

THE LIMITING DISTRIBUTION OF EXTREMAL EXCHANGE RATE RETURNS

MARTIEN C. A. B. HOLS

Nederlands Economisch Instituut, 3006 Rotterdam, The Netherlands

AND

CASPER G. DE VRIES

Center for Economic Studies, Katholieke Universiteit Leuven, B-3000 Leuven, Belg...

SUMMARY

Several nonnested fat-tailed distributions have been advocated for modelling exchange rate returns. Instead of directly estimating these nonnested distributions we investigate the extremal distribution of the returns. The advantage is that the parameter which characterizes the amount of tail fatness can be estimated without maintaining a specific distribution, and hence enables one to test hypotheses. The parameter of the limit law is estimated by employing nonparametric procedures based on order statistics. The appropriateness of these procedures is assessed. Given this estimate one can derive bounds on the returns for very low probabilities on an excess. Such information is useful in evaluating the volatility of exchange rates.

1. INTRODUCTION

Tests of economic theory are predominantly couched in terms of averages, given the prevalence of regression analysis. Nevertheless, some analysis is directed towards extreme occurrences. An example is the qualitative response model. McFadden (1973) shows that the multinomial logit model can be derived from a latent variable model with a stochastic utility function, where the random components follow an extreme value distribution. Actually, economic theory is often formulated in terms of seeking an extreme, i.e. the paradigm of describing economic behaviour as a problem of constrained optimization. Therefore, casting the empirical analysis in terms of statistical extremal analysis can be fruitful; cf. Roy (1952) and Mandelbrot (1963a). In view of this, the aims of the paper are two-fold. First, we draw attention to the uses of extremal analysis for empirical issues in economics. This is done by showing how these statistical techniques can be exploited to shed light on the volatility of foreign exchange rates. Second, we apply recently developed nonparametric techniques that are based on order statistics for estimating the distribution of extremes.

Exchange rates are one of the economic variables for which we believe the theory of extremes has much to offer. The unit root feature of nominal exchange rates notwithstanding, cf. Meese and Rogoff (1983), one can still try to exploit the statistical properties of the residuals to answer interesting economic questions. For example, a much-debated question is whether exchange rates are excessively volatile. Some have interpreted this as excessive volatility relative to the fundamentals, see e.g. Levich (1985) and Marston (1987). Due to, for

example, overshooting phenomena, however, the exchange rate may be more volatile than its determinants, so that excessiveness is hard to judge. Moreover, it is not clear what the volatility of the fundamentals is, as it is unclear how to model these (see Frankel and Meese, 1987). Rather than asking whether exchange rates are excessively volatile relative to their causes, it may be worthwhile to evaluate the volatility relative to its economic consequences. For example, it is often said that excessive exchange rate volatility disrupts trade. Given the other uncertainties in running a business, and the limited possibilities for risk diversification, what matters here are the magnitudes of the changes rather than the average variations. A one-time extremal change in the exchange rate, or a combination of several smaller changes, may drive a firm out of business. One is therefore interested in the distribution of the largest order statistic, M_n say, which gives the likelihood of an extreme realization. For this purpose it is sensible to focus on the tail of the distribution. From the tails one can derive the probability on the occurrence of the maximum exchange rate change which the firm can sustain before going bankrupt. McCulloch (1981) has conducted an analysis along these lines with respect to the interest risk of commercial banks. McCulloch presumes a specific distribution for this risk, from which bankruptcy probabilities are calculated. In an early contribution Roy (1952) proposed the more robust safety-first criterion, by not relying on a specific distribution but by using the Tchebycheff bound. With a little more information one can improve considerably on this bound by exploiting the limit law for the distribution of M_n , which is the approach taken in this paper.

There are good reasons for not relying on a single class of distributions but to use limit laws. There is agreement about the general properties, e.g. that the distribution of exchange rate returns is fat-tailed. With respect to the specific class the distribution belongs to, however, there is no agreement; cf. the survey by Boothe and Glassman (1987). The three fat-tailed alternatives that have received most of the attention are the stable and Student's t distributions and the ARCH process.¹ A comparison between these models is hampered by their nonnestedness. If one concentrates on the distribution of the extremes, however, the main parameter of concern, i.e. the so-called tail index, is nested. The tail index is a good indicator for the mass in the tails, and therefore provides a direct measure for how leptokurtic exchange rate returns are. Hence the estimation of the tail index of the extremal distribution provides information about the underlying distribution and thereby allows for discrimination between the alternative hypotheses. Moreover, given the value of the tail index one can calculate bounds on the returns for low probabilities of an excess (bankruptcy).

In order to be able to pursue the above analysis the tail index and the bounds which are associated with certain low probabilities of an excess have to be estimated. Recently, several estimators have been advanced in the probability literature that are based on order statistics. These estimators have appeal because they explicitly exploit the extremity of the observations, while not presuming that the extremal distribution holds exactly. In addition, the estimators are relatively easy to calculate. To the best of the authors' knowledge these estimators are novel to the economics literature.² This paper's intention is therefore to make extremal analysis available to a wider audience, and to introduce evaluate new estimators that are straightforward to implement. The next section provides the necessary theoretical background, the third section contains the empirical analysis.

¹ Sometimes a thin-tailed distribution is advocated; see, however, the discussion in Section 2.

² Akgiray, Booth, and Seifert (1988) briefly apply maximum-likelihood methods to black market exchange rates. Koedijk, Schafgans, and de Vries (1990) provide an application to some EMS issues, and Kofman and de Vries (1990) study the Dutch Potato Futures Market.

2. THEORY

Consider a stationary sequence X_1, X_2, \dots, X_n of random variables (RV) with a distribution function $F_i(x)$ (d.f. F_i). Suppose one is interested in the probability that the maximum

$$M_n = \max(X_1, X_2, \dots, X_n) \tag{1}$$

of the first n RVs is below a certain level x . This probability is given by

$$P\{M_n \leq x\} = P\{X_1 \leq x, X_2 \leq x, \dots, X_n \leq x\}. \tag{2}$$

If the X_i are independent and identically distributed (i.i.d.) with common d.f. F , then

$$P\{M_n \leq x\} = F^n(x).$$

Extreme value theory studies the limiting distribution of the order statistic M_n appropriately scaled for both i.i.d. and dependent but stationary RVs. That is, one is interested under which conditions there exist suitable normalizing constants $a_n > 0, b_n$, such that

$$P\{a_n(M_n - b_n) \leq x\} \xrightarrow{w} G(x); \tag{3}$$

and in case the X_i are i.i.d.

$$F^n((x/a_n) + b_n) \xrightarrow{w} G(x),$$

where $G(x)$ is one of the three asymptotic distributions that are defined below, and w stands for weak convergence. If (3) holds, we shall say that F belongs to the domain of attraction of G , and write $F \in D(G)$. Define the class of limiting laws which may appear in (3) as follows:

Definition 1. A nondegenerate d.f. G is called *max-stable* if there exist real constants $A_n > 0$ and B_n such that for all real x and $n = 1, 2, \dots$

$$G^n(A_n x + B_n) = G(x).$$

One can show that if (3) obtains, then G is max-stable. Thus for i.i.d. RVs X_i with common d.f. G, M_n scaled has G itself as its limit law. For stationary X_i with d.f. F_i unequal to G, G may still be the limit law of M_n scaled if (3) holds. The main result is the extremal types theorem, which shows that G takes one of the following three parametric forms.

