

EXTREME DEPENDENCE IN THE NASDAQ AND S&P 500 COMPOSITE INDEXES

John W. Galbraith* and Serguei Zernov**
version of: 25. 5. 2006

Abstract

Dependence among large observations in equity markets is usually examined using second-moment models such as those from the GARCH or SV classes. Such models treat the entire set of returns, and tend to produce very similar estimates on the major equity markets, with a sum of estimated GARCH parameters, for example, slightly below one. Using dependence measures from extreme value theory, however, it is possible to characterize dependence among only the largest (or largest negative) financial returns; these alternative characterizations of clustering have important applications in risk management. In this paper we compare the NASDAQ and S&P in this way, and implement tests which can be used for the null hypothesis of the same degree of extreme dependence. Although GARCH-type characterizations of second-moment dependence in the two markets produce similar results, the same is not true in the extremes: we find significantly more extreme dependence in the NASDAQ returns. More generally, the study of extreme dependence may reveal contrasts which are obscured when examining the conditional second moment.

Key words: extremal index, extreme dependence, NASDAQ composite index, S&P 500 composite index, tail index

JEL Classification numbers: G10, G18

The authors thank the *Fonds québécois de la recherche sur la société et la culture* (FQRSC), the Social Sciences and Humanities Research Council of Canada (SSHRC) and the *Institut de Finance Mathématique de Montréal (IFM2)* for support of this research. We are grateful to John Maheu for valuable comments.

* Department of Economics, McGill University, 855 Sherbrooke St. West, Montreal H3A 2T7, and *Centre Interuniversitaire de recherche en analyse des organisations* (CIRANO).

** Department of Economics, McGill University, and TD Asset Management.

1. INTRODUCTION

A great deal is known about the second moments and second moment dependence of financial asset returns. While this knowledge implies some information about extreme events in these markets—periods of high conditional volatility are relatively likely to generate extreme movements—relatively little is known about the extremes *per se* of these asset return distributions. In learning about dependence in extreme circumstances, it is an advantage to have techniques that are focused purely on extreme movements, and are not influenced by the degree of temporal dependence in more routine circumstances. The resulting information about the degree to which extreme losses are more likely following earlier extreme losses is particularly valuable in the measurement of risk over fixed time intervals.

There is a substantial theoretical literature concerning technique for modelling the extremes of distributions. Two key concepts in this literature are the *tail index* and the *extremal index*, characterizing respectively the rate of decay of probability mass in the tails of a distribution, and (the inverse of) the degree of clustering of extreme events. Low values of the tail index correspond with relatively slow decay of probability mass in the tail, relatively frequent extreme events, and therefore (in the context of financial returns) a relatively high expected loss conditional on a return in the lower $\alpha\%$ tail of the distribution. Low values of the extremal index correspond with a relatively high degree of clustering of events exceeding a particular threshold magnitude, so that the extremal index provides a dependence measure more directly targeted to the extremes of a distribution than is the conditional variance measure provided by second-moment models. Recent advances in estimation make it possible to obtain extremal index estimates that do not depend upon a *priori* specification of a parameter which separates clusters of extreme events.

Although these concepts from extreme value theory have been much less widely applied in financial contexts than have conditional variance models, some literature has addressed these tail features. For example, Longin (1996) and Longin and Solnik (2001) studied properties of extremes of equity market returns; Longin (1999) and Cotter (2001) considered futures markets, and in particular margin requirements and exceedances. Galbraith and Zernov (2004) studied changes in the tail index of equity returns in the context of evolving elements of market architecture such as circuit breakers. A similarly substantial literature considers extremes in foreign exchange markets; Koedijk et al. (1990) and Quintos et al. (2001) are two examples. The extremal index, that is, the clustering measure, has evidently had fewer applications to date; Longin (2000) is an exception, which we will describe below.

The present paper examines returns on two major equity market indices, the NASDAQ

composite index and the S&P 500. The comparison of these two sets of returns provides an example of a type of comparison which may be of interest in numerous contexts where little information is available on either the extremes of the return distribution or dependence in the extremes. Of the two markets considered here, the NASDAQ market is typically considered riskier than the S&P 500, and this view is justified by a comparison of unconditional second moments of the returns. However the dependence structures embodied in, for example, the parameters of GARCH characterizations of the conditional second moments are similar for these returns; nonetheless the second-moment dependence implied by GARCH characterizations is determined by the entire set of returns, and may not provide a good characterization of properties of the series that we can uncover by studying the tails alone.

In approaching equity markets with the measures from extreme value theory, we are able to ask whether the rate of decay of probability mass in the lower tail (and therefore the interpretation of value at risk measures) is similar in the two equity markets, and whether patterns of dependence of extreme events, as opposed to second moment dependence, are similar across the markets. We will be able to provide some evidence on each of these points. In doing so we are able to provide a more focused characterization of the risk characteristics of these equity markets than is available in examining second moment features alone, and an indication of the implications of the results for management of large risks. We find that, although unconditionally the relative frequency of extreme events is similar, the degree of dependence in these extremes appears to differ across the markets.¹

In the next section we define the main parameters of interest and the methods used in their estimation and comparison across markets. Section 3 implements the methods and describes the empirical results comparing characteristics of the two markets. Section 4 briefly describes application of the clustering measures in characterizing risk over fixed time intervals, and the final section summarizes the conclusions.

2. METHODS AND PARAMETERS OF INTEREST

2.1 THE TAIL INDEX

A key definition for understanding the significance of the tail index is that of *regular variation at infinity*. The distribution function $F_X(x)$ (assumed to be common to all observations) of a random variable $X \in \mathcal{R}$ is said to be regularly varying at ∞ if there exists $\alpha > 0$ such that

$$(1 - F_X(hx))/(1 - F_X(x)) \rightarrow h^{-\alpha} \text{ as } x \rightarrow \infty, \forall h > 0; \quad (2.1.1)$$

¹Extreme events are defined with respect to each market, as the largest negative (or positive) returns on that market.

α is referred to as the tail index, or index of regular variation; see Hsing (1991).

From (2.1.1), $(1 - F_X(x)) = O(x^{-\alpha})$ as $x \rightarrow \infty$, where $1 - F_X(x) = P(X > x)$. Therefore α describes the rate of decay of probability mass in the tail of the distribution (this definition refers to a tail at $+\infty$, but may equally well be applied to, for example, a tail at $-\infty$). The characterization (2.1.1) is valid for a broad class of distributions, so that the tail index is a meaningful quantity in a large set of cases.

In financial contexts, the tail index is useful not only for the direct information that it provides on the relative frequency of extreme values in a return distribution, and a bound on the existence of moments of the distribution, but also in interpreting risk measures such as the value at risk. For example, consider two assets A and B with return distributions F_A and F_B respectively, each with 1% VaR of $q_{.01}$ (that is, the lower 1% quantile of the return distribution is at $q_{.01}$). Let the tail indices of these distributions be such that $\alpha_A < \alpha_B$. Then conditional on observing a return in the lower 1% tail of the distribution, i.e. a loss exceeding $-q_{.01}$, A will tend to display more extreme losses than B, as density decays less quickly in the tail of F_A . Consequently the expected loss in A conditional on a loss exceeding $q_{.01}$ will exceed the expected loss in B under the same circumstances. Define the expected shortfall² of a random variable $X \in \mathcal{R}$

$$\mu_q = E(-X|X \leq q); \tag{2.1.2}$$

then for asset returns X_A and X_B and sufficiently small q , $\mu_q^A > \mu_q^B$.

