MEASURING RISK WITH EXTREME VALUE THEORY

Richard L Smith [1]

1 INTRODUCTION

As financial trading systems have become more sophisticated, there has been increased awareness of the dangers of very large losses. This awareness has been heightened by a number of highly publicised catastrophic incidents —

• *Barings.* In February 1995, the Singapore subsidiary of this long-established British bank lost about $1.3 billion because of the illegal activity of a single trader, Nick Leeson. As a result the bank collapsed, and was subsequently sold for one pound.

• *Orange County.* Bob Citron, the Treasurer of Orange County, had invested much of the county's assets in a series of derivative instruments tied to interest rates. In 1994, interest rates rose, and Orange County went bankrupt, losing $1.7 billion.

• *Daiwa Bank.* A single trader, Toshihide Iguchi, lost $1.1 billion of the bank's money over a period of 11 years, the losses only coming to light when Iguchi confessed to his managers in July 1995.

• *Long Term Capital Management.* In the most spectacular example to date, this highly-regarded hedge fund nearly collaped in September 1998. LTCM was trading a complex mixture of derivatives which, according to some estimates, gave it an exposure to market risk as high as $200 billion. Things started to go wrong after the collapse of the Russian economy in the summer of 1998, and to avoid a total collapse of the company, 15 major banks contributed to a $3.75 billion rescue package.

These and other examples have increased awareness of the need to quantify probabilities of large losses, and for risk management systems to control such events. The most widely used tool is *Value at Risk* (henceforth, VaR). Originally started as an internal management tool by a number of banks, it gained a higher profile in 1994 when J.P. Morgan published its RiskMetrics system [2]. Subsequent books aimed at financial academics and traders (Jorion 1996, Dowd 1998) explained the statistical basis behind VaR. Despite the complexity of financial data management that these systems need, the statistical principles behind them are quite simple.

According to the most usual definition, we have to fix a time horizon T and a failure probability α. A common value for T is ten trading days, while α is often set to be .05 or .01. The VaR is then defined to be the largest number x such that the probability of a loss as large as x over the time horizon T is no more than α. Since it is widely accepted that, conditionally on the current volatility σ, the daily log returns ($Y_t = 100 \log(X_t/X_{t-1})$ where X_t is the price on day t) are independent normally distributed with standard deviation σ, the VaR becomes a routine calculation of normal probabilities. When the joint behaviour of a large number of assets is considered, as is needed to calculate the VaR of a portfolio, it is usual to adopt a multivariate normal distribution, though much work goes into the computation of the variances and covariances required. For instance, it is common to use some variant of either principal components analysis or factor analysis to reduce the dimensionality of the statistical estimation problem.

What has been outlined is the simplest approach to VaR estimation. There are at least three competing approaches, none of them so reliant on distributional assumptions. The *historical data* approach uses historical market movements to determine loss probabilities in a statistically nonparametric way. The disadvantage of this is that historical data may not adequately represent current market conditions, or may not be available in sufficient quantity to allow reliable risk calculations to be made. The *stress testing* approach puts much less emphasis on the assessment of small probabilities, instead relying on computing losses under various scenarios of unlikely but plausible market conditions. Finally there is the approach discussed in the present paper, using *Extreme Value Theory* (EVT) to characterise the lower tail behaviour of the distribution of returns without tying the analysis down to a single parametric family fitted to the whole distribution.

The use of EVT in financial market calculations is a fairly recent innovation, but there is a much longer history of its use in the insurance industry. The excellent recent book by Embrechts *et al.* (1997) surveys the mathematical theory of EVT and discusses its applications to both financial and insurance risk management. In Section 2 of the current paper, I outline some of the statistical techniques used in EVT and illustrate them with a recent example of insurance data. However, I also highlight some aspects of financial data — specifically, the presence of variable volatility — that makes direct application of such methods to financial data inappropriate.

In subsequent sections, I outline some current areas of theoretical development that have strong potential for applicability in the insurance and financial industries —

• Bayesian methods (Section 3) as a device for taking account of model uncertainty in extreme risk calculations,

• Multivariate EVT (Section 4) as an alternative approach to risk assessment in high-dimensional systems,

• A random changepoint model (Section 5) as one approach to long-term stochastic volatility.

The overall message of the paper is that EVT contains rich possibilities for application to finance and insurance risk management, but that these areas of application also pose many new challenges to the methodology.

2 OUTLINE OF EXTREME VALUE THEORY

The mathematical foundation of EVT is the class of extreme value limit laws, first derived heuristically by Fisher and Tippett (1928) and later from a rigorous standpoint by Gnedenko (1943). Suppose $X_1, X_2, ...,$ are independent random variables with common distribution function $F(x) = \Pr\{X \leq x\}$ and let $M_n = \max\{X_1, ..., X_n\}$. For suitable normalising constants $a_n > 0$ and b_n, we seek a limit law G satisfying

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = F^n(a_n x + b_n) \to G(x) \tag{2.1}$$

for every x. The key result of Fisher-Tippett and Gnedenko is that there are only three fundamental *types* of extreme value limit laws [3]. These are

$$\text{Type I}: \qquad \Lambda(x) = \exp(-e^{-x}), \quad -\infty < x < \infty,$$

$$\text{Type II}: \qquad \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\alpha}), & x > 0, \end{cases}$$

$$\text{Type III}: \qquad \Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\}, & x \leq 0, \\ 1, & x > 0. \end{cases}$$

In Types II and III, α is a positive parameter. The three types may also be combined into a single *generalised extreme value* distribution, first proposed by von Mises (1936), of form

$$G(x) = \exp\left\{-\left(1 + \xi\frac{x - \mu}{\psi}\right)_+^{-1/\xi}\right\}, \tag{2.2}$$

where $y_+ = \max(y, 0)$, $\sigma > 0$ and μ and ξ are arbitrary real parameters. The case $\xi > 0$ corresponds to Type II with $\alpha = 1/\xi$, $\xi < 0$ to Type III with $\alpha = -1/\xi$, and the limit $\xi \to 0$ to Type I.

Classical EVT is sometimes applied directly, for example by fitting one of the extreme value limit laws to the annual maxima of a series, and much historical work was devoted to this approach (Gumbel 1958). From a modern viewpoint, however, the classical approach is too narrow to be applied to a wide range of problems.

---

[3] Two probability distributions G_1 and G_2 are said to be *of the same type* if they may be related by a location-scale transformation, $G_1(y) = G_2(Ay + B)$ for some $A > 0$ and B. Thus, in saying that there are only three types, we mean that any extreme value limit law may be reduced to one of the three given forms by a location-scale transformation.

An alternative approach is based on *exceedances over thresholds* (Smith 1989, Davison and Smith 1990, Leadbetter 1991). According to this approach, we fix some high threshold u and look at all *exceedances* of u. The distribution of *excess values* is given by

$$F_u(y) = \Pr\{X \le u + y \mid X > u\} = \frac{F(u+y) - F(u)}{1 - F(u)}, \quad y > 0. \tag{2.3}$$

By analogy with classical EVT, there is a theory about the asymptotic form of $F_u(y)$, first given by Pickands (1975). According to this, if the underlying distribution function F is such that a classical extreme value distribution (2.1) exists, then there are constants $c_u > 0$ such that as $u \to \omega_F$ [4]

$$F_u(c_u z) \to H(z), \tag{2.4}$$

where

$$H(z) = \begin{cases} 1 - \left(1 + \frac{\xi z}{\sigma}\right)_+^{-1/\xi}, & \xi \ne 0, \\ 1 - e^{-z/\sigma}, & \xi = 0, \end{cases} \tag{2.5}$$

where $\sigma > 0$ and $-\infty < \xi < \infty$. This is known as the *generalised Pareto distribution* (GPD). There is a close analogy between (2.5) and (2.2), because ξ is the same and there are also mathematical relations among μ, ψ and σ (Davison and Smith 1990).