Theorem 1. Every max-stable distribution is of one of the following types

$$\begin{aligned} \text{Type I: } & G(x) = \exp(-e^{-x}) & -\infty < x < \infty; \\ \text{Type II: } & G(x) = 0 & x \leq 0, \\ & = \exp(-x^{-\alpha}) & x > 0; \\ \text{Type III: } & G(x) = \exp(-(-x)^\alpha) & x < 0, \\ & = 1 & x \geq 0; \end{aligned} \tag{4}$$

with the index $\alpha > 0$. This index α is called the *tail index*, and for notational convenience its inverse $\gamma = 1/\alpha$ is used as well.

Proofs of the theorem and the other claims can be found in Leadbetter, Lindgren, and Rootzén (1983), which provides a comprehensive treatment, and also discusses the case of dependent variates. In the latter case the scaling parameter a_n has to be modified by a constant multiplicative factor $\theta, 0 < \theta \leq 1$, called the *extremal index*, but the tail index α is unaffected.

Consider, for example, the simple infinite MA process (AR 1 process): $X_t = \sum_{j=0}^{\infty} \lambda^j \varepsilon_{t-j}$, where the ε_t are i.i.d. symmetric stable variates with characteristic exponent β , $1 < \beta < 2$, and $0 < \lambda < 1$. The stable RVs, (not to be confused the max-stable variates defined above) are defined by the property that their d.f. are invariant under addition. As is shown below, for i.i.d. stable RVs the tail index α equals the characteristic exponent β . The effect of the dependency is that the larger values tend to come in clusters. This affects the norming constants but not the tail index. The extremal index thins the observations to make the series comparable to an i.i.d. series with the same tail index. Specifically, the limit law in (4) becomes $\exp(-\theta x^{-\alpha})$, where $\theta = 1 - \lambda^\alpha$, and $\alpha = \beta$ (cf. Leadbetter *et al.* 1983, ch. 3). As we focus primarily on α , most of the results below do carry over to dependent variates, but this is not made explicit everywhere for considerations of space.

The limit in (3) explicated in (4) is most easily interpreted by analogy with the central limit theorem. The difference is the focus on order statistics rather than averages, but its usefulness is the same as no detailed knowledge of the d.f. F is needed to apply the asymptotic theory. A complication is the fact that there are three limit laws. Usually economic theory is not informative about the specific d.f. F that applies. However, the qualitative characteristics of the economic process may point to the relevant limit law. Consider the following two necessary conditions from De Haan (1976).

Condition 1. If $F \in D$ (type I $G(x)$) and $F(x) < 1$ for all x , then $\int_1^\infty t^k dF(t)$ is finite for all k .

Condition 2. If $F \in D$ (type II $G(x)$), then $F(x) < 1$ for all x and $\int_1^\infty t^k dF(t)$ is finite for $k < \alpha$ and infinite for $k > \alpha$.

The intuition behind these conditions is as follows. Loosely speaking, the tail of the distribution is either declining exponentially or by a power. In the first case all moments exist, but in the second case the higher moments do not exist as t^k does not decay rapidly enough when 'weighted' by the tail probabilities, i.e. the d.f. F is fat-tailed. This explains the appearance of the double exponential in the type I limit law and the single exponential format of the type II law as well. The third limit law is characterized by the fact that it has a finite upper endpoint. Anticipating on the next section, given that exchange rate returns are strongly fat-tailed, and unbounded in principle, the type II limit law is the relevant one, if the maximum yield distribution converges at all.

In the case of an i.i.d. sequence a sufficient condition on the d.f. F for the type II limit to obtain is:

Condition 3. It is sufficient for $F \in D$ (Type II $G(x)$) that it has no finite upper endpoint, and for each $x > 0$ and some $\alpha > 0$

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}.$$

The latter condition boils down to regular variation at infinity, see e.g. Feller (1971, ch. VIII.8); see also Mood, Graybill, and Boes (1974, page 261, th. 16).

The following discussion shows how the above conditions may be employed in specific cases. Return to the d.f. values F that have received most of the attention in the literature on exchange rates. Essentially two classes of d.f. values have been employed for modelling the (un-)conditional d.f. of exchange rate returns. The first class contains variations of the normal

d.f.; e.g. discrete mixtures of the normal d.f., mixed diffusion jump processes, and the power exponential d.f. In comparison with the normal d.f. these models all exhibit a higher kurtosis, but nevertheless possess all moments. Hence, Condition 2 cannot be satisfied, while Condition 1 could apply. In fact by Theorem 1.6.1 in Leadbetter, Lingren, and Rootzén (1983), it can be shown that all these models are in the domain of attraction of the type I limit law, just like the normal d.f. itself. It follows, *inter alia*, that the higher than normal kurtosis does not necessarily imply fat tails. Therefore these alternatives seem unfit for modelling exchange rate returns.

Now turn to the other class of d.f. which have been used in the literature. These are the Student's t and stable d.f. and the ARCH process. As each of the three models may not possess all unconditional moments, Condition 2 could apply. A sufficient condition for the i.i.d. sequences is regular variation. To verify that the Student's t class satisfies Condition 3 is straightforward (see, for example, Mood, Graybill, and Boes (1974, page 262) for a proof. Note that the degrees of freedom ν are equal to the tail index α in the type II d.f. of equations (4). The proof that the fat-tailed stable distribution satisfies Condition 3 takes some effort, as only in specific cases does a closed form solution for its density $f_\beta(t)$ exist (i.e. the normal, the Cauchy, and the inverted chi-square, for which the so-called characteristic exponent β is respectively 2, 1 and 1/2). However, Ibragimov and Linnik (1971) provide the following asymptotic formula for large t in the case $0 < \beta < 2, \beta \neq 1$:

$$f_\beta(t) = \frac{-1}{\pi t} \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \sin(\frac{1}{2}n\pi\beta)\Gamma(n\beta + 1)t^{-n\beta}. \tag{5}$$

Note, the β in (5) equals the α in (4). Condition 3, and using de l'Hôpital's rule, show that $\lim x f_\beta(tx)/f_\beta(t) \rightarrow x^{-\beta}$ as $t \rightarrow \infty$. The Cauchy case, i.e. $\alpha = 1$, follows directly from the discussion of the Student's t d.f.

