

What information does the distribution of firm sizes in the FT-SE 100 Index provide about long term equity returns?

David Bowie

Department of Actuarial Mathematics and Statistics
Heriot-Watt University, Edinburgh, EH14 4AS
Scotland, Tel: +44 131 451 3204
E-Mail: D.C.Bowie@ma.hw.ac.uk

Abstract

In this paper extreme value distributions are fitted to the market capitalisations of the constituents of the FT-SE 100 Index. It is argued that the type of extreme value distribution fitted provides information about the stochastic process which governed the past growth in capital values of the shares and ultimately resulted in the observed cross-sectional distributions. The evidence suggests that the second moment of the long term growth process is finite and hence that growth processes based on the non-normal Pareto-Levy stable distributions might be inappropriate for modelling stock returns over the longer term. An additional insight is that over the eleven years 1984–1994, the distribution of market capitalisations within the FT-SE 100 became more equitable (in the sense of a more uniform distribution of capital across the constituent companies).

Keywords: extreme value distributions, FT-SE 100 Index, long term stock return distributions, market capitalisation

What information does the distribution of firm sizes in the FT-SE 100 Index provide about long term equity returns?

David Bowie

1. INTRODUCTION

One of the difficulties of trying to establish the distribution of long term firm growth is that for any one long-surviving firm the available data about its size will be reliably recorded for only a very small number of 'long periods'. This paper attempts to model aspects of firm growth in an alternative, indirect manner. The basic premise is that the cross-sectional firm sizes in the market at a point in time can be regarded as a set of possible outcomes for a single hypothetical firm at that time. In other words, if we assume that all firms start at the same size at the same time and are subject to identical independent stochastic growth processes then the cohort of observed firm sizes over time will represent a set of sample paths for the growth of the hypothetical company.

The above assumptions are clearly not valid in the actual market. However, if we consider only the large firms then it might be argued that they will (almost) invariably have been in existence¹ for a long time and so the differences in starting sizes will have less impact on the current size than the growth experienced during their existence. This will imply that there will be very little confounding of cohort effects and growth effects when analysing the cross-sectional sizes.

The assumption that firms will have followed identically distributed growth processes is also difficult to justify. It is well-known that smaller firms will have different growth processes (at least in the short term) from those of larger firms. Furthermore, firms in different industries are subject to different risks and hence different pricing uncertainties. Again, a focus solely on large firms does obviate the difficulties associated with these factors somewhat: the firms will be reasonably homogenous with respect to size and, consistent with the earlier assumption that the large firms will have been in existence for a reasonably long time, they will have survived a wide variety of market conditions. Provided the wide variety of market conditions implies that each industry will have encountered favourable and unfavourable periods equally frequently, we can assume that all the firms have experienced the same spectrum of risks in their lifetimes.

The assumption that the growth processes are independent is also not a trivial assumption! Indeed, the assumption of independence is possibly most pertinent in the case of the large firms. The large firms will have been competing for capital which will have induced a correlation between the growth processes. Specifying what the form of the correlation is likely to have been is beyond the scope of this analysis and we therefore have to interpret any conclusions drawn about the growth processes in a conditional sense, i.e. the inferred growth process is conditional on the correlation structure in the market.

The main thrust of this paper is therefore to use the cross-sectional market capitalisation data from constituents of the FT-SE 100 Index² to generate information about the distribution of long-term growth. Because of the focus on the largest firms, the information provided is concentrated on the tail area of the distribution. A natural framework for exploring the tail area of distributions is provided by extreme value theory (see Coles (1995) or Longin (1993)).

Even if the basic hypothesis that the distribution of firm sizes provides information about the distribution of stock returns is not strictly true, the analysis will still be useful as a consideration of business concentration within the UK stock market. Much research in the past has focused on income inequality and concentrations of resources within the economy and the analysis presented here will provide some comparable empirical evidence about unevenness of capitalisations at the largest end of a formal market structure. (Previous work in this area includes Hart & Prais (1956) and Hart (1960).)

The structure of the paper is as follows: Section 2 is a very brief review of distribution fitting to stock returns; Section 3 contains a brief discussion of extreme value methodology for analysing distributions; Section 4 contains a brief graphical analysis of the distribution of the market capitalisations; Section 5 contains the results of fitting the extreme value distributions to the capitalisations of the FT-SE 100 stocks; Section 6 contains a summary and some concluding remarks.

2. DISTRIBUTIONS FOR STOCK RETURNS AND PRICES

The search for suitable distributions for stock returns and stock accumulations³ can be separated into two streams. The first is defined by the aim of finding a family of distributions which best matches the observed distributions; the second of finding the best fitting distribution which exhibits a set of theoretical properties required for a particular application of financial theory to hold true. Examples of theory driving the second methodology include the desire to have additivity in the distribution so that the distribution of returns should be closed under portfolio formation⁴, and under temporal aggregation⁵. Another desirable feature would be to find a distribution such that the separation theorems of portfolio selection theory hold, see Ross (1978) or Chamberlain (1983).

The earliest work in the area is usually attributed to Fama (1965) and Mandelbrot (1963). The normal and Stable (Levy-Pareto) distributions investigated by them have received continued attention mainly because they are closed under addition, e.g. Walter (1995). Empirical evidence, however, has indicated clearly that the tails of stock return distributions are heavier than the normal distribution. Although the Stable distributions allow for more frequent extreme values than the normal, they are generally difficult distributions to work with as their density functions cannot be written down explicitly (except in a few special cases) and because their variances are generally not finite. Other authors have considered mixtures of normal and Stable distributions (e.g. Fielitz & Rozelle (1983)) or the t -distribution (e.g. Praetz (1972)).

