SP500	6.01%	3.18%	1.28%	0.60%	4.62%	2.11%	0.71%	0.34%
	JPY/USD	6.26%	3.01%	1.34%	0.92%	5.23%	2.31%	0.86%	0.19%
	WTI	6.27%	2.82%	1.15%	0.48%	4.68%	2.30%	0.99%	0.63%
	TSE2	5.39%	2.91%	1.29%	0.69%	4.76%	2.33%	0.89%	0.21%
EVT	TOPIX	5.23%	2.68%	1.18%	0.69%	5.38%	2.90%	1.30%	0.72%
	SP500	5.61%	3.11%	1.27%	0.69%	5.55%	3.12%	1.27%	0.70%
	JPY/USD	5.51%	2.98%	1.31%	0.95%	6.43%	3.03%	1.53%	0.81%
	WTI	5.08%	2.94%	0.87%	0.48%	5.32%	2.78%	1.23%	0.63%
	TSE2	5.45%	2.72%	1.23%	0.72%	5.75%	2.99%	1.38%	0.78%

Notes: Length of Test: TOPIX=12,679, SP500=11,133, JPY/USD=3,593, WTI=2,521, TSE2=6,666. Each cell contains the ratio of VaR violations to total number of observations. The correct ratio is in the top row.

returns is symmetric. This has two drawbacks for risk forecasting. First, it implies that the upper and lower tails are identical. However, this is clearly not the case. It is well known that return distribution is not symmetric, with the upper tail typically thinner than the lower tail. This is also the case with our data (see Table 6). In addition, with the normal GARCH model, the thinner tail will be more important in the estimation, due to the exponential kernel. Hence, the thicker tail, typically the lower, will have a relatively lower impact on the estimation. We see that VaR predictions with the GARCH model bear this out. There is considerable asymmetry in the GARCH results. The VaR levels in the upper tail are overpredicted, while those in the lower tail are underpredicted.

C. VaR Volatility

As discussed above, VaR volatility is of considerable concern. Table 5 shows statistics on VaR predictions for a portfolio of ¥1 billion of the various models under consideration. We first focus on the entire sample period. The normal GARCH model suggests that the largest one-day 1 percent VaR is ¥280 million, while the largest one-day loss on the TOPIX was only ¥150 million with an unconditional probability of 0.007 percent, clearly indicating the implausibility of the GARCH results. Attempting to rectify the forecast by the use of a Student-*t* GARCH model does not help much, since its largest predicted loss is ¥220 million. A completely different picture emerges from the EVT results. The VaR forecasts are stable, and do not have any of the wild VaR forecasts produced by the GARCH models. The standard error of the normal GARCH model VaR forecasts is 1.04, for the Student-*t* GARCH model 1.09, but only 0.70 for the EVT forecast.

A similar picture emerges for the first and second quarters of 1999. (Q1 and Q2) We see that the normal GARCH model predicts VaR ranging from ¥18 million to ¥41 million. A factor of three change in VaR in three months is problematic for a financial institution using VaR for either external or internal risk management, but this is a typical result from conditional volatility models. Similar results obtain for a Student-*t* GARCH model, with VaR ranging from ¥20 million to ¥42 million. In Q1 of 1999, the EVT model had VaR forecasts which ranged from ¥31 million to ¥33 million, in line with expectations.

D. Violation Clustering

We also consider VaR clustering in the forecasts from the models by use of the extreme clustering test discussed above. These results are presented in Table 8. We find that, when considering the entire dataset, there is still considerable clustering in EVT forecasts, and, as expected, the dependence in the GARCH models results in much less global clustering. However, when considering local clustering, another picture emerges. Figure 1 shows both GARCH normal and EVT forecasts of TOPIX VaR. We see that the GARCH method lags behind the returns, but overcompensates by producing an extremely large VaR forecast. The EVT method is hardly affected by the event. The overly sharp reaction in the GARCH model, combined with a rapid

Table 8 Violation Dependence

		Lower tail				Upper tail			
Model	Data	5%	2.5%	1%	0.5%	5%	2.5%	1%	0.5%
Normal	TOPIX	***	***	***	***	***	***	***	***
	SP500	***	***	***	***	***	***	***	***
	JPY/USD	***	***	**	***	**	**		*
	WTI	***	***	***	***	***	***	***	***
	TSE2	***	***	***	***	***	***	***	***
GARCH	TOPIX	***	***	*					
	SP500	***	***						
	JPY/USD								
	WTI		***						
	TSE2	***	***		*	*			
GARCH-*t*	TOPIX	***	***						
	SP500	***	**				**		
	JPY/USD								
	WTI								
	TSE2	***	*			**			
EVT	TOPIX	***	***	***	***	***	***	***	***
	SP500	***	***	***	***	***	***	***	***
	JPY/USD	***	***	***	***	***		***	
	WTI	***	***	***		***	***	***	***
	TSE2	***	***	***	***	***	***	***	***

Notes: ***indicates significance at the 1% level, **at the 2.5% level and *at the 5% level.

decrease, suggests that the long memory models would be much preferable to the GARCH model, and in the absence of a long memory model, the EVT method is preferred.

