JUUSO TÖYLI, KIMMO KASKI and ANTTI KANTO

Variability of Hex Return Distribution

ABSTRACT

This paper discusses the shape of the stock return distribution using the all shares index of Helsinki Stock Exchange. Non-parametric kernel density estimation and power exponential family of distributions are used to model the shape of the return distribution. The parameters of power exponential distribution are estimated with Bayesian approach. The findings suggest that the shape of the distribution does not vary from one weekday to another. However, substantial deviations over time are observed while there are also temporal periods when it is reasonable to assume the return generating process as normal. Furthermore, the results indicate that the return distribution approaches normal when the time interval used to calculate returns is increased.

Keywords: *Finance, Bayesian inference, power exponential distribution.*

1. INTRODUCTION

Volatility clustering, which results partly from the time-dependency of variance, has been widely studied and reported. This paper shows that the shape of the unconditional distribution for the

JUUSO TÖYLI, Laboratory of Computational Engineering
Helsinki University of Technology • e-mail: juuso.toyli@hut.fi
KIMMO KASKI, Laboratory of Computational Engineering
Helsinki University of Technology • e-mail: kimmo.kaski@hut.fi
ANTTI KANTO, Department of Economics and Management Science
Helsinki School of Economics • e-mail: kanto@hkkk.fi

stock market index returns varies over time. On the other hand, the shape of the distribution seems *not* to vary from one weekday to another as it has been reported in case of the variance and the mean. This paper also concludes that time periods, when the distribution is roughly normal, do exist, but substantial deviations from normality are also observed. Thus, part of the contradiction in the previous results concerning the shape of the distribution might be explained by the variability of the shape of the distribution over time.

The shape of the asset return distribution is crucial for asset pricing models, mean-variance portfolio theory, and risk measurement. The behaviour of the variance has a significant effect on the prices of contingent claims. The classical theory of finance is mainly based on the assumed normality having its roots on Bachelier's original work at the beginning of last century (see Bachelier 1900). Thus, it is common to assume that returns behave like random walkers, which implies that the returns are identically and independently distributed showing a zero expectation value and constant variance. The normality assumption is usually then added and the process is defined to be Brownian motion (see Osborne 1959). However, it was Mandelbrot (1963) and Fama (1965) who first reported the fundamental deviation from the normality: price distributions are leptokurtic and have longer tails than normal distributions. Many others have thereafter observed similar deviations. Few studies also indicate that the return distributions are skewed (Kon 1984, Fielitz–Rozell 1983), but this skewness is not generally regarded as a problem. When the returns of individual stocks are considered, bimodality, i.e., the distribution has two peaks, is also occasionally observed (see Kanto et al. 1998). In addition, the distributional properties of Finnish stock returns have earlier been explored by Booth et al. (1992).

Previous research has suggested that the standard deviation of Monday's returns is higher than that of other days (Fama 1965). A statistically significant difference in the mean return of Mondays compared to that of other days has been found (French 1980; Gibbons–Hess 1981) and non-stationary mean-excess returns have been reported in the first trading week of January (Keim 1983). In addition, the turn of the month effect – i.e., the mean return for stocks is positive only for days immediately prior to and during the first half of calendar months and indistinguishable from zero for days during the last half of the month – has been found out (Ariel 1987). Asset returns are not usually found to be auto-correlated like the squared returns, variances, and volumes are (Schwert 1989). These findings suggest that the return generating process might be *non-linear* and that returns might be *dependent*. This possible dependency does not necessarily imply market inefficiency since it is enough for efficiency that the return-generating process can be represented as a martingale (Fama 1970).

The question whether the variance converges to a finite limiting value is related to the time dependency. The stable distributions lead to models that assume infinite variance. However, it has been indicated that the variance of asset returns is finite but time dependent in a complex non-linear manner (Perry 1983; see also Tucker 1992). This time dependency is able to explain the volatility clustering observed in financial time series and the leptokurtic unconditional return distributions. The models assuming time dependent behaviour can be divided in three categories: autoregressive heteroscedastic (ARCH) models (Engle 1982, Bollerslev 1986, 1987), stochastic volatility (SV) models (see Taylor 1985, Taylor 1994), and models based on chaos theory leading to complex dynamics.

There are three distinct approaches to the modelling of unconditional distribution of asset returns. *First*, by modelling the assumed return generating stochastic process that results in a distribution of returns. (Epps–Epps 1976; Oldfield et al. 1983; Tauchen–Pitts 1983; Osborne 1959; Akgiray–Bouth 1987; Kon 1984) *Second*, by seeking a distribution that empirically fits into the observed data. (Mandelbrot 1963; Fama 1965; Blattberg–Gonedes 1974, Hagerman 1978) *Third*, by modelling the behaviour of individual agents who act on the market (Arthur et al. 1997).

The modelling problem can be divided in two parts by asking whether the parameters of the distribution are stationary and what the functional shape of the distribution is. The past research concentrated initially on the shape of the distribution and only recently on the time dependency of the parameters, especially of the variance. The results are contradictory and, although deviations from normality are commonly accepted, there is no commonly accepted model. This might be partly related to the possible variability in the shape of the distribution over time, which has mainly been ignored in the past research and needs to be studied further.

The purpose of this paper is to explore the possible variability in the shape of the unconditional distribution. This problem is studied with the help of the following questions: does the shape of distribution vary from on weekday to another, is the shape of distribution constant over time, and does the shape of the distribution vary along the time horizon used to calculate the returns? The nature of this paper is descriptive and it seeks *not* to give any substance or modelling based *explanation* of the possible time variation of the shape of return distribution.

The results indicate that there are significant variations in the shape of the distribution over time while periods of normality are also observed. This finding suggests that part of the contradiction in previous results concerning the shape of the distribution might be related to the different time-periods used in the analyses. The results of this study imply that the reported time-dependency of variance might not be enough to completely explain the observed distributions. In contrast, it seems that the shape of the distribution might also be time-dependent. However, differences in the shape of the distribution between weekdays are not observed, but the tendency towards normality, when the time interval is increased, is observed. Monthly returns seem to be roughly normally distributed.

This paper is constructed so that first the data is described and then the models used here are reviewed. These include the non-parametric kernel density estimate and parametric power exponential family of distributions. The estimation and a few computational facts are discussed in connection to that. Then, the results are reported and, finally, the paper is shortly summarised.

2. DATA

The Helsinki Stock Exchange all shares daily return index (HEX) from 2.1.1991 to 30.12.1997 was used in this study. The sample contained 1751 data points. The HEX all shares index includes all stocks listed on the Helsinki Stock Exchange and the shares are weighted with their market capitalisation. The base date of the index is 28 December 1990 and the base number is 1000. In the international scale, Helsinki Stock Exchange is small and rather volatile stock exchange. The full data set, data split by weekdays, data split by years, and 251 data points' rolling windows were used in daily analyses. The window was rolled over the data by forwarding it every time by one data point and its length (251 data points) was set to reflect the typical length of a trading year. Since previous research has generally concluded that monthly returns are approximately normally distributed, the evolution of the shape of return distribution was studied in different holding periods from one day to thirty days. The logarithm of daily HEX all shares return index is shown in Figure 1.

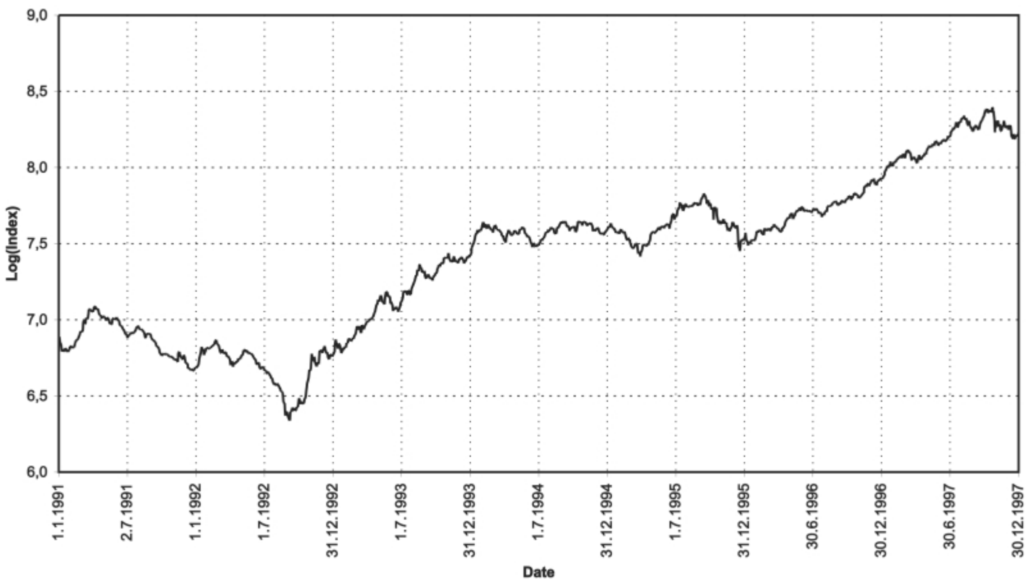

FIGURE 1: Logarithm of hex daily return index from 2.1.1991 to 30.12.1997.