On the purely empirical side of the fitting literature, Bookstaber & McDonald (1987) have suggested the generalised beta distribution of the second kind (GB2) as a flexible

family of distributions suitable for stock accumulations. The family includes a wide variety of familiar distributions which have been proposed in the literature as models for stock accumulations, including the log-normal distribution. One drawback which Bookstaber & McDonald pointed out is that the distribution, in general, does not have the property of closure under multiplication⁶.

An additional motivation for considering the GB2 distribution is that McDonald and others (e.g. McDonald (1984) and McDonald & Butler (1987)) have offered evidence that it is also suitable for modelling other economic variables such as the distribution of income and the duration of unemployment. Given the secondary motivation for this work as being the description of inequality of capitalisation in the stock market, it is sensible to use distributions from a family which is known to model such variables adequately.

In order to put many of the distributions encompassed within the GB2 family in perspective it is useful to give a 'family tree'. The GB2 distribution is a four-parameter family of distributions. Bookstaber & McDonald show that at least four important three-parameter daughter distributions can be obtained from the GB2 distribution either as special cases or as limiting cases of the four-parameter parent. These distributions are commonly referred to as the log- t , the generalised or transformed gamma, the generalised Pareto (beta distribution of the second kind) and the Burr (type 12 or Singh Madalla) distributions. They can in turn give rise to special- or limiting case two parameter families of distributions: for example, the log-normal distribution arises as a limiting case of the generalised gamma distribution or of the log- t distribution; the Weibull distribution can be treated either as a limiting case of the Burr distribution or a special case of the generalised gamma; the gamma distribution is (as may be expected) a special case of the generalised gamma distribution, but can also be described as a limiting case of the generalised Pareto. The GB2 family also arises naturally as a mixture of two distributions from within the family. For example, the mixture of the inverse generalised gamma distribution with the generalised gamma distribution gives the GB2 family. More details can be found in Appendix B.

Although it would clearly seem desirable to model stock accumulations or firm sizes directly by the GB2 family, it turned out to be extremely difficult to estimate the four parameters using standard statistical methodologies, in particular maximum likelihood and minimum Cramer-von Mises distance estimators (details of the methods are provided in Appendix A). The difficulty arises mainly because the data in this study are somewhat non-standard: the data correspond to the year-end market capitalisations of the companies in the FT-SE 100 index and are certainly not a random selection of companies. One approach is to treat the market capitalisations of the FT-SE companies as the order statistics from the conceptual sample of all the companies in existence at any one time. The conceptual sample is of imprecise but large size. The FT-SE stocks provide the largest 90 from this 'sample'⁷. Even fitting the special cases of the GB2 distribution which have fewer parameters (e.g. Weibull) is less than satisfactory because a total sample size has to be assigned to the conceptual sample described above.⁸

Because of the problems associated with fitting conventional distributions, the methodology employed in this paper is to use the extreme value theory (see Coles, 1995) to fit

what are known as the generalised extreme value (GEV) distributions to the FT-SE 100 constituents. Based on asymptotic theory, it has been shown that for all distributions in a domain of attraction, the largest observations in a very large sample will follow one of the GEV distributions (which come in Gumbel, Frechet or Weibull flavours). In other words, the GEV distributions fulfil for the largest values of samples a corresponding role to that performed by the normal distributions for the sample means (in the central limit theorems). Depending on the type of GEV which fits the data, we will be able to draw some tentative conclusions about the long-term nature of firm growth, or at least have some potentially useful information about the cross-sectional distribution of firm size among the largest companies.

3. SOME MORE DETAILS ABOUT THE GEV DISTRIBUTIONS

For convenience of reference, a brief precis of the description of the GEV distribution as given by Coles (1995) is provided in this section.

The distribution function of the GEV is given by

$$G(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \quad (1)$$

defined on $\{x : 1 + \xi(x - \mu)/\sigma > 0\}$. The so-called type II (Frechet) and type III (Weibull) classes of extreme value distribution correspond to the cases $\xi > 0$ and $\xi < 0$ in the above parameterisation. The type I (Gumbel) class arises in the limit as $\xi \rightarrow 0$. The parameter ξ is referred to as the shape parameter while μ and σ are location and scale parameters, respectively. Estimates of extreme quantiles are easily obtained by inverting equation (1):

$$x_p = \mu - \frac{\sigma}{\xi} \left[1 - \{-\log(1 - p)\}^{-\xi} \right] \quad (2)$$

where $G(x_p) = 1 - p$. When the shape parameter, ξ , is negative, the quantiles will be bounded above for all p , in contrast with $\xi \geq 0$ where the quantiles are unbounded. The parameters of the GEV distribution function can be estimated by maximising the likelihood:

$$l(\mu, \sigma, \xi) = \sum_{i=1}^k \left\{ -\log \sigma - (1 + 1/\xi) \log \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right] - \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}. \quad (3)$$

One problem with using maximum likelihood for estimating the parameters of the GEV family is that the end-point of the GEV distribution is parameter dependent ($= \mu \pm \sigma/\xi$). Nevertheless, Coles (1995, section 1.7) indicates that provided $\xi > -0.5$, the maximum likelihood estimators exist and are completely regular. For $-1 < \xi < -0.5$, maximum

likelihood estimators exist, but are non-regular and for $\xi < -1$ maximum likelihood estimators do not exist.