The threshold approach is most usually applied by fitting the GPD to the observed excesses over the threshold. One advantage of this method over the annual maximum approach is that since each exceedance is associated with a specific event, it is possible to make the parameters σ and ξ depend on covariates. This has been done, for instance, in assessing the probability of a high-level exceedance in the tropospheric ozone record as a function of meteorology (Smith and Shively 1995). Other aspects of the method are the selection of a suitable threshold, and treatment of time series dependence. In environmental applications, the latter aspect is often dealt with by the simple procedure of restricting attention to *peaks* within clusters of high exceedances (Davison and Smith 1990), though as we shall see, such a simple-minded approach does not appear to work for handling stochastic volatility in financial time series.

There are other approaches to extreme value modelling, based on variants of the theoretical results already discussed. One approach extends the annual maximum approach to the joint distribution of the k largest or smallest order statistics in each year — this was first developed statistically by Smith (1986) and Tawn (1988), though the underlying probability theory is much older (see, for example, Section 2.3 of Leadbetter *et al.* 1983). This method is not used much, but we shall see an example of it in Section 3.

A more substantial variant is to take the point-process viewpoint of high-level exceedances, which again has been very well developed as a probabilistic technique (e.g. the

---

[4] $\omega_F = \sup\{x : \ F(x) < 1\}$, the right-hand endpoint of the distribution, usually but not necessarily assumed to be $+\infty$.

books by Leadbetter *et al.* (1983) and Resnick (1987) both use it, though from quite different viewpoints) and was developed as a statistical technique by Smith (1989). According to this viewpoint, the exceedance times and excess values of a high threshold are viewed as a two-dimensional point process (Fig. 1). If the process is stationary and satisfies a condition that there are asymptotically no clusters among the high-level exceedances, then its limiting form is non-homogeneous Poisson and the intensity measure of a set A of the form $(t_1, t_2) \times (y, \infty)$ (see Fig. 1) may be expressed in the form

$$\Lambda(A) = (t_2 - t_1) \cdot \left(1 + \xi \frac{y - \mu}{\psi}\right)_+^{-1/\xi}. \tag{2.6}$$

Here, the interpretation of the parameters μ, ψ and ξ is exactly the same as in (2.2) — indeed, if the time scale in (2.6) is measured in years then the corresponding version of (2.2) is precisely the probability that a set $A = (t_1, t_1 + 1) \times (y, \infty)$ is empty, or in other words, that the annual maximum is $\leq y$. However, one can also derive the GPD as a consequence of (2.6) and hence tie this view of the theory with the peaks over threshold analysis.

A more general form of the model allows for time-dependent behaviour by replacing the fixed parameters μ, ψ, ξ with functions μ_t, ψ_t, ξ_t where t denotes time. In particular, we consider models of this form in Section 5, and equation (5.1) gives the generalisation of (2.6) in this case. In this way, dependence on covariates or other time-dependent phenomena may be incorporated into the model.
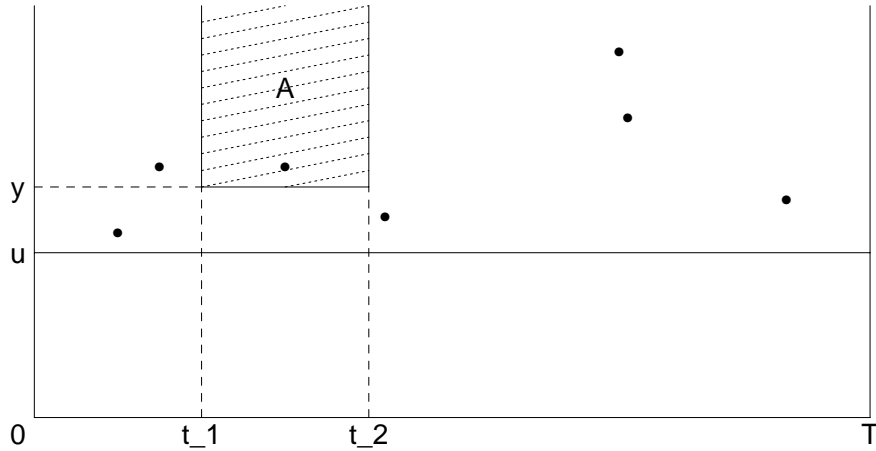

Fig. 1. Illustration of high-level exceedances represented as a two-dimensional point process.

5

A number of diagnostic techniques have been devised to test whether these assumptions are satisfied in practice. Among these are the mean excess plot (Davison and Smith, 1990), which is a plot of the mean of all excess values over a threshold u against u itself. This is based on the following identity: if Y is a random variable with distribution function (2.5), provided $\xi < 1$, then for $u > 0$,

$$\mathrm{E}\{Y - u \mid Y > u\} = \frac{\sigma + \xi u}{1 - \xi}.$$

Thus, a sample plot of mean excess against threshold should be approximately a straight line with slope $\xi/(1 - \xi)$. This is a useful tool in selecting the threshold.

In practice, the plot can be hard to interpret because for large u there are few exceedances and hence very high variability in the mean, but its real purpose is to detect significant shifts in slope at lower thresholds. As an example, Fig. 2(a) shows the negative log daily returns for Standard and Poors index (S&P 500), 1947–1987. The values are negated because our interest in this discussion is in the possibility of very large losses, so the values of interest appear as large positive values in the plot. In particular, the spike at the right hand end of the plot is the October 19, 1987 value. A mean excess plot (Fig. 2(b)) shows an apparent "kink" near $y = 3.8$, so it would seem unwise to include values below that threshold. (In fact this discussion is too simple because we have not taken variable volatility into account, but we return to that point later.)
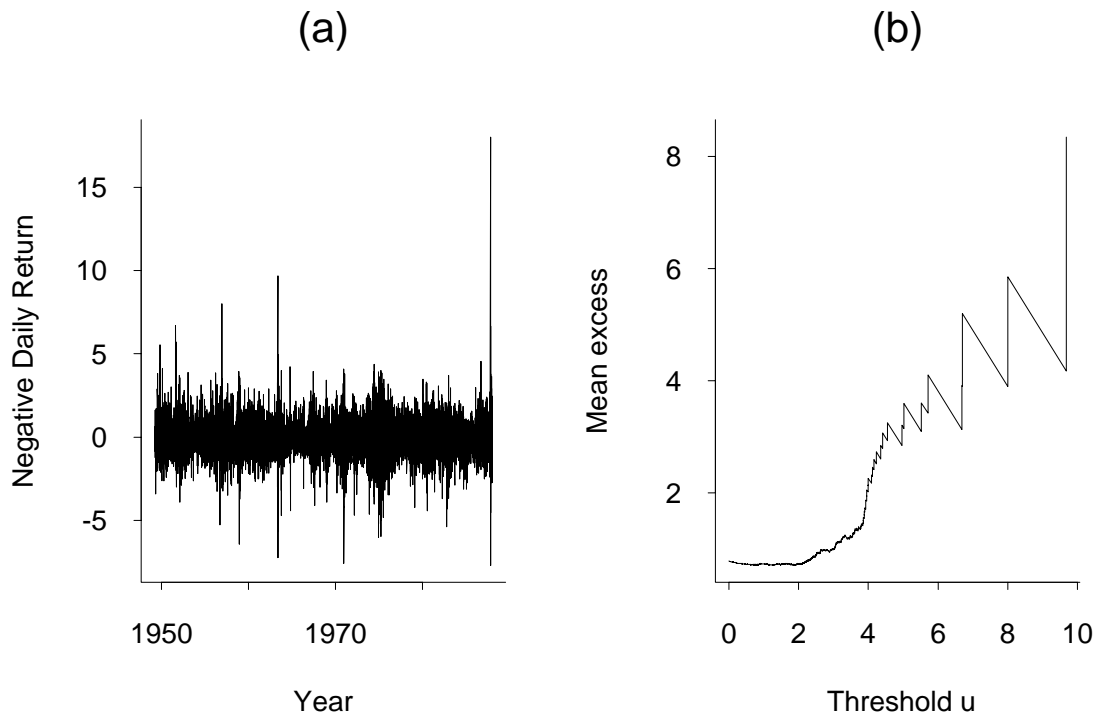

Fig. 2. Negative daily returns from the S&P 500 (a), and a Mean Excess Plot based on these data (b).

