# FINANCIAL DATA ANALYSIS WITH TWO SYMMETRIC DISTRIBUTIONS

BY

WERNER HÜRLIMANN

## ABSTRACT

The normal inverted gamma mixture or generalized Student t and the symmetric double Weibull, as well as their logarithmic counterparts, are proposed for modeling some loss distributions in non-life insurance and daily index return distributions in financial markets. For three specific data sets, the overall goodness-of-fit from these models, as measured simultaneously by the negative log-likelihood, chi-square and minimum distance statistics, is found to be superior to that of various "good" competitive models including the log-normal, the Burr, and the symmetric $a$-stable distribution. Furthermore, the study justifies on a statistical basis different important models of financial returns like the model of Black-Scholes (1973), the log-Laplace model of Hürlimann (1995), the normal mixture by Praetz (1972), the symmetric $a$-stable model by Mandelbrot (1963) and Fama (1965), and the recent double Weibull as limiting geometric-multiplication stable scheme in Mittnik and Rachev (1993). As an application, the prediction of one-year index returns from daily index returns is discussed.

## KEYWORDS

Claim size data, financial market data, index return, normal inverted gamma mixture, generalized Student t, symmetric double Weibull, goodness-of-fit.

## 1. INTRODUCTION

The fitting of probability distributions to financial data is a statistical subject with a long tradition in both actuarial and financial literature. The detailed analysis of the available models leads to many unsolved problems of theoretical and practical importance, and this field of research always generates new challenges. The present contribution is a further piece of this big puzzle.

The proposed models belong, after appropriate transformation, to the class of symmetric distributions. Let us argue in favor of such a seemingly severe

restriction. First of all, applying an adequate transform $T(X)$ to a random variable $X$ often reveals approximate symmetry in the sense that

$$T(X) = \mu + c \cdot Z, \tag{1.1}$$

where $(\mu, c)$ are *location* and *scale* parameters, and $Z$ is a *symmetric* random variable with mean zero. The ubiquitous transform in this respect is the logarithmic transform $T(X) = \ln(X)$. Another motivation for considering symmetric distributions for $Z$ in (1.1) is the desire to measure the departure from a normal random variable. Besides its practical appeal, the latter working hypothesis finds some theoretical foundation (e.g. Efron (1982)). Empirical arguments are also available. Important financial data for which this approach has been considered adequate include in particular daily returns in equity markets (e.g. Taylor (1992), p. 45). Furthermore, the logarithm of non-life claim sizes has often a low skewness, and can therefore be modeled using the device (1.1). A short outline of our study follows.

Section 2 presents the method applied to determine the unknown parameters and the goodness-of-fit statistics used to assess the overall fit of an estimated distribution. Sections 3 and 4 introduce the proposed symmetric distributions. The required formulas to do all calculations for the comparative distributions used in our study are summarized in the Appendix. The results of our extensive data analysis are exposed in Sections 5 and 6. Finally, to illustrate the potential use of the proposed models, we show in Section 7 how one-year index returns can be predicted from the distributions of daily index returns.

## 2. Estimation method and goodness-of-fit statistics

Given a restriction to two and three parameter distributions, the distribution of $Z$ in (1.1) is either parameter-free or contains one *shape* parameter. The location and scale parameters $\mu$ and $c$ in (1.1) are throughout estimated with the maximum likelihood method. The theoretical justification of this procedure lies in asymptotic statistics, and is explained in many of the modern statistical textbooks. A recent unification result about the maximum likelihood estimation of location and scale parameters is presented in Hürlimann (1998a). A remaining shape parameter $\alpha$ is either included in a three parameter maximum likelihood estimation or it is treated as nuisance parameter. In the latter case, it is chosen to minimize individually or simultaneously some of the goodness-of-fit statistics presented below. Maximum likelihood estimators are denoted $\hat{\mu}, \hat{c}, \hat{\alpha}$. The value of a nuisance parameter is simply denoted by $\alpha$ (without a "hat"). The estimation procedure for the shape parameter is motivated as follows. As our experience has shown, a simultaneous three parameter maximum likelihood estimation often causes numerical difficulties, and does not always lead to an overall best fit. The latter point is illustrated in the text with the NIG ranked 5 in Table 6.3.

In the practical analysis, it is assumed that the data sets consisting of $n$ observations are grouped into $m$ classes with boundaries $\xi_0, \xi_1, \ldots, \xi_m$. The

only available information are the frequencies $\lambda_1$, $\lambda_2$, ..., $\lambda_m$ of the corresponding classes $(\xi_0, \xi_1]$, $(\xi_1, \xi_2]$, ..., $(\xi_{m-1}, \xi_m]$. Financial results are supposed to occur at the average values in $(\xi_{i-1}, \xi_i]$, say at $m\xi_i$, i=1, ..., $m$. Often, in particular in case the average value is not known, we set by convention $m\xi_i = \frac{1}{2}(\xi_{i-1} + \xi_i)$, $i = 1$, ..., $m$. Suppose that the data are observations from a random variable $X$ with a survival distribution $S(x) = S_x(x; \theta)$, $\theta = (\theta_1, ..., \theta_p)$ an unknown parameter vector, and that the data are truncated at $\xi_0$. Of interest is thus the truncated random variable $X_0 = (X | X > \xi_0)$ with survival distribution

$$S_0(x) = \begin{cases} 1, & x \le \xi_0, \\ \dfrac{S(x)}{S(\xi_0)}, & x > \xi_0. \end{cases} \tag{2.1}$$

In case the data is not truncated at $\xi_0$, we assume that $S(\xi_0) = 1$ and our subsequent analysis remains valid with $S_0(x) = S(x)$.