For the ARCH process Condition 3 is of no avail, due to the interdependency between the realizations. The case of ARCH is of theoretical interest because even though the building blocks are normal variates, yet the unconditional distribution of the realizations is fat-tailed. The intuition behind this fact can be obtained from the simplest ARCH scheme: $\varepsilon_t = X_t(\beta + \lambda\varepsilon_{t-1}^2)^{1/2}$, where $\{X_t\}$ are i.i.d. $N(0, 1)$ RVs, $\beta > 0, 0 < \lambda < 1$. Squaring this relation gives the first-order stochastic difference equation $\varepsilon_t^2 = \beta X_t^2 + \lambda X_t^2 \varepsilon_{t-1}^2$. By a result in Kesten (1973), if there exists a $\chi > 0$ such that $E[\lambda X^2]^\chi = 1$, and some other regularity conditions are satisfied, then the RV S which is the stationary solution to the random difference equation has a Pareto-type tail $cs^{-\chi}$ for large s . Using appropriate mixing conditions for tail events, which are considerably weaker than e.g. strong mixing,³ one can then show, (see Theorem 2.1 of De Haan *et al.*, 1989), that $\exp(-\tilde{\theta}x^{-\chi})$ and $\exp(-\theta x^{-\alpha})$, where $\alpha = 2\chi$, are the limit laws for the scaled M_n from the process $\{\varepsilon_t^2\}$ and the ARCH(1) process $\{\varepsilon_t\}$ respectively. Closed form expressions for the extremal indices $\tilde{\theta}$ and θ are unknown, but the values can be obtained numerically. The tail index α directly follows from the condition $E[\lambda X^2]^\chi = 1$, i.e. from solving $\Gamma(\chi + 1/2) = \pi^{1/2}(2\lambda)^{-\chi}$ for χ . Table I provides some λ, θ, α combinations. Note the inverse relation between λ on the one hand and θ, α on the other hand. Formal generalizations to higher-order ARCH or GARCH processes are as yet nonexistent, but some generalizations can be easily obtained. For example if the innovations $\{X_t\}$ are fat-tailed as well, and are Student's- t distributed with $\nu > 2$, then it is straightforward to verify that Theorems 1.1 and 2.1 of De Haan *et al.* [1989] do hold.

³ Leadbetter, Lindgren, and Rootzén's (1983, page 53) condition $D(u_n)$ is used. Condition $D(u_n)$ applies if for a real sequence $\{u_n\}$ and for any integers $1 \leq i_1 < \dots < i_p < j_1 < \dots < j_k \leq n$ for which $j_1 - i_p \geq m$, we have $|F_{i_1 \dots i_p, j_1 \dots j_k}(u_n) - F_{i_1 \dots i_p}(u_n)F_{j_1 \dots j_k}(u_n)| \leq \alpha_{n,m(n)}$, where $\alpha_{n,m(n)} \rightarrow 0$ as $n \rightarrow \infty$ for some sequence $m(n) = o(n)$.

Table I. ARCH parameters

λ	θ	$\alpha = 2\lambda$
0.1	0.999	26.48
0.5	0.835	4.73
0.9	0.612	2.30
0.99	0.571	2.02

From the above discussion it is obvious that the competing fat-tailed d.f. are nested within their limit law $G(x)$, and are distinguished by different values for the tail index α . Specifically, the leptokurtic stable hypothesis requires $\alpha < 2$, while the Student's t and ARCH(1) process allow for $\alpha \geq 2$. The idea is now to estimate α directly without prior commitment to either hypothesis.

Broadly speaking, the estimation procedures for α fall into two categories. A traditional approach uses 'yearly maxima' and assumes that each period's maximum exactly follows one of the three limit laws. If the type II limit law applies, direct estimation by maximum-likelihood is consistent. A drawback of this method is that the excesses are assumed to follow the limit law exactly, whereas this is only approximately the case (cf. Akgiray and Booth, 1988). Therefore maximum-likelihood may not be the most efficient procedure, (see e.g. Smith, 1987); in addition, it has the drawback that an extra scale parameter has to be estimated. Recently, some nonparametric estimators have been proposed based on the largest-order statistics, which require only that the distribution generating these observations is in a sense well-behaved. This implies that the remaining estimation error can be solely attributed to the use of finite samples. For example, regular variation at infinity is often a sufficient condition. For this reason the focus here is on these methods.

Suppose X_1, X_2, \dots, X_n is a stationary sequence such that M_n has a type II $G(x)$. Define $X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(n)}$ as the descending order statistics from the sample X_1, \dots, X_n . Table II presents three alternative estimators for γ based on the largest-order statistics $X_{(i)}$. Recall that $\gamma = 1/\alpha$.

Table II. Tail index estimators

$$\hat{\gamma}_P = \left[\log \frac{X_{(m)} - X_{(2m)}}{X_{(2m)} - X_{(4m)}} \right] / \log 2$$

$$\hat{\gamma}_H = \frac{1}{m-1} \sum_{i=1}^{m-1} \log X_{(i)} - \log X_{(m)}$$

$$\hat{\gamma}_R = [\log X_{(1)} - \log X_{(m)}] / \log m$$

The following properties of the estimators have been established for i.i.d. sequences. Pickands (1975) first proposed $\hat{\gamma}_P$ and obtained its weak consistency. If $m(n)$ increases suitably rapidly, Dekkers and De Haan (1989) obtain strong consistency and derive the asymptotic normality of $(\hat{\gamma}_P - \gamma)m^{1/2}$ with zero mean and variance $\gamma^2(2^{2\gamma+1} + 1)/\{2(2^\gamma - 1)\log 2\}^2$. The statistic $\hat{\gamma}_H$ first appears in Hill (1975). Mason (1982) proves that if Condition 3 is satisfied, $\hat{\gamma}_H$ is a consistent estimator for γ . By results in Hall (1982) and Goldie and Smith (1987), it follows that $(\hat{\gamma}_H - \gamma)m^{1/2}$ is asymptotically normal with mean zero and variance γ^2 . De Haan and Resnick (1980) derive $\hat{\gamma}_R$ as a consistent estimator for γ . For non-i.i.d. but stationary sequences, $\hat{\gamma}_H$ and $\hat{\gamma}_R$ can be shown to be weakly consistent if the mixing condition $A(u_n)$ of

Watts, Rootzén, and Leadbetter (1982) holds, and the marginal d.f. satisfy Condition 3, by exploiting Theorem 2.4 of Watts *et al.*

Pickand’s estimator is more general than the other two. It provides an estimator of the extremal index for all three type limit laws, whereas the other two are only applicable in case the type II limit law holds. The intuition behind these estimators is best illustrated by discussing the third estimator $\hat{\gamma}_R$. Let $F_n(x)$ denote the empirical d.f. For the m th descending order statistic the left limit of $F_n(X_{(m)})$ equals $(n - m)/n$. By the Glivenko–Cantelli Theorem we may replace the probability in equation (3) by $(n - m)/n$ and combine this with the type II limit law in expression (4) with x appropriately scaled. Note, for a type II $G(x)$ one can always set $b_n = 0$. Upon taking logarithms twice, one gets for the m th-order statistic

$$\log \left[-\log \frac{n - m}{n} \right] = c - \alpha \log X_{(m)}, \tag{6}$$

where c is a constant depending on a_n, θ, α and $b_n = 0$. For large n and small m/n we may approximate the left-hand side of equation (6) by $\log m$. Estimate the slope α by taking two points on this line, say $m = 1$ and $n > m > 1$. Simple trigonometry then generates $\hat{\gamma}_R$ of Table II.

