Analysing Financial Returns by Using Regression Models Based on Non-Symmetric Stable Distributions

Article *in* Applied Statistics · February 1999

DOI: 10.1111/1467-9876.00161 · Source: RePEc

2 authors:

Philippe Lambert
University of Liège

James Lindsey
University of Liège

Appl. Statist. (1999)
48, *Part* 3, *pp.* 409–424

Analysing financial returns by using regression models based on non-symmetric stable distributions

Philippe Lambert and J. K. Lindsey

Université de Liège, Belgium

[Received August 1997. Final revision January 1999]

Summary. The daily evolution of the price of Abbey National shares over a 10-week period is analysed by using regression models based on possibly non-symmetric stable distributions. These distributions, which are only known through their characteristic function, can be used in practice for interactive modelling of heavy-tailed processes. A regression model for the location parameter is proposed and shown to induce a similar model for the mode. Finally, regression models for the other three parameters of the stable distribution are introduced. The model found to fit best allows the skewness of the distribution, rather than the location or scale parameters, to vary over time. The most likely share return is thus changing over time although the region where most returns are observed is stationary.

Keywords: Extreme values; Fourier transform; Infinite variance; Regression model; Skewness; Stable distribution

1. Introduction

Processes in economics often include considerable 'noise' as they evolve. Share prices in stock-markets, even if they tend to show a global upward, constant or downward trend, are usually extremely variable. The daily closing price y_t (in pence) of the shares of the (British-based bank) Abbey National between July 31st and October 8th, 1991 (Buckle, 1995), reproduced in Table 1, displays such a behaviour, as can be seen in a plot of the share relative returns (circles and 0s) against time t in Fig. 1. (Note that Buckle considered as a datum a closing price for the last Monday of August 1991, the 26th, a bank-holiday in England. In fact, this was just the closing price of the previous trading day, Friday, yielding a zero return for that Monday.)

This behaviour, resulting from a large number of external (usually uncontrollable) independent influences, might be modelled by using simple distributions such as the normal distribution, if it were not for some rather common extreme observations (such as those indicated by a 0 in Fig. 1). These 'outliers', although they may be of no direct interest in studying the global evolution of share prices, cannot be discarded. They do not result from recording errors: they were truly produced in some way by the stochastic process under study.

Various questions can be asked when studying the evolution of relative returns.

(a) We might be interested in understanding and forecasting extreme events to develop a short-term investment strategy for example.

Address for correspondence: Philippe Lambert, Faculté d'Economie, de Gestion et de Sciences Sociales, Université de Liège, Sart-Tilman B31, B-4000 Liège, Belgium.
E-mail: plambert@ulg.ac.be

 0035–9254/99/48409

Table 1. Prices and relative returns of Abbey National shares between July 31st and October 8th, 1991

Day	*Price (pence)*	*Relative return*	*Day*	*Price (pence)*	*Relative return*	*Day*	*Price (pence)*	*Relative return*
July 31st	296	—	August 23rd	307	0.0066	September 17th	295	0.0000
August 1st	296	0.0000	August 26th	—	—	September 18th	293	−0.0068
August 2nd	300	0.0135	August 27th	304	−0.0098	September 19th	292	−0.0034
August 5th	302	0.0067	August 28th	303	−0.0033	September 20th	297	0.0171
August 6th	300	−0.0066	August 29th	304	0.0033	September 23rd	294	−0.0101
August 7th	304	0.0133	August 30th	304	0.0000	September 24th	293	−0.0034
August 8th	303	−0.0033	September 2nd	309	0.0164	September 25th	306	0.0424
August 9th	299	−0.0132	September 3rd	309	0.0000	September 26th	303	−0.0098
August 12th	293	−0.0201	September 4th	309	0.0000	September 27th	301	−0.0066
August 13th	294	−0.0034	September 5th	307	−0.0065	September 30th	303	0.0066
August 14th	294	0.0000	September 6th	306	−0.0033	October 1st	308	0.0165
August 15th	293	−0.0034	September 9th	304	−0.0065	October 2nd	305	−0.0097
August 16th	295	0.0068	September 10th	300	−0.0132	October 3rd	302	−0.0098
August 19th	287	−0.0271	September 11th	296	−0.0133	October 4th	301	−0.0033
August 20th	288	0.0035	September 12th	301	0.0169	October 7th	297	−0.0133
August 21st	297	0.0312	September 13th	298	−0.0100	October 8th	299	0.0067
August 22nd	305	0.0269	September 16th	295	−0.0101			

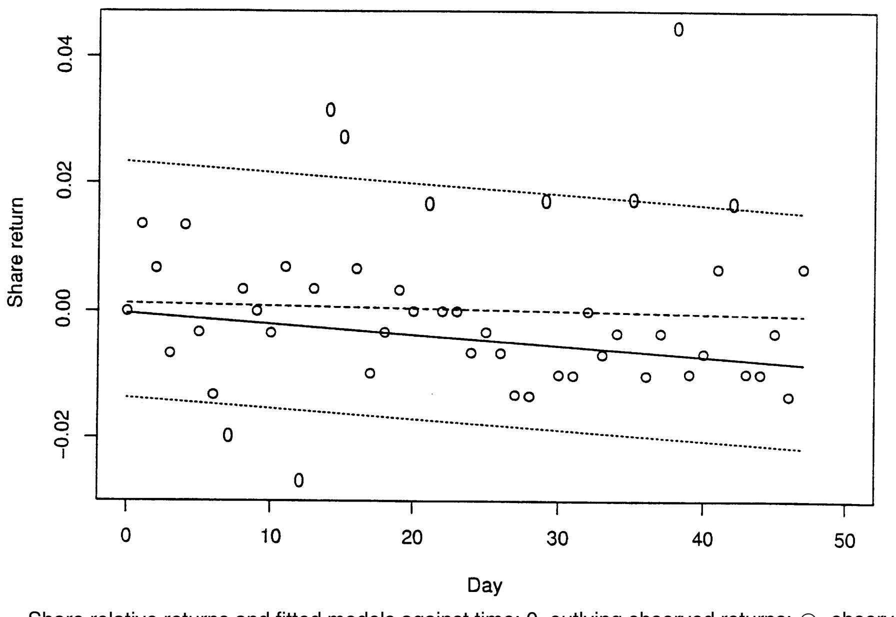

Fig. 1. Share relative returns and fitted models against time: 0, outlying observed returns; ○, observed returns; ———, fitted mode under the stable model (d); ··········, 10% relative density contours under the stable model (d); - - - - -, mean under the normal model (l)

(b) We might be concerned by changes in the shape of the region where most returns are observed, occasional extreme responses being of no direct interest. Hence, these observations should not be influential in an analysis undertaken to answer such a question.