The quality of fit of the various models will be measured using 5 goodness-of-fit statistics, the first 3 of which have a well-known theoretical justification. The other 2 ad hoc statistics are based on reliability measures, and have been used and motivated by some actuaries (see e.g. Hogg and Klugman (1984), pp. 108-111). To assess the *overall goodness-of-fit*, a decision under multiple criteria is necessary. Our simple overall rank is based on the first 3 theoretical criteria. A fitted distribution is ranked before another one if two of the negative log-likelihood, chi-square and minimum distance statistics have a smaller value. One should emphasize that the defined criterion is merely another decision rule, which helps to select good models. It cannot replace a *formal statistical test* like the chi-square goodness-of-fit test or the Kolmogorov-Smirnov test for the ultimate validation or rejection of a model (consult Klugman et al. (1998), Section 2.9, for further discussion on this). In particular, any informal decision rule is necessarily a subjective judgement, which may lead to inappropriate conclusions. A significant illustration of this phenomenon is provided in Section 6. We do not include the other empirical measures in our overall goodness-of-fit criterion. There are two reasons for this. First, our examples show that the LE- and ME-statistics defined below are quite sensitive to changes in parameter values. Second, it is possible to find low LE- and ME-values even if the 3 theoretical criteria do not attain at all their minimal values (e.g. the lnNIG ranked 5 in Table 5.3, the NIG ranked 5 and the log-normal in Table 6.3). A decision including the LE- and ME-statistics appears thus *inconsistent* with our estimation method, at least with respect to the negative log-likelihood and chi-square criteria.

## 2.1. The negative log-likelihood statistic

The *negative log-likelihood* of $X_0$ reads

$$-\ln L = n \cdot \ln S(\xi_0) - \sum_{k=1}^{m} \lambda_k \cdot \ln\left[ S(\xi_{k-1}) - S(\xi_k) \right], \tag{2.2}$$

and the goal is to minimize this quantity. This is achieved through application of the *scoring method* (e.g. Hogg and Klugman (1984), chap. 3.7 and 4.3, or Klugman et al. (1998)). Define

$$P_i = P_i(\theta) = \frac{S(\xi_{i-1}) - S(\xi_i)}{S(\xi_0)}, \quad i = 1, ..., m, \tag{2.3}$$

and consider the information matrix $A = A(\theta)$ with elements

$$a_{rs} = n \cdot \sum_{i=1}^{m} \frac{\partial P_i}{\partial \theta_r} \cdot \frac{\partial P_i}{\partial \theta_s} \cdot \frac{1}{P_i}, \quad r, s = 1, ..., p, \tag{2.4}$$

and the score vector $S = S(\theta)$ with elements

$$S_r = \frac{\partial \ln L}{\partial \theta_r} = \sum_{i=1}^{m} \lambda_i \cdot \frac{\partial P_i}{\partial \theta_r} \cdot \frac{1}{P_i}, \quad r = 1, ..., p. \tag{2.5}$$

Given a preliminary estimate $\theta_0$, then an iterative method to get the maximum likelihood estimate $\hat{\theta}$ of $\theta$ is described by the recursion

$$\theta_k = \theta_{k-1} + A(\theta_{k-1})^{-1} \cdot S(\theta_{k-1}), \quad k = 1, 2, ... \tag{2.6}$$

In case this sequence converges to $\hat{\theta}$, insertion in (2.2) will yield a numerical approximation to the desired minimum value of $-\ln L$. Even if the sequence does not converge, it is possible to obtain with this method estimates $\hat{\theta}$ with a comparatively small practical value of $-\ln L$.

## 2.2. The chi-square statistic

With grouped data the quality of fit is often measured using *Pearson's goodness-of-fit statistic*

$$\chi^2 = \chi^2(\theta) = \sum_{i=1}^{m} \frac{(\lambda_i - nP_i(\theta))^2}{nP_i(\theta)}, \text{ with} \tag{2.7}$$

A comparatively small value $\chi^2$ of is an indicator of an acceptable fit according to the following elegant theory (e.g. Hogg and Klugman (1984), p. 107). Suppose $\theta_{\min}$ solves the minimization problem

$$\chi^2_{\min} = \chi^2(\theta_{\min}) = \min_{\theta} \left[ \chi^2(\theta) \right]. \tag{2.8}$$