The estimation procedures require $m(n) \rightarrow \infty$, but for a finite sample it is not known how to choose m optimally.⁴ A heuristic procedure is to compute $\hat{\alpha}$ for different m and to select an m in the region over which $\hat{\alpha}$ is more or less constant. There exists such a region, because when one uses insufficient order statistics, then α will vary heavily with m due to inefficiency. In the opposite case the curvature of the distribution F generating the data weighs too heavily, i.e. only the tail probabilities are well approximated by the limit distribution G . In the empirical section we conduct a Monte-Carlo study to select m . Due to the asymptotic normality of $\hat{\gamma}$ in two out of the three cases, the MSE criterion may be used for selecting an optimal m for given sample size n and d.f. F . This procedure is not entirely satisfactorily as the optimal m varies (inversely) with the true α value. But for at least two reasons it is best to choose m conservatively: first, to ensure consistency and asymptotic normality m should not be taken too high, (see Hall, 1982); second, Dekkers and De Haan (1989) suggest thinning the number of observations employed for estimation in order to reduce the effects of dependency; see also footnote 8, below.

Once the index α is estimated it can be used to establish extreme returns that rarely occur. Fix a high return level (quantile) $x_p, x_p > 0$. We say an *exceedance* occurs if $X_i > x_p$, and define $Y_i = X_i - x_p$ as the corresponding *excess*. The subscript p of x_p refers to the probability of an exceedance. Precisely formulated, we ask which return level x_p corresponds to a given low probability p of an exceedance, or equivalently, of an excess. Again, the procedures for establishing such levels can be classified as to whether or not it is assumed that the extremes exactly follow one of the limit laws. It is more elegant not to assume that the extreme observations exactly follow the extreme value distribution. This is in the same spirit as the given estimators. The following procedure requires regular variation, an i.i.d. sequence X_1, \dots, X_n , and uses order statistics to estimate the \hat{x}_p such that

$$P\{X_1 \leq \hat{x}_p, \dots, X_k \leq \hat{x}_p\} = F^k(\hat{x}_p) = 1 - p, \tag{7}$$

for small p and given k .

Recall that $F_n(x)$ denotes the empirical d.f. Define the functions (the ‘-1’ superscript

⁴This is the case for both the maximum-likelihood and the order statistics methods.

denotes the inverse function)

$$U(x) = \left[\frac{1}{1-F} \right]^{-1}(x), \quad U_n(x) = \left[\frac{1}{1-F_n} \right]^{-1}(x), \quad a_k(p) = \frac{1}{1-(1-p)^{1/k}}.$$

From equation (7) and the definition of U it follows that $x_p = U(a_k(p))$. Also, for p small and by taking logarithms it may be seen that $a_k(p) \approx k/p$. Hence, for small p , $U(k/p) = x_p$. For the r th ascending-order statistic $X_{(r)}$, from $X_{(1)} \leq X_{(2)} \dots \leq X_{(n)}$, $F_n(X_{(r)}) = r/n$, so that $X_{(n-r)} = U_n(n/r)$. Assume that $F \in D$ (type II $G(x)$), then

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{U(ty) - U(t)} = \frac{x^\gamma - 1}{y^\gamma - 1},$$

see for example De Haan (1984). For $p < 1/n$, the idea is to extend the empirical distribution function by using this limit. Consider the identity

$$x_p = U(a_k(p)) = \frac{U(a_k(p)) - U(n/r)}{U(n/r) - U(n/2r)} (U(n/r) - U(n/2r)) + U(n/r).$$

Replace the first term on the right-hand side as $n/r \rightarrow \infty$, by the two limits above, and use the empirical counterpart of U_n . We arrive at the following consistent estimator

$$\hat{x}_p = \frac{(kr/pn)^{\hat{\gamma}} - 1}{1 - 2^{-\hat{\gamma}}} (X_{(n-r)} - X_{(n-2r)}) + X_{(n-r)}, \quad (8)$$

where n is the number of observations, k is the time period considered, $r = m/2$, m is the lowest descending-order statistic $X_{(m)}$ used in computing $\hat{\gamma}$, p is the probability of an excess, and $\hat{\gamma}$ is a consistent estimator of γ . The proof of consistency of \hat{x}_p when $\gamma < 1/2$ is given in Dekkers and De Haan (1989). No results are available at present for non-i.i.d. sequences.

An heuristic interpretation of formula (8) is as follows. The pattern of the empirical d.f. $F_n(x)$ as signified by the level $X_{(n-r)}$ and step size $(X_{(n-r)} - X_{(n-2r)})$ is extrapolated outside its domain by using the way the limit law extends. The latter is represented through the multiplication factor in front of the step size. For $p > 1/n$ the inverse of the empirical d.f. $F_n(x_p)$ is a good estimator for x_p , due to its mean squared error consistency (see Mood, Graybill, and Boes, 1974, page 507). But for $p < 1/n$, $F_n(x)$ is of no avail, and the above is a device to extend $F_n(x)$ beyond $1/n$ and use it for estimating x_p .⁵

3. EXTREMAL FOREIGN EXCHANGE RATE RETURNS

In this section the theory of extremes is applied to foreign exchange rate data. The data consist of weekly Thursday closing quotations of the Canadian–United States dollar spot exchange rate. The sample covers the period starting the last week of 1973 and ending the fourth week of 1983; a total of 475 observations. By now there is consensus about the empirical regularities of nominal exchange rates. In the short run returns rather than exchange rate levels are the variable of interest, because investors focus on returns. Survey articles by Mussa (1979) and Levich (1985) collect the following stylized facts: (i) the logarithm of the spot rate is nonstationary, while the first differences are stationary; (ii) sample moments of the empirical return distribution indicate absence of skewness, but the kurtosis points to fat tails; (iii) periods of quiescence and turbulence tend to cluster together. In the literature on the unconditional distribution of the returns the fat tail property has attracted most of the

⁵ A less precise alternative method, given the fat tail phenomenon, consists in using the Techebycheff bound.