In contrast, Fig. 3(a) shows 15 years of insurance claims data from a well-known multinational company (Smith and Goodman 2000), and the corresponding mean excess

plot in Fig. 3(b). In this case the series is dominated by two very large claims in the middle of the plot, which together account for 35% of all claims in the series, but in spite of this apparent evidence of outliers, the mean excess plot is surprisingly stable. Repeated fitting of the model (2.6), to a variety of thresholds (Table 1), shows comparatively little variation in the parameters μ, ψ and ξ, which is another indication that the model is a good fit.
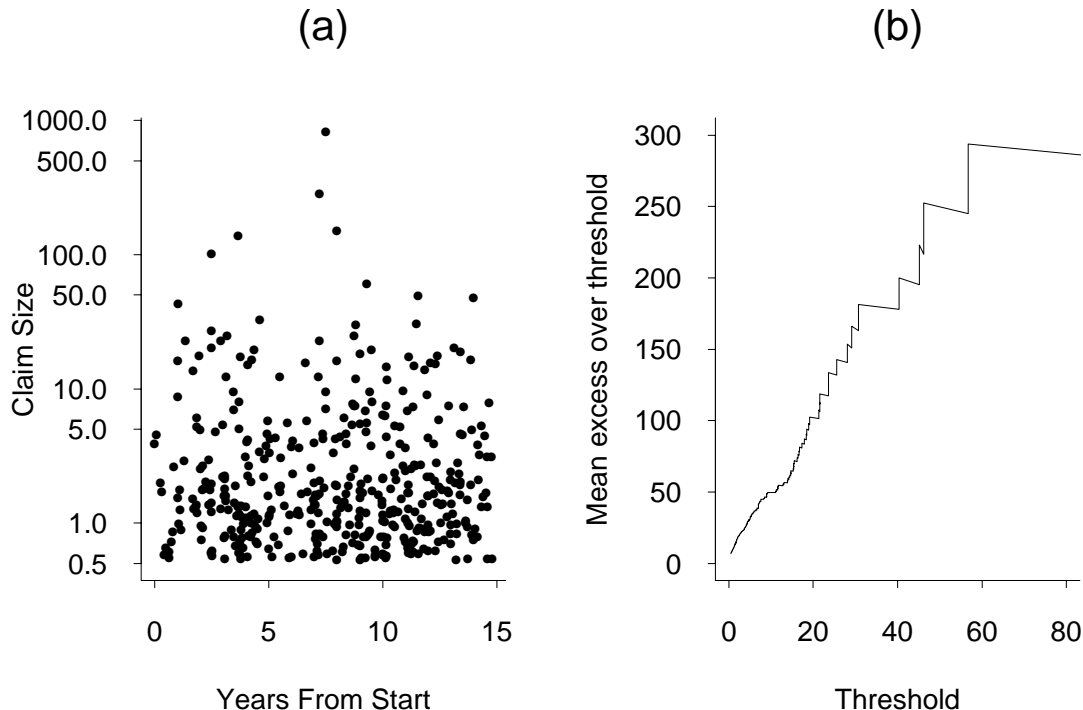

Fig. 3. Scatterplot of large insurance claims against time (a), and a Mean Excess Plot based on these data (b).

Threshold	Num. of exceedances	μ	$\log \psi$	ξ
0.5	393	26.5	3.30	1.00
2.5	132	26.3	3.22	0.91
5	73	26.8	3.25	0.89
10	42	27.2	3.22	0.84
15	31	22.3	2.79	1.44
20	17	22.7	3.13	1.10
25	13	20.5	3.39	0.93

Table 1. Parameter estimates for the insurance claims data based on a variety of thresholds.

Other diagnostics may be derived from the fitted point process. For example, under the model (2.6), the one-dimensional point process of exceedance times of a fixed threshold

7

u is a nonhomogeneous Poisson with parameter $\lambda = \{1+\xi(u-\mu)/\psi\}^{-1/\xi}$. As noted already following (2.6), in general we may permit the parameters μ, ψ, ξ to be functions of time and in that case the constant λ is replaced by a time-dependent intensity λ_t. For this model, with constant λ as a special case, if the observations begin at time T_0 and the successive exceedance times are at $T_1, T_2, ...$, the variables

$$Z_k = \int_{T_{k-1}}^{T_k} \lambda_t dt, \quad k \geq 1, \tag{2.7}$$

should be independent exponentially distributed random variables with mean 1. This may be tested graphically, for example, via a QQ [5] plot of observed order statistics versus their expected values under the independent exponential assumption.

We can also test the marginal distribution of excesses in similar fashion. In this case the appropriate test statistic is

$$W_k = \frac{1}{\xi_{T_k}} \log \left[1 + \xi_{T_k} \left\{ \frac{Y_{T_k} - u}{\psi_{T_k} + \xi_{T_k}(u - \mu_{T_k})} \right\} \right], \tag{2.8}$$

Y_{T_k} being the observed value of the process at time T_k and the notation indicating that the parameters μ, ψ and ξ are all dependent on time in the most general form of the model. Once again, if the assumed model is correct then the $\{W_k\}$ are independent exponentially distributed random variables with mean 1, and this may be tested in various ways, for example, through a QQ plot of the order statistics. The plots based on the Z and W statistics were first suggested by Smith and Shively (1995).

As an example, Fig. 4 shows the Z and W plots for the insurance data of Fig. 3, in the case that the extreme values parameters μ, ψ, ξ are assumed constants independent of time. In this case, both plots look quite close to a straight line of unit slope, indicating an acceptable fit to the model. As a standard for later comparison, we calculate the R^2 for regression $(1 - \sum(y_i - x_i)^2 / \sum(y_i - \bar{y})^2$ where (x_i, y_i) are the coordinates of the i'th point in the plot). The R^2 values in this example are .992 for the Z plot and .980 for the W plot.

Fig. 5, based on the S&P index data, is more problematic. The W plot ($R^2 = .971$) is a good fit except for the very largest value, which is of course the October 19, 1987 market crash, so this is easily understood if not so easily explained. However, the Z plot ($R^2 = .747$) is completely wrong, and no obvious variant on the methodology (such as changing the threshold, transforming the response variable, or adding simple time trends to the model) will do anything to correct this. The explanation is that variation in volatility results in substantial variation in the mean time between exceedances over the threshold, and no simple modification of the model can account for this.

---

[5] quantile-quantile

(a) Plot of Z values

(b) Plot of W values

Fig. 4. Z-plot and W-plot for the insurance data, all exceedances over threshold 5.

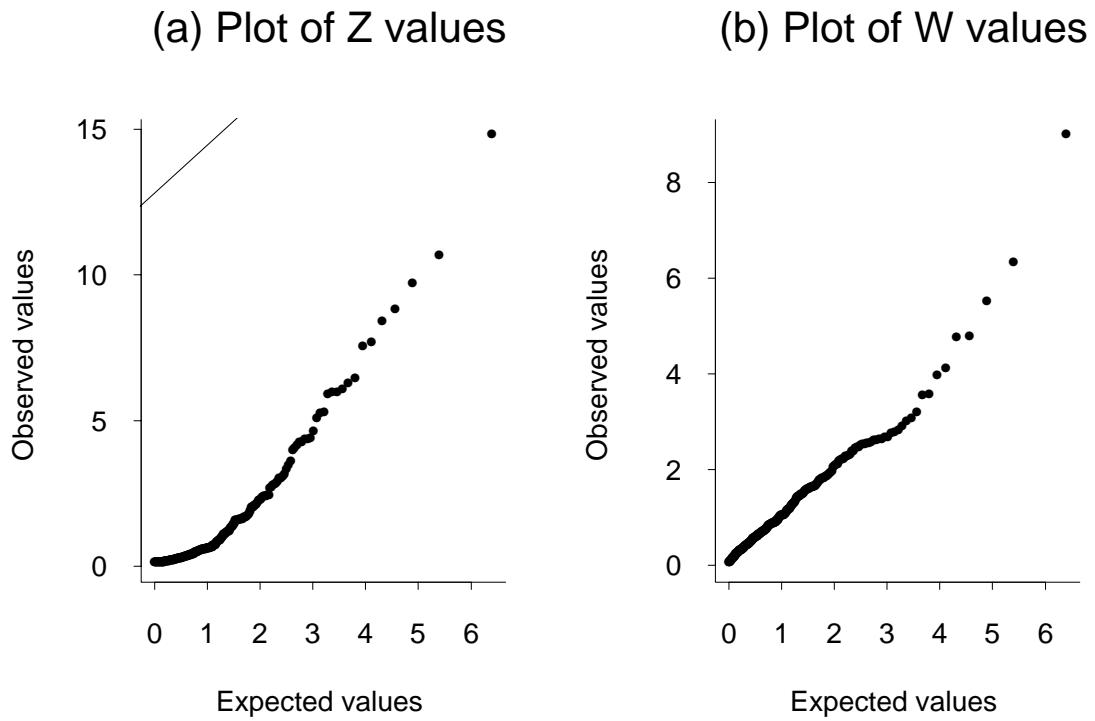## (a) Plot of Z values

(b) Plot of W values

Fig. 5. Z-plot and W-plot for the S&P 500, all exceedances over threshold 2.