Then the statistic $\chi^2_{\min}$ has an approximate chi-square distribution with $m - 1 - p$ degrees of freedom (e.g. Cramér (1946), Fisz (1973), p. 512-513). If $\chi^2_{\min}$ is sufficiently small, then one accepts $S_0(x; \theta_{\min})$ as a reasonable model. However,

if $\chi^2_{\min} \geq z_\alpha$, where $\Pr(\chi^2(m-1-p) \geq z_\alpha) = \alpha$, then one rejects the model $S_0(x;$ $\theta_{\min})$ at the $\alpha$ significance level. Even if the minimum chi-square estimate $\theta_{\min}$ is actually too tough to be calculated, the $\chi^2$-statistic is very useful. For example, if the maximum likelihood estimate or another estimate $\hat{\theta}$ is substituted into $\chi^2(\theta)$ instead of $\theta_{\min}$, then $\chi^2(\hat{\theta}) \geq \chi^2_{\min}$. Therefore, using $\chi^2(\hat{\theta})$ instead of $\chi^2_{\min}$, a model will be rejected at a somewhat larger significance level as that required. Another justification for this substitution is the fact that the maximum likelihood and the minimum chi-square estimators are *asymptotically equal* in case the same class boundaries $\xi_0, \xi_1, \ldots, \xi_m$ are used (e.g. Cramér (1946)).

## 2.3. The minimum distance statistic

With grouped data, another important measure of the quality of fit is the *weighted Cramér-von Mises statistic*

$$K = \sum_{i=1}^{m} \frac{m}{F_0(\xi_i) \cdot S_0(\xi_i)} \cdot \left[ F_i - F_0(\xi_i) \right]^2, \qquad (2.9)$$

with $F_i = \frac{1}{n} \sum_{j=1}^{i} \lambda_j$, $i = 1, \ldots, m$, the empirical distribution function (e.g. Hogg and Klugman (1984), p. 135). For the "true" parameter vector $\theta$, each term has a chi-square distribution with one degree of freedom, which justifies this statistic for empirical testing. Though substitution of an estimate $\hat{\theta}$ for $\theta$ will destroy the chi-square property, the K-statistic is an appealing measure. Each term makes an equal contribution to the total, and the weights $w(\xi_i) = m \cdot \left[ F_0(\xi_i) \cdot S_0(\xi_i) \right]^{-1}$ are largest at the ends of the distribution. In particular, the K-statistic is useful for the analysis of long-tailed data.

## 2.4. The mean excess distance statistic

Consider the *mean excess function* of $X_0 = (X | X > \xi_0)$, that is

$$e(x) = E \left[ X_0 - x | X_0 > x \right] = \frac{\pi_X(x)}{S_X(x)}, \quad x > \xi_0 \text{ ,where} \qquad (2.10)$$

$$\pi_X(x) = \int_x^\infty S_X(t)\,dt \qquad (2.11)$$

is the stop-loss transform of $X$, and its empirical counterpart

$$\hat{e}(\xi_i) = \frac{\sum_{j=i+1}^{m} (m\xi_j - m\xi_i) \cdot \lambda_j}{n - \sum_{j=1}^{m} \lambda_j}, \quad i = 1, \ldots, m-1. \qquad (2.12)$$

In general, the mean excess plot of (2.12) exhibits an increasing slope for long-tailed data, a constant plot for exponential distributions, and a decreasing slope for short-tailed data. As our financial market data sets have shown, a convex plot may also occur quite frequently. Due to scarce observations in the tails, there may be a large uncertainty about the true behavior of $e(x)$, especially in the tails of the distribution. The best fit in this respect might not always lead to the best actuarial decision (Hogg and Klugman (1984), chap. 4). Despite of these and other shortcomings (lack in sampling distribution theory, see however Carriere (1992)), it seems useful to consider the *mean excess distance statistic*

$$ME = \sum_{i=1}^{m-1} \left[ \frac{e(\xi_i) - \hat{e}(\xi_i)}{e(\xi_i)} \right]^2, \tag{2.13}$$

which should be as small as possible for a good fit.

## 2.5. The limited expected value distance statistic

In some situations it is impossible to calculate the mean excess function, for example when the mean of $X$ does not exist, or it is impossible to compute the empirical mean excess function, for example when the data are censored. It is then useful to consider the *limited expected value function* of $X_0$ at $x$ defined as the mean of $X_0$ censored at $x$ through the expression

$$LE(x) = E\left[\min(X_0, x\right] = \xi_0 + \frac{\int_{\xi_0}^{x} S(t)dt}{S(\xi_0)} \tag{2.14}$$

If the mean excess function exists, one has the relationship

$$LE(x) = \xi_0 + e(\xi_0) - S_0(x) \cdot e(x). \tag{2.15}$$

The empirical counterpart of (2.15) is (e.g. Hogg and Klugman (1984), p. 151)

$$L\hat{E}(\xi_i) = \frac{1}{n} \sum_{j=1}^{i} \lambda_j \cdot m\xi_j + \xi_i \cdot [1 - F_1], \quad i = 1, ..., m. \tag{2.16}$$

As a goodness-of-fit measure one uses the *limited expected value distance statistic*

$$LE = \sum_{i=1}^{m} \left[ \frac{LE(\xi_i) - L\hat{E}(\xi_i)}{LE(\xi_i)} \right]^2, \tag{2.17}$$

which should be as small as possible.