In the analyses, logarithmic returns were used. The logarithmic returns are generally favoured in financial studies of this kind for three reasons. First, the change in logarithmic price is the yield, with continuous compounding, from holding a security for the period in questions. Second, variability of a price change is an increasing function of the price level, but the use of logarithms neutralises most of this effect. Third, for small changes the change in logarithmic price is very close to the percentage price change.

The daily returns indicated modest first autocorrelation (0.1835) but this autocorrelation vanished when longer lags are being considered. This observation is in line with the typical finding that financial data shows first autocorrelation that is lower than 0.2 while other autocorrelations are considered as zero. However, this might result in increased probability mass on the tails of the observed return distribution, which in turn is likely to lead to biased inference about the shape of the distribution and its variability over time. The possible effect of the autocorrelation was studied by transforming the data with Cochrane-Orcutt correction defined as follows: $y_t = x_t - rx_{t-1}$ where r is the estimated first autocorrelation, x_t the original return at time t, and y_t the corrected series. This correction should remove the first order autocorrelation. The estimate r = 0.1835 was obtained from the full data set and the correction was applied to the full data. All analyses were made to the corrected as well as to the raw data. Since it turned out that the estimated first autocorrelation varied substantially in different rolling windows, the Cochrane-Orcutt correction was also applied in rolling manner. However, the inference remains the same whether the rolling correction or correction to full data set is used. It needs to be pointed out that Cochrane-Orcutt correction does not remove the possible higher order dependencies form the data. Such dependencies could, at least partially, explain the observed non-normality of the asset return distribution. This issue needs to be studied further before conclusive comments can be presented.

Descriptive statistics from Cochrane-Orcutt corrected and raw data are given in Table 1. The data shows only minor differences. The Cochrane-Orcutt correction seems to remove skewness and to increase kurtosis when all returns are analysed, but, in case of returns grouped by weekday, this correction seems to have no systematic impact.

TABLE 1: Descriptive statistics from the data.

	N	Mean		Std.dev.	Skewness		Kurtosis	
Data*	**Statistic**	**Statistic**	**Std.Error**	**Statistic**	**Statistic**	**Std.Error**	**Statistic**	**Std.Error**
ALL RETURNS								
Raw	1750	0.0767	0.0300	1.2537	–0.1753	0.0585	3.8244	0.1169
Corrected	1749	0.0636	0.0295	1.2323	–0.0048	0.0585	4.3668	0.1170
MONDAY								
Raw	351	–0.0269	0.0645	1.2080	–0.2751	0.1302	1.6280	0.2597
Corrected	351	–0.0455	0.0635	1.1892	–0.2625	0.1302	1.8196	0.2597
TUESDAY								
Raw	359	0.0525	0.0691	1.3092	–0.4979	0.1287	5.9339	0.2568
Corrected	359	0.0586	0.0671	1.2717	–0.2502	0.1287	4.5125	0.2568
WEDNESDAY								
Raw	355	0.1161	0.0703	1.3253	0.7007	0.1295	3.3801	0.2582
Corrected	355	0.1088	0.0694	1.3084	1.1815	0.1295	6.6353	0.2582
THURSDAY								
Raw	345	0.1363	0.0637	1.1833	–0.3841	0.1313	1.9454	0.2619
Corrected	344	0.1164	0.0642	1.1903	–0.6347	0.1315	2.8862	0.2622
FRIDAY								
Raw	340	0.1077	0.0669	1.2327	–0.6291	0.1323	5.3054	0.2638
Corrected	340	0.0810	0.0647	1.1934	–0.4513	0.1323	4.8105	0.2638

* Returns were multiplied by 100 for representational clarity

3. METHODS

3.1. Models

In order to model the temporal variation of the shape of stock return distributions, we need a model that is flexible enough in capturing various shapes possibly appearing in the distributions. Since the traditional model is normal distribution, the model family should include this as a special case. Furthermore, the model should be able to capture the observed excess kurtosis and long tails in empirical security return distributions. The remaining issues to be considered are: possible bimodality, skewness, and finiteness or infiniteness of variance. These will be discussed next.

Bimodality is seldom observed in analyses of individual stocks and has not been reported in analyses that concentrate on stock indices or on stock portfolios (see Kanto et al. 1998). Since the preliminary tests with non-parametric kernel estimate showed no evidence of bimodality, the unimodality can be accepted without a major threat to the validity.

The possible skewness is a more complicated problem since it has been reported (Kon 1984, Fielitz–Rozell 1983). As a matter of fact the skewness might substantially affect, e.g., the measurement of the downside risk when value-at-risk methodology is used. (See Simons 1996 for summary of value-at-risk). Thus, the model should at least allow the addition of skewness parameter at a later stage of the study if needed.

An assumption that is even more crucial in developing a model is the distinction between models having finite and infinite second moment. This issue has been intensively discussed since Mandelbrot's (1963) and Fama's (1965) seminal works that suggested a stable distribution, which does not have a finite second moment, as a model of security returns. The later research has provided substantial evidence that does *not* support the stable paretian hypothesis. (Blattberg–Gonedes 1974, Hagerman 1978, Officer 1972, Hsu et al. 1974, Perry 1983, Tucker 1992). In contrast, Mantegna and Stanley have recently reached the opposite result again, but this result may partially be due to the use of differences instead of logarithmic returns and perhaps more specially it may be due to the use of high frequency transaction level data in the analysis (Mantegna–Stanley 1995, see Kullman et al. 1999). In general, it is usually accepted that the variance is finite although time-dependent in complex manner when daily data is being analysed (Perry 1983, Tucker 1992, see also Fama–French 1988). Based on these observations, it seems reasonable to assume that the second moment is finite and that the model needs to be built accordingly. This is also a convenient assumption from the theory development point of view since it allows the use of techniques, which take advantage of the second moment. This is crucial, for example, in option pricing models.

Based on the above discussion, two models were selected in this study. Preliminary tests were established with the help of non-parametric kernel density estimate and the main inference was done based on the power exponential distribution (PED) family. No asymmetry parameter was included in PED since the preliminary tests and statistics suggested that the sample involved no substantial skewness. However, it is possible to introduce a skewness parameter, if needed in the future, although this will substantially increase the computer time needed for solving the model. In addition, the PED family does not have the favourable property of stability under addition but its use guarantees that all the moments exist. There is also a multivariate generalisation available (see Gómez et al. 1998).

Neither of the applied models takes into account the possible time dependencies in the data. This should not be a problem since our study applies an approach where the data is prewhitened. However, more research is needed before conclusive comments can be given about this. Next we discuss the models.

Kernel density estimate is a non-parametric alternative to the parametric approaches of density estimation. The kernel estimator with kernel K is defined by:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{3.1}$$

where X_i is the stock return on day *i,* the parameter *n* is the number of points, and *h* is the smoothing parameter. This kernel estimator can be considered to be a sum of bumps placed at the observations where the kernel function *K* defines the shape of the bumps and the smoothing parameter *h* defines their width. If the kernel function *K* is a probability density function (pdf), it follows from the definition that $\hat{f}$ is also a pdf (Silverman 1986).

Here we have applied the Gaussian kernel because the inference is not very sensitive to the selection of kernel function (Silverman 1986, 43). It can be shown that under Gaussian kernel the minimum approximate mean integrated square error is achieved when the smoothing parameter is as follows:

(3.2) $$h_{opt} = \left(\frac{4}{3}\right)^{\frac{1}{5}} \sigma n^{-\frac{1}{5}}$$

This works well when the underlying population is normally distributed but it over-smooths if the population is multimodal. Consequently, a more robust estimator can be achieved by using an adaptive estimate of the spread of distribution instead of σ in Eq. (3.2). The adaptive estimate is defined by:

(3.3) A = min(standard deviation, interquartile range/1.34)

In the case of normal distribution, the two arguments are theoretically equal. Silverman (1986, 48) has argued that the results could be improved if the factor (4/3)^(1/5) = 1.0592 in (3.2) is reduced slightly, for example to 0.9.

Power exponential family of distributions[1] (PED) is the second model we use in this study. This model is parametric and its density function is given in Eq. (3.4).