3 BAYESIAN STATISTICS FOR RISK ASSESSMENT

So far, our statistical viewpoint has implicitly been classical frequentist, including maximum likelihood estimation for the parameters of the extreme value model. In this section, I argue that there may be substantial advantages of application and interpretation by taking a Bayesian approach to the analysis. The approach taken is not specifically tied to the subjective viewpoint of probability theory, since there may also be substantial advantages to the proposed approach from a frequentist viewpoint, though the evidence on the latter point is still unclear at the present time.

To illustrate the ideas, I take an example from a quite different field to the ones discussed so far. Fig. 6(a) shows the five best performances by different athletes in the women's 3000 metre track race for each year from 1972 to 1992, together with the remarkable new world record established in 1993 by the Chinese athlete Wang Junxia. Many questions have been raised about Wang's performance, including the possibility that it may have been assisted by drugs, though no direct evidence of that was ever found. The present discussion is based on analysis by Robinson and Tawn (1995) and Smith (1997a).
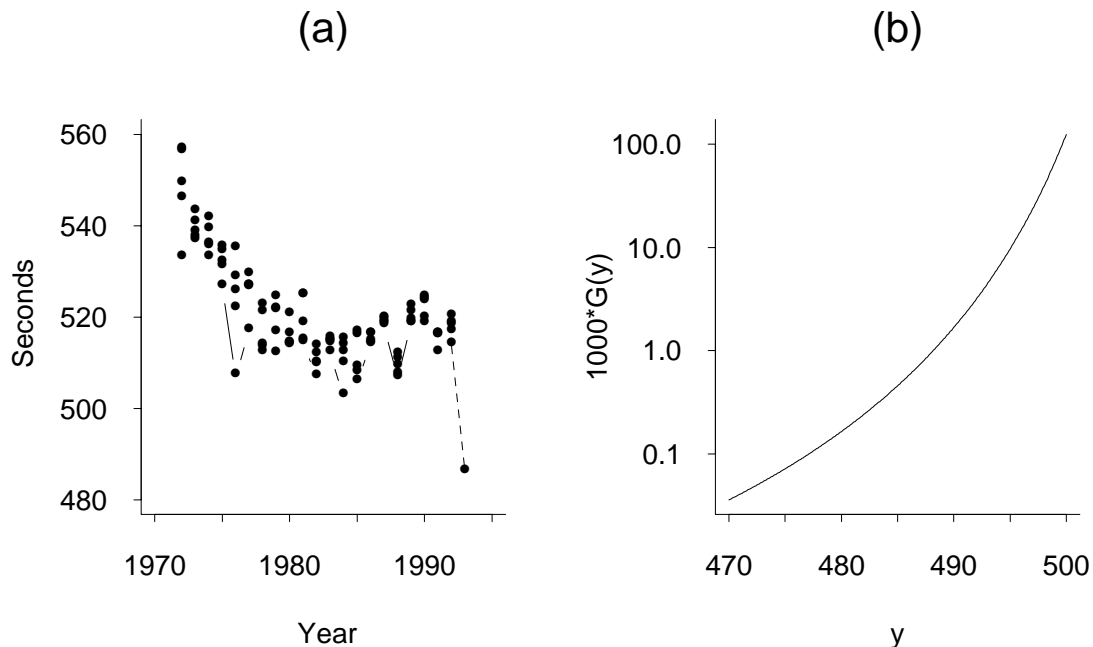

Fig. 6. (a) Plot of best five performances by different athletes in each year from 1972–1992, together with Wang Junxia's performance from 1993. (b) Plot of predictive conditional probability distribution given all data up to 1992.

The natural extreme value model for this problem is a Type III or Weibull distribution, which implies a finite lower bound β on the distribution of running times. The basic model

10

for the distribution function of an annual maximum is (2.2). This is applied to running times multiplied by −1, so as to convert minima into maxima. When $\xi < 0$, the distribution function (2.2) has a finite upper bound (for which $G(x) = 1$) at $x = \mu - \psi/\xi$. Thus when this is applied to −1 times the running times, there is a finite *minimum* running time at $\beta = -\mu + \psi/\xi$. We therefore concentrate on estimation of this parameter, finding the maximum likelihood estimate and likelihood-based confidence intervals for β, based on the data up to 1992. If Wang's actual performance lay outside this confidence interval, that could be interpreted as evidence that something untoward had taken place. As noted briefly in Section 2, the actual estimate was based on the k best performances in each year, where in this analysis, $k = 5$.

In one analysis, Smith (1997a) analysed the data from 1980, ignoring the trend in the early part of the series, and obtained a 95% confidence interval for β of (481.9, 502.4) (seconds). Wang's actual record was 486.1, so while this lies towards the lower end of the confidence interval, the analysis does not definitively establish that there was anything wrong. Earlier, Robinson and Tawn (1995) gave a number of alternative analyses based on various interpretations of the trend in Fig. 6(a), but all led to the same conclusion, that Wang's record lay within a 95% confidence interval for β.

However, Smith (1997a) went on to argue that obtaining a confidence interval for β was solving the wrong problem. Consider the situation as it appeared at the end of 1992. A natural question to ask is: what is the probability distribution of the best performance that will be observed in 1993? This is a question about the *predictive* distribution of an as yet unobserved random variable. As a partial protection against the obvious selection bias associated with the choice of year, the paper proposed that the predictive probability be calculated conditionally on the event that a new world record be set.

There is no known frequentist solution to this problem that adequately takes account of the fact that the model parameters are unknown [6], but a Bayesian solution is straightforward. If the required conditional predictive distribution is denoted $G(y; \theta)$, this being the probability that the best performance in 1993 will be smaller than y, given that it is better than the existing record, as a function of model parameters θ, then the Bayesian solution is based on the estimate

$$\tilde{G}(y) = \int G(y; \theta)\pi(\theta \mid \mathbf{X})d\theta, \tag{3.1}$$

---

[6] A naïve solution is to substitute a point estimator of the unknown parameters, such as the maximum likelihood solution, into the predictive distribution: in the notation of (3.1), $\tilde{G}(y) = G(y; \hat{\theta})$ where $\hat{\theta}$ is the MLE. In the present example, the MLE $\hat{\beta}$ based on the data up to 1992 is greater than the actual time run by Wang, so such an approach would automatically lead to the value 0 for the predictive probability. However we can see this this approach is too simplistic, because as has already been pointed out, a 95% confidence interval for β includes Wang's record.

$\pi(\theta \mid \mathbf{X})$ denoting the posterior density of the parameters θ given past data $\mathbf{X}$. Writing $\theta = (\mu, \sigma, \xi)$ and x in place of $-y$, $G(y; \theta)$ is given by (2.2). As already noted, the transformation from x to $-y$ was made to convert minima into maxima.

Using a vague prior for θ and a Monte Carlo integration to evaluate (3.1), a predictive probability of .00047 (since slightly revised to .0006) was attached to the actual record run by Wang. The complete curve of $\tilde{G}(y)$ is shown in Fig. 6(b). This calculation seems definitively to establish that her performance was inconsistent with previous performances in the event. It does not, of course, provide any direct evidence of drug abuse.