Table III. Unit root tests*

	τ	τ_μ	$Z(l=2)$	$Z(l=10)$	$Z^*(l=2)$	$Z^*(l=10)$
Log-levels	0.823	-0.068	0.672	1.194	-1.005	-0.849
Returns	-12.106	-12.618	-19.201	-18.664	-19.377	-19.065

The test statistics are the Dickey-Fuller statistics τ , τ_μ for the models $x_t = c + x_{t-1} + \beta(x_{t-1} - x_{t-2})$ with c zero or unrestricted, and the Phillips statistics $Z(l)$, $Z^(l)$ for the model $x_t = c + x_{t-1}$ with c zero or unrestricted and truncation lag l . The variable x_t refers to either the log exchange rate level or first differences.

Table IV. Unit root simulations*

Test statistic	Distribution	Probability of a smaller value	
		0.025	0.975
τ	Normal	-2.26	1.66
τ	Cauchy	-1.98	1.36
τ_μ	Normal	-3.12	0.23
τ_μ	Cauchy	-3.85	0.28

*The table corresponds to Table 8.5.2 in Fuller (1976) for the normal distribution and sample size $n = 474$; in addition it gives the critical values if the innovations are Cauchy distributed. The table is based on 10,000 replications.

attention. Because of the additivity of returns, think of the triangular arbitrage condition, and the observed fatness in the tails, Mandelbrot (1963a,b) proposed the class of stable distributions for modelling return data. This class is the only class whose distributions are invariant under addition and fat-tailed.⁶ If one is less strict about the additivity requirement, other distribution classes become available. For example, the Student's t distributions with finite degrees of freedom ν have been used, as these are heavy-tailed as well but do possess a finite mean and variance for $\nu > 2$. The advantage is that traditional CAPM analysis, for example, relies on finite variances and is therefore compatible with the Student's t distributions. The disadvantage is the loss of additivity. A feature which has not been exploited in the economics literature, however, is that additivity in the tails may nonetheless be preserved. The following result is due to Feller (1971, ch. VIII.8).

Theorem 2. Consider an i.i.d. sequence X_i with a common d.f. $F(x)$. If $1 - F(x)$ varies regularly at infinity, i.e. satisfies Condition 3, then the M_n from $F(x)$ or any finite convolution of $F(x)$ follow the same limit law.

Hence, if only summability for large returns is imposed, then the Student's t distribution is a reasonable alternative. Note that Theorem 2 complements the central limit law. If $\nu > 2$, normed sums of i.i.d. Student's t variates tend to be normal distributed, while normed maxima of just sums of Student's t variates follow a type II $G(x)$ limit law.

The empirical properties of our data set were investigated in a number of ways. We first

⁶The normal distribution is the exception, as it is not fat-tailed. All other stable distributions only have a finite mean, i.e. if $\beta > 1$, or no moments at all if $\beta \leq 1$.

investigated the stationarity of the log exchange rate levels and their first differences (returns). Dickey–Fuller test statistics, (see Fuller, 1976), and Phillips’ tests (see Phillips, 1987), reported in Table III all indicate that nonstationarity for the levels cannot be rejected, while the returns appear to be stationary. Given that we want to allow for infinite variances, because of the stable model, one wonders whether these tests are appropriate. Chan and Tran (1989), however, show that the Dickey–Fuller test procedure is appropriate for stable variates, although the critical values of Table 8.5.2 in Fuller (1976) may be too conservative. Therefore we simulated with the Cauchy d.f. to obtain accurate critical values; results are reported in Table IV. For the normal d.f. our results closely correspond to those reported by Dickey and Fuller. Note that the confidence interval for Cauchy variates is tighter for the model without intercept, i.e. the H_0 model, and larger for the misspecified model with intercept in comparison with normal variates. The intuition is that with more extreme realizations the estimates of a correctly specified model converge more rapidly, while under misspecification the extremes deter convergence. Phillips (1989), moreover, has recently shown that his (1987) $Z(I)$ statistic also applies to stable variates. Given this latter result and the simulations reported in Table IV, it follows that the evidence reported in Table III leads to the same conclusions whether the innovations are normally or nonnormal stable distributed.

We also calculated a number of summary statistics to investigate symmetry and kurtosis in the returns. These confirmed the usual findings, such as reported in Boothe and Glassman (1987), that exchange rate returns are by and large symmetric and fat-tailed, and are omitted for considerations of space. To enhance the reader’s intuition for the fat-tail property, however, we plotted the first two sequential moments of the returns in Figures 1 and 2 (see Mandelbrot (1963b) for this technique). While the mean clearly stabilizes, the variance tends to wander away as the sample size increases. This suggests that the variance could be unbounded.