(3.4) $$p(y \mid \theta, \sigma, \beta) = \frac{\left\{\Gamma\left[\frac{3(1+\beta)}{2}\right]\right\}^{\frac{1}{2}}}{(1+\beta)\left\{\Gamma\left[\frac{(1+\beta)}{2}\right]\right\}^{\frac{3}{2}} \sigma} \exp\left[-\left\{\frac{\Gamma\left[\frac{3(1+\beta)}{2}\right]}{\Gamma\left[\frac{(1+\beta)}{2}\right]}\right\}^{\frac{1}{(1+\beta)}} \left|\frac{y-\theta}{\sigma}\right|^{\frac{2}{(1+\beta)}}\right]$$

where the parameters are $-\infty < y < \infty$, $\beta > -1$, $-\infty < \theta < \infty$, and $\sigma > 0$. The power exponential family of distributions is *normal* when $\beta = 0$, *Laplace* or *double exponential* when $\beta = 1$, and approaches *uniform* when $\beta \to -1$. In literature it seems that a common practice is to limit β as

1 With a different parameterisation, the Power Exponential Family is called generalised error distribution (GED). This model has been considered by Taylor 1994 and Hsu 1982.

follows $-1 < \beta \leq 1$ (see Welsh 1996, Box–Tiao 1973). In this study, β is not limited to value less than unity because, if it is allowed to be at the maximum unity, then the maximum excess kurtosis would be three. This is inadequate for financial data because larger kurtosis and even heavier tails than in double exponential distribution ($\beta = 1$) are often observed. Fortunately, there is no mathematical constrains to limit β to be unity, although this would simplify the integration and handling of β's priori. As a result, $\beta > -1$ is assumed in this study. This leads to a situation, in which the maximum likelihood estimates for location and scale parameters do not necessarily exist although the shape parameter is known (see Agró 1995). In Figure 2 we show a few density functions with varying values of the shape parameter β.

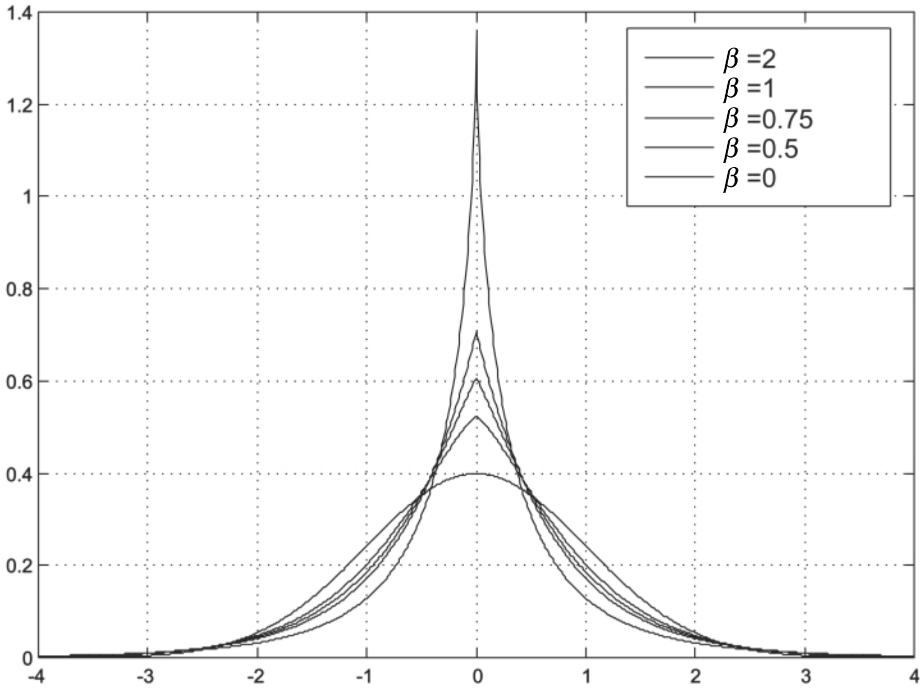

FIGURE 2: Power exponential density function for five values of β.

3.2. Estimation

The estimation of kernel density is straightforward and evident from the definition. However, the estimation of PED's parameters, especially the shape parameter β, is difficult since the maximum likelihood estimates exist for scale and location parameters only if the shape parameter is known and it is unity or less. In addition, Agrò (1995) has suggested a method to numerically minimise the log(likelihood) function, but this procedure assumes that $-1 < \beta \leq 0$. Consequently, such framework was not applicable in this study since the leptokurtic distributions are of main interest.

In order to overcome the inconvenience related with the parameter estimation, we have applied the Bayesian paradigm (Bayes 1763; Laplace 1774; 1812), which is based on the notion that after the data is collected it is known for sure. Thus, the inference is based on the conditional distributions of the parameters given the data. This conditional distribution is defined by

$$p(\theta \mid y) \propto L(\theta \mid y)\, p(\theta) \tag{3.5}$$

where $\theta' = (\theta_1,\ldots,\theta_k)$ is a vector of k parameters and $y'=(y_1,\ldots,y_n)$ is a vector of n observations. The likelihood function $L(\theta \mid y)$ plays a significant role since it is the function through which the data y modifies prior knowledge of θ included in its prior distribution $p(\theta)$. This formulation leads as such to an adaptive model that is capable of learning from experience.

One of the main problems in Bayesian approach is the selection of an adequate prior distribution for the parameters but this problem is not significant in this study since the large amount of data points used in estimation reduced the weight of the prior distribution close to the zero. However, the purpose of this study is to explore the behaviour of the shape of the distribution and a proper starting point is to construct the model so that as little as possible is assumed about the shape *a priori* and then let data to modify this view. The approach used here is based on Box and Tiao's (1973) work and the modifications to their framework were minor. The assumptions behind the resulting model are discussed next.

The selection of the prior distribution for θ and σ is based on Jeffreys' rule resulting in non-informative priors. This use of Jeffreys' prior leads to identical results independent of parameterisation. The location parameter θ is assumed to be independent of σ and β *a priori* and $\log(\sigma)$ is assumed locally uniform and independent of β *a priori*. This does not, however, imply that the parameters are independent *a posteriori*. The prior distribution of the non-normality parameter β was not fixed, but it seemed reasonable to use uniform prior for it. Based on these assumptions, the joint distribution of the location parameter θ and shape parameter β can be written as follows:

$$P(\theta,\beta \mid y) \propto \frac{\Gamma\left(1+\frac{1}{2}n(1+\beta)\right)P(\beta)}{\Gamma\left(1+\frac{1}{2}(1+\beta)\right)^{n}\left[\sum_{i=1}^{n}\left|y_i-\theta\right|^{\frac{2}{1+\beta}}\right]^{\frac{n(1+\beta)}{2}}} \tag{3.6}$$

where $P(\beta)$ is the prior distribution of β. By integrating θ out from Eq. (3.6), the marginal distribution of β given the data is as follows:

$$P(\beta \mid y) = \frac{P_u(\beta \mid y)P(\beta)}{\int_{-1}^{\infty} P_u(\beta \mid y)P(\beta)d\beta}$$

(3.7) *where*

$$P_u(\beta \mid y) \propto \frac{\Gamma\left(1+\frac{1}{2}n(1+\beta)\right)}{\Gamma\left(1+\frac{1}{2}(1+\beta)\right)^n} \int_{-\infty}^{\infty} \frac{1}{\left[\sum_{i=1}^{n} \left|y_i - \theta\right|^{\frac{2}{1+\beta}}\right]^{\frac{n(1+\beta)}{2}}} d\theta$$

in which $-1 < \beta < \infty$, $-\infty < \theta < \infty$, and $n = 1,2,3\ldots$ Since the integrals in Eq. (3.7) cannot be solved analytically, numerical methods were applied. The first and second moments of the marginal β were also defined as integrals[2] that were solved numerically. The percentiles were calculated numerically using the Newton-Raphson method with bracketing.

The location parameter θ is of minor interest and it is also rather common practice to assume it to be zero in this kind of studies. However, this practise was not applied for two reasons. First, it was of interest to avoid the effect of the possible miss-treatment of the location parameter on the inference about the shape of the distribution. Second, the elimination of location parameter by assuming it to be zero would not have significantly simplified the estimation of shape parameter since, if the location parameter were assumed zero, the nuisance parameter σ would have had to be integrated out numerically. The formula for marginal θ is based on the assumptions discussed earlier and is given as follows:

$$p(\theta \mid y) = \frac{p_u(\theta \mid y)P(\theta)}{\int_{-\infty}^{\infty} p_u(\theta \mid y)P(\theta)d\theta}$$

(3.8) *where*

$$p_u(\theta \mid y) \propto \int_{-1}^{\infty} \left[\sum_i \left|y_i - \theta\right|^{\frac{2}{1+\beta}}\right]^{-\frac{n(1+\beta)}{2}} \Gamma\left(\frac{2+n(1+\beta)}{2}\right)\Gamma\left(\frac{3+\beta}{2}\right)^{-n} d\beta$$

The percentiles were again calculated using Newton-Raphson method with bracketing.