As Coles points out, fitting GEV models is enormously inefficient. The data required to fit the models are the maxima of samples, with the rest of the data being ignored. An alternative method is to take the r -largest order statistics, or all observations greater than a certain threshold level and use these to fit the model. These approaches can be asymptotically motivated by using a point process characterisation. The basic idea is that a set of data can be recast as a point process indexed by 'time' (observation number) and magnitude. We denote the point process for sample size n by P_n :

$$P_n = \left\{ \left(\frac{i}{n+1}, X_i \right) : i = 1, \dots, n \right\} \quad (4)$$

where $\{X_i\}_{i=1}^n$ are a random sample drawn from an unknown distribution, F . Asymptotically, the process, P_n , away from its lower boundary behaves as a non-homogenous Poisson process. We can characterise the process by an intensity measure $\Lambda(A) = E(\text{number of points in region } A)$. If we can assume that the numbers of points in non-overlapping regions are independent, it follows from the Poisson property that for a region $A = \{[0, 1] \times (x, \infty)\}$

$$\begin{aligned} \exp\{-\Lambda(A)\} &= \Pr\{\text{no points in } A\} \\ &= \Pr\{M_n \leq x\} \\ &\approx \exp\left\{-\left[1 + \xi \frac{x - \mu}{\sigma}\right]^{-1/\xi}\right\} \end{aligned} \quad (5)$$

where the last equality holds approximately for large n by virtue of the asymptotic GEV distribution function and where the notation M_n is used to denote the largest element in the sample. Thus, provided n is large and we consider only elements from the sample which are larger than some threshold, P_n will be (approximately) a Poisson process with intensity function given by:

$$\Lambda\{([0, 1] \times (x, \infty))\} = \left[1 + \xi \frac{x - \mu}{\sigma}\right]^{-1/\xi} \quad (6)$$

The point process approach is actually not needed in the analysis reported in this paper because the time-series nature of the observations is ignored. In cases such as this, it is useful instead to use a slightly older methodology based on the generalised Pareto distribution (GPD). Using the notation established in the discussion of the GEV distribution, it has been shown that, asymptotically,

$$\Pr\{X_i > u + x | X_i > u\} = \left[1 + \xi \frac{x}{\sigma}\right]^{-1/\xi} \quad (7)$$

for a threshold u which is sufficiently large. Note that the μ parameter which appears in the GEV formulation is lost because of the conditioning in the GPD case.

The estimation of the parameters σ and ξ can be performed directly by differentiating equation (7) to form the likelihood, and then maximising it over the two parameters. A

faster method of obtaining the maximum likelihood estimates is to reparameterise the likelihood by putting $\lambda = \xi/\sigma$. The maximum likelihood estimator of σ can then be expressed as a simple function of λ , which in turn enables us to write the likelihood in terms of the one variable, λ . The numerical part of the maximisation is thus considerably facilitated.⁹

The reparameterisation also enables us to gain some insight into when we are likely to obtain reasonable estimates of the parameters, in particular of ξ . In order for the probabilities in equation (7) to be positive,

$$\lambda > -1/M_n \Rightarrow \xi > -\sigma/M_n. \quad (8)$$

In particular, if the variability (σ) is larger than the maximum observation (M_n) then it reasonably follows that any estimators based on that sample will be poor; this remark corresponds with the earlier reference to the fact that maximum likelihood estimators of ξ will not exist for $\xi < -1$. In practical terms what can be inferred from an estimate of ξ which is large and negative is that the largest firm size is extremely close to the largest possible firm size and we might infer that the firm size distribution is bounded. These comments have a bearing on the interpretation of the results presented later in Section 5.

Some further remarks on the properties of the GPD will also be useful in the later discussion. In particular, the k th moment about the origin of a random variable drawn from the GPD will exist only if $1/\xi > k$ or if $\xi < 0$. The first condition arises as a well known property of the Pareto distributions and the second condition follows because the distribution is bounded above when $\xi < 0$ and so all moments will be finite. The case where $\xi = 0$ equates to the exponential distribution for which all moments exist and are finite. In particular, the second moment (and hence variance) will be finite provided $\xi < 0.5$. This observation is fairly important because of all the symmetric distributions which are closed under addition, the normal is the only one with finite second moment. The value of ξ could therefore provide important information about whether the non-normal stable Pareto-Levy distributions are sensible distributions for modelling stock returns over the long term.

4. DESCRIPTION OF THE DATA AND MEAN RESIDUAL ANALYSIS

The data in this study comprise the market capitalisations as recorded by Brumwell (1984–1994) of the year-end constituents of the FT-SE 100 Index for the eleven year 1984 to 1994. The constituents and market capitalisations are taken to be those immediately after the review which occurs at the year-end. At each year-end the FT-SE 100 Index contains 100 stocks which correspond roughly to the largest 100 firms listed on the London stock exchange.

A useful graphical technique for identifying what sorts of distributions are likely to provide good models for the tails of samples is the mean residual life plot (see Hogg & Klugman (1984, p109)). The sample mean residual life function, $e_n(x)$, is defined as the difference between the average of all items in the sample greater than or equal to x and x itself. Plots of the sample mean residual life functions of the year-end market

capitalisations for the years 1984, 1986, 1988, 1990, 1992 and 1994 are illustrated in Figure 1. By comparing the sample functions with the theoretical functions derived from particular models, an assessment of the suitability of the various models can be made. In this case the shapes of the sample functions resemble those for Weibull or log-normal distributions (as given in Hogg & Klugman, p109). Of course this method of model identification is a little crude and more formal fitting procedures need to be considered.