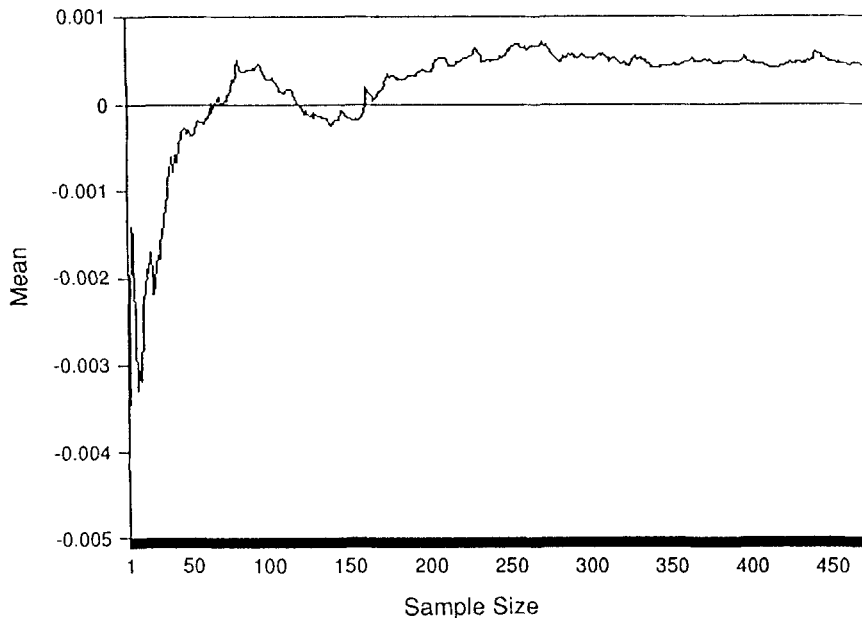


Figure 1. Sequential mean

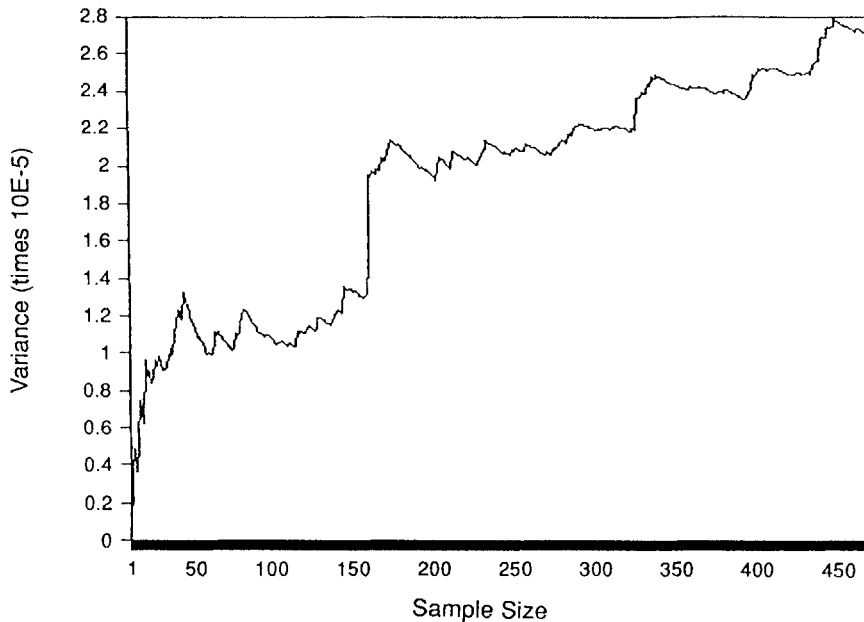


Figure 2. Sequential variance

The sequential moment plots and summary statistics point to a fat-tailed d.f. For the volatility debate the interesting problem is to determine in a robust manner how fat-tailed the d.f. is. In order to achieve this we propose to apply extreme value analysis and to estimate the inverse tail index γ . The estimators of Table II are conditional upon the portion m/n of the sample used for calculating the statistic $\hat{\gamma}$. As it is not known how to choose m in finite samples, we conducted a Monte-Carlo experiment in order to establish the m -level, conditional upon a sample size n , for which the MSE is minimal. This criterion seems appropriate given the asymptotic normality of $\hat{\gamma}$. As our sample return data comprise 474 observations, we set $n = 474$. For the i.i.d. case we simulated with three different d.f. values: the Student's t distribution with 1, 2 and 3 degrees of freedom respectively.⁷ In all cases an intermediate range of m -levels was found where the MSE attained its minimum, confirming the predicted U-shaped relationship. Table AI (in the Appendix) gives the minimizing MSE m -levels. The m -levels are seen to vary inversely with the true tail index α . The reason is that the lower α is, the fatter are the tails of the distribution, and hence the more 'outliers' are available for estimation. The MSE differ considerably per estimator. In particular, $\hat{\gamma}_H$ is the superior estimator while $\hat{\gamma}_P$ appears inferior. The dominance of $\hat{\gamma}_H$ over $\hat{\gamma}_R$ can be understood from a result by Hall and Welsh (1984), who show that the optimum asymptotic convergence rates of the two estimators are algebraic and logarithmic respectively (see also Hall, 1982, page 37). Inspection of plots of $\hat{\gamma}_P$ against m revealed that it varies wildly, due to the fact that only four dispersed-order statistics are used in the calculations *vis-à-vis* the other statistics. Theoretically, moreover, it is easily shown that the MSE of $\hat{\gamma}_H$ is always below the MSE of $\hat{\gamma}_P$ for any given m and positive γ . *Inter alia*, this also holds for the maximum-likelihood procedure which is less efficient than the $\hat{\gamma}_H$ by a factor $(1 + \alpha)^2$ (cf. Dumouchel, 1983, page 1028). To investigate the effects of dependency we also conducted a Monte-Carlo experiment for the ARCH(1)

⁷Note that the first distribution is the Cauchy distribution, which is also a member of the stable d.f.

process (see Table AII). The results are similar to the i.i.d. case; the optimal m -levels do not vary appreciably (given the same α values). Because previous estimates for α indicated values in the interval (1, 5), and for the reasons mentioned in Section 2, we decided to use the lower m -values. As the m -levels also vary with the type of estimator, we selected the following three m -values 30, 20 and 50 for $\hat{\gamma}_P$, $\hat{\gamma}_H$ and $\hat{\gamma}_R$ respectively.

The tail index estimates and its inverse, i.e. α and γ respectively, are reported in Table V. While all point estimates for α are above 2, they are not significantly different from values below 2 at the 5 per cent significance level, i.e. γ above 1/2 cannot be rejected. However, if one disregards $\hat{\gamma}_P$ and $\hat{\gamma}_R$ because these estimators are less efficient than $\hat{\gamma}_H$, then the one-sided hypothesis $H_0: \gamma \geq 0.5$ is rejected at the 5 per cent level on basis of $\hat{\gamma}_H$. It follows that of the three nonnested alternatives the evidence weakly favours a Student's t or ARCH(1) model *vis-à-vis* the stable model. (Note that for the ARCH(1) model $\alpha = 3$ implies a $\lambda = 0.73$). The conclusion is that the unconditional d.f. of the return data is heavily fat-tailed, but that the second moment may just be defined.⁸

Although predictions about future exchange rates seem misplaced due to the unit root feature, one can try to exploit information about the distribution of the returns. In this respect extreme value theory has a lot to offer apart from the test between the Student's t and ARCH models *vis-à-vis* the stable model. In the short run, only large disturbances evoke interest. Large swings cause large gains and losses, and impair normal business. For this reason most central banks stand ready to intervene in order to provide for 'orderly market conditions', private banks delimit their open positions, and futures markets clearing houses protect themselves by a marking to market settlement of accounts on a daily basis. The focus is therefore on large movements, and extreme value theory provides information about the likelihood of excess returns. Table VI gives excess probabilities p of high returns x_p on the basis of formula (8). Specifically, it gives the probability of observing a weekly return above a level x_p over a period of $k = 100$ weeks. As we did not observe any weekly returns in our sample above the 3 per cent, the empirical d.f. is not informative for returns exceeding this level. Here, extreme value theory can be usefully employed.

Table VI should be interpreted in the following way. For example, observing at least 1 week with a yield above the 8 per cent in any 2-year period ($k = 100$) carries a probability of about

Table V. Estimates of the tail index*

Estimators	$\hat{\gamma}$ -values	95% confidence interval for γ	$\hat{\alpha}$ -values
$\hat{\gamma}_P$	0.14	-0.48-0.90	6.97
$\hat{\gamma}_H$	0.31	0.21-0.54	3.28
$\hat{\gamma}_R$	0.39	not available	2.58

* Estimates are conditioned on $m = 30, 20$ and 50 for $\hat{\gamma}_P, \hat{\gamma}_H$ and $\hat{\gamma}_R$ respectively.