The relevance of this example to risk assessment in finance and insurance is threefold:

1. There is a clear distinction between *inference* about unknown *parameters* and *predictive distributions* about future *variables*. Many risk applications, including VaR itself, revolve around questions of the form "What is the probability that I will lose a certain amount of money over a certain period of time?" These are questions about prediction, not inference.

2. In evaluating predictive distributions, account must be taken of the fact that model parameters are unknown.

3. Bayesian methods provide an operational solution to the problem of calculating predictive distributions in the presence of unknown parameters. There are pure frequentist solutions based on asymptotic theory (for example, Barndorff-Nielsen and Cox (1996)), and it remains an open question just how well Bayesian solutions to these kinds of problems perform from a frequentist point of view, but the evidence currently available is encouraging, provided proper account is taken of the loss function in a decision-theoretic formulation of the problem (Smith 1997b, 1999).

As an example of the possible application of these ideas to risk assessment problems, suppose we want to calculate the predictive distribution of the largest loss over a future one-year time period, based on the data in Figure 3 and assuming a constant distribution. Fig. 7 shows a plot of the Bayes posterior median (solid curve) of the probability of exceeding a given level y, for each of a series of y values represented on the vertical axis. In this plot we represent the probability of exceedance as $1/N$, and the value of N is represented on the horizontal axis. Also shown on the plot are 50% and 95% posterior probability intervals for the probability of exceedance, defined by the dashed lines and the dotted lines respectively. In the more detailed analysis of this data set, Smith and Goodman (2000) have provided a number of alternative analyses taking account of alternative features of the data. In particular, the data included a "type of claim" indicator, and when this is taken into account, the predictive distribution changes substantially, but that lies beyond the scope of the present discussion.
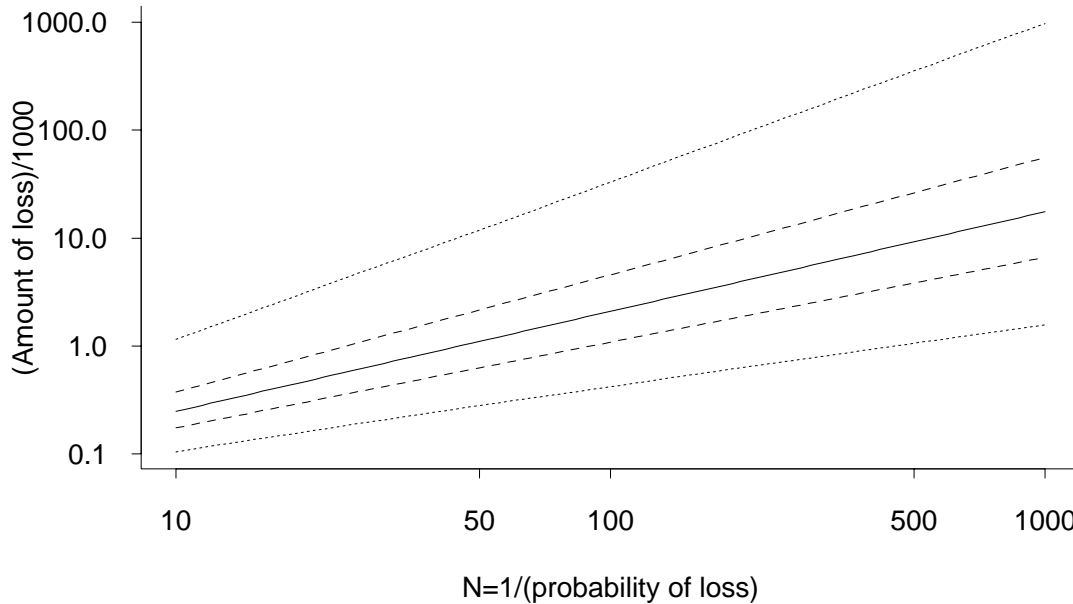
Fig. 7. Median posterior loss curve with 50% and 95% probability bounds for insurance data, one-year losses, based on all exceedances over threshold 5.

4 MULTIVARIATE EXTREMES

So far, our discussion has been entirely about extreme value theory for a single variable. However, it is more usual for VaR calculations to be made about a portfolio of assets rather than a single asset. In this context, a portfolio is simply a linear combination of individual asset prices. If the composition of the portfolio is held fixed, then it may be possible to assess the risk using univariate EVT, by simply treating the portfolio price as the variable of interest. However, the real rationale for doing this is often to help design the portfolio — for example, one may want to do this to maximise the expected return subject to some constraint on the VaR of the portfolio. To solve a problem of this nature, in which the weights on the different assets have to be determined, it is essential to consider the joint distribution of the asset prices. Conventional VaR theory is based on an assumption of multivariate normality for the joint distribution of log returns, but it is highly questionable whether such an assumption is appropriate for the calculation of extreme tail probabilities.

One approach to this problem is through multivariate EVT. Limiting relations such as (2.1) and (2.4) may be generalised to vector-valued processes, and for any $p \geq 1$, lead to a class of p-dimensional multivariate extreme value distributions (MVEVDs) and their threshold equivalents. There are numerous mathematically equivalent representations of MVEVDs, but one convenient form, due to Pickands (1981), is as follows. We may without loss of generality assume that all the marginal distributions have been transformed to the "unit Fréchet" distribution function $e^{-1/x}$, $0 < x < \infty$; the joint distribution function of

13

$x = (x_1, ..., x_p)$ is then of the form

$$G(x) = \exp\left\{-\int_{S_p} \max_{1 \le j \le p}\left(\frac{w_j}{x_j}\right) dH(w)\right\}, \tag{4.1}$$

where $S_p = \{(w_1, ..., w_p) : w_1 \ge 0, ..., w_p \ge 0, \sum w_j = 1\}$ is the unit simplex in p dimensions and H is some non-negative measure on S_p satisfying

$$\int_{S_p} w_j dH(w) = 1, \quad j = 1, ..., p. \tag{4.2}$$

Resnick (1987) is an excellent source of information about MVEVDs. The difficulty for statistical applications is that when $p > 1$, the class of MVEVDs does not reduce to a finite-dimensional parametric family, so there is potential explosion in the class of models to be considered. Most approaches to date have focussed either on simple parametric subfamilies, or on semiparametric approaches combining univariate EVT for the marginal distributions with nonparametric estimation of the measure H. Some example papers representing both approaches are Coles and Tawn (1991, 1994), Smith (1994) and de Haan and Ronde (1998). Recently, it has even been suggested that multivariate EVT may not be a rich enough theory to encompass all the kinds of behaviour one would like to be able to handle, and alternative measures of tail dependence have been developed. The main proponents of this approach so far have been Ledford and Tawn (1996, 1997, 1998); the last paper, in particular, contains an application to foreign exchange rates.

As I see it, the main difficulty with the application of this approach to VaR is in how to extend the methodology from the joint extremes of a small number of processes to the very large number of assets in a typical portfolio. Most of the papers just cited are for $p = 2$; some have considered extensions to $p = 3, 4, 5, ...$, but the model complexity increases greatly with p and there seems to be no hope of applying multivariate EVT directly to large portfolios in which p may be of the order of hundreds.

Recently, Smith and Weissman (1999) have proposed some alternative representations of extreme value processes aimed at characterising the joint distribution of extremes in multivariate time series of the form $\{X_{ij}, i = 1, 2, ..., 1 \le j \le p\}$. As in the preceding discussion, there is no loss of generality in assuming that unit Fréchet marginal distributions apply in the tail, because we may use univariate EVT to estimate the marginal tail distributions and then apply a probability integral transformation to each component. Smith and Weissman then defined a class of *multivariate maxima of moving maxima* (M_4 processes for short) by the equation

$$X_{ij} = \max_{\ell \ge 1} \max_{-\infty < k < \infty} a_{\ell k j} Z_{\ell, i-k}. \tag{4.3}$$

where $\{Z_{\ell, i}\}$ are a two-dimensional array of independent unit Fréchet random variables and the constants $\{a_{\ell k j}\}$ satisfy

$$a_{\ell k j} \ge 0, \quad \sum_\ell \sum_k a_{\ell k j} = 1 \quad \text{for all } j = 1, ..., p. \tag{4.4}$$

The main focus of the paper by Smith and Weissman is to argue that under fairly general conditions, extremal properties of a wide class of multivariate time series may be calculated by approximating the process by one of M_4 form. The fundamental ideas behind representations of this form are due to Deheuvels (1978, 1983), and they can be regarded as an alternative approach to those based on the representation (4.1).

