QUANTITATIVE FINANCE
RESEARCH CENTRE

UNIVERSITY OF
TECHNOLOGY SYDNEY

QUANTITATIVE FINANCE
RESEARCH CENTRE

UTS

THINK.CHANGE.DO

QUANTITATIVE FINANCE RESEARCH CENTRE

Research Paper 194 **March 2007**

Empirical Evidence on Student-*t* Log-Returns of Diversified World Stock Indices

Eckhard Platen and Renata Sidorowicz

ISSN 1441-8010 www.qfrc.uts.edu.au

Empirical Evidence on Student-t Log-Returns of Diversified World Stock Indices

Eckhard Platen [1] and **Renata Sidorowicz**[1]

March 6, 2007

Abstract. The aim of this paper is to document some empirical facts related to log-returns of diversified world stock indices when these are denominated in different currencies. Motivated by earlier results, we have obtained the estimated distribution of log-returns for a range of world stock indices over long observation periods. We expand previous studies by applying the maximum likelihood ratio test to the large class of generalized hyperbolic distributions, and investigate the log-returns of a variety of diversified world stock indices in different currency denominations. This identifies the Student-t distribution with about four degrees of freedom as the typical estimated log-return distribution of such indices. Owing to the observed high levels of significance, this result can be interpreted as a stylized empirical fact.

1991 *Mathematics Subject Classification:* primary 90A12; secondary 60G30, 62P20.
JEL Classification: G10, G13
Key words and phrases: diversified world stock index, growth optimal portfolio, log-return distribution, Student-t distribution, generalized hyperbolic distribution, likelihood ratio test

[1]University of Technology Sydney, School of Finance & Economics and Department of Mathematical Sciences, PO Box 123, Broadway, NSW, 2007, Australia

1 Introduction

There is a long standing interest in the financial literature in the identification of the typical log-return distribution of financial instruments. The importance of a systematic study of log-returns of financial securities was highlighted by the 1987 stock market crash. In this instance, the typical assumption that the distribution of log-returns of financial indices was Gaussian failed to be any longer acceptable. It became clear that the probabilities of extreme values of log-returns were much larger than those supported by the standard Gaussian assumption. Moreover, this invalid assumption turns out to be rather dangerous for a range of applications in risk management. Many studies after the crash strongly rejected the normality of log-returns for indices, stocks and exchange rates. Their results were often based on the observation of excessively large kurtosis, or the better fit of various leptokurtic log-return densities.

It has now been established that log-return densities of financial indices exhibit heavier tails and are more peaked than the Gaussian assumption would permit. However, a distribution that will generally fit log-returns of particular classes of financial securities has yet not been agreed upon in the literature. This paper aims to resolve this problem for the case of diversified stock market indices.

As will be explained below, the most promising instruments for which one may identify a particular type of log-return density, are diversified stock indices. Studies on log-returns for indices include the two papers by Markowitz & Usmen (1996a, 1996b) analyzing S&P500 log-returns in a Bayesian framework. These authors considered the rich family of Pearson distributions and identified the Student-t distribution with about 4.5 degrees of freedom as the best fit to daily log-return data of the S&P500. Independently, Hurst & Platen (1997) reached a similar conclusion by studying daily log-returns of the S&P500 and other regional stock market indices. Their research was focused on a large family of normal-variance mixture distributions, [see Clark (1973)], which included the log-return distributions generated by several important models proposed in the literature. These distributions included among others the normal, [see Samuelson (1957) and Black & Scholes (1973)]; the alpha-stable, [see Mandelbrot (1963)]; the Student-t, [see Praetz (1972) and Blattberg & Gonedes (1974)]; the normal-inverse Gaussian, [see Barndorff-Nielsen (1995)]; the hyperbolic, [see Eberlein & Keller (1995) and Küchler et al.(1999)]; the variance gamma, [see Madan & Seneta (1990)]; and the symmetric generalized hyperbolic distribution, [see Barndorff-Nielsen (1978) and McNeil, Frey & Embrechts (2005)]. In Hurst & Platen (1997) the Student-t distribution with 3.0-4.5 degrees of freedom was determined as the best fit to daily, regional stock market index log-returns. This confirmed and generalized Markowitz's and Usmen's findings by the use of an alternative statistical methodology and a wider range of stock market indices. Recently, Fergusson & Platen (2006) employed a maximum likelihood ratio test, [see Rao (1973)], in the class of symmetric generalized hyperbolic distributions. They studied the log-return distribution of a world stock index, whose constituent weights were determined by market capitalization, and considered different currency denominations of such an index. These authors concluded, at a high level of significance, that the log-returns of their index exhibited a Student-t behavior with approximately four degrees of freedom.

Finally, we should mention that for multivariate log-returns there exists an advanced statistical methodology for identifying particular generalized hyperbolic distributions as described in McNeil, Frey & Embrechts (2005). These authors showed in the application of their results to indices, exchange rates and stocks that Student-t type log-return distributions are often likely to fit the data. However, they did not quantify any level of significance.

The aim of this study is to extend the analysis in Fergusson & Platen (2006) by (a) using a range of different purposely constructed, diversified world stock indices, and (b) by searching among the larger class of generalized hyperbolic distributions for a best fit. The use of diversified world stock indices, for studying properties of log-returns, instead of stock prices or exchange rates, is motivated by the *benchmark approach* developed in Platen (2002, 2004) and Platen & Heath (2006). Such diversified indices approximate the *growth optimal portfolio* (GOP), which has a number of outstanding properties. The GOP maximizes expected logarithmic utility from terminal wealth. It is almost surely the portfolio with the maximum long term growth rate and is also known as Kelly portfolio, [see Kelly (1956)]. In our study we partly follow Le & Platen (2006), who proposed a general construction methodology for *diversified portfolios* (DPs). By a *diversification theorem*, presented in Platen (2005), a DP can be considered to be a good approximation of the GOP under certain assumptions. The *equally weighted index* (EWI) can be regarded as one of the most diversified portfolios and thus, as a likely proxy for the GOP, [see Platen & Heath (2006)]. Moreover, we will demonstrate that the purposely constructed equally weighted index EWI104s, which is based on 104 world industry sector indices, is one of the best performing world stock indices when its historical long term growth is considered. This supports the view adopted in this paper that the EWI104s may be an excellent proxy for the GOP.

We analyze a number of differently constructed, diversified stock indices by a maximum likelihood methodology, and estimate the distribution of their log-returns. The estimated distribution of the log-returns of the EWI104s turns out to be the *Student-t* distribution with approximately *four* degrees of freedom in almost all currency denominations when searching among the large family of *generalized hyperbolic* (GH) distributions. Since this result will be established at a very high level of significance, it can be interpreted as a stylized empirical fact. Although all other DPs show a Student-t feature, the EWI104s exhibits it most clearly. Consequently, any advanced model for a diversified world stock index should generate log-returns whose estimates appear to be Student-t distributed with about four degrees of freedom.

This paper is organized as follows. First, in Section 2 we present a methodology for the construction of diversified stock indices. This is followed by a comparison of different world stock indices with respect to their performance as investment portfolios. In Section 3, we introduce the wide class of generalized hyperbolic distributions and describe the maximum likelihood ratio test for this class. The application of this methodology to index data in Section 4 leads to test statistics accompanied by the corresponding estimated parameters and high levels of significance.

2 Index Construction

This section focuses on a methodology for the construction of diversified world stock indices. Such indices are usually formed in order to measure the general market performance and general market risk, [see Basle (1996)]. They are widely used as benchmarks in investment management. Some of the following indices also have a more theoretical motivation; we construct them to be self-financing portfolios.

The data selected for the $d \in \mathcal{N} = \{1, 2, \dots\}$ constituents of the indices consist of daily data for the period from 1973 to 2006. We construct world stock indices from regional stock market indices and from world sector market indices by using similar methodologies. The regional stock market indices represent market capitalization weighted stock indices as constructed and provided by Thomson Financial. The world sector indices are also constructed and provided by Thomson Financial, and reflect the worldwide evolution of respective industries. The regional or the world sector indices are used in our study as constituents of the newly constructed indices.

Portfolio Generating Functions. We emphasize in this study four main types of indices, *market capitalization weighted indices* (MCIs), *diversity weighted indices* (DWIs) as described by Fernholz (2002), *equally weighted indices* (EWIs), and some type of *world stock indices* (WSIs) introduced in Le & Platen (2006).