⁸ The estimators were also evaluated at different choices for m . While this hardly affected $\hat{\gamma}_H$ and $\hat{\gamma}_R$, $\hat{\gamma}_P$ changed drastically. For example taking $m = 25$, we calculated $\hat{\gamma}_P = 0.49$, but for $m = 100$ the $\gamma_H = 0.56$ and $\gamma_R = 0.42$. To be able to use the asymptotic confidence intervals one needs the i.i.d. assumption with regard to the highest observations. To ensure this, we followed the informal procedure suggested in Dekkers and De Haan (1989) and removed the two observations adjacent to each extreme. Such a procedure leads to a thinning of the data, which is what the extremal index θ signifies in the reformulation of Theorem 1 for dependent sequences provided condition $D(u_n)$ holds (see footnote 3). This gave the following tail estimates: 3.29, 3.70 and 2.59 using the estimators of Table II and m -values of 27, 19 and 43 respectively. It seems that the second and third estimator are robust against possible dependency in the data.

Table VI. The probability on an excess return

Method	Probabilities p			Level \hat{x}_p
	$\hat{\gamma}_P$	$\hat{\gamma}_H$	$\hat{\gamma}_R$	
	0·0208	0·0805	0·2071	0·04
	0·0026	0·0230	0·0783	0·06
	0·0005	0·0093	0·0387	0·08
	0·0001	0·0046	0·0223	0·10
	0·0000	0·0005	0·0039	0·20

* Each of the three point estimates of γ reported in Table IV were used in formula (8) to generate (p, \hat{x}_p) combinations. The other two parameters were set at $k = 100$ and $r = m/2$. Calculations are based on the presumption that the higher-order statistics are approximately i.i.d.

0·1 on the basis of Hill's estimate of the tail index. There is a notable difference between the first column in Table VI and the other two. We experimented with different r -values in formula (8), but this hardly affected the generated probabilities. On the other hand, the outcomes were very sensitive to variations in $\hat{\gamma}$, explaining the differences in Table VI. Still, all columns carry the message implied by fat-tailed distributions. To get a feeling for the effects of fat-tailed distributions, we computed the (p, \hat{x}_p) combinations maintaining that the d.f. F generating the yields is normal, using the sample mean (0·000452) and sample variance (0·000027). We know that normality is untenable due to the high kurtosis (5·931), but still the assumption of normality is widely used. For a 1·3 per cent return the associated probability of an excess is 0·61. For a 5 per cent return this probability is already less than 4×10^{-10} ! Hence, the tail fatness weighs heavily and the assumption of normality severely underestimates the probability on large yields. (On the other hand the Tchebycheff bound severely overestimates these probabilities; e.g. for $x_p = 0·20$ the bound on p is 0·0678). This may prove to be useful information to managers of foreign funds, among whom the assumption of normality is commonly made.

In the introduction we briefly discussed the risk incurred by a firm which has some foreign exchange exposure. To be more specific, consider the problem faced by a banker who has to decide on the maximum period over which he allows his foreign currency dealers to hold an open position. Such limits are set in order to delimit possible losses. Similarly, on futures markets positions are settled on a daily basis to circumvent the possibility of insolvency at the maturity date. One might argue that, instead of one heavy loss, bankruptcy can arise through a sequence of smaller losses. As the example of the futures market shows, however, the clearing house can cancel a contract on any day a party is unable to meet her margin requirement. The clearing house then takes over the contract and reverses her open position the next trading day. This prevents a possible accumulation of losses if settlement were less frequent. The question is how to choose this frequency or how to determine the maximum holding period. Suppose the commercial bank can sustain a certain loss associated with a return x without impairing its normal business. Moreover, the management accepts there is always a possibility that actual losses may be larger, but it allows this only to happen with a very small probability p . Table VII provides the information on the maximum period length, given \hat{x}_p and p .

Table VII. High return levels for different periods*

Probability	Level \hat{x}_p		
	Week	Fortnight	Month
$p = \frac{1}{500}$	0·030	0·047	0·069
$p = \frac{1}{5000}$	0·063	0·100	0·145

* In computing these levels we used $\hat{\gamma}_H = 0\cdot305$; $r = 10$; $n = 474$ for the weekly, $n = 237$ for the fortnightly, and $n = 118$ for the monthly periods: k was set equal to 1. Calculations are based on the presumption that the higher-order statistics are approximately i.i.d.

Table VII gives high return levels for given p such that

$$1 - p = P\{X_1 \leq \hat{x}_p\}, \text{ or}$$

$$1 - p = P\{X_1 + X_2 \leq \hat{x}_p\}, \text{ or}$$

$$1 - p = P\{X_1 + X_2 + X_3 + X_4 \leq \hat{x}_p\},$$

and where the X_i are the weekly yields. To arrive at the biweekly and monthly order statistics, we formed nonoverlapping biweekly and monthly yields. The importance of Theorem 2 should now become clear, as it suggests we may use the same $\hat{\gamma}$ for the different frequency data. Hence, $\hat{\gamma}$ does not need to be re-estimated for the lower frequency data which necessarily contain less observations.

To interpret Table VII, suppose one of the bank's dealers holds an open position of one million dollars. Moving from a maximum holding period of 1 week to 2 weeks adds, with probability $1 - p = 0\cdot998$, at the most a loss of 1·7 per cent or 17,000 dollars in nominal value. But moving to a month would increase the possible losses to 3·9 per cent or 39,000 dollars. The former increase may be regarded acceptable, the latter not.

5. CONCLUSIONS

This paper addresses three topics: (i) the use of extremal analysis for analysing exchange rate volatility, (ii) nonparametric estimation of the tail index, and (iii) estimating bounds on the returns which are rarely exceeded. By focusing on extremal exchange rate returns one may use the limiting law G instead of the d.f. F that generates these yields. On the one hand this implies a loss, on the other hand it constitutes a gain as the different hypotheses about F considered in the literature now appear as nested hypotheses about the tail index α . This index, which indicates the amount of tail fatness in the returns, was estimated by means of robust techniques which are based on order statistics. These have the advantage that the extreme observations do not have to follow the limit law exactly, and, moreover, they are easy to compute. Given an estimate of the tail index α , we computed probabilities on extremal yields which are outside the range of the observed frequencies. This may be more informative than reporting a summary statistic like the variance, as it is the magnitude of the exchange rate changes that has economic implications. Lastly, we computed return levels for different frequency data, given a prespecified low probability on an excess. Such information may be used, for example, to fix a maximum period for holding open positions.

ACKNOWLEDGEMENTS

We are extremely grateful to Laurens de Haan for his expertise, to Guido Imbens, Tony Lancaster, Adrian Pagan, and the anonymous referees for their perceptive comments. Philip Stork and Geert Gielens provided able research assistance. We benefited from seminars given at Indiana University, Purdue University, Texas A&M University, Brown University, and the European Meetings of the Econometric Society of Bologna. The data for this paper were kindly supplied to us by David Longworth from the Bank of Canada. The paper contains the first results of our project on extreme value analysis. Support for this work by Erasmus Universiteit Rotterdam is gratefully acknowledged.