### 3. The normal inverted gamma mixture or generalized Student t distribution.

The following distribution has been proposed to model financial returns by Praetz (1972) (see also Blattberg and Gonedes (1974), Kon (1984), Taylor (1992), Section 2.8). Its potential usefulness in actuarial science has been pointed out in Hürlimann (1995a).

If $(X|\theta)$ is conditional on $\theta$ normally distributed with mean $\mu$ and variance $1/\theta$, and $\theta$ follows a conjugate gamma prior $\Gamma(\frac{1}{2}c^2, a)$, $a > 0$, then $X$ has the unconditional density (e.g. Hogg and Klugman (1984), p. 52-53, Heilmann (1989), example 3.7):

$$f_X(x) = \frac{\Gamma(a + \frac{1}{2})}{\sqrt{\pi} \cdot \Gamma(a)} \cdot \frac{1}{c} \left[ \frac{c^2}{c^2 + (x - \mu)^2} \right]^{a + \frac{1}{2}}, \tag{3.1}$$

with $\Gamma(x)$ the gamma function. We say that $X$ has a normal inverted gamma mixture with parameters $\mu$, $c$, $a$, abbreviated $NIG(\mu, c, a)$. The location-scale transform $Z = \frac{X - \mu}{c}$ has a Pearson type VII density (e.g. Johnson et al. (1995), Section 28.6)

$$f_z(z) = \frac{1}{B(a, \frac{1}{2}) \cdot (1 + z^2)^{a + \frac{1}{2}}}, \tag{3.2}$$

where $B(a, b) = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a + b)}$ is a beta coefficient. This can be viewed as a generalized Student t distribution because if $a = \frac{v}{2}$, $v = 1, 2, 3, \ldots$ is an integer, the random variable $\sqrt{v} \cdot Z$ has a Student t with $v$ degrees of freedom. In particular, $a = \frac{1}{2}$ is a Cauchy and $a = 1$ is a Bowers distribution (for the latter see Hürlimann (1993/95a/97/98b) among others). The substitution $t = \frac{z^2}{1 + z^2}$ shows the integral identity

$$\int_0^x \frac{dz}{(1 + z^2)^{a + \frac{1}{2}}} = \frac{1}{2} \int_0^{\frac{x^2}{1 + x^2}} t^{-\frac{1}{2}} (1 - t)^{a - 1} dt, \tag{3.3}$$

from which it follows that the survival distribution satisfies the equivalent expressions

$$S_z(z) = \begin{cases} \frac{1}{2} \left[ 1 - \beta \left( \frac{1}{2}, a; \frac{z^2}{1 + z^2} \right) \right], & z \geq 0, \\ \frac{1}{2} \left[ 1 + \beta \left( \frac{1}{2}, a; \frac{z^2}{1 + z^2} \right) \right], & z \leq 0, \end{cases} \tag{3.4}$$

and

$$S_z(z) = \begin{cases} \frac{1}{2} \beta \left( \frac{1}{2}, a; \frac{1}{1 + z^2} \right), & z \geq 0, \\ 1 - \frac{1}{2} \beta \left( \frac{1}{2}, a; \frac{1}{1 + z^2} \right), & z \leq 0, \end{cases} \tag{3.5}$$

where $\beta(a,b;x) = 1 - \beta(b,a;1-x) = \dfrac{1}{B(a,b)} \displaystyle\int_0^x t^{a-1}(1-t)^{b-1}\,dt$ is a beta density. While the mean and skewness of $Z$ are zero, the variance (if $a>1$) and kurtosis (if $a>2$) are given by

$$\sigma_z^2 = Var[Z] = \frac{1}{2(a-1)}, \quad \gamma_{2,z} = \frac{E[Z^4]}{Var[Z]^2} = 3 \cdot \left[\frac{a-1}{a-2}\right]. \tag{3.6}$$

The kurtos is takes values in $[3, \infty)$, and is therefore capable to model leptokurtic data. The scoring method for maximum likelihood estimation requires the knowledge of the partial derivatives of the survival distribution found in the Appendix.