We assume that all the constituents of our constructed indices are capable of unbounded positive jumps and negative jumps leaving constituents arbitrarily close to zero. This reflects the fact that, in principle, any of the constituents S_t^j, $j \in \{1, 2, \dots, d\}$, can almost default at any time for one reason or another. Therefore, since we consider only *strictly positive portfolios* in our study, their fractions of wealth always need to remain nonnegative.

In order to eliminate the possibility of short sales, and have a systematic way of generating nonnegative fractions using a wide range of methods; we introduce the notion of a *portfolio generating function* (PGF), which has been inspired by a similar construct described in Fernholz (2002). More precisely, a PGF $A : \Re^d \to [0,1]^d$ maps a given vector of fractions $\boldsymbol{\pi}_{\delta,t} = (\pi^1_{\delta,t}, \pi^2_{\delta,t}, \dots, \pi^d_{\delta,t})^\top \in \Re^d$, into a

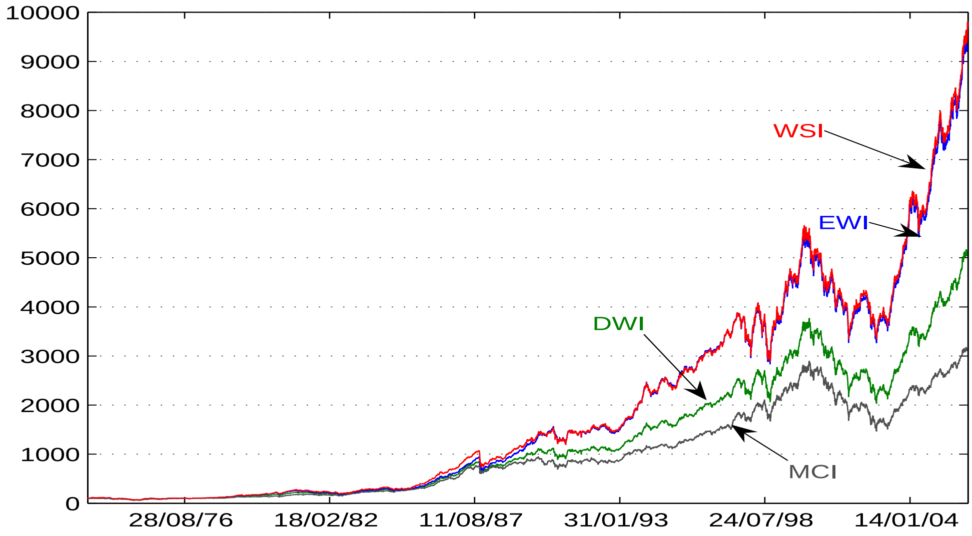

Figure 1: Indices constructed from regional stock market indices

vector of nonnegative fractions

$$\tilde{\boldsymbol{\pi}}_{\delta,t} = (\tilde{\pi}^1_{\delta,t}, \tilde{\pi}^2_{\delta,t}, \ldots, \tilde{\pi}^d_{\delta,t})^\top = A(\boldsymbol{\pi}_{\delta,t}) \in [0,1]^d \tag{2.1}$$

such that $\sum_{j=1}^d \tilde{\pi}^j_{\delta,t} = 1$ for all $t \in \Re^+$. Note that the given vector of fractions $\boldsymbol{\pi}_{\delta,t}$ may contain negative components. These components can be obtained by any kind of method, including the use of statistical estimates of optimal fractions. Estimates provided by experts or economically based theoretical predictions can also be used. These fractions are then translated into nonnegative fractions by a PGF. Note that we do not include the savings account in a PGF, as is also typical for most commercial indices.

Market Capitalization Weighted Indices. For an MCI we define the fraction of wealth held in the j-th constituent at time t as follows:

$$\pi^j_{\delta_{MCI},t} = \frac{\delta^j_t S^j_t}{\sum_{i=1}^d \delta^i_t S^i_t}, \tag{2.2}$$

$j \in \{1, 2, \ldots, d\}$. Here δ^j_t is the number of units of the jth constituent of the portfolio S^δ_t at time t, which is typically held constant over certain periods of time.

Diversity Weighted Index. The so called *diversity weighted indices* (DWIs) are theoretically and practically interesting indices, which were proposed in Fernholz (2002). Here the PGF is a function of the MCI fractions $\boldsymbol{\pi}_{\delta_{MCI},t}$ given in

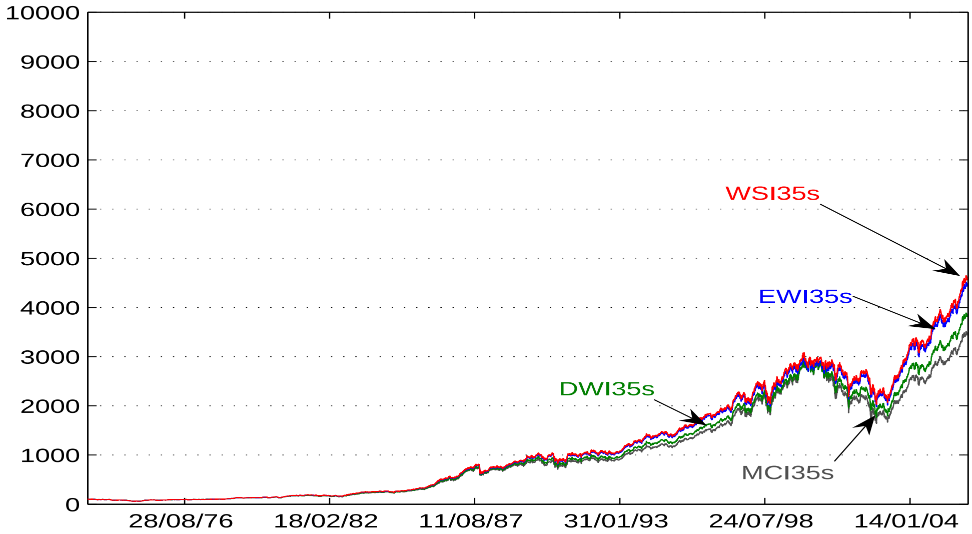

Figure 2: Indices constructed from sector indices based on 35 industries

(2.2) and has the form

$$\tilde{\pi}^j_{\delta,t} = \frac{(\pi^j_{\delta_{MCI},t})^p}{\sum_{l=1}^d (\pi^l_{\delta_{MCI},t})^p} \tag{2.3}$$

for some choice of a real number $p \in [0, 1]$ and $j \in \{1, 2, \ldots, d\}$, $t \in \Re^+$. The DWI has been designed to outperform the market portfolio, that is the MCI, [see Fernholz (2002)].

Equally Weighted Indices. An almost ideally diversified index is obtained by setting all fractions equal. The jth fraction of the EWI is then simply given by the constant

$$\pi^j_{\delta_{EWI},t} = \frac{1}{d} \tag{2.4}$$

for all $j \in \{1, 2, \ldots, d\}$, where d is the number of constituents. The main advantage of this index is that it forms the best diversified portfolio and does not need the calculation of its fractions from data or other sources. We will show in this paper that EWIs does not only exhibit excellent long term performance but also have very clear distributional properties. These distributional features, as well as their excellent long term growth rate, make EWIs important tools for theoretical investigations and practical applications.

A Family of World Stock Indices. There have been many attempts in the literature and in real life to construct investment portfolios with outstanding returns. It is evident from estimation theory, [see for instance Kelly, Platen & Sørensen (2004)], that, in principle, hundreds of years of data are necessary to estimate risk premia with any reasonable level of significance. Such long data

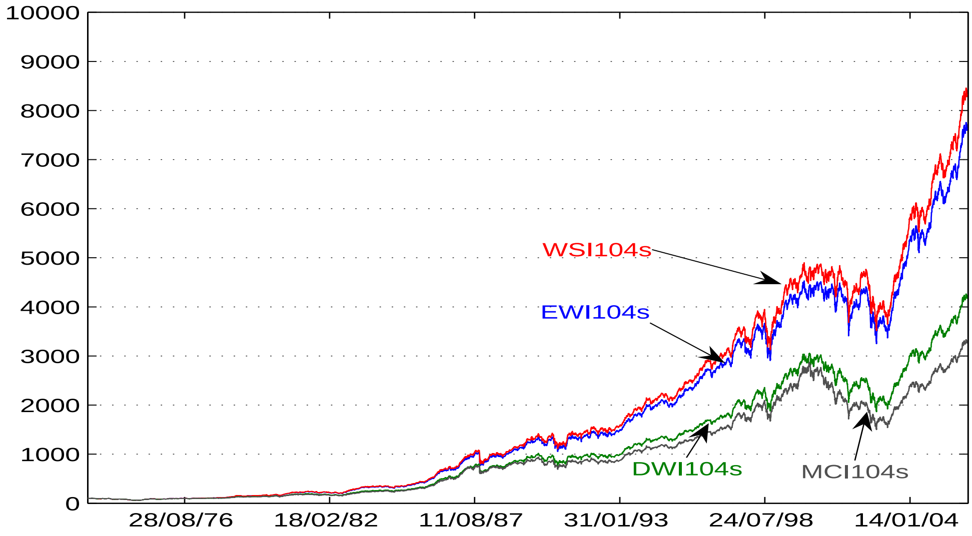

Figure 3: Indices constructed from sector indices based on 104 industries

sets are not available, and the risk premia cannot be expected to remain constant over sufficient long periods of time. This makes it very difficult to use any statistical method successfully in investment portfolio construction. There may exist strategies that outperform the diversified indices described above. However, it is unlikely that this superiority is systematic and can be sustained over long periods. Despite the empirical difficulties, various stock indices were recently studied by Le & Platen (2006) with the aim of approximating the GOP by also using standard statistical estimates of risk premia and volatilities. No significant advantage was reported for these constructed indices, as will be confirmed in this paper.