There are several possible methods of estimation of α of which that of Hill (1975), despite its well-known limitations, is well suited for our present comparative purpose. The estimator uses the m largest (smallest) order statistics of the sample for estimation on the upper (lower) tail. Define the order statistics from the original sample $\{X_i\}_{i=1}^T$ as $\{X_{(1)}^T, X_{(2)}^T, \dots, X_{(T)}^T\}$ such that $X_{(1)}^T \leq X_{(2)}^T \leq \dots \leq X_{(T)}^T$. The largest m of these, $X_{(T-m+1)}^T$ to $X_{(T)}^T$, are used for estimation on the right tail of the distribution, or the smallest m values multiplied by -1, for estimation on a left tail at $-\infty$. In either of these cases Hill's estimate of the tail index (the inverse of the 'Hill index') is then

$$\hat{\alpha} = \left[m^{-1} \sum_{i=1}^m \ln X_{(T-i+1)}^T - \ln X_{(T-m+1)}^T \right]^{-1}. \tag{2.1.3}$$

²The *tail conditional expectation* is similarly defined. Note that this is not in general the same as the first lower partial moment; the n th lower partial moment of $X \in \mathcal{R}$ at q is defined as $E((q - X)^n | X \leq q)$.

While this estimator is afflicted with substantial problems arising from bias and sensitivity to nuisance parameter selection, these will be less important in the case of a comparison of α across two distributions from which we have approximately the same sample sizes; the bias will affect the estimates from different data sets similarly when values of the tail index are similar, and we will compare the data sets on an entire sequence of nuisance parameters, so that we are not required to select a single value.³ The advantages of the estimator for our present purpose are robustness to dependence (see in particular Hsing 1991) and the existence of formal inferential methods. Rather than attempt to select an optimal value of m , we will produce a sequence of Hill estimates (the ‘Hill plot’) on a large set of values of m , and a corresponding set of point tests of $H_0 : \alpha_A = \alpha_B$ conditional on each value of m . We are able to bypass the nuisance parameter selection problem in the sense that the key inferential results hold across a wide range of values of m .

The observations $\{X_i\}_{i=1}^T$ will be daily returns on the NASDAQ and the S&P 500 indices. The samples are described below.

2.2 THE EXTREMAL INDEX

The key feature of the tails of the return distribution that we investigate is the time dependence, or clustering, of extreme events.⁴ Such clustering is an important feature in understanding risk over given time intervals, as we will describe below. It is well known that financial returns tend to display clustering of large changes; this clustering is almost invariably described by models of the second moment (e.g. GARCH or stochastic volatility), which are functions of the full set of observations and are therefore heavily influenced by the large number of events which, while representing substantial absolute returns, would not be considered extreme. An alternative approach is possible which addresses behaviour in the extreme tails only, and which therefore characterizes dependence in a small subset of returns taken from the tails, or a single tail, of the return distribution.

Clustering of such extreme events can be measured by the *extremal index*, which can be interpreted as the *inverse of the mean number of extreme events in a cluster*. The extremal index has a value of unity in independently distributed data; in data for which extreme events display some clustering, the parameter may range down to zero. We will follow Weissman and Novak (1998) and Embrechts et al. (1997) in this exposition.

³Where the absolute level of the tail index is the key object of interest, as in Section 2.3 below, methods such as that of Huisman et al. (2001) provide substantial bias reductions while retaining many of the desirable features of Hill estimation.

⁴Note that this is distinct from cross-sectional dependence in extreme events across markets.

Let the sequence of random variables $\{X_i\}_{i=1}^T$ on the sample of size T be such that a sequence of thresholds $\{u_T\}$ can be chosen to ensure $T(1 - F(u_T)) \rightarrow \tau > 0$ as $T \rightarrow \infty$. It follows that $P(M_T \leq u_T) \rightarrow e^{-\theta\tau}$ as $T \rightarrow \infty$ for $M_T = X_{(T)}^T = \max_{i \leq T}(X_i)$, i.e. the maximum on the sample of size T (the largest order statistic of the sample). The parameter θ is called the extremal index of the distribution of X , and $\theta \in (0, 1]$. As we have noted, θ^{-1} is the mean number of events in a cluster which exceed u_T .

Estimation of θ has usually been carried out by one of two methods, each of which estimates a mean number of observations in a cluster. Using either of these two methods requires a definition of a cluster, so that distinct clusters may be identified and separated, and of an associated de-clustering parameter, r , which once chosen allows separation of the sample into clusters. The first method (Leadbetter 1983) divides the sample into k blocks of size r , where $r = \frac{T}{k}$ or $\text{int}(\frac{T}{k})$. On each block, compute the maximum,

$$M_r^{(i)} = \max(X_{(i-1)r+1}, X_{(i-1)r+2}, \dots, X_{ir}).$$

The extremal index can be estimated by the sample analogue of $r^{-1} \left[\frac{P(M_r^{(i)} > u)}{P(X_i > u)} \right]$, or

$$\hat{\theta}_B = \frac{\sum_{i=1}^k I(M_r^{(i)} > u)}{\sum_{i=1}^T I(X_i > u)} = \frac{n_B}{n}, \quad (2.2.1)$$

where $I(\cdot)$ is the indicator function, u is the threshold, n_B is the number of blocks showing at least one value in excess of the threshold, and n is the total number of ‘exceedances’ of the threshold on the sample. The interpretation of θ as the inverse of the mean cluster size (i.e. the inverse of $\frac{n}{n_B}$) is then natural, where each block showing at least one exceedance is defined to be a cluster.

The second method based on a de-clustering procedure is that of *runs* (O’Brien 1974), in which a cluster is defined by using a given number of observations below the threshold to separate clusters; in this case the de-clustering parameter r is this number of consecutive observations below the threshold u required to define separation. Define a cluster ending at i by the condition C_i , which is deemed to hold when an exceedance at i is followed by at least r observations below the threshold: $X_i > u, X_{i+1} \leq u, \dots, X_{i+r} \leq u$. Then the number of clusters is the number of times that the condition holds on the sample, $n_C = \sum_{i=1}^{T-r} I(C_i)$, and the mean cluster size is $\frac{n}{n_C}$, with associated estimator of the extremal index

$$\hat{\theta}_R = \frac{\sum_{i=1}^{T-r} I(C_i)}{\sum_{i=1}^T I(X_i > u)} = \frac{n_C}{n}. \quad (2.2.2)$$

For each of these estimators we must have $r = r(T)$ such that $r \rightarrow \infty$, $\frac{r}{T} \rightarrow 0$.

Limit theory is available for both estimators. Consistency of the blocks and runs estimators was established by Leadbetter (1983) and O'Brien (1974) respectively, and consistency and asymptotic normality are established under different sets of conditions by Hsing (1991b) and Weissman and Novak (1998) for the blocks estimator, and Weissman and Novak (1998) for the runs estimator. The limit theory does not however extend to expressions for asymptotic variance.