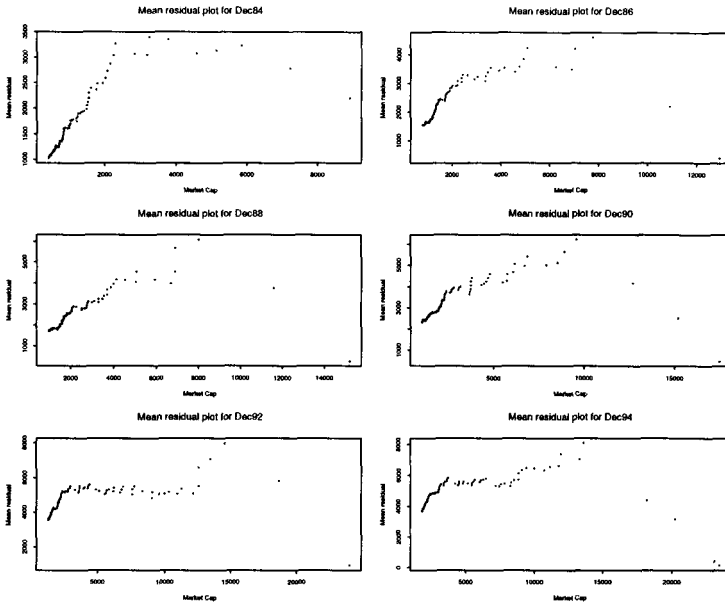


Figure 1: *Sample mean residual life functions for 1984, 1986, 1988, 1990, 1992 and 1994*

5. RESULTS FROM FITTING GPD TO MARKET CAPITALISATIONS

The GPD distribution has been fitted to the data using two approaches. The first is to consider each calendar year separately and to compare the fitted parameters over time. The second is to combine the data from all eleven years into one large sample to which the model is fitted. The results from the two approaches are presented in the following two subsections.

5.1 Results using the year by year data

Table 1 summarises the estimates of ξ obtained from a fit of the GPD to each year of

	Dec84	Dec85	Dec86	Dec87	Dec88	Dec89	Dec90	Dec91	Dec92	Dec93	Dec94
$r=10$	-0.174	0.2771	-0.3123	-0.0247	0.1424	NA	NA	0.0615	0.2731	-0.6382	NA
$r=12$	-0.1053	0.1976	-0.1419	0.3364	-0.0231	-0.6203	-0.1139	0.3539	0.1499	-0.6177	NA
$r=15$	0.3216	0.25	0.003	0.1366	0.2343	0.0625	0.2692	0.089	0.1722	-0.175	-3.2643
$r=20$	0.3985	0.2739	0.0246	0.4039	0.5143	0.5375	0.1875	-0.001	-0.0627	0.0553	0.1082
$r=30$	0.592	0.4247	0.2835	0.3967	0.2753	0.0577	0.1049	0.1579	-0.0245	0.0854	0.0223
$r=40$	0.4844	0.4979	0.3773	0.3923	0.3852	0.434	0.3181	0.1237	-0.015	0.0618	-0.0519
$r=50$	0.6791	0.6908	0.6203	0.5567	0.4959	0.3207	0.512	0.3199	0.0364	0.1273	0.1283
$r=60$	0.6903	0.5544	0.742	0.5341	0.6727	0.5292	0.4142	0.5881	0.2503	0.2054	0.1759
$r=70$	0.6027	0.5204	0.6878	0.55	0.6928	0.5066	0.4483	0.4934	0.3495	0.3266	0.3777
$r=80$	0.6048	0.5889	0.5322	0.5343	0.4427	0.5297	0.4922	0.5669	0.5314	0.4203	0.5138
$r=90$	0.5917	0.4703	0.5229	0.3809	0.4293	0.5617	0.477	0.6325	0.6378	0.5508	0.6736

Table 1: *Fitted ξ parameters for each year using the r -largest firms in each year.*

market capitalisations. Each estimate is based on the market capitalisations of the r -largest firms in each year. Thus, for example, we can see that fitting a GPD model to the 10-largest firms in 1984 resulted in an estimate of ξ of -0.174 . Scanning down each column we note that, in broad terms, the estimates of ξ increase as the number of firms used increases. For example, in 1984, the estimate of ξ starts at -0.174 when the largest 10 firms are used and increases to around 0.60 as the number of firms used is increased to over about 50. What this seems to indicate from an application-specific perspective is that the firm sizes ‘flatten off’ at the very upper end of the market, even though they increase steeply over the index as a whole. Note that the ‘NA’ entries in the table represent cases when the GPD model could not be fitted using maximum likelihood. These cases therefore also correspond to the situation mentioned earlier where the variation in the largest r firms is greater than the magnitude of the largest firm, i.e. the largest observed firm size is close to the largest possible firm size attainable under the estimated model.¹⁰

From a statistical perspective, the systematic trend in the estimates as r is increased represents the change in trade-off between the reduced estimating error involved in using more firms to estimate the parameter and the increased bias in using a lower threshold (which makes the asymptotic nature of the GPD model less likely to be applicable).