(c) Finally, we might wish to describe how the most probable relative return is changing with time, informing us about trends in likely future returns. Again, extreme values should not influence this description as they are occasional and thus rather unlikely.

In this paper, we shall propose answers to the last two questions for the Abbey National shares data set.

The traditional normal regression model (on time) is clearly an inadequate tool as the slope of the fitted mean (plotted as a broken line in Fig. 1) is very sensitive to the occasional extreme (0) observations. Therefore, it is important to use models that are based on distributions with heavier tails than the normal distribution when the extreme observations are considered to be a reality, but of no direct interest. In this context, stable distributions provide an attractive alternative approach.

Such distributions have been used in various fields besides economics (Mandelbrot, 1963; Buckle, 1995), such as physics (Janicki and Weron (1994), pages 115–116) and survival analysis (Hougaard, 1986).

Stable distributions are defined in Section 2 by using their characteristic function. The main properties of stable processes are then presented and some well-known members described. In Section 3, we discuss appropriate inference procedures. Most of the methods used in the literature put a constraint on the skewness parameter, restricting discussion to symmetric stable distributions. The direct likelihood inference procedure that we use does not rely on such an assumption and is thus more general. It even allows direct modelling of this skewness parameter.

In Section 4, stable generalized regression models for the location parameter are suggested as a generalization of the well-known normal linear regression. When parameters other than the location are held constant, this is shown to induce an implicit regression model for the mode, the most probable response, a model that is easy to interpret. Finally, the model is further generalized to regression equations for any or all of the four parameters of the distribution.

In Section 5, the Abbey National share relative returns are analysed by using these stable generalized regression models. In the last section, we discuss the implications of our results.

The data that are used in this paper can be obtained from

```
http://www.blackwellpublishers.co.uk/rss/
```

2. Stable distributions

We first review the definition and the basic properties of stable distributions. A systematic review of their theoretical aspects can be found in Samorodnitsky and Taqqu (1994).

2.1. Definition

A random variable Y is said to have a stable distribution if and only if it has *a domain of attraction*, i.e. if there is a sequence of independent and identically distributed random variables $\{Z_1, Z_2, \ldots, Z_n\}$ and constants $\{a_n\}$ and $\{b_n\}$ such that

$$\frac{Z_1 + Z_2 + \ldots + Z_n}{a_n} + b_n \qquad \forall n$$

converges in distribution to Y. In the special case where the Z_is have a finite variance, it can be shown that Y is normally distributed.

The common way to specify a stable distribution is by its characteristic function $\phi(t)$, its density function $f(y)$ not being available in an explicit form, except in three special cases that we discuss in Section 2.2. Until recently, this has been a major justification for not considering stable distributions in practice, although their interesting theoretical properties indicate their potential importance in applied statistics. Indeed, they provide the only possible limiting distributions for the sum of independent, identically distributed random variables. This property, which was originally used to derive the form of their characteristic function, generalizes the central limit theorem to the case of infinite variance. Therefore, when an observed process results from a large number of external (usually uncontrollable) independent influences, as is often the case in economics, the family of stable distributions is an alternative that is worth considering.

The logarithm of the characteristic function of the four-parameter family of stable distributions can be written

$$\begin{aligned}\log\{\phi(t)\} &= \mathrm{i}\gamma t - \delta|t|^{\alpha}\{1 + \mathrm{i}\beta\,\mathrm{sgn}(t)\,\omega(t, \alpha)\} \qquad (1)\\ &= \log(\mathcal{F}_t^{+} f) = \log\left\{\int_{\mathbb{R}} \exp(\mathrm{i}ty)\, f(y)\,\mathrm{d}y\right\}\end{aligned}$$

where

$$\omega(t, \alpha) = \begin{cases} \tan(\pi\alpha/2) & \text{if } \alpha \neq 1,\\ (2/\pi)\,\log|t| & \text{if } \alpha = 1,\end{cases}$$

$\mathcal{F}_t^{+} f$ denotes the positive Fourier transform of $f(\cdot)$, $\gamma \in \mathbb{R}$ is a *location* parameter, $\delta \in \mathbb{R}^{+}$ is a *scale* parameter, $\alpha \in\,]0, 2]$ is the *characteristic exponent* determining the type of stable distribution, especially the thickness of the tails, and $\beta \in [-1, 1]$ is an index of *skewness*. The distribution is respectively left skewed or right skewed when $\beta > 0$ or $\beta < 0$, and it is symmetric when $\beta = 0$. Note that this has the opposite interpretation in terms of the sign compared with the traditional coefficient of skewness based on third moments.

When a random variable Y has such a characteristic function, we shall write

$$Y \sim S_{\alpha}(\gamma, \delta, \beta).$$

The distribution of Y can be standardized by using

$$Z = \frac{Y - \gamma}{\delta^{1/\alpha}} \sim S_{\alpha}(0, 1, \beta).$$

The properties of any stable distribution can be deduced from the standardized stable distribution with the same values of α and β.

2.2. Special cases

In special circumstances, the characteristic function in equation (1) corresponds to a density function that can be written in an explicit form.

(a) $\alpha = 1$ and $\beta = 0$ yields the symmetric (about γ) Cauchy distribution with density

$$f(y) = \frac{1}{\pi\delta\{1 + (y - \gamma)^2/\delta^2\}}.$$

(b) $\alpha = 2$ yields the normal distribution $N(\gamma, 2\delta)$.

(c) $\alpha = \frac{1}{2}$ and $\beta = 1$ yields the Lévy distribution with density

$$\frac{\delta}{\surd(2\pi)}\,(y - \gamma)^{-3/2}\,\exp\left\{-\frac{\delta^2}{2(y - \gamma)}\right\} \qquad \text{with } y > \gamma.$$

In all other situations, the density must be generated numerically. However, as Hoffmann-Jørgensen (1994), pages 406–411, pointed out, the characteristic function can be inverted and expressed in terms of incomplete hypergeometric functions, which are generally only defined as the sum of infinite series that can be numerically approximated.