Each of these two methods requires specification of a de-clustering parameter, the block size or run length. A third principle for estimation, that of Ferro and Segers (2003), is based on times between exceedances rather than on an estimate of mean cluster size, and has the substantial advantage that it does not require *a priori* specification of a de-clustering parameter. The asymptotic distribution of interexceedance times depends on only one parameter, the extremal index θ ; estimation of θ is therefore possible from the set of interexceedance times, which depend only on the threshold of interest, and not on a run length. Let s_1, \dots, s_n be the observation numbers or times of the n exceedances, and $\tau_1, \dots, \tau_{n-1}$ be the lengths of time between these exceedances. Ferro and Segers propose two estimators based on these intervals,

$$\hat{\theta}_\tau = \min \left[\frac{2 \left(\sum_{i=1}^{n-1} \tau_i \right)}{(n-1) \sum_{i=1}^{n-1} \tau_i^2}, 1 \right] \quad \text{and} \quad \hat{\theta}_\tau^* = \min \left[\frac{2 \left(\sum_{i=1}^{n-1} (\tau_i - 1) \right)}{(n-1) \sum_{i=1}^{n-1} (\tau_i - 1)(\tau_i - 2)}, 1 \right] \quad (2.2.3)$$

for cases in which $\max_{1 \dots n-1} \tau_i \leq 2$ and > 2 respectively. A de-clustering parameter having the same interpretation as that in the runs estimator can then be estimated (rather than specified *a priori*) from $\hat{\theta}, n$, and the set of interexceedance times. Note that these estimators do not require IID data; consistency has been established for m -dependent sequences (that is, sequences for which dependence exists up to any finite order m).

2.3 INFERENCE ON DIFFERENCES IN EXTREMAL INDICES

A number of procedures are available for inference on the tail index; see in particular Loretan and Phillips (1994) and Quintos et al. (2001). For the extremal index, by contrast, drawing inferences on differences across populations is not straightforward, because the asymptotic variances of the estimates are not known.⁵ In order to conduct inference on differences in

⁵The asymptotic variance of θ , which would facilitate inference, does not seem to have been established. Weissman and Novak (1998) do not propose consistent estimators for the asymptotic variances appearing in the asymptotic normal distributions of $\hat{\theta}_B$ and $\hat{\theta}_R$, but indicate that such estimators are under investigation. Ferro and Segers (2003) suggest a bootstrap procedure for inference on $\hat{\theta}_\tau$ or $\hat{\theta}_\tau^*$.

extreme dependence in the two equity markets, we will therefore consider two types of standard test procedure: non-parametric tests using the sequence of estimated clusters, and bootstrap tests based on the estimated values of θ . In the first case we test a null of identical distributions of cluster size in the two markets, versus a difference in location. Since the extremal index is asymptotically the inverse of the mean cluster size, this is indirectly a test for equality of the extremal index across the markets, although the typical cluster size itself may be an object of direct interest for risk management. In the second case we conduct inference directly on the extremal index, with similar results.

The blocks and runs methods imply different clusters of exceedances of a threshold, which suggest different forms of test. In the case of blocks estimation, these clusters are specified *a priori* by the block size parameter: given our observations on a common set of dates in the two markets, a common set of beginning and ending dates for blocks can be constructed for the NASDAQ and S&P 500. These clusters then form a set of date-matched pairs to which non-parametric matched-pair comparisons can be applied. For k blocks of size r as defined above, denote by the two sequences $\{g_i^N\}_{i=1}^k$ and $\{g_i^S\}_{i=1}^k$ the numbers of exceedances in each of the k blocks or clusters, for returns on the NASDAQ and S&P 500 composites respectively.

The null hypothesis of identical cluster-size distributions could in principle be tested against the alternative of a location difference by a statistic such as the Wilcoxon matched-pair signed-rank statistic which uses the set of cluster-size differences, $\{d_i\}_{i=1}^k \equiv \{g_i^N - g_i^S\}_{i=1}^k$; see e.g. Lehmann (1975). Such a test has however substantial disadvantages arising from the fact that it is based on blocks estimation. The block size must be specified *a priori*, and for matched-pair comparison must in any event be the same for the two sequences. The two distributions can differ in the mean number of exceedances per (non-empty) cluster, but cannot differ in cluster length. The tests will therefore have no power against differences in average length of the true clusters. As well, specification of a fixed grid of blocks does not correspond well with the pattern of randomly-occurring, random-length periods of volatility clustering typically observed in financial data. For these reasons, we do not implement this test, but instead use two tests based on runs de-clustering.

(i) Rank-sum statistic based on runs

De-clustering based on runs does not constrain runs to be of equal length either within or across sequences. Our first test uses this method, with the de-clustering parameter estimated by the Ferro and Segers (2003) procedure, implying a sequence of random-length clusters for each data set. The number of clusters is not constrained to be equal across the two data sets.

To implement this test, we obtain sequences $\{h_i^N\}_{i=1}^{\ell_N}$ and $\{h_i^S\}_{i=1}^{\ell_S}$ of cluster sizes for

the extreme returns on the two markets. Under the assumption of independence of clusters across the two markets, we can test the null of identical cluster size distributions, without requiring the same number of clusters in each market. Let $\{h_j^*\}_{j=1}^{\ell_N+\ell_S}$ be the merged set of cluster sizes sampled from the two markets, and sorted in ascending order. Assign the ranks $1, 2, \dots, \ell_N + \ell_S$ to these values, and compute the sum of the ranks corresponding with the observations from either distribution. For IID data (here, IID clusters) the statistic is asymptotically normal with mean (for the statistic based, e.g., the NASDAQ ranks) of $\ell_N(\ell_N + \ell_S + 1)/2$ and variance $\ell_N\ell_S(\ell_N + \ell_S + 1)/12$; the statistic is standardized to yield an $N(0,1)$ quantity. This variance can be adjusted in the case of tied ranks, for which mid-ranks can substituted; alternatively randomization can be used to assign neighbouring ranks, and we use the latter method, averaging over a large number of randomizations to obtain the test statistic.

A disadvantage of this test is that it assumes independence of the two sets of cluster sizes, which unlikely to hold literally. However, independence may be a good approximation. By definition the runs de-clustering parameter is a value that produces independent runs of exceedances, so that dependence across the two sets of returns should be limited to those runs occurring at the same, or overlapping, historical dates; since the run lengths and exceedances differ, however, the two series of clusters do not occur at the same sets of historical dates (as they would in the blocks method, with a common block size). Moreover, since positive dependence between the two time series reduces the variance of differences, we would expect that any induced departure from the nominal distribution would be in the direction of lower variance, implying a conservative test.

We obtain p -values for the test both from the asymptotic normal distribution and from 1999 bootstrap replications of the statistics. The p -values from each of these methods are presented in Figure 6a (described below) for a set of numbers of exceedances; those for the standardized Normal distribution use as a variance the *larger of* 1 or the empirical variance in the bootstrap replications.

(ii) Bootstrap inference based on inter-exceedance times

A second test that we implement uses a bootstrap distribution, also based on runs de-clustering, using the inter-exceedance times. This test also, therefore, allows different run lengths within and across sequences. A non-parametric bootstrap procedure for statistics of interest, based on the bootstrap distribution, is described by Ferro and Segers (2003); by resampling from the sets of inter-cluster times and exceedances, bootstrap replications of the sample are constructed.