2 $\mu = \int_{-\infty}^{\infty} x f(x)dx$ and $\delta^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$

3.3. Computational aspects

In Bayesian analyses, the nuisance parameters are integrated out from the joint distributions, but usually these integrals cannot be calculated analytically. Thus, the numerical integration plays an essential role and it has been discussed in the Bayesian context by Evans and Swartz (1995) and Shaw et al. (1996).

In this study, the number of dimensions was quite low so the curse of dimensions did not occur as a major problem. The applied approach was the basic Romberg integration with a modification, i.e., midpoint rule, which does not require the integrand to be evaluated at the ends of the integration interval. (See Press et al. 1992 for the algorithms). The integrals were truncated so that even a substantial increase (at least 15%) in the interval did not change the results. This approach was computationally not very efficient since it would have been more efficient to use an algorithm that does not require a uniform integration step. However, this approach led to accurate results within a reasonable computer time.

The formulations represented in the previous section cannot be used directly in the numerical calculation because, if they are implemented, the resulting algorithm tends to overflow when the sample size is larger than sixty points and the algorithm also becomes relatively inaccurate. The first related problem is that the Gamma function overflows with quite modest values. However, since a large value of Gamma is divided by another large value, this problem can be easily solved by coding the equation as subtraction of logarithms resulting in perfectly ordinary value. The second difficulty is related to the integral over θ that tends to underflow or overflow depending on the number of points in the sample resulting in some inaccuracy. This difficulty can be removed by scaling the data with an appropriate constant while in this study the scaling factor ten was used.

By implementing the above improvements, it is possible to establish an algorithm that works fine in samples containing approximately three hundred or less data points. With a very large number of points in the sample, the algorithm inevitably suffers from overflow, which can be solved by adding a constant into the equations. The form of $P_u(\beta \mid y)$ used when numerically calculating the marginal β of Eq. (3.7) is as follows:

$$P_u(\beta \mid y) \propto \int_{-\infty}^{\infty} e^{\ln\left\{\Gamma\left(1+\frac{1}{2}n(1+\beta)\right)\right\} - n\ln\left\{\Gamma\left(1+\frac{1}{2}(1+\beta)\right)\right\} - \frac{n(1+\beta)}{2}\ln\left\{\left[\sum_{i=1}^{n}\left|y_i-\theta\right|^{\frac{2}{1+\beta}}\right]\right\} \pm \ln(c)} d\theta \tag{3.9}$$

in which c is some arbitrary positive constant. Unfortunately, there is no formal method to find the best value for the parameter c or for the scaling constant. The value for c was set by guestimating it so that the not-normalised marginal β did not overflow at all and underflows were not encountered within the relevant range. Despite of that, with very large values of β as well as with values very close to minus one, the underflows cannot be avoided but this did not significantly affect the results. The scaling factor ten seemed to work well with all sample sizes needed and, therefore, it was used throughout the analyses.

By following a similar logic, $P_u(\theta \mid y)$ in Eq. (3.8) can be written in the following form:

$$\text{(3.10)} \quad p_u(\theta \mid y) \propto \int_{-1}^{\infty} e^{-\frac{n(1+\beta)}{2}\ln\left(\left[\sum_i |y_i-\theta|^{\frac{2}{1+\beta}}\right]\right)+\ln\left(\Gamma\left(\frac{2+n(1+\beta)}{2}\right)\right)-n\ln\left(\Gamma\left(\frac{3+\beta}{2}\right)\right)\pm\ln(c)} d\beta$$

where c is some arbitrary positive constant.

The major computational threats to the validity of this study are the possible round off errors and inadequate truncation of integrals. However, the algorithms developed seemed not to suffer from round off errors and the results remained virtually unchanged although the truncation points were selected differently.

4. RESULTS

4.1. Kernel density estimate

The data sets were transformed to have unit variance and zero mean prior the construction of kernel density estimates because the different dispersions and locations might bias the inference based on visual comparisons. This should result in plots where the differences in shapes can be visually observed. There also seems to be two inherent problems in the kernel density estimates. First, the effect of the window width and, second, the fact that a quite limited amount of data is available. Fortunately, it seems that the inference is robust in relation to the selection of the window width parameter. For example, the various ways, suggested by Silverman, to obtain the window width do not affect the inference at all (visually the estimates are virtually identical). As a result, the estimates based on adaptive window width are reported in the subsequent discussions. The limited amount of data available is in turn a substantial problem as shall be shown.

In Figure 3 we show the kernel density estimates which were calculated from the entire HEX index by using the raw data and Cochrane-Orcutt corrected data. Normal distribution is plotted as a benchmark. The data in Figure 3 suggests that there are no observable differences between raw data, and Cochrane-Orcutt corrected data. Both estimates indicate modest skewness. However, it seems that there are too few points available to construct a reliable density estimate in this way. It is quite likely that the density estimates are biased so that the central region of the true distribution is extensively over-smoothed. For example, if some of the well-known parametric models of asset returns (mixture normal distribution, Lévy distribution, mixed diffusion jump, Student t) are fitted (See Gillemot et al. 2000 for the algorithms) to the data and plotted in the same figure, the resulting parametric density estimate has higher peak around the mean. Along with these features, the tails of the kernel density estimate are very messy. This leads to a situation where any inference concerning them is difficult. On the basis of this, it seems that kernel density estimates cannot be used as a model of asset return distributions when tail probabilities are of interest; e.g., in value-at-risk analysis. In summary, these findings seem to indicate that the Kernel density estimate does not give a correct picture about the true distribution given the sample size.

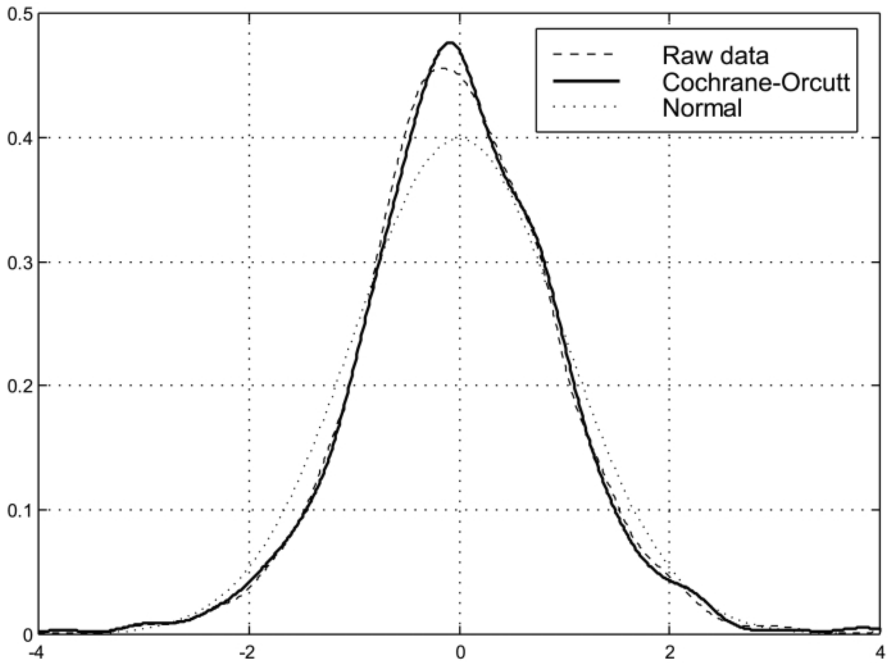

FIGURE 3: Kernel density estimates from HEX.

The above-discussed possible biases are even more significant in case of returns grouped by weekdays. Figure 4 shows the kernel density estimates based on raw data grouped by weekdays. In general, these estimates are quite messy even in the central region of the distribution. Thus, it is not possible to do reliable inference based on these estimates. Nevertheless, one minor feature can be observed: the asset returns on Wednesdays might be slightly different from other weekdays. However, this has never before been reported and it is likely because of the bias in the density estimates.