In principle, (4.3) is simpler to handle than (4.1). Moreover it is a more general result, dealing directly with the case of multivariate time series and not just of independent multivariate observations. Another feature which makes (4.3) more directly interpretable for financial time series is that it represents the process in terms of an independent series of "shocks" – in essence, large values among the $\{Z_{\ell,i}\}$ (the shocks) determine the pattern of extremes among the $\{X_{ij}\}$ and this has an obvious interpretation for the financial markets. On the other hand, estimating a three-dimensional array of unknown constants is a challenging problem in itself, and it is likely that some restrictions to specific classes will be necessary before this is feasible. Another difficulty with models of this form is that they suffer from degeneracies — the joint density of a set of random variables defined by (4.3) is typically singular with respect to Lebesgue measure and this causes problems for maximum likelihood techniques. However, this difficulty can be avoided by adding some additional noise to the observations and research is continuing into ways in which this might be done.

5 A CHANGEPOINT MODEL FOR STOCHASTIC VOLATILITY

We have seen that the standard extreme value methods do not appear to apply to the S&P 500 data. The explanation is non-constant volatility: it is apparent from simple inspection of the data in Fig. 2(a) that the variance of the series is much bigger in some years than in others, and consequently there is substantial variation in the rate in which any high threshold is exceeded. This problem is near-universal in financial time series: every other example which I have tried has exhibited problems similar to those with the Z-plot in Fig. 5.

There is by now a rich literature of models for financial time series taking into account changes in volatility. These divide broadly into two categories: models of the GARCH family, in which the variance of the process at time t, usually denoted σ_t, is expressed deterministically as a function of past values σ_s, $s < t$, and of the observations themselves; and models in which the volatility is treated as a stochastic process estimated by some form of state space model analysis. An excellent review of developments in both approaches is the paper by Shephard (1996).

It therefore seems worthwhile to develop extensions of the extreme value statistical methodology to take into account variable volatility. So far, very few attempts have been made to do this. McNeil and Frey (1998) have taken an approach built around the standard GARCH model, but in which the innovations, instead of being normally distributed as in the usual GARCH approach, are allowed to be long-tailed and estimated by methods

similar to those presented earlier in this paper, but taking account of the variation in σ_t estimated for the GARCH process. In another recent paper, Tsay (1999) has used methods similar to those of the present paper, but allowing the extreme value parameters to depend on daily interest rates.

The aim of the present section is to suggest an alternative approach which is not tied to GARCH or to any particular model of volatility, but which simply assumes that the extreme value parameters change from one period to another according to a random changepoints process. Only an outline will be presented here; a fuller description is being prepared for publication elsewhere.

To describe the basic model, we first generalise (2.6) to

$$\Lambda(A) = \int_{t_1}^{t_2} \left(1 + \xi_t \frac{x - \mu_t}{\psi_t} \right)_+^{-1/\xi_t} dt \tag{5.1}$$

in which the notation explicitly reflects that the parameters μ, ψ and ξ are time-dependent.

The model is of hierarchical Bayesian structure, and is defined as follows. We assume that the process is observed over a time period $[0, T^*]$.

Level I. At the top level of the hierarchy, we define hyperparameters m_μ, s_μ^2, m_ψ, s_ψ^2, m_ξ, s_ξ^2 with a prior distribution (to be specified later).

Level II. Conditionally on the parameters of level I, let the number of changepoints K have a Poisson distribution with mean νT^*. Conditionally on K, let the individual changepoints $C_1, ..., C_K$ be independent uniform on $[0, T^*]$, and then ordered so that $0 < C_1 < ... < C_K < T^*$. (An equivalent description is that the random changepoints form a realisation of a homogeneous Poisson process with intensity ν.) For convenience we also write $C_0 = 0$, $C_{K+1} = T^*$. Also, let $\mu^{(1)}, ..., \mu^{(K+1)}$ be independently drawn from the $N(m_\mu, s_\mu^2)$ distribution, $\log \psi^{(1)}, ..., \log \psi^{(K+1)}$ independently drawn from $N(m_\psi, s_\psi^2)$, $\xi^{(1)}, ..., \xi^{(K+1)}$ independently drawn from $N(m_\xi, s_\xi^2)$.

Level III. Conditionally on the parameters in Level II, suppose that for each k between 1 and $K + 1$, the exceedance times and values over a threshold u on the time interval $C_{k-1} < t \leq C_k$ are defined by the Poisson process with cumulative intensity given by (5.1), in which $\mu_t = \mu^{(k)}$, $\psi_t = \psi^{(k)}$, $\xi_t = \xi^{(k)}$.

For the prior distributions at level I, we assume that (m_μ, s_μ^2) are of "gamma-normal" type: let τ_μ be drawn from the gamma distribution with density proportional to $\tau_\mu^{\alpha-1} \exp(-\beta\tau_\mu)$, $0 < \tau_\mu < \infty$, and then define $s_\mu^2 = 1/\tau_\mu$, $m_\mu \sim N(\eta, \frac{1}{\kappa\tau_\mu})$. This model may be summarised by the notation $(m_\mu, s_\mu^2) \sim GN(\alpha, \beta, \eta, \kappa)$. Similarly, we assume the pairs (m_ψ, s_ψ^2), (m_ξ, s_ξ^2) are independently drawn from the same distribution. We fix $\alpha = \beta = \kappa = 0.001$ and $\eta = 0$ to represent a proper but very diffuse prior distribution.

16

The treatment of the prior parameter ν is somewhat problematic in this set-up. It might be thought desirable to put a vague hyperprior on ν, but this is not possible because an improper prior leads to an improper posterior (and, in the practical implementation of the algorithm, the number of changepoints grows to ∞). Instead, therefore, I have specified a value for ν. In different runs, the values $\nu = 20$, 25 and 30 have all been tried, with some differences in the posterior distribution of the number of changepoints (Fig. 8(a) below) but fortunately these were not too great.

The actual algorithm uses the reversible jump Markov chain Monte Carlo sampler in manner very similar to that of Green (1995): indeed, my whole approach was very much motivated by Green's treatment of a famous data set concerned with coal mining disasters. However, in the present paper I omit all details of the algorithm, which essentially consists of iteratively updating all the parameters of the models using a reversible jump sampler to take care of the fact that the number of changepoints, and hence the dimension of the model to be estimated, are *a priori* unknown.

Fig. 8 shows the outcome of one run of this analysis, based on a total of 100,000 iterations of the reversible jump sampler with every 100'th iteration recorded and used to construct the plots. Thus, for example, the histogram in Fig. 8(a) is based on 1,000 sampled values of the number of changepoints. The posterior distribution of the number of changepoints has (in this run) a mean of 23.9 and a standard deviation of 2.5. Figs. 8(b) and 8(c) show the Z and W plots computed from the posterior means in the changepoint model; in this case, $R^2 = .992$ for the Z plot and .981 for the W plot. There is still some concern about the very largest values in the W plot but otherwise the fit in the model seems much better than in the earlier discussion of Fig. 5.

Fig. 9(a) shows the posterior mean estimate of the crossing rate of the threshold 2, as it varies across time. This shows very clearly the effects of stochastic volatility, with periods when there is a high probability of crossing the threshold such as around 1971, during 1973-1974 or the early 1980s, interspersed with periods when the probability of crossing the threshold is much lower, such as the late 1970s or mid 1980s. Fig. 9(b) shows a similar plot for the posterior mean of the ξ_t parameter as t ranges over the data set. Many, though not all, of the rises and falls in this plot match the rises and falls in the threshold crossing rate.