We will study *world stock indices* (WSIs) as special cases in a family of indices which also include the MCI, DWI and EWI. The PGF used for the construction of this general family of indices is given by

$$\tilde{\pi}^j_{\delta,t} = \frac{(\pi^j_{\delta,t} + \mu_t)^p}{\sum_{l=1}^d (\pi^l_{\delta,t} + \mu_t)^p}, \tag{2.5}$$

for all $j \in \{1, 2, \ldots, d\}$ and $t \in \Re^+$, where $p \in [0, 1]$ is some real number. This construction is slightly more general than what has been suggested in Fernholz (2002), [see (2.3)]. Essentially, the above PGF keeps the ranking of the fractions $\pi^j_{\delta,t}$ intact and transforms the original fractions into positive fractions. We obtain the fractions of a DWI if $\pi^j_{\delta,t} = \pi^j_{MCI,t}$ and $\mu_t = 0$ for all $t \in \Re^+$. The fractions $\pi^j_{\delta,t}$ can be chosen to approximate those of the GOP on the basis of whatever information is available. Theoretically one obtains the fractions of the GOP of

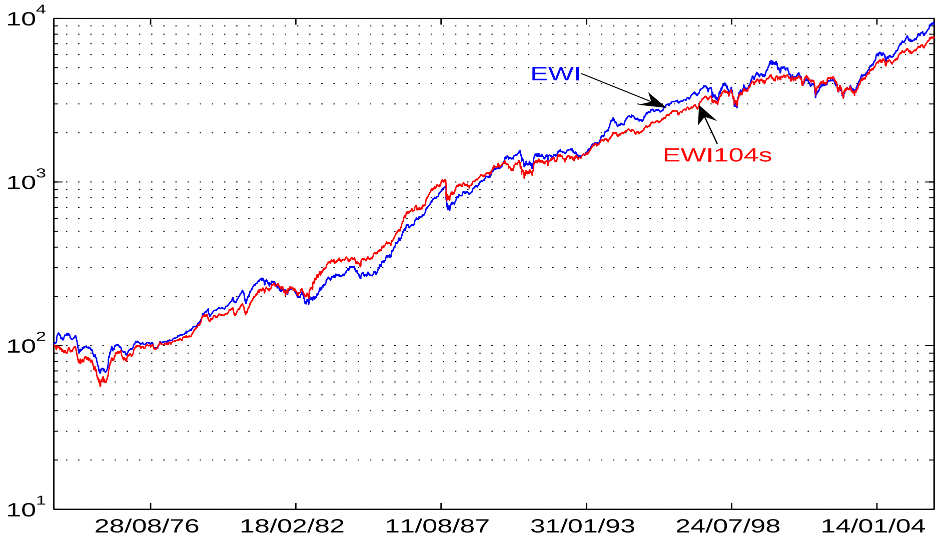

Figure 4: The regional EWI and sector EWI104s indices in log-scale

the stock market in the form

$$\boldsymbol{\pi}_{\delta_*,t} = \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{a}_t - r_t\mathbf{1}) \tag{2.6}$$

for $t \in \Re^+$, see [Merton (1973), Platen & Heath (2006) or Filipoviċ & Platen (2007)]. Here $\boldsymbol{\Sigma}_t$ denotes the covariance matrix of returns and $\boldsymbol{a}_t$ the vector of expected returns of the constituents, while r_t is the short rate.

It is reasonable to estimate the covariance matrix $\boldsymbol{\Sigma}_t$ from the observation of daily returns, say, of the most recent one year period. Despite our reservations about the low significance of any estimates for the appreciation rate vector $\boldsymbol{a}_t$, we may nevertheless estimate it in a standard manner from daily returns over the same one year period. This is also what, in principle, is often performed in active fund management. The resulting suggested fractions of the GOP vary in an extreme manner and can be largely negative. To make the fractions positive via a PGF we set

$$\mu_t = \left|\inf_j \pi^j_{\delta,t}\right| + \mu, \tag{2.7}$$

for some choice $\mu \geq 0$, $t \in \Re^+$.

Comparison of Constructed Indices. The indices described above are graphed in Figs. 1-3. They are constructed from 38 regional stock market indices, 35 industry sector indices or 104 industry sector indices, respectively. For all three types of constructed indices we observe that MCIs always perform worse than DWIs, which again perform worse than EWIs and WSIs. This is a common feature of both regional and industry sector based indices. In particular, the EWI and the EWI104s perform extremely well. It can be noticed that the performance of well diversified indices is generally much better than that of less diversified indices. The diversification theorem in Platen (2005) provides an explanation for this phenomenon by stating that well diversified portfolios are likely to be better proxies for the GOP. Since the GOP is known almost surely to have the best long term growth rates, this appears to be supported by the constructed indices.

Furthermore, we note that the EWI and the WSI are almost identical when setting $p = 0.5$. This is due to the fact that the resulting WSI fractions are very close to those of the EWI. On the other hand, the fractions of the WSI are still flexible and make it possible for the WSI to outperform the EWI slightly. This, however, comes with a significant computational cost and a decreasing level of diversification. The real question is whether the suggested fractions for the GOP have some statistical information that is readily exploitable. Unless this is the case, it is unlikely that a WSI outperforms an EWI significantly in the long run.

In Fig. 4, we now plot on a logarithmic scale the best performing indices, which are the EWI, based on 38 regional stock market indices, and the EWI104s, with 104 world industry sectors as constituents. This graph shows that historically sometimes the EWI and on other occasions the EWI104s performs slightly better. They are both well diversified indices and, therefore, can be expected to be good proxies of the GOP. This raises the question as to whether there are empirical features that make one of these the better proxy of the GOP. We clarify this question by studying the distribution of their log-returns.

3 A Class of Log-return Distributions

Let us introduce a framework for a sufficiently rich class of log-return distributions. The models we consider mix the normal distribution with different stochastic means and variances, resulting in a class of normal mean-variance mixture distributions.

The random variable X is said to have a normal mean-variance mixture distribution if

$$X = m(W) + \sqrt{W}\sigma Z, \tag{3.1}$$

where $Z \sim N(0,1)$ is standard Gaussian. Here $W \geq 0$ is a nonnegative random variable which is independent of Z; $\sigma \in \Re$ is a constant and $m : [0,\infty) \rightarrow \Re$ is a measurable function. A possible concrete specification for the function $m(W)$ is the affine function $m(W) = \mu + W\gamma$, where μ and γ are parameters in $\Re$. Note that this specification of normal mean-variance mixture distributions allows skewness. The skewness parameter here is $\gamma \in \Re$. When the skewness vanishes, with $\gamma = 0$, and the mean $\mu = 0$ the normal mean-variance mixture distribution is simplified to its symmetric case; this is the normal variance-mixture distribution. Note that in this particular case the random variable X, given by (3.1), becomes $X = \sqrt{W}\sigma Z$, using the same notation as before.

Generalized Hyperbolic Distributions. The *generalized hyperbolic* (GH) distribution belongs to the family of mean-variance normal mixture distributions. Its density function can be obtained by assuming a *generalized inverse Gaussian* (GIG) distribution for the mixing density, that is $W \sim \mathrm{GIG}(\lambda, \chi, \psi)$.