Looking across the years (columns) in Table 1, it is interesting to note that for most of the values of r , there exists a downwards trend, e.g. the estimate of ξ for $r = 30$ in 1984 is 0.592 compared with $\hat{\xi} = 0.2753$ in 1988 and $\hat{\xi} = 0.0223$ in 1994. The decrease across time would indicate a general flattening over the period of this investigation of the distribution of firm sizes (among the largest firms). For the largest values of r , i.e. $r \geq 80$, the decrease in the estimates of ξ are not nearly as marked, implying that, taken as a whole, the FT-SE stock sizes still have the same distributional shape.

The picture is therefore one of a fairly steep increase in the size of firms from smallest to largest, with a flattening of the size increments for the very largest stocks. Furthermore, over the eleven years in the sample the ‘flatter’ portion of the size gradient has extended to less large stocks. One interpretation of this flattening of the distribution indicates that firms have shared available capital increasingly fairly over the period 1984-1994. An alternative explanation to the ‘increased equality’ hypothesis is that the number of firms has increased over time, so that the amount of capital available for *all* firms has been

r	1	2	3	4	5	6	7	8	9	10	12	15	20	30	40
ξ	0.05	0.27	-0.35	-0.53	-0.49	-0.43	-0.36	-0.36	-0.29	-0.21	-0.17	-0.02	0.08	0.19	0.27
μ	9992	9688	9766	9883	9796	9625	9489	9494	9377	9291	9267	9290	9331	9643	10134
σ	846	833	1449	1743	1706	1719	1725	1724	1780	1888	1957	2315	2618	3184	3839

Table 2: *The fit of the point process representation of the GEV distribution using various numbers of firms for each year, but including all years and allowing repetitions of firms within the sample.*

more thinly spread.

For $r \leq 40$ only two estimates of ξ are greater than 0.5 and for $r > 40$, the estimates range between 0.5 and 0.7 up until about 1989 and are nearly all less than 0.5 thereafter. On balance, the evidence from this analysis would seem to indicate that $\xi < 0.5$. Further comment will be held back until after we have more information regarding the likely standard errors of the estimates.

5.2 Results obtained by combining years

The other way of using the available data is to pool the years, thereby potentially increasing the number of 'large' firms. There are a couple of additional considerations involved. First, the use of data over eleven years necessitates some 'scaling' of the raw data to allow for the growth in the firm sizes from year to year. There exist many ways of deflating the capitalisations; the one adopted in this paper is to divide all the firms sizes by the All Share Index. The effect of dividing by the index is roughly equivalent to using the residual components from the index model (i.e. the market model with unit beta and zero alpha).¹¹

The second major consideration is that the same firms will appear in the largest r companies for many successive years. Provided that the scaled data are used, the multiple appearances of particular companies should not be a problem in the estimation procedure. However, any inference based on the estimated parameters will be distorted because of biased estimates of the standard errors of the estimates.

The results of the 'ignore the repetition problem, but scale using the All Share Index' approach are summarised in Table 2. The features of the estimated parameters are that the shape parameters, ξ , are negative for small values of r (except for $r = 1$ and $r = 2$ when sample sizes are still very small), but increase and become positive as r is increased. As noted earlier, a negative ξ parameter implies a bound for quantiles of the distribution whereas a positive value of ξ indicates an unbounded distribution. Thus it would appear that when we take the largest five, say, firms, the implied asymptotic distribution is of the Weibull (bounded) GEV type, whereas the firms smaller than that ranked from, say, six would indicate a Fréchet (unbounded) GEV type.

The μ parameter is fairly constant over time. Indeed adding in a linear time effect provided virtually no (we cannot really say 'insignificant' because of the violation of independence of the sample items) improvement in the fit of the model. Indeed, the lack of change is quite surprising given that many smaller securities are added to the sample by increasing r . The σ parameter on the other hand does increase as expected as the

r		10	12	15	20	30	40	50
Regular	mean	-0.3509	-0.5184	-0.1427	0.059	0.1218	0.2423	0.3509
	s.d.	1.5029	1.2378	0.6934	0.2487	0.1332	0.1168	0.1026
All	mean	-1.6828	-1.1473	-0.5009	-0.1294	0.1176	0.2479	0.3512
	s.d.	2.2338	1.686	1.1531	0.6953	0.1351	0.1223	0.1006

r		60	70	80	90	110	130	150
Regular	mean	0.4467	0.5705	0.6305	0.6428	0.6637	0.6598	0.4396
	s.d.	0.0995	0.121	0.1151	0.0941	0.0909	0.0703	0.0401
All	mean	0.4514	0.5681	0.6282	0.6441	0.6609	0.6546	0.4392
	s.d.	0.0968	0.121	0.1151	0.0924	0.0882	0.0693	0.0385

Table 3: *Summary statistics of fit of GPD model to random samples of all firms in the FTSE in the period 1984-94.*

number of securities is increased.

An additional study was performed in order to estimate an ‘average’ ξ parameter and a standard deviation of the estimate. The analysis is a bootstrap-type of analysis with samples constructed as follows. Each share which appeared at some time in the FTSE 100 index was listed along with its scaled market capitalisations at the date of each December review for which it was in the index. One of the sizes associated with the firm was selected at random to be included in the sample. The sample therefore comprised 167 (the total number of different firms in the index at the December reviews; see also Appendix C) firm sizes. The sampling procedure was then repeated 50 times to give 50 samples each of size 167.¹²

For each of the samples thus produced, the GPD model was fitted using maximum likelihood. For some samples the estimated value of ξ was less than minus one, i.e. the estimate was not completely regular as indicated earlier. Table 3 is a summary of the estimated ξ parameters for various values of r . The means and standard deviations of the estimates across samples are given in two forms: one using only those estimates for which the maximum likelihood estimates were regular and the second including all estimates, whether regular or not. Contrasting first the use of the ‘regular’ and ‘all’ estimates, we note that the ‘regular’ estimates are generally larger than the ‘all’ estimates for small values of r . As the number of firms included is increased, i.e. r is increased, so the difference diminishes.