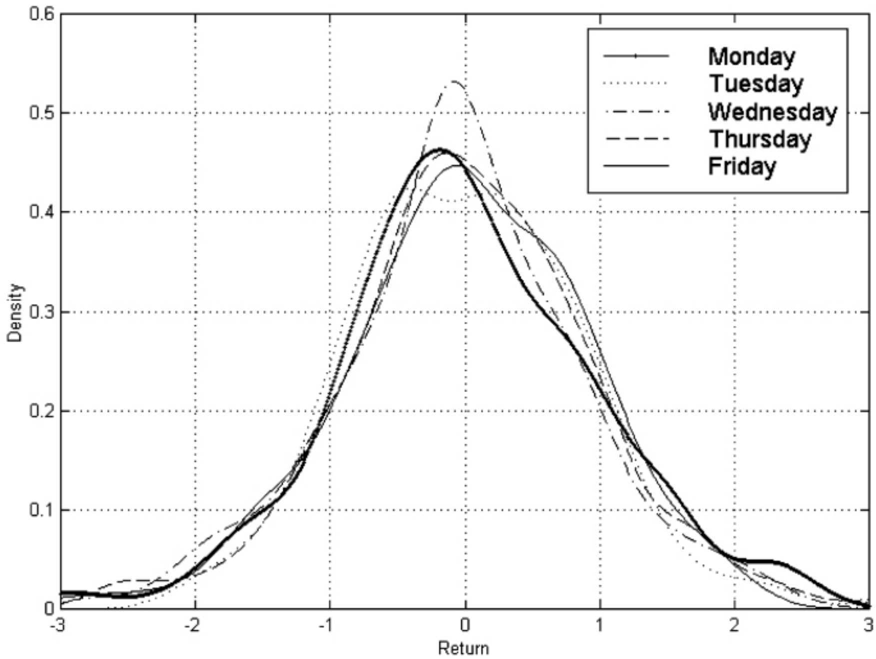

FIGURE 4: Kernel density based on HEX data grouped by weekdays.

4.2. Power exponential

4.2.1. Joint distribution of θ and β

Figure 5 shows the joint distribution of θ (location parameter) and β (shape parameter) given the full raw data set. This preliminary result suggests that the distributions of θ and β are quite symmetric. By maximising numerically this joint density in the two-parameter space, the resulting values for θ and β are 0.0067 and 0.57 respectively. As indicated in Section 3.3, returns were scaled with multiplier 10 for computational convenience. This scaling affects θ's level but β is not affected.

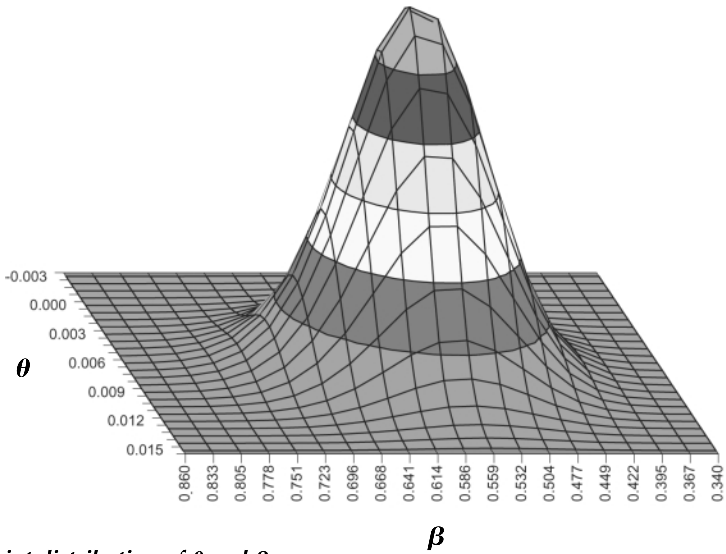

FIGURE 5: Joint distribution of θ and β.

4.2.2. Marginal θ

In order to exclude the possible effect of the location, and its variability, on the inference concerning the shape of the distribution, it is studied first. The marginal distributions of θ using all data points are shown in Figure 6. There seems to be no difference between raw data and

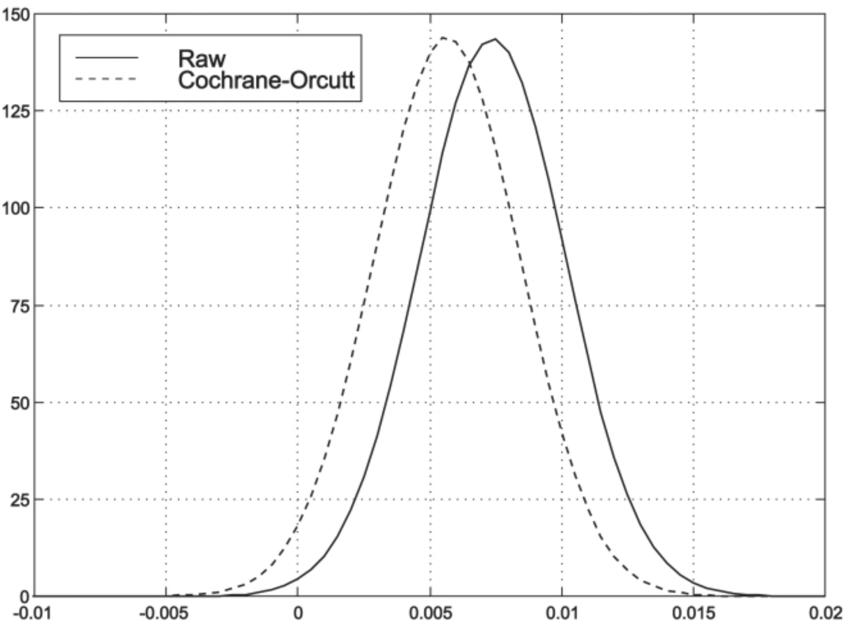

FIGURE 6: Normalised densities of marginal θ calculated from all data points.

Cochrane-Orcutt corrected data while this applies to both the full data and the annually split data (the plot is not shown). In general, the long run θ calculated from all data points seems to be slightly positive, which makes sense intuitively.

The distributions of marginal θ seem to be symmetric as was expected on the basis of the fact that θ's marginal distributions can be well approximated with Student t distribution when β is between –0.75 and 0.75 (Box–Tiao 1973, 170). This symmetry can also be clearly observed from semi-logarithmic plots (not reported).

The moments, modes, and percentiles of θ for the full data and for the annually split data are given in Table 2. The value of θ seems to deviate significantly from zero and that it is not constant over time. However, the observed deviations are so small that given the transaction costs it is hardly possibly to develop profitable trading rules based on these deviations. This result also implies that in case of risk measurement, e.g., in value-at-risk framework it might be justifiable to assume θ as zero. In addition, the variance of the marginal distribution seems to decrease when the sample size is increased.

TABLE 2: Summary statistics from location parameter.

Year*	Correction	N	Mean	Std.dev.	90%	Median	10%	Mode
1991–97	None	1749	0.0068	0.0026	0.0101	0.0068	0.0034	0.0068
	Cochrane-Orcutt	1749	0.0051	0.0027	0.0086	0.0051	0.0017	0.0052
1991	None	248	–0.0112	0.0055	–0.0041	–0.0114	–0.0181	–0.0112
	Cochrane-Orcutt	248	–0.0076	0.0053	–0.0011	–0.0073	–0.0147	–0.0065
1992	None	251	–0.0021	0.0074	0.0073	–0.0020	–0.0114	–0.0020
	Cochrane-Orcutt	251	–0.0033	0.0070	0.0055	–0.0032	–0.0124	–0.0026
1993	None	251	0.0261	0.0082	0.0366	0.0261	0.0155	0.0261
	Cochrane-Orcutt	251	0.0226	0.0079	0.0327	0.0226	0.0125	0.0226
1994	None	251	0.0042	0.0068	0.0129	0.0041	–0.0045	0.0040
	Cochrane-Orcutt	251	0.0023	0.0068	0.0111	0.0023	–0.0064	0.0023
1995	None	249	–0.0008	0.0081	0.0096	–0.0009	–0.0111	–0.0012
	Cochrane-Orcutt	249	–0.0013	0.0084	0.0094	–0.0014	–0.0120	–0.0013
1996	None	250	0.0195	0.0058	0.0269	0.0195	0.0120	0.0196
	Cochrane-Orcutt	250	0.0159	0.0059	0.0236	0.0160	0.0083	0.0160
1997	None	249	0.0238	0.0088	0.0354	0.0236	0.0127	0.0236
	Cochrane-Orcutt	249	0.0223	0.0090	0.0339	0.0224	0.0106	0.0224

*For computational convenience returns were multiplied by 10

The previous analysis suggested that θ is not constant over time. The next point of interest is whether there are deviations between weekdays since the previous research has generally suggested anomalies related to weekdays. The marginal distributions of θ calculated from the data grouped by weekdays are given in Figure 7. Since major deviations between corrected and raw data are not observed, the results from raw data are only reported. Along with that, the Monday returns seemed to slightly differ from other weekdays. The Monday expectation and 50% percentile for marginal θ are negative while this expectation and 50% percentiles are positive for other weekdays. Nevertheless, these differences are not significant but, interestingly, the expectation and 50% percentile tend to somewhat increase towards the end of the week.