A series expansion of the density is also available in the symmetric case ($\beta = 0$; see Fama and Roll (1968)) when $\alpha > 1$. These series can be integrated term by term to yield an approximation to the cumulative density of symmetric standardized stable random variables.

2.3. Properties

A random variable, $Y \sim S_\alpha(\gamma, \delta, \beta)$, generally takes its values on $\mathbb{R}$. Notable exceptions to this are the Lévy distribution mentioned above and distributions with $\alpha < 1$ and $\beta = -1$, these having zero density on $(-\infty, \gamma]$. By definition, stable distributions are invariant under addition, i.e. the sum of independent variables with the same characteristic exponent α will still be stable with that characteristic exponent. Moments of order greater than or equal to α do not exist, unless $\alpha = 2$, in which case all the moments are finite. From this, we see that all the stable distributions have an infinite variance, except the normal distribution. Moreover, their mean is defined if and only if $\alpha \in (1, 2]$. Therefore, stable distributions are potential candidates to model heavy-tailed processes. All stable distribution functions are unimodal (Yamazato, 1978) and bell shaped (Gawronski, 1984).

3. Inference

Various methods have been suggested to estimate the parameters of a stable distribution $S_\alpha(\gamma, \delta, \beta)$ when the corresponding random variable Y is assumed to have generated a sample $\{y_1, \ldots, y_n\}$. We review some of these techniques in Section 3.1. Then, in Section 3.2, we show how to compute the likelihood function by a numerical inversion of the Fourier transform in equation (1) to obtain the density function at any point, and hence the likelihood. This evaluation of the likelihood function is made possible by efficient non-linear optimizers. We also show how traditional linear, and non-linear, regression techniques can be implemented practically.

3.1. Review of current techniques

Fama and Roll (1968) integrated series expansions of the density to construct tables of cumulative density functions (CDFs) for standardized symmetric ($\beta = 0$) stable distributions at various values of α in [1, 2]. The location and scale parameters are estimated with sample quantiles and used to standardize the observations. The corresponding empirical CDF is then compared with the theoretical CDFs in these tables to obtain an estimate for α.

Press (1972) suggested alternative methods that are also suitable for asymmetric distributions. The resulting parameter estimates are based on the empirical characteristic function

$$\hat{\phi}(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(\mathrm{i} t y_j).$$

By noting that $|\phi(t)| = \exp(-\delta|t|^\alpha)$, and considering two distinct values t_1 and t_2 for t, Press (1972) suggested, in the case $\alpha \neq 1$, 'moment' estimates for α and δ, obtained by simultaneously solving

$$\tilde{\delta}|t_i|^{\tilde{\alpha}} = -\log|\hat{\phi}(t_i)| \qquad i = 1, 2.$$

An analogous argument based on the imaginary part of the logarithm of the characteristic function also leads to moment estimates for the location and skewness parameters. Similar expressions are also available in the case $\alpha = 1$. Press (1972) mentioned the problem of developing a strategy for choosing the t_is, but he did not propose a solution although this should affect the precision of the estimates. He did, however, point out that most applications in economics involve large data sets (e.g. daily observation of a share over several years) which, owing to the consistency of $\hat{\phi}(t)$, lead to reasonable estimates of the parameters. The asymptotic distribution of these estimators can also be derived in the symmetric case. This can be used to construct confidence intervals.

Paulson *et al.* (1975) also used the empirical characteristic function to develop a method of estimation. Their idea, which was already partially developed by Press (1972), is to find the parameters α, γ, δ and β that minimize

$$\int_{-\infty}^{\infty} |\hat{\phi}(t) - \phi(t)| \exp(-t^2) \, \mathrm{d}t. \tag{2}$$

It is legitimized by the coincidence of distributions sharing the same characteristic function. They illustrated their technique by analysing security prices. Koutrouvelis (1980) also used the empirical characteristic function to estimate stable distribution parameters in a 'regression-type' approach.

On an empirical basis, Fielitz and Smith (1972) suggested using asymmetric stable distributions, instead of the often-recommended symmetric distributions, to model changes in stock prices. Leitch and Paulson (1975) also showed that 'symmetry is definitely the exception, not the rule' when modelling changes in log(stock prices). They illustrated their claim by using the empirical characteristic function method of Paulson *et al.* (1975) discussed above. This procedure gives better results than that of Fama and Roll (1968, 1971), particularly with parameters affecting the tail.

DuMouchel (1973) reminded us that stable distributions are not the only alternative to model long-tailed (infinite variance) processes. He suggested that, in some situations, extreme observations might be generated by a different process than that producing the main body of the data. Therefore, he proposed a mixture model of a (short-tailed) normal distribution with a symmetric (long-tailed) stable distribution (with $\alpha < 2$). This enables us to assess (partially) the validity of the usual 'stability hypothesis'. (Note that Teichmoeller (1971) has shown that a mixture of normal distributions does not properly model stock price changes, a typical example of a heavy-tailed process.) DuMouchel (1973) suggested computing the Kullback–Leibler information to discriminate between symmetric stable and mixture models. The same method could be used with alternative long-tailed distributions. DuMouchel (1973) also warned that

> 'arguments for applying a stable model to data based on appeals to the generalized central limit theorem have much less force in the infinite-variance case than in the finite-variance normal theory case'.

Thus, in practice, the stable family of distributions should certainly not be considered as the ultimate answer, even to modelling share prices, as often suggested in the literature.

In a second paper on stable distributions, DuMouchel (1975) showed how to approximate the Fisher information matrix about the stable distribution parameters in both the symmetric and the asymmetric cases. This can be used to construct confidence intervals for the maximum likelihood estimates (MLEs). He also provided some examples where the efficiency of estimators, such as those proposed by Fama and Roll (1968, 1971), is assessed: their estimation of γ is very efficient, whereas the estimators for δ and β do not perform as well, although they can be used as initial values in an iterative estimating strategy.

Chambers *et al.* (1976) showed how to simulate stable random variables. Buckle (1995) estimated the parameters in stable distributions by using Bayesian arguments and Monte Carlo simulation.

Stable distributions were also considered in econometry: see for example Liu and Brorsen (1995) (and the references therein) where a generalized autoregressive conditional heteroscedastic process with residuals having a conditional stable distribution was proposed.

Although this review has not been comprehensive, it clearly shows that the likelihood has not often been considered for estimating stable distribution parameters. In the next section, we show how the Fourier inversion formula can be used to compute a stable density, making likelihood inference possible.