The stop-loss transform of this statistical model reads (trivial exercise)

$$\pi_X(d) = \frac{c^2 + (d-\mu)^2}{2a-1} \cdot f_X(d) - (d-\mu) \cdot S_X(d). \tag{3.7}$$

Setting $t_a = \sqrt{2(a-1)} \cdot Z$ if $a>1$, a standardized $NIG(0,1,a)$, one gets with $c = \sigma \cdot \sqrt{2(a-1)}$, $\sigma = \sigma_X$, that

$$\pi_X(d) = \frac{2(a-1)\sigma^2 + (d-\mu)^2}{2a-1} \cdot \frac{1}{\sigma} f_{t_a}\left[\frac{d-\mu}{\sigma}\right] - (d-\mu) \cdot S_{t_a}\left[\frac{d-\mu}{\sigma}\right]. \tag{3.8}$$

As $a \to \infty$ one knows that $t_a$ is a standardized normal random variable, hence (see also Hürlimann (1995a) for a special case)

$$\lim_{a\to\infty} \pi_X(d) = \sigma \cdot \varphi\left[\frac{d-\mu}{\sigma}\right] - (d-\mu) \cdot \overline{\Phi}\left[\frac{d-\mu}{\sigma}\right], \tag{3.9}$$

with $\Phi(x)$ the standard normal distribution, $\overline{\Phi}(x) = 1 - \Phi(x)$ and $\varphi(x) = \Phi'(x)$. This is the stop-loss transform of a normal $N(\mu, \sigma)$ random variable. Using that the stop-loss transform uniquely determines the distribution function (e.g. Gerber (1979), Müller (1996), Hürlimann (2000)), one sees that a $NIG(\mu, c, a)$ is asymptotically normally distributed as $a \to \infty$.

Of interest is also the logarithmic version of the above. The random variable $X$ such that $\ln(X) = \mu + c \cdot Z$, with $Z$ a $NIG(0, 1, a)$ random variable, defines the *logarithmic normal inverted gamma mixture*, abbreviated $lnNIG(\mu, c, a)$. Its density and survival distribution are given by

$$f_X(x) = \frac{1}{cx} f_Z\left[\frac{\ln(x)-\mu}{c}\right], \quad S_X(x) = S_Z\left[\frac{\ln(x)-\mu}{c}\right]. \tag{3.10}$$

Since $E[e^{cz}] = \infty$ the mean and stop-loss transform do not exist. However, since $NIG(0, 1, a)$ converges asymptotically as $a \to \infty$ to a normal distribution, the $lnNIG(\mu, c, a)$ is a valuable alternative to the log-normal. This situation occurs in case the mixture is used as a Bayesian prediction model as in

Hürlimann (1995a). Indeed, if $D_n = (x_1, ..., x_n)$ is a sample of $n$ observations from $X$, then the up-dated parameters of the predictive distribution $\ln NIG(\mu, c_n, a_n)$ are (see Hürlimann (1995a), (2.2))

$$c_n = \sqrt{c^2 + \sum_{i=1}^{n} [\ln(x_i) - \mu]^2}, \ a_n = a + \frac{n}{2}. \tag{3.11}$$

For $n$ sufficiently large, the predictive distribution will be very close to a lognormal.

## 4. The double Weibull distribution

Recently the Weibull distribution has received much attention in the modeling of financial returns (e.g. Mittnik and Rachev (1993)). This is due both to its theoretical capability to model the complexity of financial market data as well as its competitiveness in empirical fitting.

Applying the general location-scale transform $T(X) = \mu + \sigma \cdot Z$ with symmetric $Z$ about zero, we are interested in the standardized double Weibull distribution with parameter $a > 0$, abbreviated $SDW(a)$, whose density and survival distribution are given by

$$f_Z(z) = \tfrac{1}{2} a \lambda_a |z|^{a-1} \exp(-\lambda_a |z|^a), \ \lambda_a = \Gamma(1 + \tfrac{2}{a})^{\frac{\alpha}{2}}, \tag{4.1}$$

$$S_Z(z) = \begin{cases} 1 - \tfrac{1}{2}\exp(-\lambda_a |z|^a), \ z \le 0, \\ \tfrac{1}{2}\exp(-\lambda_a |z|^a), \ z \ge 0, \end{cases} \tag{4.2}$$

Observe that the value of the parameter $\lambda_a$ is chosen such that the variance is one. The special case $a = 1, \lambda_1 = \sqrt{2}$ defines the Laplace distribution, which plays a central role in the geometric-multiplication stable scheme in Mittnik and Rachev (1993). It appears also as limiting case of the simple logarithmic modified double exponential model of financial returns in Hürlimann (1995b). The skewness of $Z$ is clearly zero, and the kurtosis is

$$\gamma_{2,Z} = E\left[Z^4\right] = \frac{\Gamma(1 + \tfrac{4}{a})}{\Gamma(1 + \tfrac{2}{a})^2}. \tag{4.3}$$

Since $\gamma_{2,Z} \in [1, 6]$ for $a \ge 1$ and $\gamma_{2,Z} \in [6, \infty)$ for $a \le 1$, this distribution covers the whole range of practical kurtosis values.

The logarithmic version of the above distribution is also considered. The random variable $X$ such that $\ln(X) = \mu + \sigma \cdot Z$, with $Z$ a $SDW(a)$, defines the *logarithmic double Weibull distribution* with parameters $\mu, \sigma, a$, abbreviated $\ln DW(\mu, \sigma, a)$. It is a simple alternative to the log-normal model with density and survival distribution

$$f_X(x) = \frac{1}{\sigma x} f_Z\left[\frac{\ln(x) - \mu}{\sigma}\right], \ S_X(x) = S_Z\left[\frac{\ln(x) - \mu}{\sigma}\right]. \tag{4.4}$$

Though the stop-loss transform cannot be expressed in closed form, it can be evaluated using series representations.