The second point to note from Table 3 is that, as in Table 1, the estimated value of ξ increases from being negative for small r to being positive for larger r . The average value of ξ stabilises for $r \geq 70$ at around 0.6 but drops again to 0.44 when $r = 150$. We also note that there is not much reduction in the sampling standard error for $r \geq 40$ until r is very much bigger (over 90, say). Remembering that the bias in the estimate of ξ caused by not using only the largest sample value will increase as r is increased, it seems reasonable to hypothesise that $\xi \approx 0.3$ (i.e. Fréchet limiting distribution with finite variance). Even then, the magnitude of the standard errors would not easily reject a null hypothesis of $\xi = 0$ (i.e. Gumbel distribution, with all moments finite).

6. SUMMARY AND CONCLUDING REMARKS

The aim of this study has been to provide evidence as to which type of extreme value distribution best fits the largest market capitalisations on the UK market. Broadly speaking, the evidence would appear to support the hypothesis that the Frechet family with finite second moments, or possibly the Gumbel family, of the extreme value distributions were most appropriate. In other words, the distribution of the (deflated) market capitalisations is at least likely to have finite variance, if not finite higher order moments as well. For the very largest firms (the largest five, say), the distribution seems to be of type Weibull, i.e. bounded above with all the moments being finite.

If we interpret the cross-sectional distribution of market capitalisations as representing a sample of firm sizes for a long-surviving hypothetical company listed in the UK, then this evidence would indicate that the growth process was similarly of finite variance. The boundedness of the distribution of the very largest firms (i.e. the decreasing marginal increase in size from one firm to the next largest) could be ascribed to competition for finite amounts of capital at the very largest end of the market, or possibly because of regulatory and legislative interference (e.g. anti-trust laws).

The implication of the above assessment of the results for long-term stochastic investment models would be that even if it were desired that jumps were included in the growth process, then the magnitude of the jumps would have to have some upper limit. Furthermore, the evidence that might be inferred from the results in favour of 'jumps' (interpreted here as infinite variance in the growth process) is also not very strong.

A brief look at the temporal pattern in the fitted distributions indicated that the bounded nature of the market capitalisations of the firms in the FT-SE 100 had become more noticeable over the years from 1984 to 1994. The apparent increased equity in the distribution of capital across the FT-SE 100 stocks might be a peculiarity of the time period, or an indication that exogenous factors were encouraging capital to be more uniformly distributed across firms.

NOTES

¹It is important to realise that 'firms' may have existed even if they were not formally incorporated or listed on the stock exchange.

²The FT-SE 100 Index is intended to comprise the hundred largest stocks on the UK market. We provide more details of the firms actually in the index in a later section.

³In this paper the phrase 'stock accumulation' is used to refer to $R_t = P_t/P_{t-1}$ where P_t is the quoted price of the stock and 'stock return' to refer to $\log R_t$. Therefore, stock returns can be considered the continuously compounded rate of growth, and stock accumulations are one plus the effective rate of price growth. Clearly finding a distribution for either the stock accumulations or returns will immediately define the distribution for the other.

⁴Portfolio formation is the linear combination of securities into a portfolio. Closure of the distribution would mean that if the securities all had stock return distributions from a particular family then the portfolio would have a return distribution from the same family.

⁵The concept of closure under temporal aggregation is that if the returns on a stock over a period of a month, say, have a distribution from a family then the returns on the stock over periods of a year, say, should follow a distribution from that same family

⁶Closure under multiplication for accumulations is equivalent to closure under addition for returns since returns are just the logarithm of the accumulations.

⁷Just after each quarterly review by the FTSEA index committee, the FT-SE companies comprise the largest 90 stocks plus ten other stocks which actually have size ranks between 91 and 110. This peculiarity arises because the committee wishes to restrict the number of times a stock moves into, or out of, the index because of only small movements.

⁸The Generalised Gamma and Weibull distributions turned out to be the best fitting of the special cases of the GB2 distribution, but details of the fit are not reported in this paper.

⁹Stan Zachary kindly provided insight, assistance and code at this point.

¹⁰Convergence of the maximum likelihood estimation routine was also achieved for the 1994 data with $r = 15$ and $\xi < -1$.

¹¹A possible complication with this approach is that a market index which is corrected to remove the effects of capital changes in its constituents (such as the All Share Index) will provide a biased comparison for the growth of an individual firm's capital. On the other hand, if we use an index which consists of, say, the total market capitalisation of the constituent securities, then it is possible that this index will behave differently from some groups of securities (e.g. the largest) which may have a different propensity to capital restructuring than the 'average' constituent security.

¹²Note that for firms which appeared in the index frequently, such as the largest firms (Glaxo, BT, Shell, etc.) there are ten possible sizes which might appear in a sample. For firms which appeared only once in the index, the same size would appear in all 50 samples. The bootstrap is therefore not completely 'symmetric' in that some values are expected to appear more often than others. I am relying quite heavily on the fact that firms which made only a few appearances do so mainly at the smaller sizes where their effect is less important on the estimator of ξ .