The distribution of $X \sim \mathrm{GH}(\lambda, \chi, \psi, \mu, \sigma, \gamma)$ is characterized by its density

$$f_X(x) = \frac{\psi^{\lambda}(\psi+\gamma\beta)^{\frac{1}{2}-\lambda}(\sqrt{\chi\psi})^{-\lambda}}{\sqrt{2\pi}\sigma K_{\lambda}(\sqrt{\chi\psi})} \times \frac{K_{\lambda-\frac{1}{2}}\left(\sqrt{(\chi+Q)(\psi+\gamma\beta)}\right)}{\left(\sqrt{(\chi+Q)(\psi+\gamma\beta)}\right)^{\frac{1}{2}-\lambda}} e^{\xi\beta} \tag{3.2}$$

for $x \in \Re$, where $\xi = x - \mu$, $\beta = \gamma\sigma^{-2}$, $Q = (x-\mu)^2\sigma^{-2}$ and $K_{\lambda}(\cdot)$ is a modified Bessel function of the third kind with index λ, [see Abramowitz & Stegun (1972)].

One can introduce an alternative parametrization to the above (λ, χ, ψ)- parametrization. For the so called $(\bar{\alpha}, \lambda)$-parametrization, one sets $\alpha = \sqrt{\psi}$, $\delta = \sqrt{\chi}$, $\bar{\alpha} = \alpha\delta$. In this case the density (3.2) reads

$$f_X(x) = \frac{(\alpha^2+\gamma\beta)^{\frac{1}{2}-\lambda}\alpha^{2\lambda}\bar{\alpha}^{-\lambda}}{\sqrt{2\pi}\sigma K_{\lambda}(\bar{\alpha})} \times \frac{K_{\lambda-\frac{1}{2}}\left(\sqrt{(\delta^2+Q)(\alpha^2+\gamma\beta)}\right)}{\left(\sqrt{(\delta^2+Q)(\alpha^2+\gamma\beta)}\right)^{\frac{1}{2}-\lambda}} e^{\xi\beta} \tag{3.3}$$

for $x \in \Re$.

Note also the resulting convenient representation in the symmetric case $\gamma = 0$, $\mu = 0$. The *symmetric generalized hyperbolic* (SGH) density function for a random variable X then has the form

$$f_X(x) = \frac{1}{\delta\sigma K_{\lambda}(\bar{\alpha})}\sqrt{\frac{\bar{\alpha}}{2\pi}}\left(1+\frac{x^2}{(\delta\sigma)^2}\right)^{\frac{1}{2}(\lambda-\frac{1}{2})} K_{\lambda-\frac{1}{2}}\left(\bar{\alpha}\sqrt{1+\frac{x^2}{(\delta\sigma)^2}}\right) \tag{3.4}$$

for $x \in \Re$, where $\lambda \in \Re$, $\alpha, \delta \geq 0$. We set $\alpha \neq 0$ if $\lambda \geq 0$ and $\delta \neq 0$ if $\lambda \leq 0$. Hence, the SGH density is a three parameter density. The parameters λ and $\bar{\alpha}$ are invariant under scale transformations and can be interpreted as the *shape parameters* for the tails of the distribution. We may define a new parameter c as the *unique scale parameter* such that

$$c^2 = \begin{cases} \frac{(\delta\sigma)^2}{-2(\lambda+1)} & \text{if } \alpha = 0 \text{ for } \lambda < 0 \text{ and } \bar{\alpha} = 0, \\ \frac{2\lambda\sigma^2}{\alpha^2}, & \text{if } \delta = 0 \text{ for } \lambda > 0 \text{ and } \bar{\alpha} = 0, \\ \frac{(\delta\sigma)^2 K_{\lambda+1}(\bar{\alpha})}{\bar{\alpha} K_{\lambda}(\bar{\alpha})} & \text{otherwise.} \end{cases} \tag{3.5}$$

Special Cases of the SGH Distribution. There are several important cases of the SGH distribution arising as log-return distributions in widely investigated asset price models. In particular, we will further consider the following special cases: *Variance Gamma*: $\bar{\alpha} = 0$ and $\lambda > 0$, [see Madan & Seneta (1990)]; *Student-t*: $\bar{\alpha} = 0$ and $\lambda < 0$, [see Praetz (1972)]; *Hyperbolic*: $\lambda = 1$, [see Eberlein & Keller

(1995)] and *Normal Inverse Gaussian*: $\lambda = -0.5$, [see Barndorff-Nielsen (1995)]. The first two special cases deserve more attention as they are limiting cases and can be described taking into account the limiting behavior of the Bessel function involved.

Variance Gamma Density. The *variance gamma* (VG) density arises for the parameter choice $\bar{\alpha} = 0$, $\alpha = \sqrt{2\lambda}$, $\delta = \sqrt{\chi} = 0$. It is a normal mean-variance mixture distribution resulting from the choice of a gamma distribution $\mathrm{Ga}(\lambda, \lambda)$ for the mixing random variable W. The normalization constant of equation (3.2) then reduces to

$$C = \frac{(2\lambda)^{\lambda}(2\lambda + \gamma\beta)^{\frac{1}{2}-\lambda}}{\sqrt{(2\pi)}\sigma 2^{\lambda-1}\Gamma(\lambda)} = \frac{\sqrt{\lambda}\left(1 + \frac{\gamma\beta}{2\lambda}\right)^{\frac{1}{2}-\lambda}}{\sqrt{\pi}\sigma 2^{\lambda-1}\Gamma(\lambda)}.$$

Passing to the limit as $\gamma \to 0$, and assuming $\mu = 0$ we obtain the *symmetric variance gamma* density

$$f_X(x) = \frac{\sqrt{\lambda}}{\sqrt{\pi}\sigma 2^{\lambda-1}\Gamma(\lambda)}\left(\frac{\sqrt{2\lambda}|x|}{\sigma}\right)^{\lambda-\frac{1}{2}} K_{\lambda-\frac{1}{2}}\left(\frac{\sqrt{2\lambda}|x|}{\sigma}\right) \tag{3.6}$$

for $x \in \Re/\{0\}$ and

$$f_X(0) = \frac{\sqrt{\lambda}}{\sqrt{\pi}\sigma 2^{\lambda-1}\Gamma(\lambda)} 2^{\lambda-\frac{3}{2}}\Gamma\left(\lambda - \frac{1}{2}\right). \tag{3.7}$$

The symmetric variance gamma density is, therefore, a two parameter density with λ as its shape parameter. Smaller values of λ indicate increasingly heavier tails. Additionally, when $\lambda \to \infty$ the variance gamma density asymptotically approaches the Gaussian density.

Student-*t* Density. A Student-*t* density has been identified to model the log-returns of financial securities, as can be seen, for instance, in Praetz (1972), Blattberg & Gonedes (1974) and Fergusson & Platen (2006). The case of a *skewed Student-t* density emerges for the parameter choice $\bar{\alpha} = 0$, $\alpha = \sqrt{\psi} = 0$, $\delta = \sqrt{\chi} = \sqrt{\nu}$, $\nu = -2\lambda$ and the *inverse gamma distribution* $\mathrm{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$ as the distribution of the mixing random variable W.

We distinguish between the following two cases $-1 < \lambda < 0$ and $\lambda \leq -1$.
1. The case $-1 < \lambda < 0$ corresponds to the degrees of freedom $\nu < 2$ for which the normalization constant diverges. This case is not relevant to the financial applications that we have in mind, and is hence skipped in what follows.
2. The case $\lambda \leq -1$ with degrees of freedom $\nu \geq 2$ is highly relevant for our study. Equation (3.2) simplifies in this case to

$$f_X(x) = \frac{2^{(1-\nu)/2}}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\sigma} \times \left(1 + \frac{Q}{\nu}\right)^{-\frac{\nu+1}{2}} \frac{K_{\frac{\nu+1}{2}}\left(\sqrt{(\nu+Q)\gamma\beta}\right)}{\left(\sqrt{(\nu+Q)\gamma\beta}\right)^{-\frac{\nu+1}{2}}} e^{\xi\beta}. \tag{3.8}$$

Passing to the limit $\gamma \to 0$ for the asymmetry parameter γ and assuming $\mu = 0$ the equation (3.8) reduces to that of the *symmetric Student-t* density

$$f_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\sigma\Gamma\left(\frac{\nu}{2}\right)} \times \left(1 + \frac{x^2}{\sigma^2\nu}\right)^{-\frac{\nu+1}{2}}. \tag{3.9}$$

Hence, the symmetric Student-t density is a two parameter density with $\nu = -2\lambda$ degrees of freedom. Note that σ is not the standard deviation of the random variable X, which is $\sigma_X = \sigma\sqrt{\frac{\nu}{\nu-2}}$. We observe an increase in the tail heaviness of this density as the degrees of freedom decrease, which implies a larger probability of extreme values. Additionally, with the increase of the degrees of freedom $\nu \to \infty$, the Student-t density converges asymptotically to the Gaussian density.