**Proposition 4.1.** The stop-loss transform exists if $a > 1$, or if $a = 1$ and $\sigma < \lambda_1 = \sqrt{2}$, and is given by

$$\pi_X(x) = \begin{cases} \frac{1}{2} e^\mu \cdot \left[ I\left(0; \sigma \lambda_a^{-\frac{1}{a}}\right) + J\left(\lambda_a (\frac{\mu - \ln x}{\sigma})^a; \sigma \lambda_a^{-\frac{1}{a}}\right) \right] - x S_X(x), & x \le e^\mu, \\ \frac{1}{2} e^\mu \cdot I\left(\lambda_a (\frac{\ln x - \mu}{\sigma})^a; \sigma \lambda_a^{-\frac{1}{a}}\right) - x S_X(x), & x \ge e^\mu, \end{cases} \tag{4.5}$$

where $I(x; y)$ and $J(x; y)$ are the infinite series

$$I(x; y) = \sum_{k=0}^{\infty} \Gamma\left(1 + \tfrac{k}{a}\right) \cdot \left[1 - \Gamma\left(1 + \tfrac{k}{a}; x\right)\right] \cdot \frac{y^k}{k!}, \tag{4.6}$$

$$J(x; y) = \sum_{k=0}^{\infty} \Gamma\left(1 + \tfrac{k}{a}\right) \cdot \Gamma\left(1 + \tfrac{k}{a}; x\right) \cdot (-1)^k \frac{y^k}{k!}, \tag{4.7}$$

and $\Gamma(\beta; x)$ is a gamma distribution with shape parameter $\beta$.

**Proof.** Consider first the case $x \ge e^\mu$. With the substitution $z = \left(\frac{\ln x - \mu}{\sigma}\right)^a$ one obtains

$$\pi_X(x) = \frac{1}{2} \cdot \int_x^\infty \exp\left[-\lambda_a (\tfrac{\ln x - \mu}{\sigma})^a\right] dx = \frac{1}{2} e^\mu \cdot \int_{(\frac{\ln x - \mu}{\sigma})^a}^\infty \frac{d}{dz}\left[e^{\sigma z^{\frac{1}{a}}}\right] \cdot e^{-\lambda_a z} dz.$$

With a partial integration one gets

$$\pi_X(x) = \frac{1}{2} e^\mu \int_{\lambda_a (\frac{\ln x - \mu}{\sigma})^a}^\infty \exp\left[\sigma \lambda_a^{-\frac{1}{a}} u^{\frac{1}{a}}\right] \cdot e^{-u} du - x S_X(x).$$

The expressions in (4.5) and (4.6) follow by noting that

$$I(x; y) = \int_x^\infty \exp\left[y \cdot u^{\frac{1}{a}}\right] \cdot e^{-u} du = \sum_{k=0}^{\infty} \frac{y^k}{k!} \int_x^\infty u^{\frac{k}{a}} \cdot e^{-u} du.$$

If $x \le e^\mu$ decompose $\pi_X(x) = \pi_X(e^\mu) + A(x)$ with

$$A(x) = \int_x^{e^\mu} S(x) dx = e^\mu - x - \frac{1}{2} \cdot \int_x^{e^\mu} \exp\left[-\lambda_a (\tfrac{\mu - \ln x}{\sigma})^a\right] dx.$$

Proceed as above to get

$$A(x) = e^\mu - x - \frac{1}{2} e^\mu \cdot \int_0^{(\frac{\mu - \ln x}{\sigma})^a} \frac{d}{dz}\left[e^{-\sigma z^{\frac{1}{a}}}\right] \cdot e^{-\lambda_a z} dz$$

$$= e^\mu - x + x[1 - S(x)] - \frac{1}{2} e^\mu + \frac{1}{2} e^\mu \cdot \int_0^{\lambda_a (\frac{\mu - \ln x}{\sigma})} \exp\left[-\sigma \lambda_a^{-\frac{1}{a}} u^{\frac{1}{a}}\right] \cdot e^{-u} du.$$

A series expansion of the last integral uses that

$$J(x;y) = \int_0^x \exp\left[-y \cdot u^{\frac{1}{a}}\right] \cdot e^{-u} \, du = \sum_{k=0}^{\infty} (-1)^k \frac{y^k}{k!} \cdot \int_0^x u^{\frac{k}{a}} \cdot e^{-u} \, du,$$

which is (4.7). Inserting further the expression $\pi_X(e^\mu) = \frac{1}{2} e^\mu \cdot \left[I\left(0; \sigma \lambda_a^{-\frac{1}{a}}\right) - 1\right]$, one obtains (4.5).                                                                                         ◊