3.2. *An alternative approach*

Before giving the technical details of our approach, let us consider an alternative, but equivalent, form for the characteristic function of stable distributions, as used by Buckle (1995):

$$\log\{\phi(t)\} = \mathrm{i}\gamma t - |t|^{\alpha}\delta'^{\alpha}\ \exp\left\{-\mathrm{i}\beta'\frac{\pi}{2}\eta_\alpha\,\mathrm{sgn}(t)\right\}, \tag{3}$$

$$\eta_\alpha = \min(\alpha, 2-\alpha) = 1 - |1-\alpha|.$$

Although the parameters play the same role, they are slightly different in this parameterization. They are related to the parameters in equation (1) by

$$\beta' = \frac{2}{\pi\eta_\alpha}\cos^{-1}\left\{\frac{\cos(\pi\alpha/2)}{\Delta}\right\},$$

$$\delta' = \left\{\frac{\Delta\delta}{\cos(\pi\alpha/2)}\right\}^{1/\alpha}$$

where

$$\Delta^2 = \cos^2(\pi\alpha/2) + \beta^2\sin^2(\pi\alpha/2),$$

$$\mathrm{sgn}(\Delta) = \mathrm{sgn}(1-\alpha),$$

$$\mathrm{sgn}(\beta') = \mathrm{sgn}(\beta).$$

The location and tail parameters, γ and α, are unchanged (see also Leitch and Paulson (1975)).

The density corresponding to this characteristic function can be computed from it by using Fourier inversion:

$$f_\alpha(y|\gamma, \beta', \delta') = \frac{1}{2\pi}\mathcal{F}_y^-\mathcal{F}_t^+ f = \frac{1}{2\pi}\mathcal{F}_y^-\ \phi(t).$$

An extended version of this expression is

$$f_\alpha(y|\gamma, \delta', \beta') = \frac{1}{\pi}\int_0^\infty \cos\left\{(\gamma - y)\frac{s}{\delta'} + s^\alpha\sin(\eta'_{\alpha,\beta'})\right\}\ \exp\{-s^\alpha\cos(\eta'_{\alpha,\beta'})\}\ \frac{\mathrm{d}s}{\delta'} \tag{4}$$

where

$$\eta'_{\alpha,\beta'} = \beta' \frac{\pi}{2} \eta_\alpha.$$

The last integral can be evaluated numerically by using for example Romberg integration which allows the prespecification of the tolerated error. In this way, we can approximate the density (and hence the likelihood) with the desired precision for fixed values of the stable distribution parameters.

This can then be optimized, yielding the MLEs for γ, δ', β' and α for a given sample. We performed this estimation of the parameters using a non-linear optimizer. Our first results were produced using procedure `OPTMUM` in GAUSS (Aptech Systems, 1992). More recently, we have developed an R (a free S-PLUS clone) package, based on the Dennis and Schnabel (1983) non-linear optimization algorithm, and allowing the interactive specification of non-linear regression models (using the Wilkinson and Rogers (1973) notation when linear) for the four parameters of the stable distribution.

Specifying reasonable starting values for the parameters is important both to ensure convergence and to allow a quick evaluation of the likelihood in the optimization procedure. Indeed, the optimizer that we use evaluates the gradient of the function numerically by taking small differences from each argument successively. The time required to evaluate the integral in equation (4) increases with the oscillation of the integrand. This oscillation is particularly important when $(y - \gamma)/\delta'$ is very large, i.e. when the model under consideration assigns a very small probability to the data. Therefore, unreasonable starting values for the parameters significantly increase the time required to compute and maximize the likelihood with the tolerated error as the optimizer tries to locate the parameter region where the MLEs are more likely to be found. The choice of these values could be made by using the techniques reviewed in Section 3.1.

The availability of the likelihood is particularly important to construct intervals of precision for the model parameters as symmetry is not the rule in practice; standard errors are not appropriate for constructing such intervals.

4. Generalized regression models

If we can compute the likelihood, nothing prevents us from modelling the location parameter in terms of covariates. If we denote by $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ the covariates associated with the observations $\{y_1, \ldots, y_n\}$ and by $g(\cdot)$ a link function, transforming the location parameter and taking values on $\mathbb{R}$, then a possible regression model would be

$$g(\gamma) = \begin{pmatrix} g(\gamma_1) \\ \vdots \\ g(\gamma_n) \end{pmatrix} = \mathbf{X}^{\mathrm{T}} \boldsymbol{\psi}^1 \tag{5}$$

where $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^{\mathrm{T}}$ and $\boldsymbol{\psi}^1$ respectively denote the design matrix and the associated regression parameter vector. The other three parameters in the stable distribution can either be fixed or simultaneously modelled if this is found necessary, as discussed below. In the special case $\alpha = 2$, we recover the traditional normal regression model. The regression in equation (5) can also be extended to models that are non-linear in $\boldsymbol{\psi}^1$.

Modelling γ may cause problems of interpretation to those accustomed to handling quantities such as the mean and the median, although γ is a location parameter in the same sense as these two (more traditional) parameters. In our view, for many scientific questions, modelling the mode as a function of the covariates will often be more appropriate than other location parameters because it is the most probable value for the response under the given model. Unfortunately, for most distributions, we do not have a closed form expression for the mode, meaning that we cannot directly model it (unless the distribution is symmetric). However, in the stable family, for fixed values of δ', β' and α, we can easily show that $\tilde{y} - \gamma$ does not vary with γ, where $\tilde{y}$ denotes the mode of the stable distribution $S_\alpha(\gamma, \delta', \beta')$. In this way, it is only a function, say $d(\delta', \beta', \alpha)$, of the other parameters. Therefore, the (generalized) regression model in equation (5) implicitly defines the regression model

$$\tilde{y}_i = g^{-1}(\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\psi}^1) + d(\delta', \beta', \alpha)$$

for the mode. In the special case where $g(\cdot)$ is the identity function,

$$\tilde{y}_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\psi}^1 + d(\delta', \beta', \alpha) \tag{6}$$

meaning that the γ and the mode regressions only differ by the intercept.