Finally, we consider the consequences of the changepoint model for the estimated extreme value parameters. Table 2 shows the estimated posterior parameters and standard errors for (a) December 1987, (b) January 1978 (chosen to represent a quiet period), (c) an overall average over all days of the series (this was calculated by sampling from the total Monte Carlo output) and (d) based on the maximum likelihood estimates for a single homogeneous model fitted to the whole series, as in Section 2. Perhaps the most interesting parameter here is ξ, which represents the overall shape of the tail. For December 1987, the posterior mean is $\hat{\xi} = .21$, representing a fairly long-tailed case (but not excessively so — values of ξ in the range .5 to 1 often occur in insurance applications, including
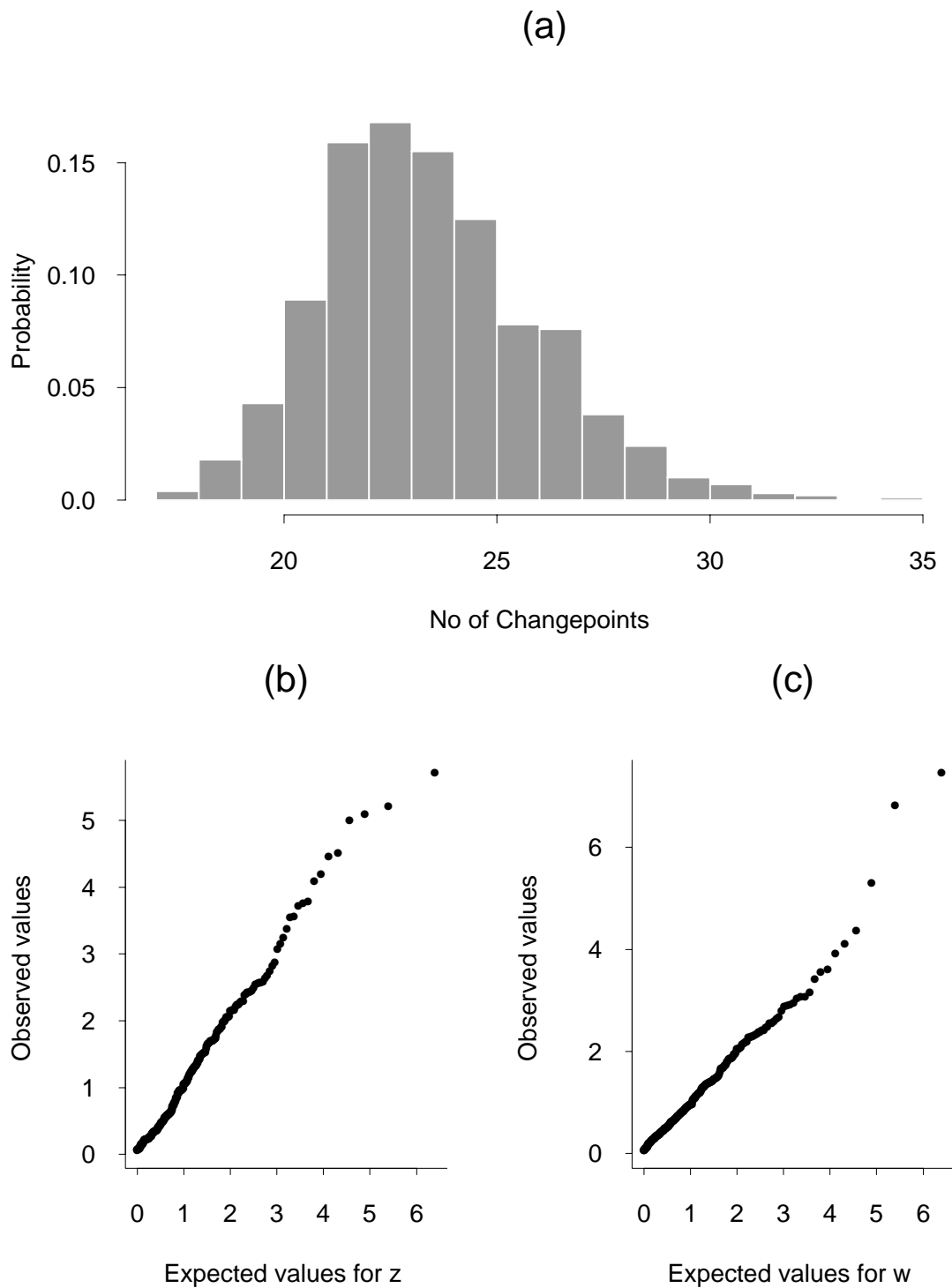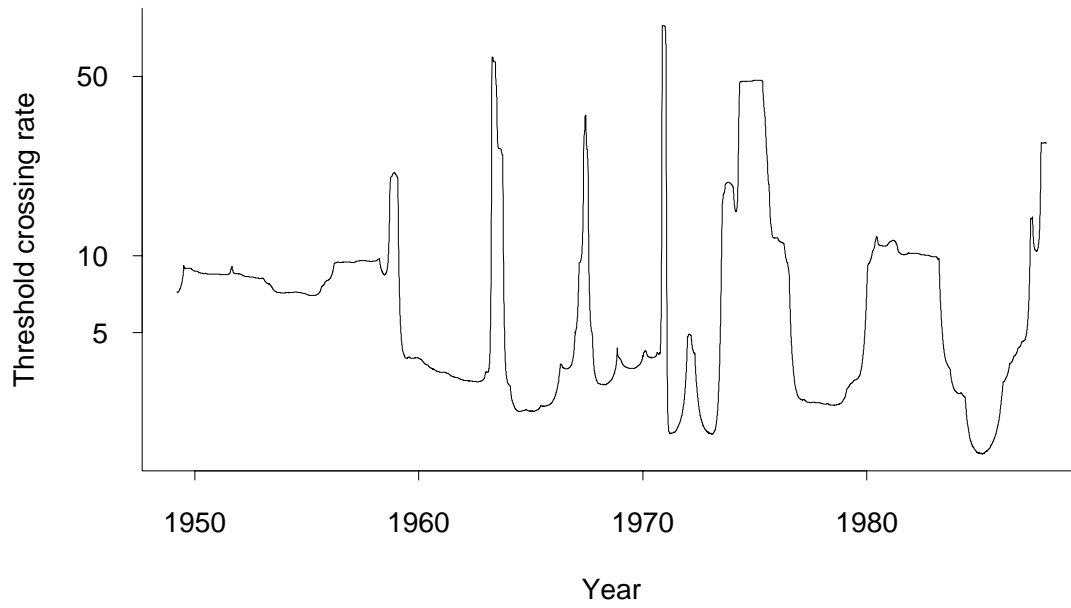
Fig. 8. Results of changepoint modelling for S&P 500 data. (a) Posterior distribution for number of changepoints. (b) Z plot. (c) W plot. Based on threshold 2, prior mean number of changepoints $\nu = 25$.
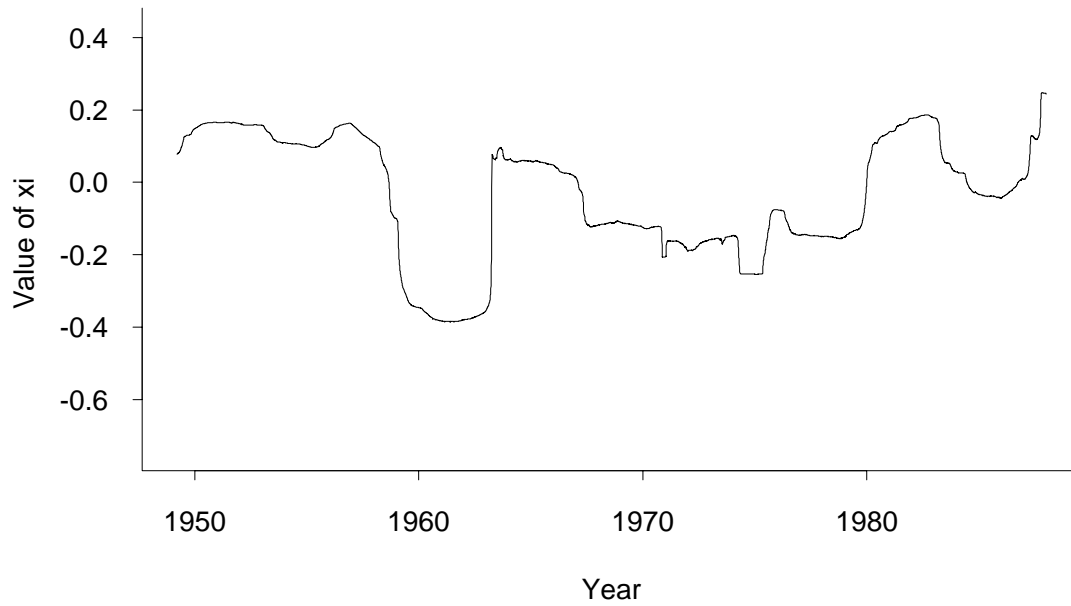
(b)