We follow Ferro and Segers in re-sampling alternately from the inter-cluster times and the *sets of* intra-cluster times, to form a bootstrap replication of the sequence of interexceedance times. We do so on each of the NASDAQ and S&P 500 data sets, via the computed interexceedance times on each, and on each bootstrap replication i compute the difference between the estimated extremal indices, $\hat{\theta}_N^{(i)} - \hat{\theta}_S^{(i)}$. This procedure is repeated 1999 times, the differences are recorded, and the p -values for $H_0 : \hat{\theta}_N - \hat{\theta}_S = 0$ are computed from the bootstrap distribution.

Finally, we note that regardless of the test, for each set of returns we choose the threshold value in order to obtain the same total number of exceedances in each market. That is, the threshold u is adjusted for the relative variances of the markets' returns, so that the set of points defined as extreme is the same proportion in each case.⁶

3. EMPIRICAL TAIL COMPARISONS, NASDAQ AND S&P

We now implement the measures and tests described above to compare the returns on two major equity market indices. Although our primary interest is in the extreme dependence measure, we begin with a background comparison of the tail decay rates in the two unconditional return distributions. We then ask whether there is evidence of different degrees of dependence (clustering) in the events in the tails of the return distributions. In each of the following sub-sections, sets of estimates and test statistics will be presented in graphical form. The results presented pertain in all cases to the lower tails of the return distributions, that is, the extreme losses. We refer in the text to the upper-tail results, which are qualitatively similar.

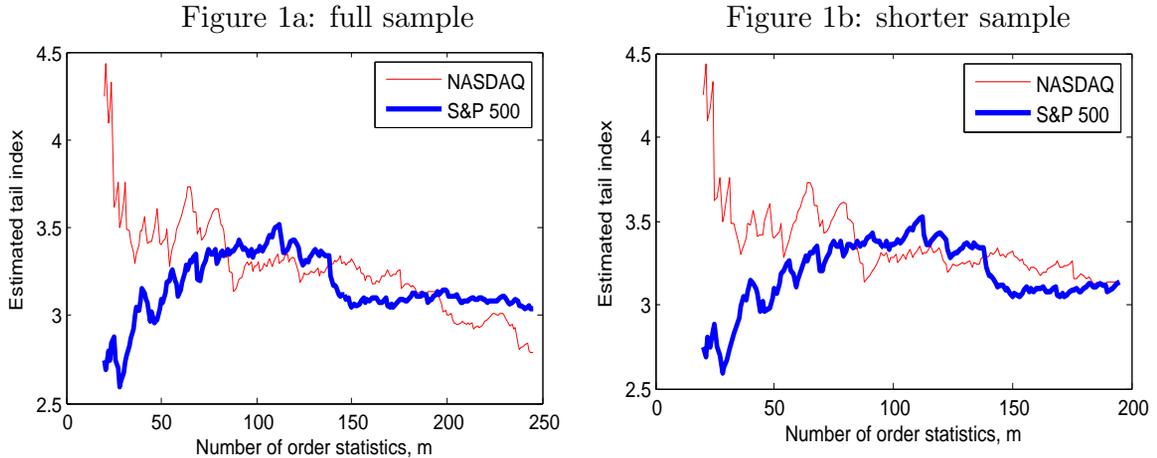
3.1 TAIL INDEX COMPARISON

Figures 1a and 1b plots the sequence of Hill tail index estimates as a function of the number of order statistics $m = 1, \dots, \bar{m}$, where \bar{m} is taken as 5% of the number of sample points (and so approximately 10% of negative-return observations). It is important to note that we are not attempting to choose a particular optimal value of m ; while this is necessary for estimation of the value of the tail index, our interest here centers on the question of whether or not the value is similar in the two markets. We therefore make comparisons for a wide range of possible values of m , corresponding with different degrees of extremity in the tail area used

⁶Note that if we were to choose a common numerical threshold instead, the higher variance of returns on the NASDAQ would imply a greater total number of exceedances, guaranteeing that the blocks method would show a greater mean cluster size for that market, since the number of blocks is the same for each market.

for comparison. Because the samples sizes are the same in the two markets, a given number of order statistics corresponds with the same proportion of the sample points in each case.

FIGURE 1
Hill plots: estimated tail index vs. number of order statistics
NASDAQ and S&P 500 composite lower tails



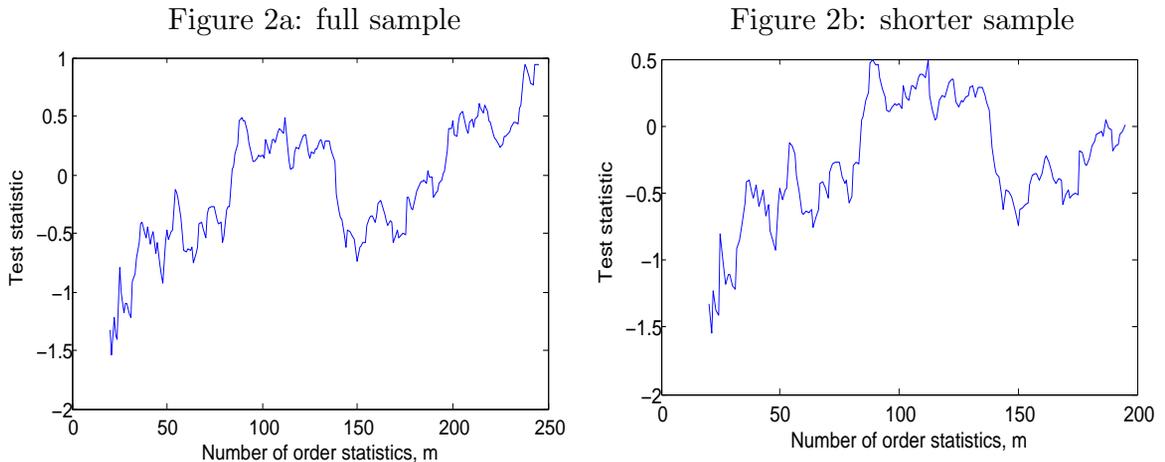
These figures use two sample sizes, the first beginning with the first available observation on the NASDAQ composite index, that is 11 October 1984, and a second shorter sample beginning 1 November 1988. The second sample is chosen for consistency with the results of Galbraith and Zernov (2004), suggesting a structural break in tail behaviour in the neighbourhood of 1985-1988, a period of relatively uncontrolled program trading followed by the introduction of the first circuit breakers in US equity markets; we therefore examine both the full sample and the shorter sample which omits this period. The two samples are of sizes 4890 and 3865 daily observations, in each of the two equity markets.⁷ We see in these plots the typical pattern of increasing bias (estimates of α tending to fall steadily) and decreasing variance as m increases.

Clearly, the tail index estimates are fairly close, and the sequences cross, suggesting that any differences are not substantial. Figure 2a/b provides some formal inferential evidence using the pointwise Loretan-Phillips (1994) statistic. We again emphasize that we cannot interpret the maxima or minima of these statistics as having any standard distribution, and to interpret large values of the maximum or minimum, we would need to rely on bootstrap inference. In these cases, however, we see that none of the values exceed even the pointwise

⁷In both the longer and shorter samples, the markets' data were matched to remove from the sample the small number of days in which trading took place for only one of the two markets.

critical values for standard inference at conventional levels (e.g. 1.96 for 5% inference, as the Loretan-Phillips pointwise tests are asymptotically standard normal).⁸ There is therefore no substantial evidence against the null of equality of the tail indices on the two equity markets' returns.⁹ This is true on each sample size.