REFERENCES

- BOOKSTABER, R.M. & J.B. McDONALD (1987), A General Distribution for Describing Security Price Returns, *Journal of Business*, **60**, 401–424
- BRUMWELL, J. (1984 to 1994), Notes on the Financial Times-Actuaries UK Equity Share Indices, *Journal of Institute of Actuaries, Transactions of the Faculty of Actuaries and British Actuarial Journal*
- CHAMBERLAIN, G. (1983), A Characterization of the Distributions that imply Mean-Variance Utility Functions, *Journal of Economic Theory*, **29**, 185–201
- COLES, S. (1995), *Extreme Value Theory and Applications*, Lancaster University course notes
- FAMA, E.F. (1965), The Behaviour of Stock-Market Prices, *Journal of Business*, **38**, 34–105
- FIELTIZ, B.D. & J. P. ROZELLE (1983), Stable Distributions and the Mixtures of Distributions Hypotheses for Common Stock Returns, *Journal of the American Statistical Association*, **78**, 28–36
- FT-SE ACTUARIES SHARE INDICES STEERING COMMITTEE (1996), *Guide to Calculation Methods*, Version 2
- HOGG, R.V. & S.A. KLUGMAN (1984), *Loss Distributions*, John Wiley and Sons
- HART, P.E. & S.J. PRAIS (1956), The Analysis of Business Concentration: a Statistical Approach (with comments), *Journal of the Statistical Society*, **119**, 150–191
- HART, P.E. (1960), Business Concentration in the United Kingdom, *Journal of the Statistical Society*, **123**, 50–58
- LONGIN, F. (1993), Booms and Crashes: Application of Extreme Value Theory to the US Stock Market, *Working paper: Institute of Finance and Accounting, London Business School*, 179
- MANDELBROT, B. (1963), The Variation of Certain Speculative Prices, *Journal of Business*, **36**, 394–419
- MCDONALD, J.B. (1984), Some Generalized Functions for the Size Distribution of Income, *Econometrica*, **52** (3), 647–663
- MCDONALD, J.B. & R.J. BUTLER (1987), Some Generalized Mixture Distributions with an Application to Unemployment Duration, *Review of Economics and Statistics*, **69**, 232–240
- PATEL, J.K., KAPADIA, C.H. & D.B. OWEN (1976), *Handbook of Statistical Distributions*, Marcel Dekker, Inc.: New York and Basel
- PRAETZ, P.D. (1972), The Distribution of Share Price Changes, *Journal of Business*, **45**, 49–55
- RAINVILLE, E.P. (1960), *Special Functions*, Macmillan: New York
- ROSS, S.A. (1978), Mutual Fund Separation in Financial Theory – The Separating Distributions, *Journal of Economic Theory*, **17**, 254–286
- WALTER, C. (1995), Levy-Stability-under-Addition and Fractal Structure of Markets: Implications for the Actuaries and Emphasized Examination of MATIF Notional Contract, *Proceedings, AFIR 1995 Congress*, **III** (B), 159–203

Appendices

A. METHODS USED FOR FITTING GB2 DISTRIBUTIONS

Two fairly common methods of fitting distributions to the data were considered: namely the minimum Cramer-von Mises distance method and the maximum likelihood approach. The motivation for implementing these two methods is firstly to provide a check on each other and secondly because they are well-known statistical methods with established asymptotic properties.

A.1 The minimum Cramer-von Mises statistic

Let X_i be a random variable denoting market capitalisation of the i th largest stock out of a total, but unobserved, sample of n stocks and let x_i be the corresponding observed sample value. Let $F(x; \theta)$ denote the distribution function for the proposed model of stock sizes with θ being a vector of parameters and $F_n(x)$ the empirical distribution function for the sample.

The minimum distance measure based on the Cramer-von Mises statistic is the solution to

$$\operatorname{argmin}_{\theta} \frac{1}{n} \sum_{i=1}^n [F_n(x_i) - F(x_i; \theta)]^2. \quad (9)$$

Because we observe only $\{x_1, x_2, \dots, x_{90}\}$ at any one point, we define the Cramer-von Mises estimator for the purposes of this investigation by:

$$\operatorname{argmin}_{\theta} \sum_{i=1}^{90} [F_n(x_i) - F(x_i; \theta)]^2. \quad (10)$$

Hogg & Klugman (1984, pp83,136) suggest weighting each term in the sum by the inverse of the asymptotic error of the estimator of the empirical distribution function. In addition they suggest replacing the empirical distribution function by $\frac{n}{n+1} F_n(x_i)$. However, because in this case the sample size is large ($n \gg 90$) the modifications will not make much difference at all.

A.2 Maximum likelihood

The joint density function for the 90 largest order statistics from a sample of size n is given by (see, e.g. Patel *et al.* (1976, p63)) the following formula, where $f(x; \theta)$ is the density function for a random variable X drawn from the model:

$$f_{1,2,\dots,90}(x_1, x_2, \dots, x_{90}) = \text{const} \times [1 - F(x_{90}; \theta)]^{n-90} \times \prod_{i=1}^{90} f(x_i; \theta) \quad (11)$$

where $\text{const} = n \times (n-1) \times \dots \times (n-89)$.

This likelihood can be maximised (or equivalently, the negative logarithm of the likelihood minimised) numerically to obtain maximum likelihood estimators of θ .