Likelihood Ratio Test. In order to test the hypothesis that a candidate log-return distribution is acceptable or not, we follow the classical maximum likelihood ratio test, [see Rao (1973)]. The *likelihood ratio* is defined by the expression

$$\Lambda = \frac{\mathcal{L}^*{}_{model}}{\mathcal{L}^*{}_{nesting\ model}}, \tag{3.10}$$

here $\mathcal{L}^*{}_{model}$ represents the maximized likelihood function of the specific nested density, while $\mathcal{L}^*{}_{nesting\ model}$ represents the maximized likelihood function of the nesting density. For example, we will later choose in some cases the SGH density as nesting model and the symmetric Student-t density as one of the nested models. Note that in the process of maximizing the likelihood, we optimize with respect to the parameters of the given parameterized distribution. Hence we obtain both the optimal parameters and the optimal value of the likelihood function. It can be shown that for increasing number of observations $n \to \infty$ the test statistic

$$L_n = -2\ln(\Lambda) \tag{3.11}$$

is asymptotically distributed as a chi-square distribution, [see Rao (1973)]. Additionally, the degrees of freedom of this chi-square distribution are determined by the difference between the number of parameters of the nesting and the nested models. Specifically, the SGH density is a three-parameter density, while the four special cases we consider: the symmetric variance gamma, Student-t, hyperbolic and generalized inverse Gaussian densities are two-parameter densities, which implies that their test statistic is chi-square distributed with one degree of freedom.

Note that asymptotically it can be shown that

$$P(L_n < \chi^2_{1-\alpha,1}) \approx F_{\chi^2(1)}(\chi^2_{1-\alpha,1}) = 1 - \alpha, \tag{3.12}$$

where $F_{\chi^2(1)}$ denotes the chi-square distribution with one degree of freedom and $\chi^2_{1-\alpha,1}$ is its $100(1-\alpha)\%$ quantile.

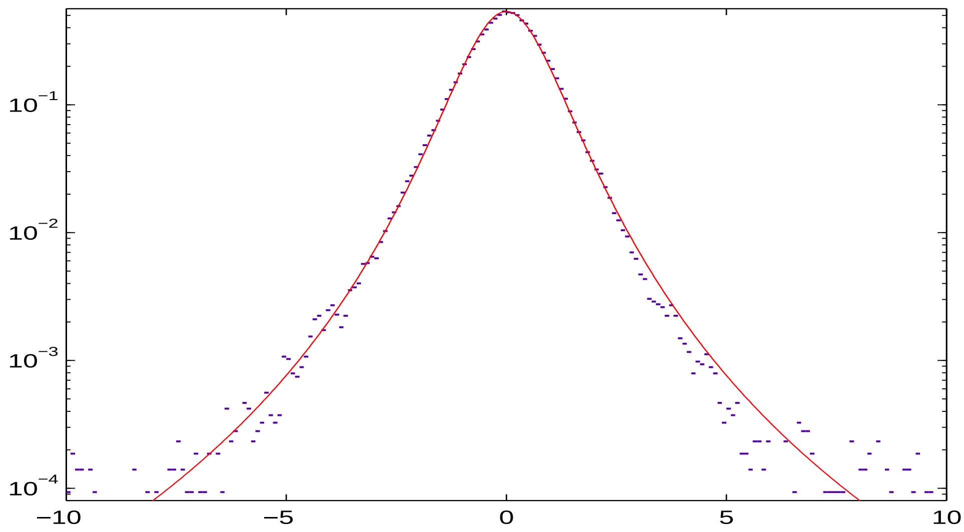

Figure 1: Log-histogram of the EWI104s log-returns and Student-t density with four degrees of freedom

We then reject the hypothesis that the suggested density is the true underlying density at the 99% level of significance if the following relation is not satisfied

$$L_n < \chi^2_{0.01,1} \approx 0.000157. \tag{3.13}$$

If we require greater precision, then the hypothesis is not rejected at the 99.9% level of significance if

$$L_n < \chi^2_{0.001,1} \approx 0.000002. \tag{3.14}$$

To conclude the above discussion, we call the density with the smallest test statistic L_n the best fit in the given family of distributions.

4 Fitted Log-Return Distributions

This section is devoted to the analysis of the log-returns of twelve world stock indices. We distinguish between the world stock indices constructed from the regional stock market indices, the world sector indices based on the 35 industry indices, and the world sector indices based on the 104 industry indices. The region based indices consist of the: MCI, DWI, EWI and WSI, while the sector based indices are represented by the: MCI35s, DWI35s, EWI35s and WSI35s, as well as, the MCI104s, DWI104s EWI104s and WSI104s. We use daily data from 1973 to 2006 provided by Thomson Financial for all the components underlying our indices. In the following, we mainly report the results for the log-returns of

Table 1: Empirical moments for log-returns of the EWI104s in various currency denominations

Country	$\hat{\mu}_y$	$\hat{\sigma}_y$	$\hat{\beta}_y$	$\underline{\hat{\kappa}}_y$
Australia	0.000565	0.008573	1.255285	38.042243
Austria	0.000423	0.008932	-0.560596	9.818562
Belgium	0.000473	0.008616	-0.572222	8.617690
Brazil	0.000774	0.010628	0.470175	15.449772
Canada	0.000519	0.007150	-0.655434	11.930284
Denmark	0.000491	0.008637	-0.493104	10.819425
Finland	0.000524	0.008646	-0.305966	9.590003
France	0.000512	0.008546	-0.522862	8.826783
Germany	0.000426	0.008627	-0.611260	8.901389
Greece	0.000867	0.009325	0.530904	27.355595
Hong Kong	0.000600	0.007379	-0.709914	16.946884
India	0.000698	0.008086	0.260450	20.710839
Ireland	0.000554	0.008863	-0.359713	35.363948
Italy	0.000622	0.008481	-0.504651	9.270888
Japan	0.000394	0.008151	-0.742019	9.827505
Korea S.	0.000593	0.009045	0.453250	36.812322
Malaysia	0.000533	0.007845	-0.664893	15.829299
Netherlands	0.000439	0.008558	-0.598923	8.979067
Norway	0.000502	0.008365	-0.431793	10.298253
Portugal	0.000714	0.009343	0.168636	13.568042
Singapore	0.000499	0.007228	-1.116746	17.150500
Spain	0.000594	0.008756	0.204594	16.586917
Sweden	0.000560	0.008372	0.256363	17.623515
Taiwan	0.000502	0.007456	-0.956096	16.691994
Thailand	0.000634	0.009012	1.861305	62.413554
UK	0.000536	0.008165	-0.593624	9.567338
USA	0.000501	0.007004	-0.819822	14.237813

Table 2: Results for log-returns of the EWI104s

	SGH	Student-t	NIG	Hyperbolic	VG
σ	0.9807068	0.7191163	0.9697258	0.9584118	0.9593693
$\bar{\alpha}$	0.0000000		0.9694605	0.7171357	
λ	-2.1629649				1.4912414
ν		4.3259646			
$\ln(\mathcal{L}^*)$	-285796.3865295	-285796.3865297	-286448.9371892	-287152.0787956	-287499.8259143
L_n		**0.0000004**	1305.1013194	2711.3845322	3406.8787696

the EWI104s when denominated in 27 currencies. A more extensive study, which also reports the results for the above mentioned indices, can be found on the website of the first author in an extended version of the paper.

A summary of the main empirical moments of the log-returns of the EWI104s when denominated in different currencies, is presented in Table 1. Note that we obtain here the average empirical mean $\hat{\mu}_y = 0.000557$, the average empirical standard deviation $\hat{\sigma}_y = 0.008437$, the average sample skewness $\hat{\beta}_y = -0.213288$ and the average sample excess kurtosis $\underline{\hat{\kappa}}_y = 17.823402$. We do not remove any extreme values as potential outliers from our data set, hence market crashes and other sudden market corrections are not discarded. Removing outliers would also not be appropriate, as the proper modeling of extreme log-returns is of great importance in risk management.