FIGURE 2
Loretan-Phillips test vs. number of order statistics
NASDAQ and S&P 500 composite lower tails



The results of this section are therefore easily summarized: there is no indication that the extreme lower tail of the NASDAQ decays less (or more) rapidly than that of the S&P 500 (note of course that this does not imply that the two markets have similar lower quantiles of returns). The unconditional tail features of the markets summarized by α appear similar.

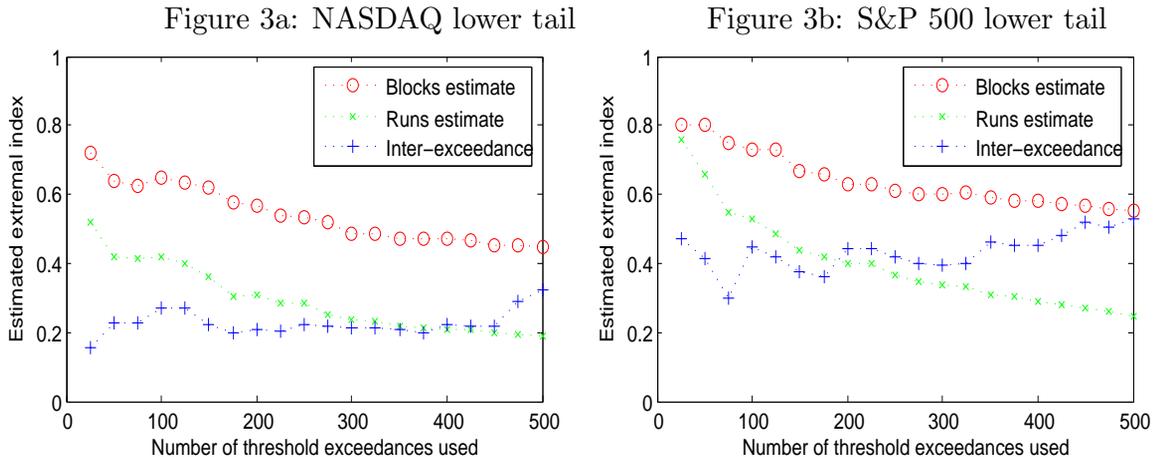
3.2 EXTREMAL INDEX COMPARISON

We now consider the conditional feature, extreme-event dependence. Figures 3a (NASDAQ) and 3b (S&P 500) present extremal index estimates for the lower tails of the return distributions, from each of the methods discussed in Section 2. Recall that relatively low values of the extremal index indicate relatively high dependence in extreme returns (relatively large mean cluster size).

⁸The pointwise tests are very liberal where the null is of a break at an unknown point in the sample; the overall critical value substantially exceeds pointwise values.

⁹It is interesting to note that there is greater evidence of difference on the upper tail, where the NASDAQ market tends to show lower estimated upper-tail index values, an effect probably attributable to the period of rapid NASDAQ price growth through early in 2000.

FIGURE 3
Extremal index as a function of number of threshold exceedances,
estimates from blocks/runs/interexceedance times



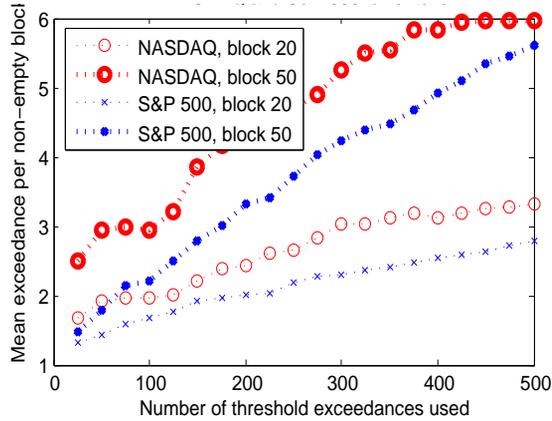
Two general features are observable. First, the blocks estimator invariably produces higher estimates of the extremal index (lower extreme dependence) than do either of the runs-based estimators. (Note that Longin’s (2000) blocks estimate of the extremal index for the S&P 500, 0.72, is similar to the blocks estimate obtained here for around 100 threshold exceedances.) Second, any of the three techniques produces higher point estimates of the extremal index on the S&P 500 than on the NASDAQ composite. Each of these conclusions holds for any of the values chosen for the number of exceedances.

The typical estimated difference between S&P 500 and NASDAQ estimates of θ is around 0.10 by the standard blocks or runs methods, and 0.20 by inter-exceedance times. These reasonably stable differences are suggestive of a lower extremal index (more dependence in extremes) in the NASDAQ returns than in the S&P 500. We now examine in more detail these point estimates, and conduct statistical inference on the measured differences.

Consider first the block estimates, which are based on exogenously-specified block sizes. Because the number of threshold exceedances used is the same for each return series, and the block sizes used are the same, the mean number of exceedances per block must also be identical for the two series. However, if one series displays greater return clustering than the other, we would expect its exceedances to be concentrated in a smaller number of non-empty blocks. Correspondingly, we would expect a lower extremal index to be reflected in a greater mean number of exceedances per *non-empty* block. Figure 4 plots this quantity for each return series and for two block sizes. The results are consistent with the point estimates of the extremal index recorded in Figure 3a/b: the NASDAQ returns show larger means on both block sizes

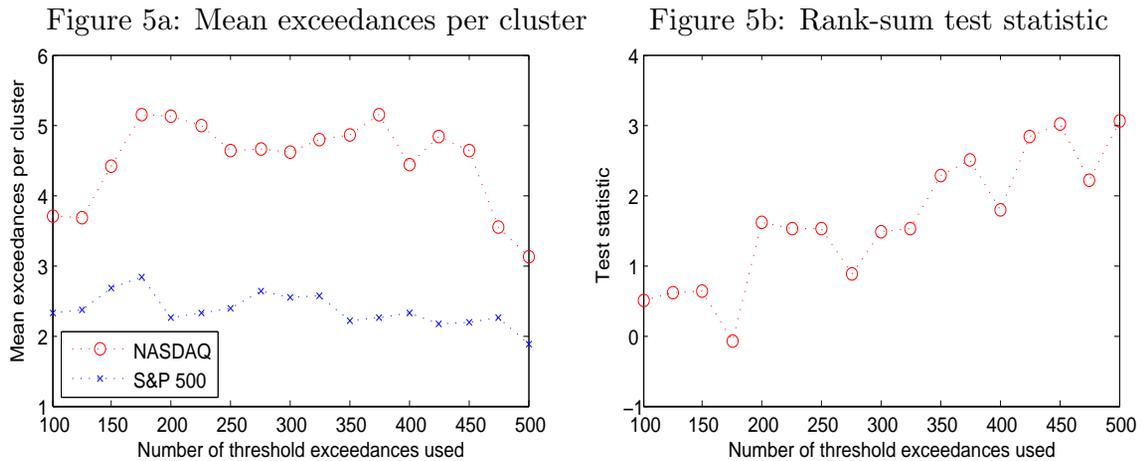
and all numbers of exceedances. Recall however that these values result from the constraint of identical block size (run length) across the two series.