Fig. 9. Results of changepoint modelling for S&P 500 data. (a) Posterior mean crossing rate of the threshold as a function of time t. (b) Posterior mean of ξ_t.

the one mentioned earlier in this paper). For January 1978, the posterior mean is $-$.02, insignificantly different from 0, which is an exponential tail (in other words, short-tailed). The *overall* average over the whole series is $-.05$, which seems to reflect that the typical behaviour is short-tailed with mainly the high-volatility periods being long-tailed. However, the maximum likelihood estimates based on a homogeneous model imply $\hat{\xi} = .22$ with a standard error of .06. This seems completely misleading, implying that long-tailed behaviour is a feature of the whole series rather than just of short high-volatility periods of it. The interpretation, I believe, is that the effect of mixing over inhomogenous periods has inflated the apparent value of ξ and has made the distribution seem more long-tailed that it really is most of the time. A similar phenomenon has also been observed for the insurance data of Section 2 (Smith and Goodman 2000), though in that case the mixing was over different types of insurance claim rather than inhomogeneous periods of time.

Time	μ	$\log \psi$	ξ
December 1987	5.12	0.24	0.21
	(.97)	(.44)	(.18)
January 1978	3.03	-0.61	-0.02
	(1.04)	(.43)	(.27)
Averaged over time	3.27	-0.51	-0.05
	(1.16)	(.56)	(.32)
Homogeneous model	3.56	-0.09	0.22
	(.10)	(.09)	(.06)

Table 2. Bayes posterior means of model parameters (posterior standard deviations in parentheses) for specific time periods. Row 1: December 1987. Row 2: January 1978. Row 3: Averaged over all time periods in the data. Row 4: Maximum likelihood estimates and standard errors based on a single homogeneous model fitted to the whole series.

6 CONCLUSIONS

The interaction between extreme value theory and the assessment of financial risk poses many exciting possibilities. Many of these seem to require new techniques. In this paper I have presented three areas in which new methodology seems to be required. Bayesian statistics is a valuable tool for the assessment of predictive distributions which is very often the real question of interest, rather than inference for unknown parameters. The possibility of applying VaR analysis to large portfolios implies the need for multivariate extreme value techniques in high dimensions, in contrast with most of the multivariate extreme value theory developed to date which has concentrated on low-dimensional problems. Finally, the last section proposed one way of dealing with the stochastic volatility problem, via a changepoint model for the extreme value parameters. However this in itself is a tentative approach; there is ample scope for exploration of alternative approaches for combining extreme value theory and stochastic volatility.

REFERENCES

Barndorff-Nielsen, O.E. and Cox, D.R. (1996), Prediction and asymptotics. *Bernoulli* **2**, 319–340.

Coles, S.G. and Tawn, J.A. (1991), Modelling extreme multivariate events. *J. R. Statist. Soc. B* **53**, 377–392.

Coles, S.G. and Tawn, J.A. (1994), Statistical methods for multivariate extremes: An application to structural design (with discussion). *Applied Statistics* **43**, 1–48.

Davison, A.C. and Smith, R.L. (1990), Models for exceedances over high thresholds (with discussion). *J.R. Statist. Soc. B* **52**, 393–442.

Deheuvels, P. (1978). Caractérisation complète des lois extrèmes multivariées et de la convergence des types extrèmes. *Publ. Inst. Statist. Univ. Paris* **23**, 1–36.

Deheuvels, P. (1983). Point processes and multivariate extreme values. *J. Multivar. Anal.* **13**, 257–272.

Dowd, K. (1998), *Beyond Value at Risk: The New Science of Risk Management.* John Wiley, Chichester.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance.* Springer, New York.

Fisher, R.A. and Tippett, L.H.C. (1928), Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.* **24**, 180–190.

Gnedenko, B.V. (1943), Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423–453.

Green, P.J. (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Gumbel, E.J. (1958), *Statistics of Extremes.* Columbia University Press, New York.

Haan, L. de and Ronde, J de (1998), Sea and wind: Multivariate extremes at work. *Extremes* **1**, 7–45.

Jorion, P. (1996), *Value at Risk: The New Benchmark for Controlling Market Risk.* Chicago: Irwin.

Leadbetter, M.R. (1991), On a basis for 'peaks over thresholds' modeling. *Statistics and Probability Letters* **12**, 357–362.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Series.* Springer Verlag, New York.

Ledford, A.W. and Tawn, J.A. (1996), Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169–187.

Ledford, A.W. and Tawn, J.A. (1997), Modelling dependence within joint tail regions. *J.R. Statist. Soc. B* **59**, 475–499.

Ledford, A.W. and Tawn, J.A. (1998), Diagnostics for dependence within time-series extremes. Preprint, Department of Mathematics and Statistics, University of Surrey.

McNeil, A.J. and Frey, R. (1998), Estimation of tail-related risk measures from heteroscedastic financial time series: an extreme value approach. Preprint, Department of Mathematics, ETH Zentrum, Zürich. Available online at http://www.math.ethz.ch.

Mises, R. von (1936), La distribution de la plus grande de n valeurs. Reprinted in *Selected Papers II*, Amer. Math. Soc., Providence, R.I. (1954), 271–294.

Pickands, J. (1975), Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119–131.

Pickands, J. (1981), Multivariate extreme value distributions. Bull. I.S.I. XLIX (Book 2), 859–878.

Resnick, S. (1987), *Extreme Values, Point Processes and Regular Variation.* Springer Verlag, New York.

Robinson, M.E. and Tawn, J.A. (1995), Statistics for exceptional athletics records. *Applied Statistics* **44**, 499–511.

Shephard, N. (1996), Statistical aspects of ARCH and stochastic volatility. In *Time Series Models: In econometrics, finance and other fields.* Edited by D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen. Chapman and Hall, London, pp. 1–67.

Smith, R.L. (1986), Extreme value theory based on the r largest annual events. *J. Hydrology* **86**, 27–43.

Smith, R.L. (1989), Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone (with discussion). *Statistical Science* **4**, 367–393.

Smith, R.L. (1994), Multivariate threshold methods. In *Extreme Value Theory and Applications*, eds. J. Galambos, J. Lechner and E. Simiu. Kluwer Academic Publishers, Dordrecht, pp. 225–248.

Smith, R.L. (1997a), Statistics for exceptional athletics records. Letter to the editor, *Applied Statistics* **46**, 123–127.

Smith, R.L. (1997b), Predictive inference, rare events and hierarchical models. Preprint, University of North Carolina.

Smith, R.L. (1999), Bayesian and frequentist approaches to parametric predictive inference (with discussion). In *Bayesian Statistics 6*, edited by J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith. Oxford University Press, 589–612.

Smith, R.L. and Goodman, D.J. (2000), Bayesian risk assessment. To be published in *Extremes and Integrated Risk Management*, edited by P. Embrechts. Risk Books, London.

Smith, R.L. and Shively, T.S. (1995), A point process approach to modeling trends in tropospheric ozone. *Atmospheric Environment* **29**, 3489–3499.

Smith, R.L. and Weissman, I. (1999), Characterization and estimation of the multivariate extremal index. Tentatively accepted for *Extremes.*

Tawn, J.A. (1988), An extreme value theory model for dependent observations. *J. Hydrology* **101**, 227–250.

Tsay, R.S. (1999), Extreme value analysis of financial data. Preprint, Graduate School of Business, University of Chicago.