To obtain expressions for the moments, first note that the mean equals (if $a > 1$, or if $a = 1$ and $\sigma < \lambda_1 = \sqrt{2}$)

$$E[X] = \pi_X(0) = \frac{1}{2} e^\mu \cdot \left[I\left(0; \sigma \lambda_a^{-\frac{1}{a}}\right) + J\left(\infty; \sigma \lambda_a^{-\frac{1}{a}}\right)\right] = e^\mu \cdot \sum_{k=0}^{\infty} \Gamma\left(1 + \frac{2k}{a}\right) \cdot \frac{\left(\sigma \lambda_a^{-\frac{1}{a}}\right)^{2k}}{2k!}. \quad (4.8)$$

Since $E[X] = e^\mu \cdot E\left[e^{\sigma Z}\right]$, the series in (4.8) is nothing else than the moment generating function $M_Z(t) = E\left[e^{tZ}\right]$ evaluated at $t = \sigma$. From this observation one gets the higher order moments of $X$ as

$$E\left[X^n\right] = e^{n\mu} \cdot E\left[e^{n\sigma Z}\right] = e^{n\mu} \cdot M_Z(n\sigma) = e^{n\mu} \cdot \sum_{k=0}^{\infty} \Gamma\left(1 + \frac{2k}{a}\right) \cdot \frac{\left(n\sigma \lambda_a^{-\frac{1}{a}}\right)^{2k}}{2k!}, \quad n = 1, 2, \ldots \quad (4.9)$$

In the attractive special case $a = 1$ (and $\sigma < \lambda_1 = \sqrt{2}$) of the log-Laplace distribution, the relevant expressions can be given in closed form. One obtains

$$S_X(x) = \begin{cases} 1 - \frac{1}{2}(e^{-\mu}x)^{\frac{\sqrt{2}}{\sigma}}, & x \le e^\mu, \\ \frac{1}{2}(e^{-\mu}x)^{-\frac{\sqrt{2}}{\sigma}}, & x \ge e^\mu, \end{cases} \quad (4.10)$$

$$\pi_X(x) = \begin{cases} e^\mu \cdot \frac{2}{2-\sigma^2} - x + \frac{1}{2}\frac{\sigma}{\sqrt{2}+\sigma}(e^{-\mu}x)^{\frac{\sqrt{2}}{\sigma}+1}, & x \le e^\mu, \\ \frac{1}{2}\frac{\sigma}{\sqrt{2}-\sigma}(e^{-\mu}x)^{-\left(\frac{\sqrt{2}}{\sigma}-1\right)}, & x \ge e^\mu, \end{cases} \quad (4.11)$$

$$E\left[X^n\right] = e^{n\mu} \cdot \frac{2}{2-(n\sigma)^2}, \quad n < \frac{\sqrt{2}}{\sigma}. \quad (4.12)$$

It is interesting to observe that the log-Laplace has Pareto tails with index $\frac{\sqrt{2}}{\sigma}$, and thus this simple special model is consistent in the tail region with Mandelbrot's Paretian hypothesis for financial returns (see Mandelbrot (1963), Fama (1963/65)). In particular, the mean excess function is linear in the tails and equals

$$e_X(x) = \frac{\pi_X(x)}{S_X(x)} = \frac{\sigma}{\sqrt{2}-\sigma} \cdot x, \quad x \ge e^\mu. \quad (4.13)$$

Since this function is increasing, and in accordance with extreme value theory (e.g. Embrechts et al. (1997)), the log-Laplace is thus susceptible to model long-tailed data. Concerning further properties and motivation, the interested reader is invited to have a look at Hürlimann (1995b).

## 5.  Fitting non-life insurance data

To start with, it appears attractive to test the goodness-of-fit of the proposed models at data sets already examined in the actuarial literature. Our analysis concentrates on the theft claim size data in Hogg and Klugman (1984), Table 4.4, and on the industrial fire insurance claim statistics in Beard et al. (1984), Table 3.5.1.

### 5.1.  Theft loss insurance data

The $n = 32451$ observations are grouped into $m = 18$ classes with boundaries, average losses and frequencies given in Table 5.1.

TABLE 5.1

THEFT LOSS DATA

| i | $\xi_i$ | $m\xi_i$ | $\lambda_i$ |
|---|---------|----------|-------------|
| 0 | 100 | 0 | 0 |
| 1 | 125 | 115 | 583 |
| 2 | 150 | 140 | 1368 |
| 3 | 156 | 154 | 280 |
| 4 | 175 | 166 | 1165 |
| 5 | 200 | 192 | 2082 |
| 6 | 211 | 206 | 631 |
| 7 | 250 | 232 | 2074 |
| 8 | 300 | 277 | 2285 |
| 9 | 350 | 327 | 1990 |
| 10 | 400 | 377 | 1646 |
| 11 | 500 | 452 | 2792 |
| 12 | 600 | 567 | 3271 |
| 13 | 850 | 713 | 4339 |
| 14 | 1100 | 972 | 2379 |
| 15 | 5100 | 1997 | 5181 |
| 16 | 10100 | 6870 | 286 |
| 17 | 25100 | 14354 | 91 |
| 18 | 50100 | 30430 | 8 |