First, to get a visual impression of the shape of the log-return density of the EWI104s, we exploit all log-returns that we observe from this index in all 27

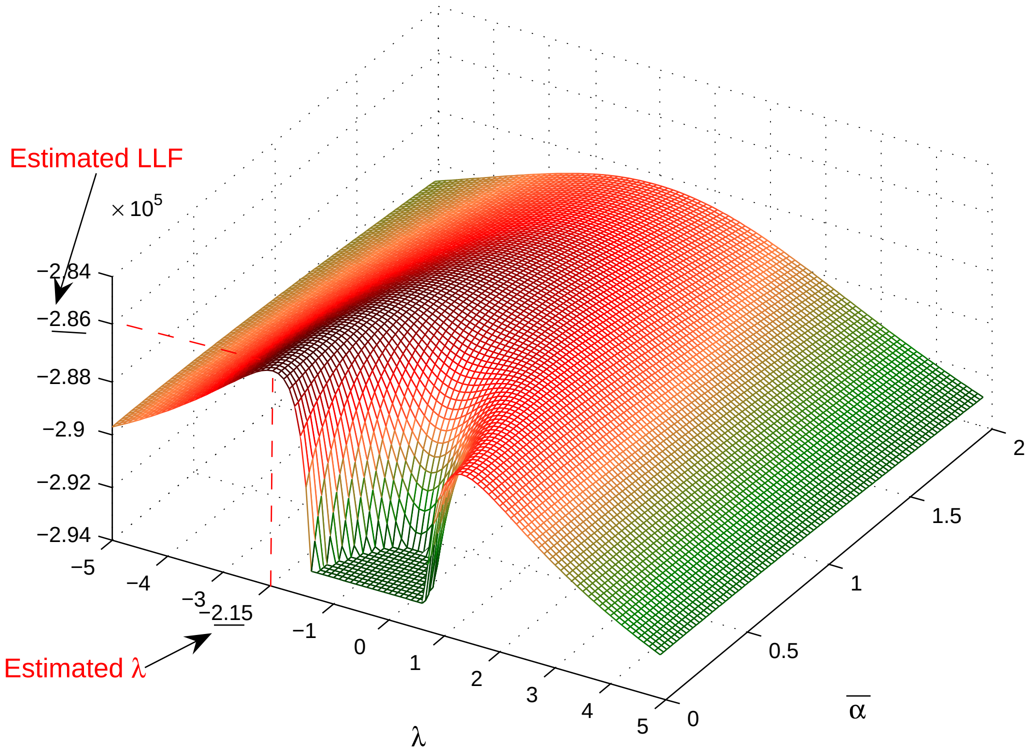

Figure 1: Log-likelihood function based on the EWI104s

currency denominations. For this purpose we combine appropriately shifted and scaled log-returns of all currency denominations. More precisely, for each currency denomination of the index we shift all the obtained log-return values so that their sample mean becomes zero, and scale them in order to obtain a sample variance of one. In Fig. 1 we present the resulting histogram of the total cohort of shifted and scaled log-returns on a logarithmic scale. Note that this histogram is based on $214,658$ observations, which makes it very reliable. Additionally in Fig. 1, we show in log-scale the theoretical Student-t density for $\nu = 4$ degrees of freedom. We observe visually an excellent fit of the log-returns of the EWI104s to the Student-t density. For the other constructed indices, a similar visual impression is obtained, with the EWI104s seeming to fit best.

The maximum likelihood methodology is then employed to estimate the parameters of the SGH density. For the same sample used to produce the histogram in Fig. 1, we exhibit the log-likelihood function for the SGH in Fig. 1 as a function of the parameters λ and $\bar{\alpha}$. One notes a clear, flat global maximum around the point $\bar{\alpha} = 0$ and $\lambda = -2$, which refers to a Student-t distribution with four degrees of freedom. We then apply the maximum likelihood method to the log-returns of the EWI104s for four special cases of the SGH distribution. These cases con-

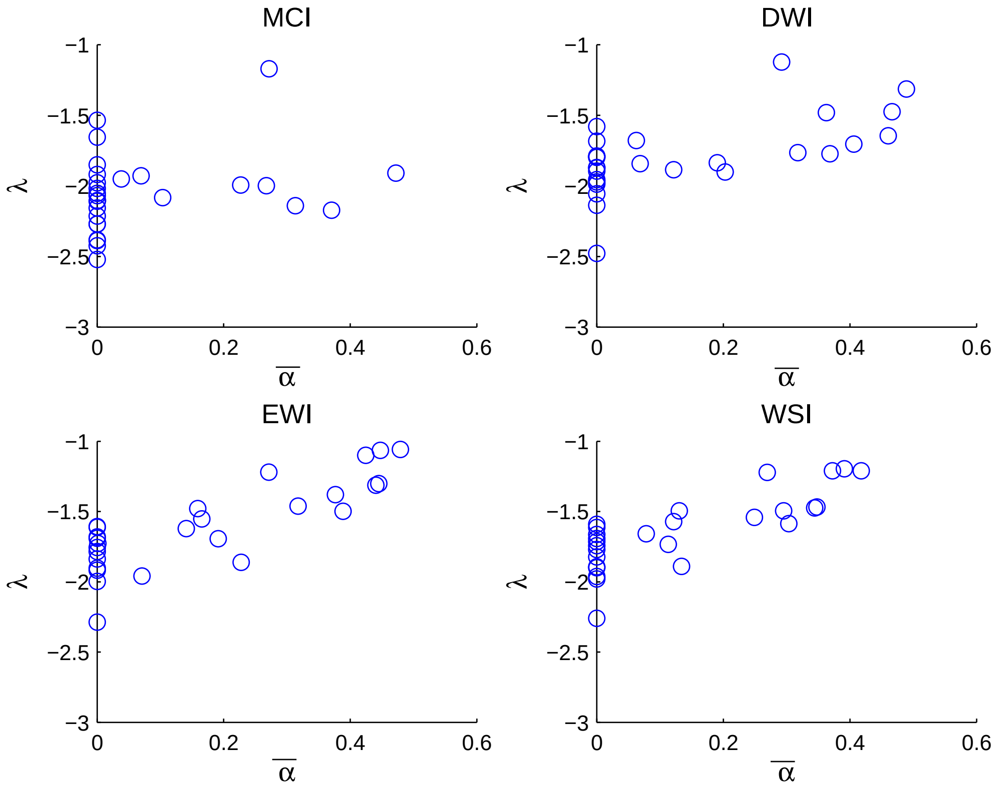

Figure 2: $(\bar{\alpha}, \lambda)$-plot for log-returns of indices in different currencies constructed from regional stock market indices as constituents

cern the Student-t density, the NIG density, the hyperbolic density and the VG density. We maximize the corresponding log-likelihood functions with respect to the parameters σ, $\bar{\alpha}$, λ and ν of the SGH, as shown in Table 2, and display their estimates in its second column. Additionally, row six of Table 2 contains the estimated parameters and maximized values of the log-likelihood functions for the SGH density and each of the four considered special cases. In the last row of Table 2 we calculated the value of the test statistic L_n. The almost zero value obtained for L_n suggests a very good fit of the Student-t distribution with $\nu \approx 4.3$ degrees of freedom to our set of data in the class of SGH distributions. The extremely small value of the test statistic, $L_n = 0.0000004$, allows us to conclude that the H_0 hypothesis of a Student-t fit to the EWI104s cannot be rejected at least on a 99.9% level of significance, since $\chi^2_{0.001,1} \approx 0.000002$. We emphasize that this is an extremely high level of significance. Note also that the estimated parameter value $\bar{\alpha} = 0$ for the SGH density suggests that a Student-t density already represents the best fit to the log-returns of the EWI104s when searching among the family of SGH densities.