FIGURE 4
Mean number of exceedances per non-empty block
NASDAQ and S&P 500 composite lower tails



For statistical inference, as noted above, we use runs de-clustering in order not to impose equal cluster size. Figure 5a plots the mean number of exceedances per cluster identified in this way. The typical measured difference is substantially higher when the constraint of equal cluster size is removed (cf. Figure 4). Figure 5b plots the associated test statistics.

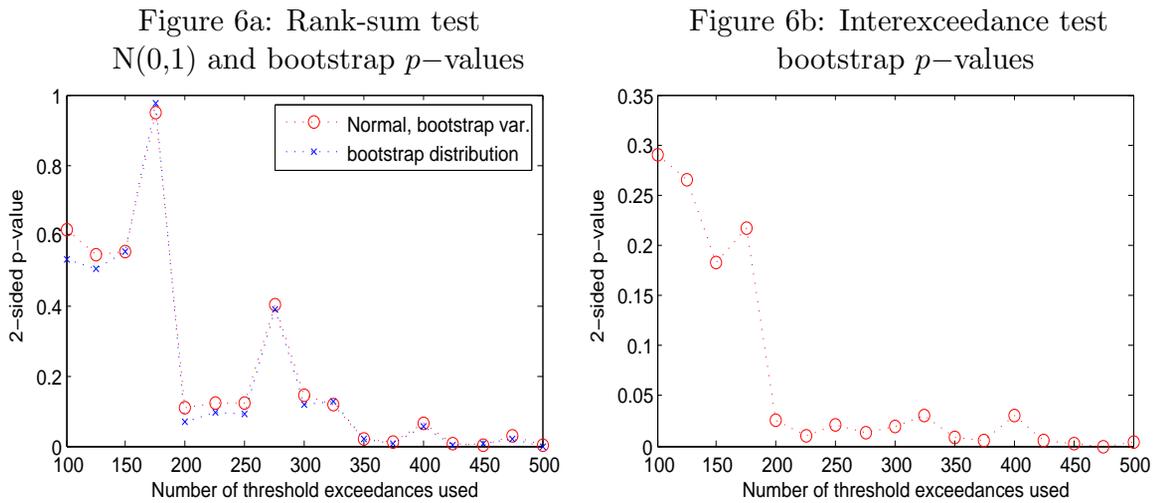
FIGURE 5
Exceedances and associated statistics
NASDAQ and S&P 500 composite lower tails



The two-sided p -values from this test are shown graphically in Figure 6a, using first the

asymptotic Normal distribution of the rank-sum statistic, and second the bootstrap p -values for the statistic. The two methods of computation give very similar results, reflecting the good conformity of the empirical distribution with the asymptotic Normal. As the sample of exceedances becomes larger, the statistic rejects (at a level of 5%, for example) the null hypothesis of equal extremal index on the two markets, although for smaller numbers of exceedances (corresponding with higher thresholds), there is not sufficient evidence to reject the null at any conventional level.

FIGURE 6
Test statistics' p -values
NASDAQ and S&P 500 composite lower tails



The second test uses bootstrap replications of the interexceedance times, which are sufficient statistics for the estimate of θ in the Ferro and Segers (2003) procedure. The two-sided bootstrap p -values are recorded in Figure 6b. At lower thresholds/larger sample sizes, the p -values become close to zero.¹⁰ The p -values tend to be somewhat lower (evidence against the null somewhat stronger) than observed in the rank-sum test. The same general pattern, of rejection of the null of equality emerging as the sample of exceedances becomes larger, emerges.

This interexceedance-time bootstrap test also lends itself readily to testing the null of no dependence in extreme observations ($\theta = 1$), via bootstrap confidence intervals for θ . The upper bound on the 99% confidence intervals for θ never exceeds 0.77 for the NASDAQ returns, and 0.82 for the S&P 500 returns, at any of the threshold values considered. This null therefore appears clearly incompatible with the data; these results are not recorded in figures.

¹⁰An estimated two-sided p -value of zero implies that no bootstrap replication of the processes yielded a statistic, $\hat{\theta}_N^{(i)} - \hat{\theta}_S^{(i)}$, as low as the hypothesized value.

These tests suggest therefore that there is substantial and statistically significant extreme dependence in both return series, and that the estimated differences in mean cluster size or extremal index across these two asset classes reflect a genuine difference in the degree of dependence in extreme observations. Repeating the tests (the results are not recorded here) on the upper tails of the two return distributions yields similar patterns and similar evidence against the null of the same degree of extreme dependence. In each case the differences are in the direction of greater extreme dependence in the NASDAQ index returns. This in turn suggests that underestimates of risk measures, resulting from ignoring extreme dependence, will be greater in the case of the NASDAQ.

4. IMPLICATIONS FOR MEASURING FINANCIAL RISK

To take an illustrative example of the importance of the extremal index in risk management, consider two insurance companies insuring against very large (relative to the firms' capitalization), but very infrequent, losses. The two are similar except in that the losses insured by the first firm are independent ($\theta = 1$), whereas those insured by the second are heavily clustered (θ near 0.) The second firm is more likely to face bankruptcy: it is more likely to have more than one extreme loss over a given short horizon, and the probability that a cumulative loss over several periods will exceed a bankruptcy threshold is increased by this clustering of extreme losses. This is related to the fact that the value at risk (VaR) is a function of clustering of extreme events.

Measures of risk based on extreme value theory allow quantification of risks which until recently were dealt with only using stress testing approaches. As well, extreme value theory does not make restrictive parametric assumptions about the dynamics of the asset prices of interest. There are different ways in which extreme value theory may be used with risk measures; for example, Martins-Filho and Yao (2006) propose an approach to evaluation of VaR and expected shortfall where dynamics of the conditional mean and conditional scale of the process are described by a non-parametric non-linear model, specifically the CHARN model of Diebolt and Guégan (1993), while the dynamics of innovations are modelled using EVT. The primary difference between the Martins-Filho and Yao method and that of Longin (2000), which we will use below, is that the former is based on a model of the conditional distribution, whereas the latter relies on the GEV distribution to approximate the marginal distribution of the series, and on the extremal index to account for dependence. Each of these studies recognizes the importance of careful treatment of extremes and of dependence when computing risk measures.

The method of Longin (2000), which we will use in the examples of this section, computes

VaR through the following steps: first, choose a frequency of returns of interest f ; the choice of this frequency may be guided by liquidity considerations, or may be imposed by regulators. Next, estimate the Generalized Extreme Value (GEV) distribution using a maximum likelihood approach by fitting block maxima (or minima) of the series. Denote the location, scale and shape parameters of the GEV distribution as μ , σ and ξ correspondingly. The chosen size of the blocks $T = fn$ should satisfy the conditions that the asymptotic GEV distribution should fit the block maxima (minima) well, and it should be reasonable to assume that the blocks are statistically independent. The goodness of fit of the GEV distribution can be tested using a Sherman test. Then the level of the VaR—a quantile of the return distribution—is chosen. Define p as the probability of a one-period return exceeding the VaR threshold (e.g. $p = 0.99$), and p_n as the probability that the minimum return observed over n periods exceeds the same threshold. Then $p_n = p^n$ in independent data, and $p_n = p^{n^\theta}$ in a dependent sequence with extremal index θ .