But we must be aware that the mode is also influenced by the skewness of the distribution. Therefore, a regression model for the mode would certainly be more delicate to interpret than a model for the location parameter γ. Indeed, this last quantity only indicates where the distribution is located: it does not inform us about the shape or the skewness of the distribution. The total independence between the four parameters of a stable distribution allows great flexibility when modelling data: this feature is particularly striking compared with exponential family distributions where the dispersion is a function of the mean.

We can also allow the scale, the skewness and the characteristic exponent parameters to vary with covariates, using appropriate link functions to guarantee that the constraints on them are fulfilled:

$$\log(\delta_i') = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\psi}^2,$$

$$\log\left(\frac{1+\beta_i'}{1-\beta_i'}\right) = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\psi}^3,$$

$$\log\left(\frac{\alpha_i}{2-\alpha_i}\right) = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\psi}^4.$$

The need to include a given covariate in one of the regression equations can be assessed by comparing the Akaike information criterion (AIC; see Akaike (1973)) of the resulting model with the AIC of the model ignoring this explanatory variable, with smaller values being preferable.

5. Analysis of the Abbey National share returns

We shall now analyse the Abbey National share data given in Table 1 and plotted in Fig. 1 by using generalized regression models. Econometricians often analyse log-returns instead of share prices with the argument that 'daily log-returns

$$\log(y_t/y_{t-1})$$

constitute a stationary process' (Embrechts *et al.* (1997), p. 403). Using Taylor's formula

$$\log\left(\frac{y_t}{y_{t-1}}\right) = \log\left(1 + \frac{y_t - y_{t-1}}{y_{t-1}}\right) \approx \frac{y_t - y_{t-1}}{y_{t-1}} = \rho_t$$

we can see that it is equivalent to relative returns.

Consider first (as suggested by Buckle (1995)) a stationary model for these relative returns (also shown in Table 1). (The definition of relative return given by Buckle (1995), in his section 3.2, is the negative of this, in disagreement with his histogram in Fig. 6, p. 611.) This model makes the assumption that the distribution of relative returns is not evolving with time. Moreover, these are assumed to be independent. Using the modelling strategy presented in Section 3.2, we find the MLEs $\hat{\gamma} = 0.00175$, $\hat{\delta}' = 0.0079$, $\hat{\beta}' = -0.822$ and $\hat{\alpha} = 1.53$. With the data set used by Buckle, these estimates become $\hat{\gamma} = 0.00173$, $\hat{\delta}' = 0.0078$, $\hat{\beta}' = -0.809$ and $\hat{\alpha} = 1.52$. Not surprisingly, they are very close to the previous estimates as the artificial closing return for the bank-holiday is not an extreme.

Under the same hypotheses and using Bayesian procedures with his data set, Buckle (1995) found the parameter estimates, as the marginal means of the posterior distributions, to be $\tilde{\gamma} = 0.00053$, $\tilde{\delta}' = 0.0079$, $\tilde{\beta}' = -0.55$ and $\tilde{\alpha} = 1.61$, whereas the marginal modes that he provided are $\tilde{\alpha} = 1.65$ and $\tilde{\beta}' = -0.80$. Except for the location parameter for which we do not have the marginal mode, these estimates are similar to the MLEs. Using the criteria proposed by Paulson *et al.* (1975), we find $\tilde{\gamma} = 0.00183$, $\tilde{\delta}' = 0.0087$, $\tilde{\beta}' = -0.99$ and $\tilde{\alpha} = 1.52$, again similar to the MLEs, except perhaps for the skewness parameter β'. Note that the estimates obtained by using the approach proposed by Press (1972) are, with our example, very sensitive to the choice of the arbitrary values t_1 and t_2 (mentioned above) at which the empirical characteristic function is set equal to its theoretical counterpart.

To obtain estimates of the precision of the maximum likelihood parameters, we can look at the normed profile likelihoods (see for example Lindsey (1996), p. 111), as plotted in Fig. 2. For example, the normed profile likelihood for β' in Fig. 2(b) is simply

$$\mathbf{R}(\beta') = \max_{(\gamma, \delta', \alpha) \in \mathbb{R} \times \mathbb{R}_0^+ \times]0,2]} \left\{ \prod_t \frac{f_\alpha(\rho_t|\gamma, \delta', \beta')}{f_{\hat{\alpha}}(\rho_t|\hat{\gamma}, \hat{\delta}', \hat{\beta}')} \right\}.$$

A likelihood interval for β' can then be obtained by considering the values of β' for which $\mathbf{R}(\beta')$ is larger than some percentage. For example, the 14.6% ($= \exp\{-\chi_1^2(0.95)/2\}$) likelihood interval corresponds to the traditional 95% confidence interval.

Because of the skewness of these (normed) profile likelihoods, we conclude that the use of standard errors as a measure of precision would be misleading.

The graph for α indicates that the normal ($\alpha = 2$) and the Cauchy ($\alpha = 1$) distributions are unlikely distributions for the data, whereas the graph for β' informs us that there is little support for symmetry ($\beta' = 0$).

The normed profile likelihood graphs obtained with the unmodified data set are nearly identical with the profiles displayed in Fig. 2. The normed profile likelihood for γ, plotted in Fig. 2(c), shows that Buckle's (1995) estimate for γ is almost as likely as our MLE. This is confirmed by the first two rows of Table 2, where the AIC values of these two models are virtually identical. (For comparability with other types of models, the likelihood function includes the Jacobian for the transformation from prices to relative returns.)

Similarly, the normed profile likelihood for β' in Fig. 2(b) indicates that the estimate of Paulson *et al.* (1975) for the skewness parameter is reasonable. This is again confirmed by the AIC values of (the corresponding) models (a) and (c) in Table 2.