The choice of n , the total size of the sample of which only the largest 90 items are observed is difficult to specify, both for the Cramer-von Mises estimator and the maximum likelihood estimators. One possible value is 6,000 which corresponds to roughly twice the number of listed companies, the increased size being used because unlisted companies would also be competing for capital.

B. SUMMARY OF GB2 DISTRIBUTIONS

The density function for a random variable from the GB2 distribution is given by:

$$f(x; a, b, p, q) = \frac{|a|x^{ap-1}}{b^{ap}B(p, q)[1 + (x/b)^a]^{p+q}} \quad (12)$$

and the distribution function by

$$\frac{y^p}{bB(p, q)} {}_2F_1(p, 1 - q, p + 1; y) \quad (13)$$

where

$$y = \frac{(x/b)^a}{1 + (x/b)^a} \quad (14)$$

and $B(\cdot, \cdot)$ is the beta function and ${}_2F_1(\cdot, \cdot, \cdot; \cdot)$ is the hyper-geometric function (see Rainville, 1960).

The model fitting routines were unable to produce satisfactory fits to the capitalisation data, i.e. convergence could not be obtained using the S-Plus *ninmb* function, for either the minimum distance or maximum likelihood estimators.

B.1 The log-*t* distribution

This log-*t* distribution is obtained from the GB2 family as the limiting case when the parameter a tends to infinity. For fitting and interpreting the model, it is more instructive to consider a reparameterisation of the density function so that the log-*t* density has parameters μ , representing the mean, σ , the scale parameter, and d , the number of degrees of freedom, of the logarithm of the random variable.

Attempts to fit this model generally failed. However, further inspection of the output from the minimisation routine indicated that the problem arose because the algorithm was attempting to increase the d parameter without bound. A log-*t* random variable with a very large d parameter in its density is equivalent to a log-normal random variable in behaviour. The two-parameter log-normal distribution would therefore appear to provide as good a fit as the three-parameter log-*t* distribution.

B.2 The generalised Pareto distribution

The three-parameter generalised Pareto distribution is obtained from the GB2 family as a special case when the parameter a equals unity. The parameterisation used in this article is as in Hogg & Klugman (1984, p223). The density is given by:

$$f(x; \alpha, \lambda, k) = \frac{\Gamma(\alpha + k)\lambda^\alpha x^{k-1}}{\Gamma(\alpha)\Gamma(k)(\lambda + x)^{k+\alpha}} \quad (15)$$

and the distribution function by

$$F(x; \alpha, \lambda, k) = B(k, \alpha; \frac{x}{\lambda + x}) \quad (16)$$

where $B(u, v; w)$ is the incomplete beta function (see Hogg & Klugman, p217), which is a special case of the hypergeometric function.

B.3 The Generalised Gamma distribution

The generalised Gamma is the three-parameter limiting case of the GB2 family obtained as the limit as the parameter q tends to infinity. The density, using the parameterisation given in Hogg & Klugman, is:

$$f(x; \alpha, \lambda, \tau) = \frac{\lambda^\tau \alpha^\tau x^{\tau\alpha-1} e^{-(\lambda x)^\tau}}{\Gamma(\alpha)} \quad (17)$$

and the distribution function by:

$$F(x; \alpha, \lambda, \tau) = G(\alpha; (\lambda x)^\tau) \quad (18)$$

where $G(\cdot; \cdot)$ is the incomplete gamma function.

B.4 Burr distribution

The Burr or Singh-Madalla distribution is the fourth and last of the three-parameter versions of the GB2 family considered. It is obtained from the general GB2 formulation by setting $q = 2$.

The density function is given by:

$$f(x; \alpha, \lambda, \tau) = \alpha \tau \lambda^\alpha x^{\tau-1} (\lambda + x^\tau)^{-\alpha-1} \quad (19)$$

and the distribution function by:

$$F(x; \alpha, \lambda, \tau) = 1 - \left(\frac{\lambda}{\lambda + x^\tau} \right)^\alpha \quad (20)$$

B.5 Two-parameter distributions

The two parameter distributions considered were

Log-normal: As mentioned above, this very standard distribution can be obtained as a limiting case of the log- t or generalised gamma families.

Pareto: This distribution is the special case of the generalised Pareto family with $k = 1$.

Gamma: Special case of the generalised gamma family, with $\tau = 1$.

Weibull: Can be specified in two ways: first, as a special case of the generalised gamma family with $\alpha = 1$ and substitute a parameter $c = \lambda^\tau$. Secondly, it can be obtained as a limiting case of the Burr family as $\alpha \rightarrow \infty$ and putting $c = -\alpha/\lambda$.

C. DYNAMICS WITHIN THE FT-SE 100

Only 17 different firms¹³ occupied the top ten ranks over the eleven years 1984 to 1994 and only 29 different firms occupied ranks 1 through 20. As reported previously, a total of 167 different firms were present in the FT-SE 100 index at some stage in the years 1984-94 which indicates that the rate of substitution in the index as a whole is fairly similar to that in the top ten and top 20 places. Indeed a similar rate of substitution is present throughout the whole index. Table 4 gives the number of different shares whose maximum rank fell into the corresponding deciles (i.e. ranks 1 to 10, ranks 11 to 20, etc.) during the period 1984-94.

decile:	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	Total
No. shares:	17	12	24	19	15	18	13	20	19	10	167

Table 4: *Numbers of securities whose maximum rank fell into the corresponding rank deciles.*