Asymptotically,

$$p_n = 1 - F_{GEV}(VaR) = \exp\left(-\left(1 + \xi\left(\frac{\mu + VaR}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right), \quad (4.1)$$

where $F_{GEV}(x)$ is the cumulative distribution function of the generalized extreme value distribution. The VaR can then be estimated using the following approximate formula:

$$VaR_{1-p} = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left(1 - \left(-n\hat{\theta} \ln(p)\right)^{-\hat{\xi}}\right). \quad (4.2)$$

It follows from equation (4.2) that the dependence of the series, captured by θ , materially affects the value at risk.

Let us now apply this method of computation to the S&P 500 and NASDAQ returns from October 1984 through December 2003. Table 1 records estimates of the parameters of the GEV distribution by maximum likelihood, fitting block maxima of returns (block minima for the left tail) using a block size of 63 days (one calendar quarter). The Hill estimator of α is also recorded in the table for comparison with the ML estimate ξ .¹¹ (Standard errors of the

¹¹Generally similar (within approximately 2%) estimates of the parameters μ and σ are obtained by first estimating the tail (shape) parameter by a bias-corrected Hill estimator (Huisman et al. 2001), and using ML for the remaining scale and location parameters of the GEV distribution; the right tail of the NASDAQ returns, however, shows substantially different estimates by this method.

parameter estimates are approximately 0.10 in S&P 500 data, 0.15 in NASDAQ data.) The fit of the GEV distribution is tested using Sherman’s statistic: the null of a match cannot be rejected in any of the cases considered.

Table 1
ML estimates of the parameters of the GEV distribution
NASDAQ and S&P 500, left and right tails

Index	tail	μ	σ	ξ	α
NASDAQ	left	2.13	1.14	0.39	0.37
NASDAQ	right	1.70	0.90	0.63	0.41
S&P 500	left	1.87	0.86	0.32	0.33
S&P 500	right	1.95	0.73	0.24	0.34

The results of the VaR computation based on (4.2) and these parameter estimates are presented graphically in Figures 7 and 8, which allows visual inspection of a range of different VaR values. In each case the dotted line indicates the points of equality between estimated (based on 4.2) and empirical left tail (Fig. 1, 3) or right tail (Fig. 2, 4) probabilities. The fit is quite good for the S&P 500, less so for the NASDAQ; in each of Figures 7a and 7b, using an estimated extremal index θ , the formula tends to overestimate the probability remaining in the tail (by a fairly small amount) and therefore the VaR. The estimate given by the formula is therefore conservative. By contrast, when extreme dependence is ignored and (4.2) is used with $\theta = 1$ (Figure 8), substantial underestimates arise of probability remaining in the tail, the VaR, or related risk measures. (Recall that the statistical inference reported in Section 3.2 suggests that the difference between the two markets, reflected in Figures 8a and 8b, is statistically significant.)

FIGURE 7
 $p - p$ plots of extreme returns

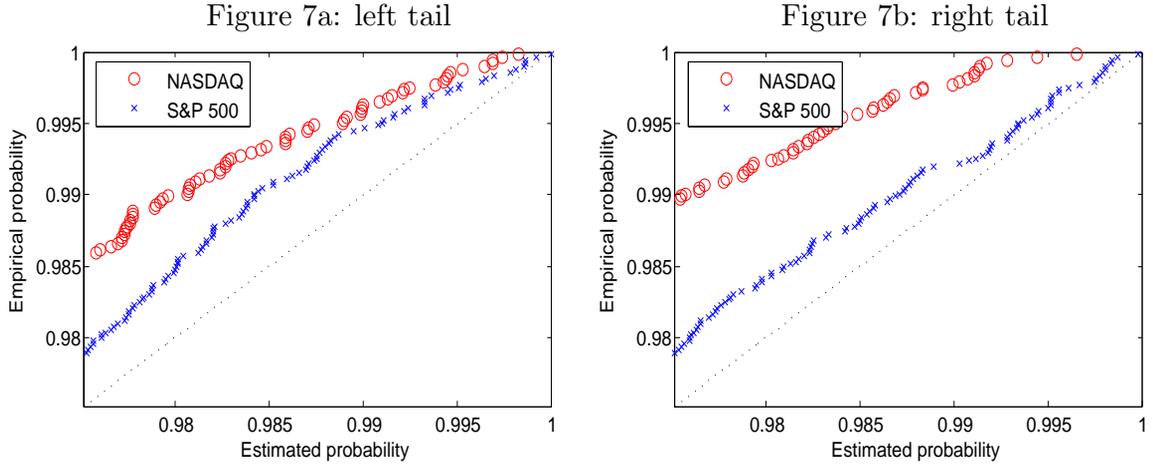
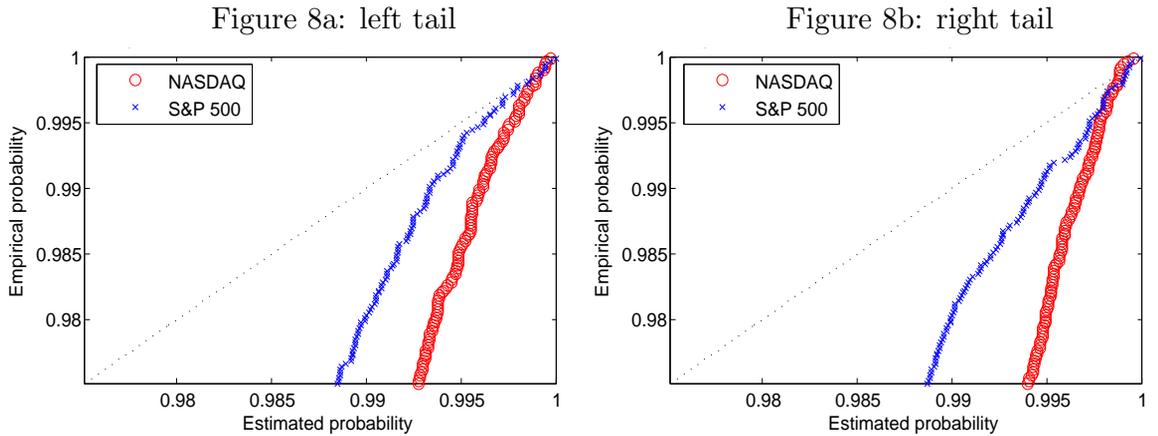


FIGURE 8
 $p - p$ plots of extreme returns, ignoring dependence



The analysis in Longin (2000) also suggests that ignoring dependence in the data leads to underestimating the VaR, but concludes that the impact of dependence is not statistically significant at the daily frequency, and that the measures of dependence are smaller again at lower frequencies, for the sample that Longin considered (S&P 500 from January 1962 to December 1993). We have already concluded that there is significant extreme dependence in both asset returns and that the dependence is significantly stronger in the NASDAQ series. This difference in the conclusions results in part from the application of recent developments in estimation, and consequent differences in estimated values of the extremal index: Longin

(2000) estimates $\hat{\theta} = 0.72$ for the daily S &P 500 series, while with interexceedance times we estimate θ to be in the neighbourhood of 0.4–0.5 for S&P 500 and of 0.2–0.3 for the NASDAQ series (again, Figures 3a and 3b).