When comparing the normed profile likelihood plots with the marginal posterior densities

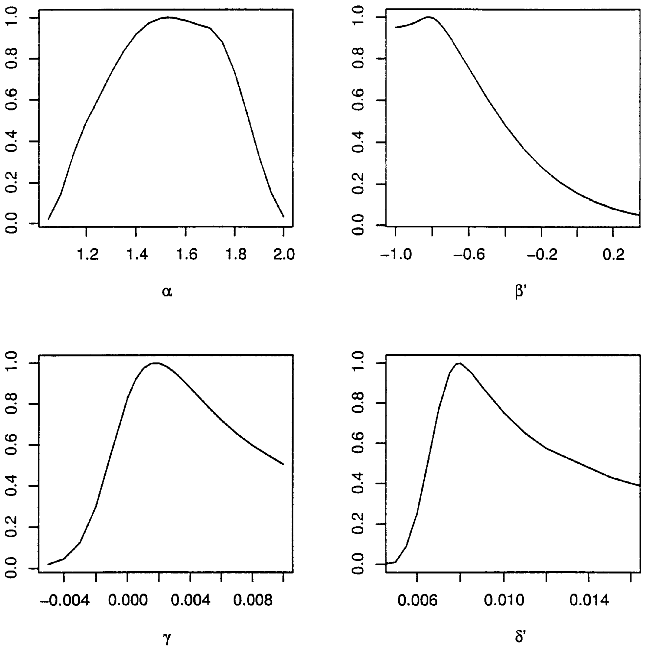

Fig. 2. Normed profile likelihood for the parameters in the stationary model

Table 2. Likelihood ($-\log(L)$) and AIC table for the Abbey National shares relative returns modelled by using stable distributions†

Model	$-log(L)$	*Number of parameters*	*AIC*
(a) Stationarity, our estimates	130.3	4	134.3
(b) Stationarity, Buckle's (1995) estimates	130.3	4	134.3
(c) Stationarity, Paulson *et al.* (1975) estimates	131.0	4	135.0
Varying location			
(d) $\gamma_t = \psi_0^1 + \psi_1^1 t$	129.4	5	**134.4**
(e) $\gamma_t = \psi_0^1 + \psi_1^1 t + \psi_2^1 t^2$	129.4	6	135.4
(f) $\gamma_t = \psi_0^1 + \psi_1^1 t + \psi_2^1 \rho_{t-1}$	129.2	6	135.2
(g) $\gamma_t = \psi_0^1 + \psi_1^1 t$ with $\beta = 0$	131.3	4	135.3
Varying shapes			
(h) $\log(\delta_t') = \psi_0^2 + \psi_1^2 t$	130.0	5	135.0
(i) $\log\left(\frac{1+\beta_t'}{1-\beta_t'}\right) = \psi_0^3 + \psi_1^3 t$	126.3	5	**131.3**
(j) $\log\left(\frac{\alpha_t}{2-\alpha_t}\right) = \psi_0^4 + \psi_1^4 t$	127.4	5	132.4
Normal distribution			
(k) Stationarity	133.7	2	**135.7**
(l) $\gamma_t = \psi_0^1 + \psi_1^1 t$	133.7	3	136.7

†The AIC values of the models selected are given in bold.

given by Buckle (1995), we see that the graphs for α and β are similar. The other two are considerably more skewed, pointing to larger values being plausible. Note that our graphs are exact, requiring no approximations either from Monte Carlo sampling or from imposing an arbitrary prior distribution. In addition, they are parameterization invariant so inferences can be drawn from these graphs about any desired transformation of the parameters. However, as for marginal posterior densities, they do summarize a four-dimensional surface so care must be taken in their interpretation. Fortunately, the parameter estimates are not highly correlated. (Asymptotic correlations among the estimates range from −0.04 between α and β' to 0.74 between γ and δ', all transformed as in the regression equations above.)

We have also considered non-stationary models as shown by models (d) and (e) in Table 2 where the location parameter is respectively modelled as a linear and a quadratic function of time. The first of these models fits about as well but a quadratic in time is not necessary. The parameter estimates are $\hat{\delta}' = 0.00804$, $\hat{\beta}' = -0.782$ and $\hat{\alpha} = 1.31$, and $\hat{\psi}_0^1 = 0.00732$ and $\hat{\psi}_1^1 = -0.000169$ for the regression parameters, the last showing that the share relative returns may be decreasing with time.

Associated with each value of t, we have the fitted stable distribution $S_{\hat{\alpha}}(\hat{\gamma}_t, \hat{\delta}', \hat{\beta}')$ describing the model structure (where $\hat{\gamma}_t = 0.00732 - 0.000169t$). The mode $\tilde{y}_t$, as the most probable value generated by this distribution, is an attractive quantity to aid in understanding the meaning of the regression model for γ_t. As we explained in Section 4, the linear regression in time for γ_t induces a linear regression model for the mode. Using equation (6), we see that $d(\delta', \beta', \alpha)$ can be determined by subtracting $\hat{\psi}_0^1$ from the mode of the fitted stable distribution $S_{\hat{\alpha}}(\hat{\psi}_0^1, \hat{\delta}', \hat{\beta}')$ at time $t = 0$. This yields the fitted linear regression model

$$\tilde{y}_t = -0.00033 - 0.000169t$$

for the mode, plotted as the full line in Fig. 1. We have also determined the interval within which the height of the density is at least 10% of its maximum (i.e. its value at the mode) for the fitted stable distribution at each time point: the extremities of these intervals are plotted as dotted lines in Fig. 1, so we have the 1.0 and 0.1 contours of (relative) density. For comparison, the regression for the mean under the normality hypothesis (model (l) of Table 2) is represented by the broken line.

The position of the borders of the 10% density intervals with respect to the mode regression line confirms that, under model (d), the estimated distribution is right skewed ($\hat{\beta}' = -0.782$) throughout the time period. Note also that the width of the interval does not vary with γ_t because δ', β' and α are assumed constant (and the link is the identity link). This can be related to the constant variance assumption in normal regression models.

No serial association is detectable in these data, as shown by the AIC for model (f) in Table 2, where conditioning on the immediately preceding share relative return ρ_{t-1} is considered in addition to the linear model in time, giving an autoregressive AR(1) process. This is not too surprising because the transformation from prices to relative returns is a form of first differencing.

From models (k) and (l) in Table 2 and the normed profile likelihood for α in Fig. 2, we can see that normality is not a reasonable hypothesis. This was already clear from Fig. 6 in Buckle (1995), p. 611, although no formal comparison of these models was proposed there. Note also that a symmetric stable model is not acceptable, as shown by the AIC of model (g) and the normed profile likelihood for β' in Fig. 2. This confirms the empirical conclusions drawn by Fielitz and Smith (1972) and Leitch and Paulson (1975) in their studies of stock price changes.

Finally, for regression involving the location parameter, we observe that the fitted normal regression line is closer to the horizontal than that for the mode under the stability hypothesis. This is due to the influence of the extreme share price relative returns, as can be seen from Fig. 1. The normal model must correct the slope of the regression line because the tails of the normal distribution do not allow 'extreme' behaviour. In contrast, the stable regression line goes through the main body of the data, nearly insensitively to the extreme observations that can be encompassed by the heavy tails of a (non-symmetric) stable distribution.