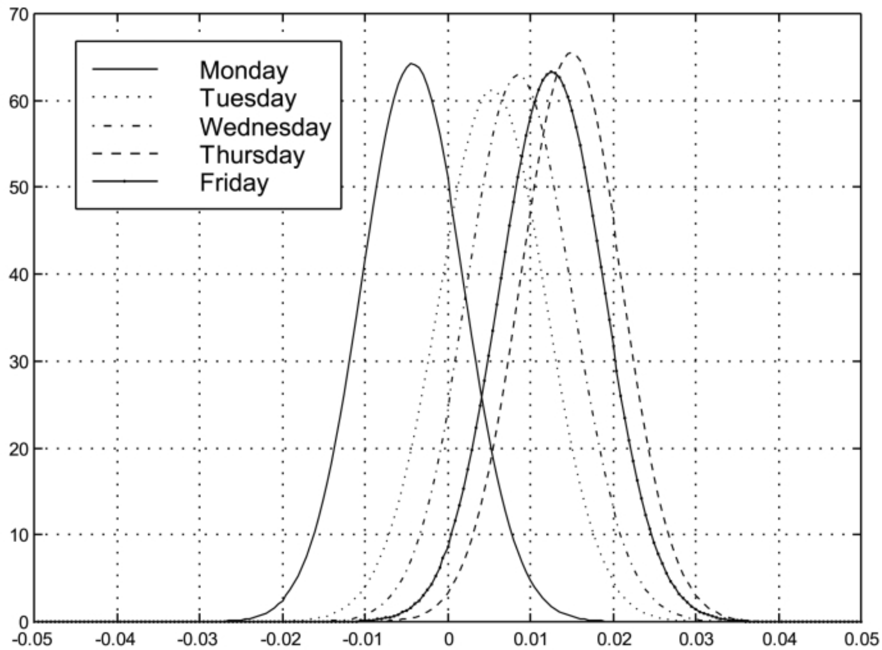

FIGURE 7: Normalised densities of marginal θ by weekdays (raw data).

4.2.3. Marginal β

The shape of the distribution is the main subject of this study and we shall discuss it in the rest of the paper. A good starting point is the long run shape, which is shown in Figure 8. The distributions resulting from raw data as well as from corrected data do not differ. The first moments are 0.5780 for raw data and 0.5522 in case of corrected data while the respective 50% percentiles are 0.5762 and 0.5505. Furthermore, there is a very low probability that the distribution is normal, i.e., $\beta = 0$. However, the marginal distribution of the shape parameter is strictly speaking not symmetric, which can be observed from the logarithmic plot (not reported). In spite of that fact, the distribution tends to be quite well approximated by a normal distribution when very small or large probabilities are not of interest.

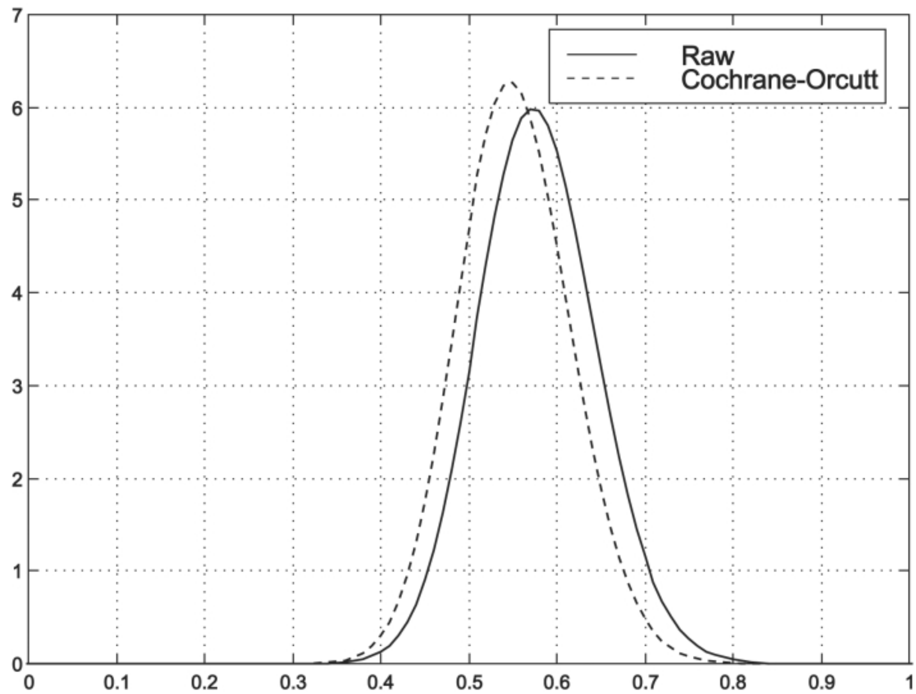

FIGURE 8: Normalised densities of marginal β calculated from all data points.

In general, several anomalies related to weekdays are observed in return series but the question whether the shape of the distribution differs between different weekdays has not been discussed with the same intensity. The possible variations between weekdays would also affect the inference about the time dependency of the shape. Figure 9 indicates the marginal distributions of the shape parameter from corrected data while the raw data yields similar re-

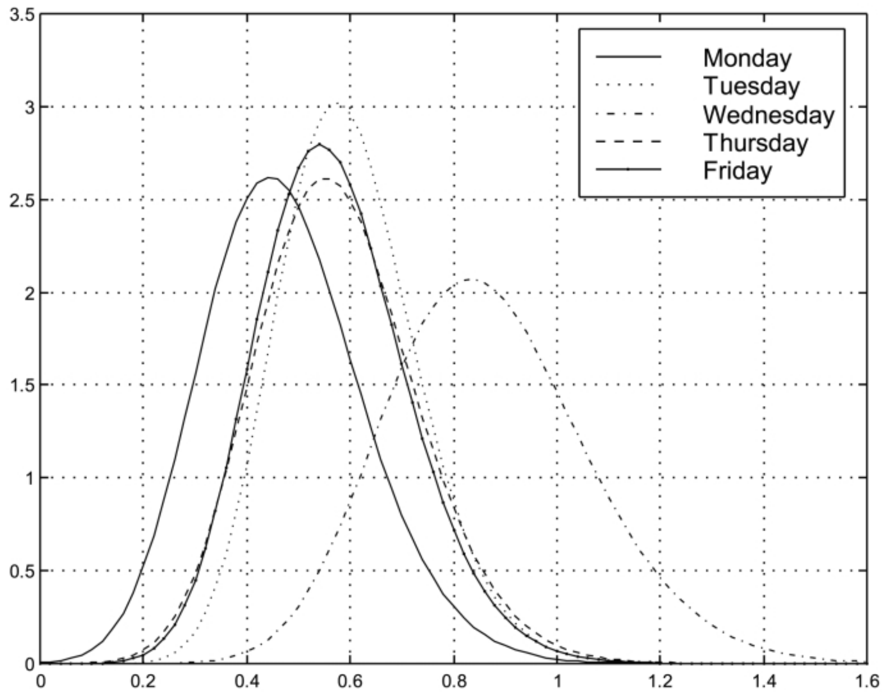

FIGURE 9: Marginal β calculated from each weekday's returns.

sults. The results show no evidence that the distributions are different between weekdays. However, in case of all weekdays, a significant deviation from normality is observed but Mondays and Fridays tend to be similar to other weekdays. The logarithms of densities (plot not reported) suggest again a slight skewness in the marginal distributions and there seem to be no systematic differences between Cochrane-Orcutt corrected and raw data. This result might not hold in other markets because the weekday effects are not generally strong on Finnish stock market (see Booth et al. 1992).

The next issue is the temporal stability of the shape of the distribution. In this case, significant variations are observed. When the marginal distributions of β based on annually split data are being analysed, there seems to be significant variations from one year to another. Furthermore, there are times when normality could be assumed do exist, i.e., $\beta = 0$. For example, the return generating process seems to be normal in the year 1993 but it is quite probably *not* normal for years 1991, 1992, 1995, and 1997. Thus, these results indicate that part of the contradictory in the previous studies might be related to the temporal variability of the shape of the distribution.

The moments, modes, 90%, 50% (median), and 10% percentiles of marginal β are given in Table 3. In addition to the above-mentioned points, the data in Table 3 suggests that the variance of marginal β is quite stable over time but depends on the sample size.

TABLE 3: Moments and percentiles of marginal β.