REFERENCES

- Akgiray, V., G. G. Booth and B. Seifert (1988), 'Distribution properties of Latin American black market exchange rates', *Journal of International Money and Finance*, 37–48.
- Boothe, P., and D. Glasmann (1987), 'The statistical distribution of exchange rates: empirical evidence and economic implications', *Journal of International Economics*, 297–320.
- Chan, N. H., and L. T. Tran (1989), 'On the first-order autoregressive process with infinite variance', *Econometric Theory*, 354–362.
- Dekkers, A. L. M., and L. de Haan (1989), 'On the estimation of the extreme-value index and large quantile estimation', *Annals of Statistics*, 1795–1832.
- Dumouchel, W. H. (1983), 'Estimating the stable index α in order to measure tail thickness: a critique', *Annals of Statistics*, 1019–1031.
- Feller, W. (1971), *An Introduction to Probability Theory and its Applications*, vol. 2, Wiley, New York.
- Frankel, J. A., and R. Meese (1987), 'Are exchange rates excessively variable?', in S. Fischer (ed.), *NBER Macro Economics Annual*, MIT Press, Cambridge, pp. 117–162.
- Fuller, W. A. (1976), *Introduction to Statistical Time Series*, Wiley, New York.
- Goldie, C. M., and R. L. Smith (1987), 'Slow variation with remainder: theory and applications', *Quarterly Journal of Mathematics* (Oxford 2nd series), 45–71.
- Haan, L. de (1976), 'Sample extremes: an elementary introduction', *Statistica Neerlandica*, 161–172.
- Haan, L. de (1984), 'Slow variation and characterization of domains of attraction', in J. Tiago d'Oliveira (ed.), *Statistical Extremes and Applications*.
- Haan, L. de, and S. I. Resnick (1980), 'A simple asymptotic estimate for the index of a stable distribution', *Journal of the Royal Statistical Society, Series B*, 83–87.
- Haan, L. de, S. I. Resnick, H. Rootzén and C. G. de Vries (1989), 'Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH-processes', *Stochastic Processes and their Applications*, 213–224.
- Hall, P. (1982), 'On some simple estimates of an exponent of regular variation', *Journal of the Royal Statistical Society, Series B*, 37–42.
- Hall, P., and A. H. Welsh (1984), 'Best attainable rates of convergence for estimates of parameters of regular variation', *Annals of Statistics*, 1079–1084.
- Hill, B. M. (1975), 'A simple general approach to inference about the tail of a distribution', *Annals of Statistics*, 1163–1173.
- Ibragimov, I. A., and Yu. V. Linnik (1971), *Independent and Stationary Sequences of Random Variables*, Wolters Noordhof, Groningen.
- Kesten, H. (1973), 'Random difference equations and renewal theory for products of random matrices', *Acta Mathematica*, 207–248.
- Koedijk, K. G., M. M. A. Schafgans and C. G. de Vries (1990), 'The tail index of exchange rate returns', *Journal of International Economics*, 93–108.
- Kofman, P., and C. G. de Vries (1990), 'Potato future returns: a tail investigation', *Review of Futures Markets*, 244–258.
- Leadbetter, M. R., G. Lindgren and H. Rootzén (1983), *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.
- Levich, R. M. (1985), 'Empirical studies of exchange rates: price behavior, rate determination and market efficiency', in R. W. Jones and P. B. Kenen (eds), *Handbook of International Economics*, vol. 2, North-Holland, Amsterdam, pp. 979–1040.

- Mandelbrot, B. (1963a), 'New methods in statistical economics', *Journal of Political Economy*, 421–440.
- Mandelbrot, B. (1963b), 'The variation of certain speculative prices', *Journal of Business*, 394–419.
- Marston, R. C. (1987), 'Exchange rate policy reconsidered', NBER Working Paper, 2310.
- Mason, D. M. (1982), 'Laws of large numbers for sums of extreme values', *Annals of Probability*, 754–64.
- McCulloch, J. H. (1981), 'Interest rate risk and capital adequacy for traditional bank and financial intermediaries', in S. J. Maisel (ed.), *Risk and Capital Adequacy in Commercial Banks*, University of Chicago Press, Chicago, pp. 223–248.
- McFadden, D. (1973), 'Conditional logit analysis of qualitative choice behavior', in P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, pp. 105–142.
- Meese, R., and K. Rogoff (1983), 'Empirical exchange rate models of the seventies: do they fit out of the sample?', *Journal of International Economics*, 3–24.
- Mood, A. M., F. A. Graybill and D. C. Boes (1974). *Introduction to the Theory of Statistics*, McGraw-Hill, New York.
- Mussa, M. (1979), 'Empirical regularities in the behavior of exchange rates and theories of the foreign exchange market', *Carnegie-Rochester Conference Series on Public Policy*, 9–57.
- Phillips, P. C. B. (1987), 'Time series regression with a unit root', *Econometrica*, 277–301.
- Phillips, P. C. B. (1989), 'Time series regression with a unit root and infinite variance errors', Cowles Foundation for Research in Economics (mimeo).
- Pickands III, J. (1975), 'Statistical inference using extreme order statistics', *Annals of Statistics*, 119–131.
- Roy, A. D. (1952), 'Safety first and the holding of assets', *Econometrica*, 431–449.
- Smith, R. L. (1987), 'Estimating tails of probability distributions', *Annals of Statistics*, 1174–1207.
- Watts, V., H. Rootzén, and M. R. Leadbetter (1982), 'On limiting distributions of intermediate order statistics from stationary sequences', *Annals of Probability*, 653–662.

Table A.I. Minimizing MSE m -levels*

Estimator	m -level	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$
$\hat{\gamma}_P$	44	0.137	0.237	0.258
	34	0.170	0.150	0.203
	26	0.153	0.187	0.168
$\hat{\gamma}_H$	73	0.013	0.031	0.048
	30	0.034	0.010	0.015
	17	0.051	0.014	0.009
$\hat{\gamma}_R$	123	0.070	0.056	0.052
	60	0.085	0.043	0.033
	45	0.095	0.045	0.032

* The Monte-Carlo experiment consisted of 100 replications of $n = 474$ draws from three Student's t d.f. values with tail index α . Subsequently the MSE was computed for each m . The minimum MSEs are along the diagonals, and the minimizing m -levels are reported in the first column.

Table A.II. Minimizing MSE m -levels for ARCH(1)*

	$\lambda = 0.1$ $\alpha = 26.48$	$\lambda = 0.5$ $\alpha = 4.73$	$\lambda = 0.73$ $\alpha = 3.00$	$\lambda = 0.9$ $\alpha = 2.30$
m -Level	6	19	32	45
MSE	0.013	0.009	0.010	0.012

* The ARCH process was initialized by 50 observations and each experiment consisted of 500 replications. See Section 2 for a discussion of the ARCH process. Results are reported for the γ_H estimator only. See the footnote to Table A.I for further explanation.