The regression models considered so far are fairly classical, except for the use of the stable family and the mode. Let us now study changes in the other parameters of the stable distribution over time. From models (i) and (j) in Table 2, we see that a regression involving either the skewness or the characteristic exponent parameter provides a better fit than any considered so far, with the former giving the best fit. We have checked that no combination of several regression equations in the same model improves the fit.

Thus, we can conclude that it is not the location, nor the scale, but the skewness of these data that is changing over time. In model (i), the skewness parameter is estimated to be changing from about 0.9 at the beginning of the series to -1 at the end. However, because the location parameter γ of the distribution is constant over time, the mode must be moving to allow the skewness to change. The estimate $\hat{\alpha} = 1.37$ indicates that the distribution has heavier tails than does the stationary model but slightly lighter than that for regression with the location parameter given above. As can be seen in Fig. 3, the mode of the distribution follows the central mass of observations remarkably well, whereas the skewness allows for the extreme values, which are negative near the beginning of the series and positive near the end.

The normed profile likelihoods for the five parameters in model (i) are plotted in Fig. 4. All are still skewed, but we see that they have changed considerably from those for the simpler model in Fig. 2. Although the final model has more parameters than the stationary model, we

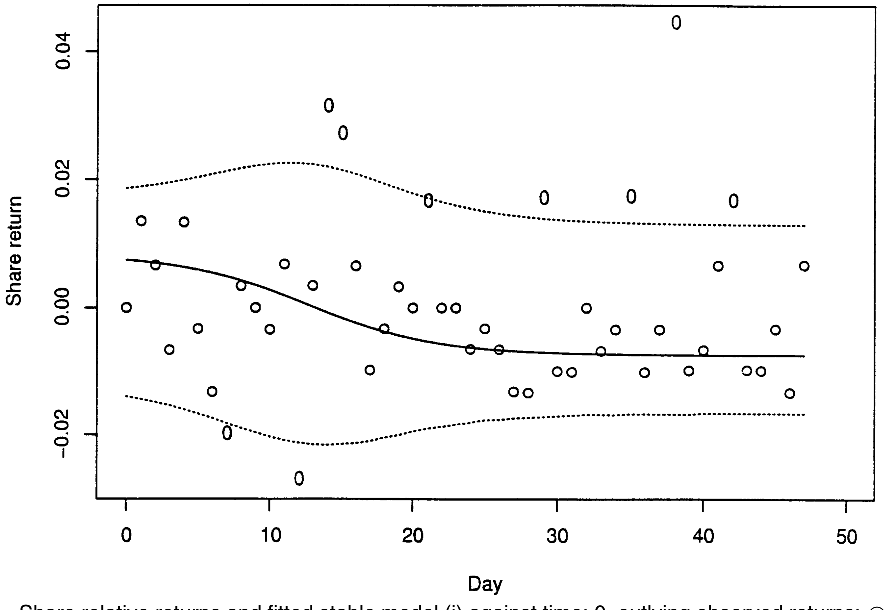

Fig. 3. Share relative returns and fitted stable model (i) against time: 0, outlying observed returns; ○, observed returns; ———, fitted mode; ··········, 10% relative density contours

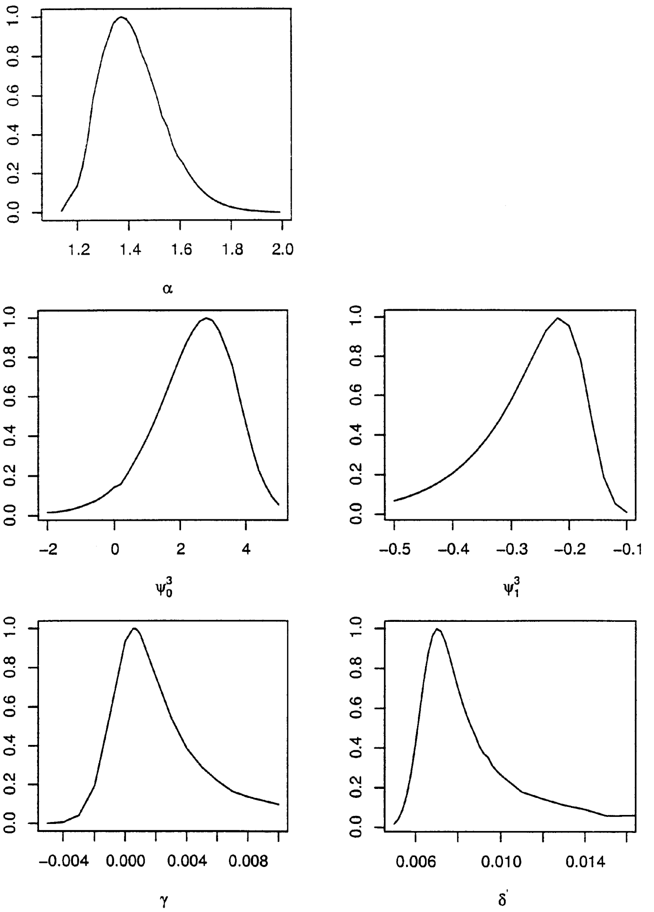

Fig. 4. Normed profile likelihood for the parameters in model (i) with varying skewness

have gained precision in the estimation of the other aspects of the distribution (as can be seen from the spread of the profile likelihoods). This clearly indicates that the final model is more reasonable. The normal ($\alpha = 2$) and the Cauchy ($\alpha = 1$) distributions remain unlikely candidates. There is strong evidence that the distribution tends to become right skewed with time (as positive values for β' are extremely unlikely). Finally note that allowing the skewness to change with time has reduced the dispersion of the fitted distributions since there is less evidence for 'large' values of δ'.

The data set just analysed is a very short time series. No substantive conclusions about share relative returns should be drawn from the model selected. With a longer series, it is quite likely that the skewness might, for example, be oscillating back and forth. But, a clear conclusion that we can draw is that, during the time period observed, the major mass of probability stayed in the same location, although fundamentally changing shape. In studying such data, among other things, we were interested in how the region in which most relative returns would lie was changing over time, and in how the most probable relative return varied over time. Our model has permitted us to answer both of these questions, as can be seen in Fig. 3. In contrast, a mean, with its standard deviation, could not have correctly answered either of them.