For all the components underlying our indices we now again use daily data from

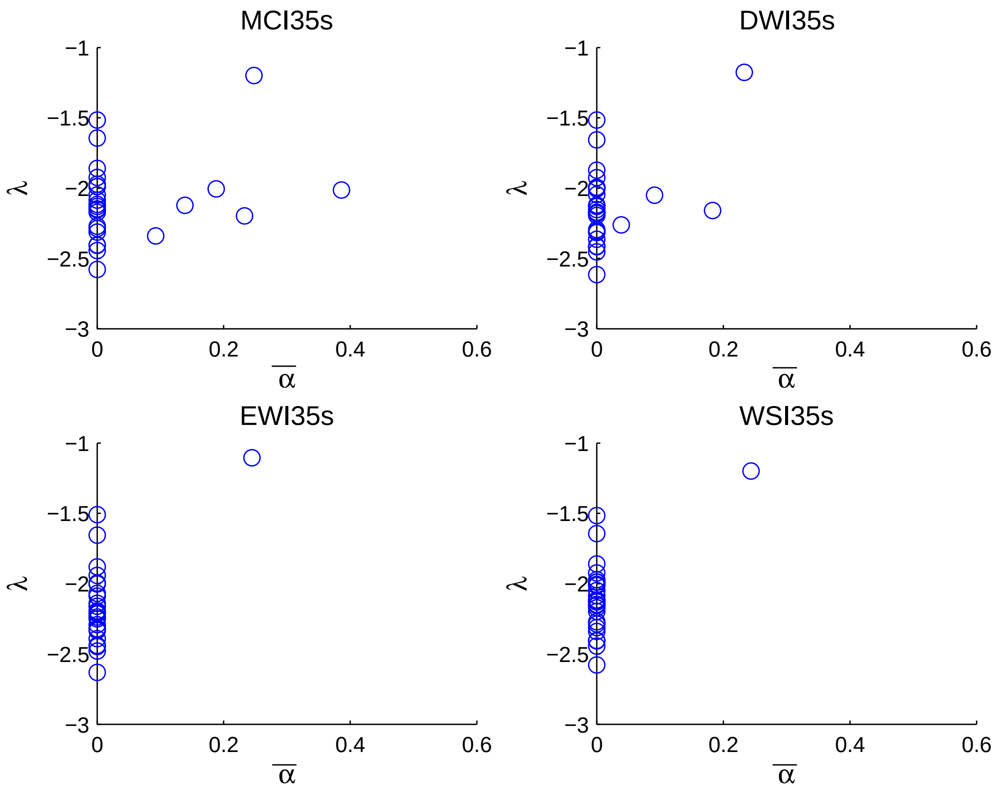

Figure 3: $(\bar{\alpha}, \lambda)$-plot for log-returns of indices in different currencies constructed from 35 sector indices as constituents

1973 to 2006, provided by Thomson Financial. Furthermore, we denominate all twelve constructed indices in 27 different currencies and study their log-returns for each of the denominations separately. Note that the denomination of a diversified world stock index in a given currency reflects in its fluctuations the general market risk with respect to this currency, [see Platen & Stahl (2003)]. We mention that the data for all exchange rates were not available from 1973. For instance, the time series for the Brazilian real starts only in 1995.

For convenience, we shift and scale the log-returns in order to obtain sample means equal to zero and sample standard deviations equal to one for all currency denominations. This does not change the generality of our analysis, but standardizes the testing procedure.

We first apply the maximum likelihood estimation methodology for both the GH and the SGH distributions for the log-returns of each index and for each currency denomination. The GH distribution contains the extra parameter $\gamma \in \Re$, which represents the level of skewness of this distribution. Our study, however, reveals that the estimated parameter values for γ are of the order 10^{-6} or less. Since γ and μ turn out to be extremely small we do not report the statistical findings for

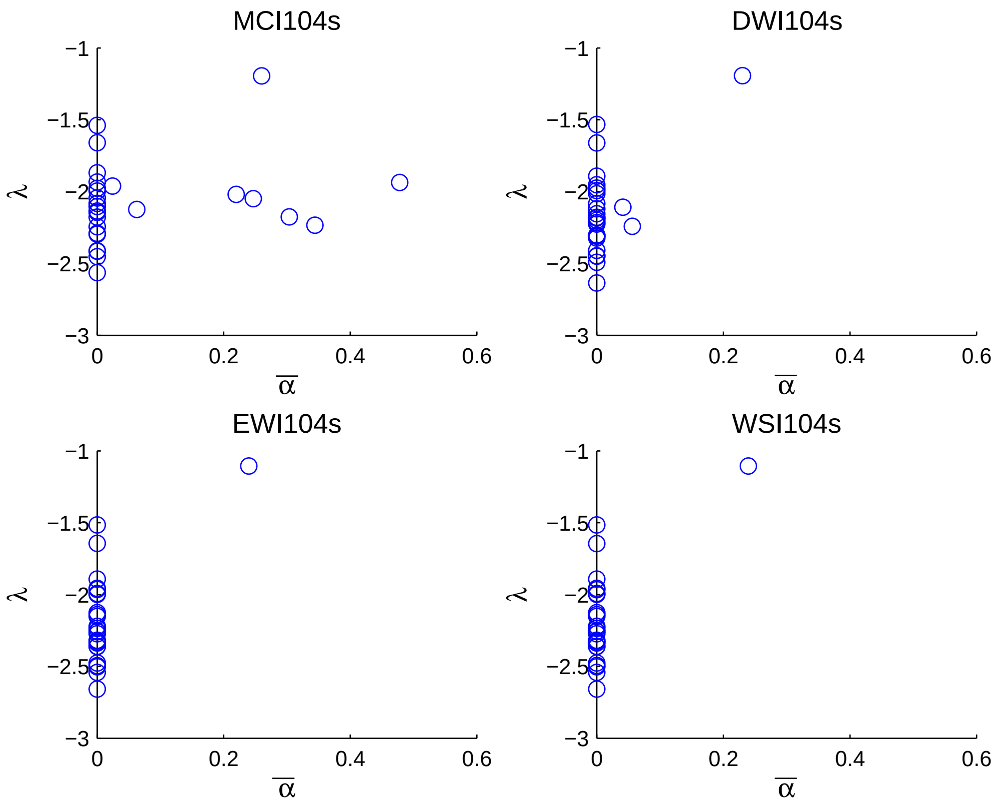

Figure 4: $(\bar{\alpha}, \lambda)$-plot for log-returns of indices in different currencies constructed from 104 sector indices as constituents

the GH distribution and just concentrate on the SGH distribution which appears to be almost identical.

In order to visualize the fitted log-return distributions of our twelve constructed indices, we plot the estimated parameter λ versus the estimated parameter $\bar{\alpha}$ in Figs. 2, 3 and 4 for each constructed index in all 27 currency denominations. Fig. 2 presents the results for each of the MCI, DWI, EWI and WSI for 27 currency denominations. Note that the estimated values of $\bar{\alpha}$ are in most cases close to zero, which already confirms the Student-t property. Furthermore, the estimated values of the parameter λ range from around -2.5 to -1.0. This indicates that the best fit can be expected for Student-t distributions with degrees of freedom $\nu = -2\lambda$ ranging from around 2 to 5. This is emphasized by the cluster of points located on the negative λ axis near -2 in Fig. 2.

Note that in the group of indices shown in Fig. 2 the log-returns of the MCI seem to fit visually the Student-t distribution best. There are just nine points which do not sit on the negative λ axis. It can be also noticed that the range of the estimated values of λ in the case of the EWI and WSI is narrower for $\bar{\alpha} = 0$.

Table 3: L_n test statistic of the EWI104s for different currency denominations

Country	Student-t	NIG	Hyperbolic	VG	ν
Australia	0.000000	76.770817	150.202282	181.632971	4.281222
Austria	0.000000	39.289103	77.505683	102.979330	4.725907
Belgium	0.000000	31.581622	60.867570	83.648470	4.989912
Brazil	2.617693	5.687078	63.800349	60.078395	2.713036
Canada	0.000000	47.506215	79.917741	104.297607	5.316154
Denmark	0.000000	41.509921	87.199686	114.853658	4.512101
Finland	0.000000	28.852844	68.677271	88.553080	4.305638
France	0.000000	26.303544	57.639325	80.567283	4.722787
Germany	0.000000	27.290205	52.667918	71.120798	5.005856
Greece	0.000000	60.432172	104.789463	125.601499	4.674626
Hong.Kong	0.000000	42.066531	100.834255	122.965326	3.930473
India	0.000000	74.773701	163.594078	198.002956	3.998713
Ireland	0.000000	77.727856	136.505582	170.013644	4.761519
Italy	0.000000	25.196598	55.185625	75.481897	4.668983
Japan	0.000000	37.630363	77.163656	102.967380	4.649745
Korea.S.	0.000000	120.904983	304.829431	329.854620	3.289204
Malaysia	0.000000	79.714054	186.013963	221.061290	3.785195
Netherlands	0.000000	26.832761	51.625813	71.541627	5.084056
Norway	0.000000	42.243851	89.012090	115.059003	4.472349
Portugal	0.000000	61.177624	137.681039	165.689683	3.984860
Singapore	0.000000	36.379685	77.600590	98.124375	4.251472
Spain	0.000000	56.694545	109.533768	138.259224	4.517153
Sweden	0.000000	77.618384	143.420049	178.983373	4.546640
Taiwan	0.000000	41.162560	96.283628	115.186585	3.914719
Thailand	0.000000	78.250621	254.590254	267.508143	3.032038
UK	0.000000	26.693076	55.937248	80.678494	4.952843
USA	0.000000	40.678242	79.617362	100.901197	4.636661