V. Conclusion

Risk management has undergone vast changes in recent years. Traditional methods, such as the GARCH class of models, have been pressed into service to provide risk forecasts, with mixed results. New techniques, such as extreme value theory (EVT), have been applied to the problem of risk forecasting. This is a dynamic field, and no single technical solution exists. One faces the classical problem of accuracy vs. complexity. In the paper, we compare and analyze the various methods for VaR forecasting of Japanese financial data.

We find that Japanese market data are well suited for advanced techniques, such as EVT. The data exhibit patterns observed in other markets, rendering risk forecasts relatively straight forward. By using EVT, we find that VaR forecasts are very accurate and stable over time. This implies that the use of EVT risk forecasting for Japanese financial institutions and other users of Japanese market data is recommended.

In contrast, we find that GARCH-type techniques are less accurate than EVT VaR forecasts, and even more worryingly, are very volatile. As a result, such models cannot be recommended for practical VaR predictions. A detailed examination of the VaR forecasts from both EVT and GARCH models demonstrated that the wild swings observed in the GARCH VaR predictions are more an artifact of the GARCH model, rather than the underlying data. Finally, we note that a long memory model would provide the best risk forecasts. However, this cannot be recommended. This is because such models are notoriously difficult to estimate and require very long estimation horizons, effectively rendering them useless for most practical risk forecasting.

APPENDIX: EXTREME VALUE THEORY

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 the results in Daníelsson and de Vries (1997a), which also provides all the proofs; the method has been applied by Daníelsson and de Vries (1997b).

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:

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

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 (A.1), is essentially the only assumption that is needed for analysis of the tail behavior of the returns x. Regular variation at infinity is a necessary and sufficient condition for distribution of the maximum or minimum 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

$$F(x)=1-ax^{-\alpha}\left[1+bx^{-\beta}+o\left(x^{-\beta}\right)\right],\quad \beta>0\quad \text{as } x\to\infty. \tag{A.2}$$

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}.

$$\frac{1}{\alpha}=\frac{1}{m}\sum_{i=1}^{m}\log\frac{X_i}{X_{m+1}} \tag{A.3}$$

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 Daníelsson and de Vries (1997a).

It is possible to use (A.2) and (A.3) 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 $1/\hat{\alpha}$. 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 (A.2) 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

$$\hat{x}_p = X_{(m+1)}\left(\frac{m}{np}\right)^{\frac{1}{\hat{\alpha}}}. \tag{A.4}$$

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 (A.2).

$$\hat{p} = \frac{m}{n}\left(\frac{x_t}{x_p}\right)^{\hat{\alpha}}. \tag{A.5}$$

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

References

Basel Committee on Banking Supervision, *Overview of the Amendment to the Capital Accord to Incorporate Market Risk,* 1996.

Bollerslev, T., "Generalised Autoregressive Conditional Heteroskedasticity," *J. Econometrics,* 51, 1986, pp. 307–327.

Daníelsson, J., "Limits to Risk Modelling," mimeo, London School of Economics, 2000a (*www.risk.is*).

———, "(Un) Conditionality and Risk Forecasting," mimeo, London School of Economics, 2000b (*www.risk.is*).

———, and C. G. de Vries, "Beyond the Sample: Extreme Quantile and Probability Estimation," mimeo, Tinbergen Institute Rotterdam, 1997a.

———, and ———, "Tail Index and Quantile Estimation with Very High Frequency Data," *Journal of Empirical Finance,* 4, 1997b, pp. 241–257.

Greenspan, A., "Discussion at Symposium: Maintaining Financial Stability in a Global Economy," at the Federal Reserve Bank of Kansas City, p. 54, 1997.

Hall, P., "Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in Nonparametric Problems," *Journal of Multivariate Analysis,* 32, 1990, pp. 177–203.

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

Longin, F., "From Value at Risk to Stress Testing: The Extreme Value Approach," Discussion Paper No. 2161, Center for Economic Policy Research, 1999.

Morgan, J. P., *Risk Metrics-Technical Manual,* Third Edition, 1995.

Nelson, D. B., "Conditional Heteroskedasticity in Asset Pricing: A New Approach," *Econometrica,* 59, 1991, pp. 347–370.

Resnick, S., and C. Stărică, "Tail Index Estimation for dependent Data," mimeo, School of ORIE, Cornell University, 1996.