6. Discussion

We have shown that, with modern computing power, the family of stable distributions, which is only known through its characteristic function, can be used, in a likelihood approach, to model continuous responses. The density at any point is determined numerically by applying a Fourier transform to the characteristic function defining the stable distribution. This allows us to maximize likelihoods, and to plot profile likelihoods, for regression models based on the stable family at a speed that makes real-time interactive modelling feasible, even on a low speed personal computer.

The availability of an efficient non-linear optimizer, such as procedure `OPTMUM` in GAUSS or the Dennis and Schnabel (1983) algorithm available in R (Ihaka and Gentleman, 1996), enables us to compute the MLEs. The AIC can then be used to compare the relative merits of (possibly non-nested) models.

We have defined a generalized regression model for the location parameter γ and shown that it induces (a closely related) regression model for the mode. This last quantity, as the most likely value of the fitted process, allows a simple interpretation of the models so constructed. We, then, generalized regression to changes in any or all of the four parameters of the stable family. Introducing regression parameters into the likelihood function does not make the estimating procedure more difficult to set up. We can also consider non-linear regression models for the various parameters without difficulty.

Finally, the capacity of stable regression models to describe heavy-tailed processes was used to analyse the evolution of the Abbey National share relative returns. The advantages of this methodology are particularly clear when 'outlying' observations, that can be encompassed by the tails of a stable distribution, have an undesirable influence on a normal distribution fit.

Because the family of generalized linear models has come to be fairly widely used, statisticians have become accustomed to regression models where the variance is not constant. However, these models impose a fixed relationship between the mean and the variance. In contrast, the stable family is particularly interesting because four different aspects of the shape of the distribution (location, scale, skewness and the thickness of the tail) can vary independently of each other. This provides a freedom of modelling that is rarely found in other families.

The results of the example have major implications for most of traditional statistical modelling. In the history of statistics, how many spuriously significant differences in means have been detected when, in fact, only the shape of the distribution (dispersion or variance, skewness or the thickness of tails) was changing? What indeed is the appropriate measure of difference in 'location' for a given problem? For example, are we interested in how the mean or the most probable response changes? All too often in statistics, means are calculated and studied without any reference to the distribution from which they are derived. But we see that a mean, or location parameter, cannot be correctly interpreted (the mean may not even exist) if the distribution from which it arises is not specified, especially if that distribution may possibly be skewed.

The R package that we have developed to produce the results in the paper is available from sites where R can be obtained, or directly from the first author.

Acknowledgements

We thank Ross Ihaka and Robert Gentleman for developing R, free software under the GNU General Public Licence. We also thank the Joint Editor for discovering the problem with the bank-holiday in the data set and both the Joint Editor and the referees for their constructive comments.

References

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In *Proc. 2nd Int. Symp. Inference Theory* (eds B. N. Petrov and F. Csàki), pp. 267–281. Budapest: Akadémia Kiadó.
Aptech Systems (1992) *The GAUSS System Version 3.0*. Maple Valley: Aptech Systems.
Buckle, D. J. (1995) Bayesian inference for stable distributions. *J. Am. Statist. Ass.*, **90**, 605–613.
Chambers, J. M., Mallows, C. L. and Stuck, B. W. (1976) A method for simulating stable random variables. *J. Am. Statist. Ass.*, **71**, 340–344.
Dennis, J. E. and Schnabel, R. B. (1983) *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. New York: Prentice Hall.
DuMouchel, W. H. (1973) Stable distributions in statistical inference: 1, Symmetric stable distributions compared to other symmetric long-tailed distributions. *J. Am. Statist. Ass.*, **68**, 469–477.
——— (1975) Stable distributions in statistical inference: 2, Information from stably distributed samples. *J. Am. Statist. Ass.*, **70**, 386–393.
Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events*. Berlin: Springer.
Fama, E. F and Roll, R. (1968) Some properties of symmetric stable distributions. *J. Am. Statist. Ass.*, **63**, 817–836.
——— (1971) Parameter estimates for symmetric stable distributions. *J. Am. Statist. Ass.*, **66**, 331–338.
Fielitz, B. D. and Smith, E. W. (1972) Asymmetric stable distributions of stock prices changes. *J. Am. Statist. Ass.*, **67**, 813–814.
Gawronski, W. (1984) On the bell-shape of stable densities. *Ann. Probab.*, **12**, 230–242.
Hoffmann-Jørgensen, J. (1994) *Probability with a View toward Statistics*, vol. I. London: Chapman and Hall.
Hougaard, P. (1986) Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, **73**, 387–396.
Ihaka, R. and Gentleman, R. (1996) R: a language for data analysis and graphics. *J. Comput. Graph. Statist.*, **5**, 299–314.
Janicki, A. and Weron, A. (1994) Can one see α-stable variables and processes? *Statist. Sci.*, **9**, 109–126.
Koutrouvelis, I. A. (1980) Regression-type estimation of the parameters of stable laws. *J. Am. Statist. Ass.*, **75**, 918–928.
Leitch, R. A. and Paulson, A. S. (1975) Estimation of stable law parameters: stock price behavior application. *J. Am. Statist. Ass.*, **70**, 690–697.
Lindsey, J. K. (1996) *Parametric Statistical Inference*. Oxford: Oxford University Press.
Liu, S. and Brorsen, B. W. (1995) Maximum likelihood estimation of a GARCH-stable model. *J. Appl. Econometr.*, **10**, 273–285.
Mandelbrot, B. B. (1963) The variation of certain speculative prices. *J. Bus.*, **36**, 394–419.
Paulson, A. S., Holcomb, E. W. and Leitch, R. A. (1975) The estimation of the parameters of the stable laws. *Biometrika*, **62**, 163–170.
Press, S. J. (1972) Estimation in univariate and multivariate stable distributions. *J. Am. Statist. Ass.*, **67**, 842–846.
Samorodnitsky, G. and Taqqu, M. S. (1994) *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. London: Chapman and Hall.
Teichmoeller, J. (1971) A note on the distribution of stock price changes. *J. Am. Statist. Ass.*, **66**, 282–284.
Wilkinson, G. N. and Rogers, C. E. (1973) Symbolic description of factorial models for analysis of variance. *Appl. Statist.*, **22**, 392–399.
Yamazato, M. (1978) Unimodality of infinitely divisible distribution functions of class *l*. *Ann. Probab.*, **6**, 523–531.