Fig. 3 illustrates the estimated parameters of the SGH distribution for the indices constructed on the basis of 35 world industry sectors as constituents. Here we used the observed log-returns of the MCI35s, DWI35s, EWI35s and WSI35s for the estimation of the log-return distribution. Similarly, as for the case of regional stock index based indices as constituents, we obtain estimates of $\bar{\alpha}$ which are close to zero and values for λ which range from approximately -1.0 to -2.5. In the case of these sector based indices we observe an even better fit of the Student-t distribution. This is, in particular, visible in Fig. 3 for the EWI35s and the WSI35s. In these two cases, only the log-returns denominated in the Brazilian real do not exhibit a proper Student-t fit, as the estimate of $\bar{\alpha}$ is not zero in this case but is approximately equal to 0.25. This is probably a consequence of the short length of our data series on the Brazilian exchange rate. It is obvious that data sets of sufficient length are necessary in order to obtain a proper fit to the underlying true distribution.

In Fig. 4 we analyzed log-returns of the MCI104, DWI104s, EWI104s and WSI104s, based on 104 world industry sector indices as constituents. We again obtain estimates for the parameters $\bar{\alpha}$ and λ, which indicate a good Student-t fit to the observed log-returns of all four indices considered. We note that the improved diversification of the indices in Fig. 4 did not greatly improve the Student-t fit when compared with Fig. 3. The best fits are here again obtained for the EWI104s and the WSI104s.

In conclusion, in all three figures the estimated $(\bar{\alpha}, \lambda)$ points are localized near the negative λ-axis which indicates an approximate Student-t density. Moreover, one can notice that the Student-t density represents a better fit for two types of the world industry sector based indices: it fits very well in the case of the EWI35s as well as the EWI104s. These fits are remarkable and constitute a stylized empirical fact.

In the second column of Table 3 we show only six digits for the Student-t test statistics, which is sufficient to decide whether these values are less than the 99.9% quantile 0.000002 of the chi-square distribution with one degree of freedom. We emphasize that the estimated degrees of freedom of the Student-t density obtained are in the range of around 3 to 5, with a concentration around 4, as can be concluded from the last column of Table 3.

One possible explanation of the above documented facts is given by the Minimal Market Model (MMM) introduced in Platen (2001) and further described in Platen & Heath (2006). The MMM models the discounted GOP by a time transformed squared Bessel process of dimension four. The squared volatility of this process is the inverse of a square root process, which has as its stationary density a gamma density with four degrees of freedom. Therefore, the mixing density for the variance of the returns of the GOP is that of the inverse of a gamma distributed random variable. Consequently, log-returns generated by the MMM, when estimated over a sufficiently long time period, appear to be Student-t distributed with four degrees of freedom.

5 Conclusions

This study has formally documented the empirical fact that log-returns of diversified world stock indices over a sufficiently long observation period appear to be Student-t distributed with about four degrees of freedom. This feature has been identified in a robust manner for a variety of diversified stock indices. Moreover, the most diversified indices, such as the equally weighted index, seem to approximate the growth optimal portfolio best. This is observable in their excellent long-term performance. Specifically, it appears that the EWI104s constructed from 104 world industry sector indices as constituents exhibits a large long-term growth rate. At the same time its log-returns fit at an extremely high level of significance the Student-t distribution with about four degrees of freedom. This makes this index an excellent benchmark for fund management and numeraire in derivative pricing using the benchmark approach.

Acknowledgements

The authors acknowledge the use of daily data provided by Thomson Financial. Furthermore, they would like to express their gratitude to Wolfgang Breymann, David Lüthi and Kevin Fergusson who made routines for the maximum likelihood estimation available. These routines, written in R, have been placed on the first author's website together with an extended version of this paper containing a complete set of statistical results for all indices and currency denominations. Finally, the authors would like to thank Joe Gani for valuable comments on the paper.

References

Abramowitz, M. & I. A. Stegun (Eds.) (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York.

Barndorff-Nielsen, O. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scand. J. Statist.* **5**, 151–157.

Barndorff-Nielsen, O. (1995). Normal-Inverse Gaussian processes and the modelling of stock returns. Technical Report 300, University of Aarhus.

Basle (1996). *Amendment to the Capital Accord to Incorporate Market Risks.* Basle Committee on Banking and Supervision, Basle, Switzerland.

Black, F. & M. Scholes (1973). The pricing of options and corporate liabilities. *J. Political Economy* **81**, 637–654.

Blattberg, R. C. & N. Gonedes (1974). A comparison of the stable and Student distributions as statistical models for stock prices. *J. Business* **47**, 244–280.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* **41**, 135–159.

Eberlein, E. & U. Keller (1995). Hyperbolic distributions in finance. *Bernoulli* **1**, 281–299.

Fergusson, K. & E. Platen (2006). On the distributional characterization of log-returns of a world stock index. *Appl. Math. Finance* **13**(1), 19–38.

Fernholz, E. R. (2002). *Stochastic Portfolio Theory*, Volume 48 of *Appl. Math.* Springer.

Filipoviċ, D. Z. & E. Platen (2007). Consistent market extensions under the benchmark approach. Technical Report QFRC 189, University of Technology, Sydney.

Hurst, S. R. & E. Platen (1997). The marginal distributions of returns and volatility. In Y. Dodge (Ed.), L_1*-Statistical Procedures and Related Topics*, Volume 31 of *IMS Lecture Notes - Monograph Series*, pp. 301–314. Institute of Mathematical Statistics Hayward, California.

Kelly, J. R. (1956). A new interpretation of information rate. *Bell Syst. Techn. J.* **35**, 917–926.

Kelly, L., E. Platen, & M. Sørensen (2004). Estimation for discretely observed diffusions using transform functions. In *Stochastic Methods and Their Applications*, Volume 41A of *J. Appl. Probab.*, pp. 99–118. Applied Prob. Trust.

Küchler, U., K. Neumann, M. Sørensen, & A. Streller (1999). Stock returns and hyperbolic distributions. *Math. Comput. Modelling* **29**, 1–15.

Le, T. & E. Platen (2006). Approximating the growth optimal portfolio with a diversified world stock index. Technical report, University of Technology, Sydney. QFRC Research Paper 184, to appear in *Journal of Risk Finance*.

Madan, D. & E. Seneta (1990). The variance gamma (V.G.) model for share market returns. *J. Business* **63**, 511–524.

Mandelbrot, B. (1963). The variation of certain speculative prices. *J. Business* **36**, 394–419. Reprinted in Cootner (1964), Chapter 15, 307-337.

Markowitz, H. & N. Usmen (1996a). The likelihood of various stock market return distributions, Part 1: Principles of inference. *J. Risk* & *Uncertainty* **13**(3), 207–219.

Markowitz, H. & N. Usmen (1996b). The likelihood of various stock market return distributions, Part 2: Empirical results. *J. Risk* & *Uncertainty* **13**(3), 221–247.

McNeil, A., R. Frey, & P. Embrechts (2005). *Quantitative Risk Management.* Princeton University Press.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica* **41**, 867–888.

Platen, E. (2001). A minimal financial market model. In *Trends in Mathematics*, pp. 293–301. Birkhäuser.

Platen, E. (2002). Arbitrage in continuous complete markets. *Adv. in Appl. Probab.* **34**(3), 540–558.

Platen, E. (2004). Modeling the volatility and expected value of a diversified world index. *Int. J. Theor. Appl. Finance* **7**(4), 511–529.

Platen, E. (2005). Diversified portfolios with jumps in a benchmark framework. *Asia-Pacific Financial Markets* **11**(1), 1–22.

Platen, E. & D. Heath (2006). *A Benchmark Approach to Quantitative Finance.* Springer Finance. Springer.

Platen, E. & G. Stahl (2003). A structure for general and specific market risk. *Computational Statistics* **18**(3), 355 – 373.

Praetz, P. D. (1972). The distribution of share price changes. *J. Business* **45**, 49–55.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications* (2nd ed.). Wiley, New York.

Samuelson, P. A. (1957). Intertemporal price equilibrium: A prologue to the theory of speculation. *Weltwirtschaftliches Archiv* **79**, 181–221.