Year	Correction	N	Mean	Std.dev.	90%	Median	10%	Mode
1991–97	None	1749	0.5780	0.0668	0.6647	0.5763	0.4937	0.5728
	Cochrane-Orcutt	1749	0.5522	0.0638	0.6349	0.5505	0.4716	0.5472
1991	None	248	0.8794	0.1960	1.1366	0.8668	0.6385	0.8420
	Cochrane-Orcutt	248	0.8942	0.1981	1.1543	0.8808	0.6514	0.8545
1992	None	251	0.6950	0.2062	0.9655	0.6812	0.4419	0.6542
	Cochrane-Orcutt	251	0.6880	0.2076	0.9604	0.6742	0.4331	0.6472
1993	None	251	0.0414	0.1489	0.2366	0.0320	–0.1419	0.0135
	Cochrane-Orcutt	251	–0.1300	0.1349	0.0470	–0.1389	–0.2958	–0.1562
1994	None	251	0.2387	0.1619	0.4516	0.2262	0.0417	0.2019
	Cochrane-Orcutt	251	0.2490	0.1569	0.4552	0.2373	0.0578	0.2143
1995	None	249	0.6164	0.1889	0.8647	0.6029	0.3855	0.5763
	Cochrane-Orcutt	249	0.6058	0.1828	0.8459	0.5932	0.3818	0.5685
1996	None	250	0.3732	0.1797	0.6092	0.3606	0.1532	0.3359
	Cochrane-Orcutt	250	0.3046	0.1622	0.5176	0.2934	0.1059	0.2713
1997	None	249	0.7773	0.1797	1.0130	0.7657	0.5563	0.7431
	Cochrane-Orcutt	249	0.7860	0.1828	1.0259	0.7737	0.5619	0.7494

In order to further study the evident time dependency of β, the rolling βs were calculated by using 251 data points' rolling window. Unfortunately, there is no formal method to select the width of this window. Thus, it was set to 251 in order to reflect the typical length of a trading year. By selecting a shorter window, the fluctuations are increased but the uncertainty results in an increased variance of marginal β that complicates the inference. The 90%, 50%, and 10% percentiles of marginal β are plotted in Figure 10. Since there seems to be no significant differences in corrected and raw data, results based on raw data are only reported (the Cochrane-Orcutt correction was also applied in rolling manner and these results lead to similar inference as the reported results). The rolling βs suggest quite strongly that the shape of the distribution is time dependent or at least it is not constant over time. It also seems that there are times when the return generating process is normal ($\beta = 0$).

It also seems that the marginal distribution of β is numerically well approximated by a normal distribution, so let us assume that the marginal β is normally distributed. When the 10% and 90% percentiles are calculated under this assumption and then compared to the directly numerically derived percentiles, the absolute average deviations are 0.00988 for 10% percentile and 0.00545 respectively for 90% percentile. In case of the first moment and 50%

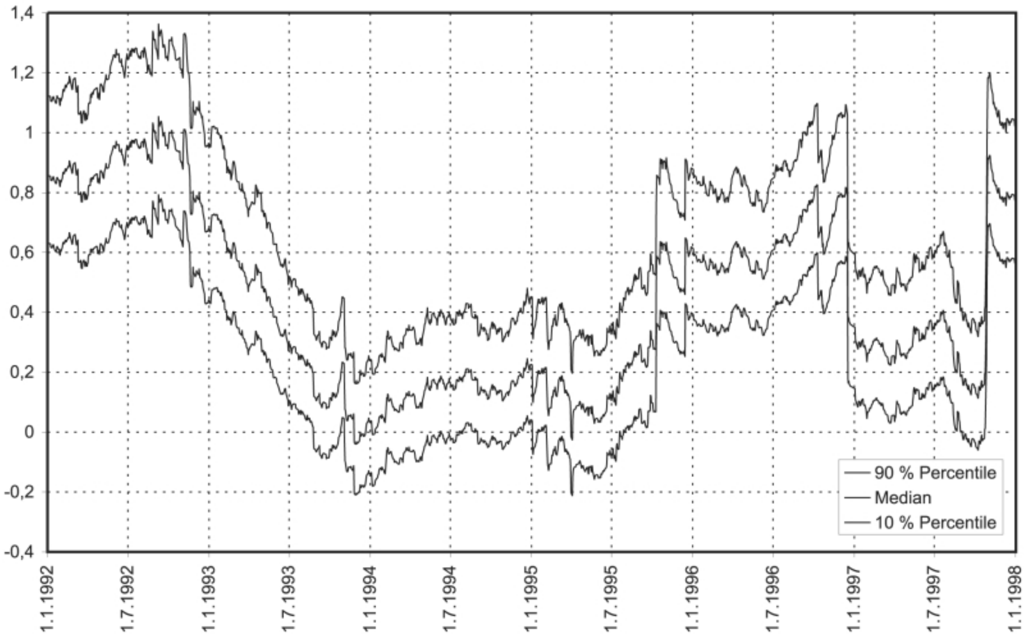

FIGURE 10: Percentiles of marginal β (raw).

percentile the absolute average deviation is slightly larger, being 0.0120. However, there is a systematic tendency in these deviations: 90% and 10% percentiles are larger than those achieved by assuming normal distribution while the 50% percentile is smaller than the first moment. However, it seems reasonable to use normal distribution as a model of marginal β, at least in numerical applications, since the deviations are very small.

When the measures of centrality, i.e., first moment, mode, and 50% percentile (median) are being considered, there are no observable differences between them. Because of the computational efficiency related to the mode of marginal β, it seems reasonable to favour it if a single measure of centrality is required. This conclusion applies to raw data as well as to corrected returns. Plots with comparisons from these statistics are not reported since the statistics cannot be visually distinguished.

The previous research has generally concluded that monthly returns are approximately normally distributed and that by increasing the time interval the return distribution approached normal distribution. This tendency was tested by calculating the first moment of the marginal β by increasing time interval used to calculate the returns. This is shown in Figure 11 and there seems to be a tendency towards normality when the time interval grows. This is perhaps

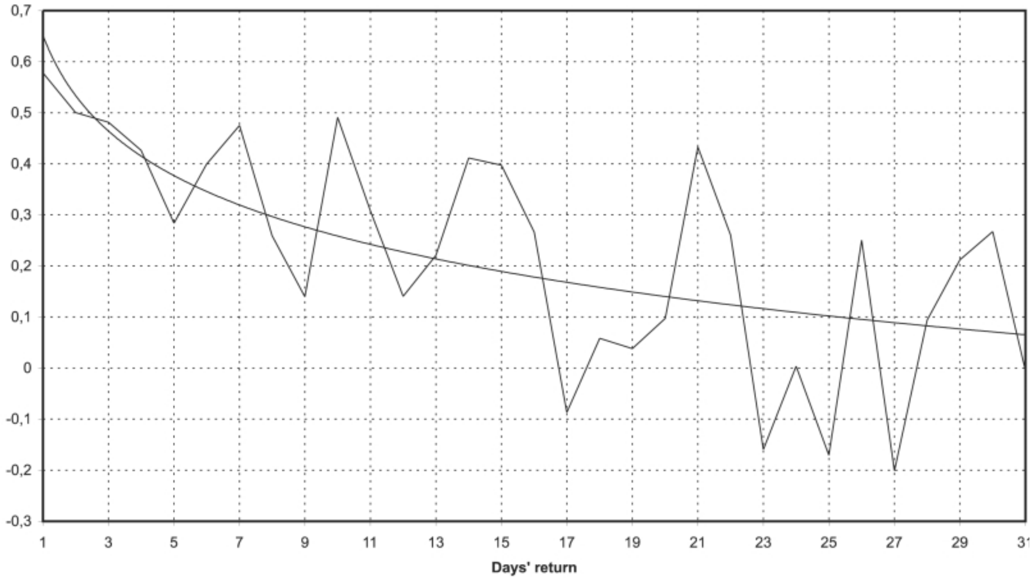

FIGURE 11: First moment of marginal β as a function of Δt.

an indication that the central limit theorem (CLT) is functioning. However, the variance of marginal β steadily increases when the number of data-points is reduced because of the increased time interval used to calculate returns. Thus, the uncertainty is substantially larger when longer than five days' returns are considered.

The estimation procedures used in this study are computationally quite intensive. However, it seems that, in numerical sense, the shape parameter can be approximated. This is done by first transforming the data to zero mean and unit variance and then assuming the respective parameters as constants. What is then left is to integrate over the shape parameter. These approximated results seem to be visually almost indistinguishable although minor deviations do exist. Consequently, it might be justifiable to use this approximation in numerical applications.

5. SUMMARY AND DISCUSSION

Here we have explored the shapes of stock return distributions. The analyses were started with the non-parametric kernel density estimator. Results indicated that return distribution is unimodal, leptokurtic, and quite symmetric. However, the limited amount of data available made the tails of the kernel density estimate very messy and, thus, no inference concerning these was possible. In addition, it seems, given the sample size, that the kernel density estimator underestimates significantly the probability mass of the central regions. These difficulties lead to the conclusion that kernel density estimation might not yield a good description of the distribution – at least the statistical sample size should be substantially larger.