If it is necessary to evaluate the VaR over a period longer than that of the sampling frequency, it is possible to apply the algorithm just described using non-overlapping returns over the period of interest. We have analyzed 10-day returns¹², and find results qualitatively similar to those reported here for one-day returns. However, using lower-frequency data when higher frequency data are available leads to a loss of efficiency. Some recent studies (see, for example, Robinson and Tawn 2000 and Hall et al. 2005) have examined the behaviour of the extremal index and of the parameters of the GEV in sub-sampled and aggregated processes; these results may prove useful in improving estimation of the value at risk, and other measures, over longer horizons.

5. CONCLUDING REMARKS

It has been observed for many years that large absolute (or squared) returns in asset markets tend to cluster. This feature is one of the factors motivating use of conditional volatility models, from for example the GARCH or SV classes, in fitting time series of these returns. Many asset return sequences have been found to be well fitted by models of these types; major equity market indices, in particular, are often found to be characterized by very similar models and parameters, suggesting similarity in the patterns of clustering of large absolute returns.

These conditional second-moment models are fitted to all observations. An alternative, explored in the present paper for two major equity indices, is to apply methods which use only the most extreme returns on a sample, thereby ignoring the centre of the distribution and describing features of the extreme observations only. In doing so we find that, while the tail index (rate of tail decay) parameter is also very similar in the two markets, the parameter characterizing clustering of extremes appears to be substantially different. This feature appears in both upper and lower tails of the distributions.

Differences in extreme clustering are important in understanding and managing risk, implying as they do differences in the probabilities of cumulative losses which exceed given thresholds over multi-period horizons. Corresponding differences are implied in risk measures such as the value at risk, which may be obscured by estimation methods which fail to account ade-

¹²A 10-day VaR is of particular interest in practice because this is the holding period that financial institutions should use in evaluating VaR according to the recommendations of the Basel Committee.

quately for extreme clustering—typically leading to underestimates of tail-based risk measures. The fairly large discrepancies in the point estimates of extreme clustering in the S&P 500 and NASDAQ composite indices suggest that investment portfolios tracing these indices will have different risk characteristics, which standard treatments using models of conditional variance may not fully reveal. More generally, the results suggest that it may be worthwhile to examine a wide variety of assets for evidence of differences in this feature of risk, which may be obscured by standard treatments using conditional variance models.

REFERENCES

- Cotter, J. (2001) Margin Exceedances for European Stock Index Futures Using Extreme Value Theory. *Journal of Banking and Finance* 25, 1475-1502.
- Davison, A.C. and R.L. Smith (1990) Models for Exceedances Over High Thresholds. *Journal of the Royal Statistical Society Ser. B* 52, 393-442.
- Diebolt, J. and D. Guégan (1993) Tail Behaviour of the Stationary Density of General Non-linear Autoregressive Processes of Order 1. *Journal of Applied Probability* 30, 315-329.
- Embrechts, P., C. Klüppelberg and T. Mikosch (1997) *Modeling Extremal Events for Insurance and Finance*. Springer, Berlin.
- Ferro, C. and J. Segers (2003) Inference for Clusters of Extreme Values. *Journal of the Royal Statistical Society Ser. B* 65, 545-556.
- Galbraith, J.W. and S. Zernov (2004) Circuit Breakers and the Tail Index of Equity Returns. *Journal of Financial Econometrics* 2, 109-129.
- Hall, A., M.G. Scotto and H. Ferreira (2005) On the Extremal Behaviour of Generalised Periodic Sub-sampled Moving Average Models with Regularly Varying Tails. *Extremes* 7, 149-160.
- Hill, B.M. (1975) A Simple General Approach to Inference about the Tail of a Distribution. *Annals of Statistics* 3, 1163-1174.
- Huisman, R., K.G. Koedijk, C.J.M. Kool and F. Palm (2001) Tail Index Estimates in Small Samples. *Journal of Business and Economic Statistics* 19, 208-215.
- Hsing, T. (1991) On Tail Index Estimation Using Dependent Data. *Annals of Statistics* 19, 1547-1569.
- Hsing, T. (1991b) Estimating the Parameters of Rare Events. *Stochastic Processes and their Applications* 37, 117-139.
- Hsing, T. (1993) Extremal Index Estimation for a Weakly Dependent Stationary Sequence. *Annals of Statistics* 21, 2043-2071.
- Kearns, P. and A. Pagan (1997) Estimating the Density Tail Index for Financial Time Series. *Review of Economics and Statistics* 79, 171-175.
- Koedijk, K.G., M.M.A. Schafgans and C.G. De Vries (1990) The Tail Index of Exchange Rate Returns. *Journal of International Economics* 29, 93-108.
- Laurini, F. (2004) Clusters of Extreme Observations and Extremal Index Estimate in GARCH Processes. *Studies in Nonlinear Dynamics and Econometrics* 8, No. 2, article 4.
- Leadbetter, M.R. (1983) Extremes and Local Dependence in Stationary Sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 64, 291-306.
- Leadbetter, M.R. and H. Rootzèn (1983) Extremal Theory for Stochastic Processes. *Annals of Probability* 16, 431-438.

- Ledford, A.W. and J.A. Tawn (2003) Diagnostics for Dependence Within Time Series Extremes. *Journal of the Royal Statistical Society Ser. B* 65, 521-543.
- Lehmann, E.L. (1975) *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, Oakland.
- Longin, F.M. (1996) The Asymptotic Distribution of Extreme Stock Market Returns. *Journal of Business* 63, 383-408.
- Longin, F.M. (1999) Optimal Margin Levels in Futures Markets: Extreme Price Movements. *Journal of Futures Markets* 19, 127-152.
- Longin, F.M. (2000) From Value at Risk to Stress Testing: the Extreme Value Approach. *Journal of Banking and Finance* 24, 1097-1130.
- Longin, F.M. and B. Solnik (2001) Extreme Correlation of International Equity Markets. *Journal of Finance* 56, 651-678.
- Loretan, M. and P.C.B. Phillips (1994) Testing the Covariance Stationarity of Heavy-Tailed Time Series. *Journal of Empirical Finance* 1, 211-248.
- Martins-Filho, C. and F. Yao (2006) Estimation of Value-at-Risk and Expected Shortfall based on Nonlinear Models of Return Dynamics and Extreme Value Theory. *Studies in Nonlinear Dynamics and Econometrics* 10, 2, article 4.
- Mikosch, T. and C. Starica (2000) Limit Theory for the Sample Autocorrelations and Extremes of a GARCH(1,1) Process. *Annals of Statistics* 28, 1427-1451.
- O'Brien, G.L. (1974) The Maximum Term of a Uniformly Mixing Sequence. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 30, 57-63.
- Quintos, C., Z. Fan and P.C.B. Phillips (2001) Structural Change Tests in Tail Behaviour and the Asian Crisis. *Review of Economic Studies* 68, 633-663.
- Resnick, S. and C. Starica (1998) Tail Index Estimation for Dependent Data. *Annals of Applied Probability* 8, 1156-1183.
- Robinson, M.E. and J.A. Tawn (2000) Extremal Analysis of Processes Sampled at Different Frequencies. *Journal of the Royal Statistical Society Ser. B* 62, 117-135.
- Smith, R.L. and I. Weissman (1994) Estimating the Extremal Index. *Journal of the Royal Statistical Society Ser. B* 56, 515-528.
- Weissman, I. and S.Y. Novak (1998) On Blocks and Runs Estimators of the Extremal Index. *Journal of Statistical Planning and Inference* 66, 281-288.