The second model, we used, was the power exponential family of distributions. Their parameters were estimated with the Bayesian approach. The location parameter θ seemed to be basically zero although, in statistical sense, small but significant deviations were observed. These deviations have little meaning in reality since, given the transaction costs, it is hardly possible to place any profitable trading rules on them. It also seems that in case of short-term inference, it would be justifiable to assume θ to be zero. On the other hand, the long-run expectation was slightly positive as was expected. No significant deviations between weekdays were observed although the expectation tends to somewhat increase towards the end of the week being negative on Mondays.

The major issue was the shape of the distribution. Since the return series generally indicate anomalies related to weekdays, the question whether the shape of the return distribution differs between weekdays was explored first. The results indicated no evidence about the differences in the shape of the distribution between weekdays. This result might not hold in other markets because the weekday effects are not generally strong on Finnish stock market.

The variability of the shape of the distribution over time was especially interesting since the analyses provided strong evidence that the shape is not constant. Along with that, periods, when the return generating process is normal, were observed. This suggests that part of the contradiction in the previous results concerning the shape of the distribution might be explained by the variability of the shape of the distribution over time. The shape of return distribution also seemed to approach steadily normal distribution when the time interval used to calculate returns was increased from one day towards thirty days' returns. This observation is also consistent result with the previous research. ■

REFERENCES

AGRO, G. (1995) Maximum Likelihood Estimation for the Exponential Power Function Parameters. *Communications in statistics. Part B Simulation and Computation.* Vol. 24:2, 523–536.

AKGIRAY, V.–BOUTH, G. (1987) Compound Distribution Models of Stock Returns: An Empirical Comparison. *Journal of Financial Research.* Vol. 10:3, 269–280.

ARIEL, R.A. (1987) A monthly Effect in Stock Returns. *Journal of Financial Economics.* Vol. 18:1, 161–174.

ARTHUR, W.B.–HOLLAND, J.H.–LEBARON, B.–PALMER, R.–TAYLER, P. (1997) Asset Pricing Under Endogenous Expectations in an Artificial Stock Market. In Arthur, W. B–Durlauf S. N.–Lane, D. A. (1997) *The Economy as an Evolving Complex System II.* Addison-Wesley: Reading, Massachusetts.

BACHELIER, L. (1900) Theorie de la Speculation. Paris: Gauthier-Villars.

BERA, A. K.–HIGGINS, M.L. (1993) ARCH models: Properties, Estimation and testing. *Journal of Economic Surveys.* Vol. 7:4, 305–366.

BLATTBERG, R.–GONEDES, N. (1974) A comparison of Stable and Student Distributions as Statistical Models for Stock Prices. *Journal of Business*. Vol. 47, 244–280.

BOLLERSLEV, T. (1986) Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*. Vol. 31:1, 307–327.

BOLLERSLEV, T. (1987) A Conditional Heteroskedastic Model for Speculative Prices and Rates of Return. *Review of Economics and Statistics*. Vol. 69:3, 542–547.

BOOTH, G.G.–HATEM, J.–VIRTANEN, I–YLI-OLLI, P. (1992) Stochastic Modeling of Security Returns: Evidence from the Helsinki Stock Exchange. *European Journal of Operational Research*. Vol. 56:1, 98–107.

BOX G.E. P.–TIAO G.C. (1973) *Bayesian inference in statistical analysis.* Addison-Wesley publishing company. Reading: Massachusetts.

ENGLE, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of UK. *Econometrica*. Vol. 50:2, 987–1007.

EPPS, T.W.–EPPS, M.L. (1976) The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distribution hypothesis. *Econometrica.* Vol. 44, 305–321.

EVANS, M.–SWARTZ, T. (1995) Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. *Statistical Science*, Vol. 10:3,254–272.

FAMA, E. (1965) The behaviour of stock market prices. *Journal of Business.* Vol. 38, 34–105.

FAMA, E. (1970) Multiperiod Consumption-Investment Decisions. *American Economic Review.* Vol. 60, 163–174.

FAMA, E.–FRENCH K. (1988) Permanent and Temporary Components of Stock Prices. *Journal of Political Economy.* Vol. 96:2, 246–273.

FIELITZ, B.–ROZELL, J. (1983) Stable Distributions and Mixtures of Distributions Hypotheses for Common Stock Returns. *Journal of the American Statistical Association*. Vol. 69, 108–113.

FRENCH, K. (1980) Stock Returns and the Weekend Effect. *Journal of Financial Economics*. Vol. 8:1, 55–70.

GIBBONS, M.–HESS, P. (1981) Day of the Week Effects and Asset Returns. *Journal Business*. Vol. 54:4, 579–596.

GILLEMOT, L.–TÖYLI, J.–KERTESZ, J.–KASKI K. (2000) Time Independent Models of Asset Returns Revisited. Forthcoming.

GÓMEZ, E.–GÓMEZ-VILLEGAS, M.A.–MARÍN, J.M. (1998) A multivariate generalization of the power exponential family of distributions. *Communications in Statistics A Theory and Methods*. Vol. 27:3, 589–600.

HAGERMAN, R.L. (1978) More evidence on the distribution of security returns. *Journal of Finance.* Vol. 33:4, 1213–1221.

HSU, D.A.–MILLER, R.–WICHERN, D. (1974) On the Stable Paretian Behaviour of Stock Market Prices. *Journal of the American Statistical Association*. Vol. 69, 108–113.

HSU, D.A. (1982) A Bayesian Robust Detection of Shift in the Risk Structure of Stock Market Returns. *Journal of the American Statistical Association*. Vol. 77, 29–39.

KANTO, A.J.–KAHRA, H.A.–BLEVINS, D.R.–SCHADEWITZ, H.J. (1998) An explanation of the Unusual Behavior of Some Market Model Residuals. *The Finnish Journal of Business Economics*. Vol. 47:3.

KEIM, D.B. (1983) Size-Related Anomalies and Stock Return Seasonality: Further empirical evidence. *Journal of Financial Economics.* Vol. 12:1, 13–32.

KON, S.J. (1984) Models of stock returns – a comparison. *The Journal of Finance*. Vol. 39:1, 147–165.

KULLMAN, L.–TÖYLI, J.–KERTESZ, J.–KANTO, A.–KASKI, K. (1999) Characteristic times in stock market indices. Physica A: Statistical Mechanics And Its Applications. Vol. 269:1, 98–110.

MANDELBROT, B. (1963) The variation of certain speculative prices. *Journal of Business.* Vol. 36, 394–419.

MANTEGNA, N.R.–STANLEY, E. (1995) Scaling behaviour in the dynamics of an economic index. *Nature*. Vol. 376:6, 46–49.

OFFICER, R. (1972) The distribution of Stock returns. *Journal of the American Statistical Association*. Vol. 67, 807–812.

OLDFIELD, F.S. JR.–ROGALSKI R. J.–JARROW, R.A. (1977) An autoregressive jump process for common stock returns. *Journal of Financial Economics*. Vol. 5, 389–418.

OSBORNE, M.F.M. (1959) Brownian Motion in the Stock Market. *Operations Research.* Vol. 7:2, 145–173.

PERRY, P.R. (1983) More evidence on the nature of the Distribution of Security Returns. *Journal of Financial and Quantitative Analysis.* Vol. 18:2, 211–221.

PRESS, W.H.–TEUKOLSKY, S.A.–VETTERLING, W.T.–FLANNERY, B.P. (1992) *Numerical recipes in C the Art of Scientific Computing.* Second Edition. Cambridge university press: USA.

SCHWERT, W.G. (1989) Why does Stock Market Volatility Change over Time? *Journal of Finance*. Vol. 44:5, 1115–1153.

SHAW, J.E.H.–GENZ, A.–MONAHAN, J.–SCHERVISH, M. J.–WASSERMAN, L.–WOLFINGER, R (1996) Comments on "Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems" by Michael Evans and Tim Swartz. *Statistical Science*. Vol. 11:1, 54–64.

SILVERMAN, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall: London.

SIMONS, K. (1996) Value at Risk – New approaches to risk management. *New England Economic Review*. September/October.

TAUCHEN, G.E.–PITTS, M. (1983) The price variability – volume relationship on speculative market. *Econometrica*. Vol. 51:2. 485–505.

TAYLOR, S.J. (1985) *Modelling Financial Time Series*. John Wiley & Sons: Chichester

TAYLOR, S.J. (1994) Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*. Vol. 4:2, 183–204.

TUCKER, A.L. (1992) A Reexamination of Finite- and Infinite-variance Distributions of Daily Stock Returns. *Journal of Business and Economic Statistics.* Vol. 10:1, 73–81

WELSH, A.H. (1996) *Aspects of Statistical Inference*. John Wiley & Sons Inc: New York.