Apart the quite good log-gamma, two parameter distributions do not seem to fit very well the present data. For example, Hogg and Klugman (1984) do not consider the log-normal, Pareto, Weibull and gamma as reasonable choices. The Benktander type I and II (see Benktander and Segerdahl (1960), Benktander (1970), Beard et al. (1984) and Embrechts et al. (1997)) defined uniquely by the mean excess functions

$$e_I(x) = \frac{\beta \cdot (1+x)}{1 + 2a\beta \cdot \ln(1+x)}, \quad e_{II}(x) = \beta \cdot (1+x)^{1-a}, \tag{5.1}$$

are similar unreasonable choices with high values of the goodness-of-fit statistics (see Table 5.3). However, note that the Benktander distributions were introduced to describe the excess losses over some higher threshold (for use in reinsurance) rather than the entire range of losses (for use in direct insurance), which is our main concern in the present study.

Among the three parameter distributions they consider, Hogg and Klugman (1984) found that a Burr provides the best fit with respect to the limited expected value criterion. The results of our parameter estimation are given in Table 5.2 and our goodness-of-fit analysis is summarized in Table 5.3.

Our implementation of the scoring method did not yield in a straightforward way the maximum likelihood estimators $\hat{\mu}$, $\hat{c}$, $\hat{a}$ for the $\ln NIG(\mu, c, a)$. Instead, and for comparisons, several possible fits were made. Each of the four lnNIG minimizes by varying $a$ approximately one of the goodness-of-fit statistics. More precisely, overall rank 2 corresponds to a minimum $\chi^2$, rank 3 to a minimum $K$, rank 4 to a maximum $\ln L$, and rank 5 to a minimum $LE$. Since maximum likelihood estimation for the $\ln DW(\mu, c, a)$ using the scoring method has been successful, such a distinction appears superfluous. The overall ranks in Table 5.2 match those in Table 5.3.

TABLE 5.2

PARAMETER VALUES OF THEFT LOSS DISTRIBUTIONS

| Overall rank | Distribution | Parameter values |
|---|---|---|
| 1 | $\ln DW(\mu, \sigma, a)$ | $\hat{a} = 1.270795$,  $\hat{\mu} = 6.013325$,  $\hat{\sigma} = 1.020931$ |
| 2 | $\ln NIG(\mu, c, a)$ | $a = 5.3$,  $\hat{\mu} = 6.044392$,  $\hat{c} = 2.966822$ |
| 3 | $\ln NIG(\mu, c, a)$ | $a = 5.1$,  $\hat{\mu} = 6.043941$,  $\hat{c} = 2.903233$ |
| 4 | $\ln NIG(\mu, c, a)$ | $a = 5.6$,  $\hat{\mu} = 6.045009$,  $\hat{c} = 3.059974$ |
| 5 | $\ln NIG(\mu, c, a)$ | $a = 4.1$,  $\hat{\mu} = 6.041057$,  $\hat{c} = 2.564565$ |
| 6 | Burr $(a, \beta, \tau)$ | $\hat{\tau} = 1.66932$,  $\hat{a} = 1.09626$,  $\hat{\beta} = 2.6691.02903$ |
| 7 | BenktanderI$(a, \beta)$ | $\hat{a} = 0.00339$,  $\hat{\beta} = 146.813$ |
| 8 | BenktanderII$(a, \beta)$ | $\hat{a} = 0.8745$,  $\hat{\beta} = 364.6117$ |

TABLE 5.3

GOODNESS-OF-FIT OF THEFT LOSS DISTRIBUTIONS

| Overall rank | Distribution | $-\ln L$ | $\chi^2$ | K | $10^3 \cdot LE$ | ME |
|---|---|---|---|---|---|---|
| 1 | lnDW | 83551 | 914 | 0.4702 | 8.94 | 1.85 |
| 2 | lnNIG | 83648.60 | 1142.79 | 0.4362 | 3.00 | – |
| 3 | lnNIG | 83649.20 | 1143.09 | 0.4359 | 2.57 | – |
| 4 | lnNIG | 83648.30 | 1143.33 | 0.4376 | 3.67 | – |
| 5 | lnNIG | 83660.33 | 1156.86 | 0.4480 | 1.29 | – |
| 6 | Burr | 83672 | 1186 | 0.4645 | 1.04 | 0.46 |
| 7 | BenktanderI | 92902 | 18762 | 67.70 | 1193 | 1006 |
| 8 | BenktanderII | 83153 | $12.8 \cdot 10^6$ | 7304.57 | 16.5 | 2.78 |

Though the fitted Burr has the lowest LE- and ME-statistics, it takes the worst overall rank among the three parameter distributions. It is also beaten by a non-optimal lnNIG with a similar LE-value. The preferred distribution is a lnDW. As the $\chi^2$-values are rather high, a formal chi-square test, which would validate one or several of the models, is not undertaken. However, as demonstrated in Section 6, such a validation is sometimes possible.

## 5.2. Industrial fire loss data

The $n = 8324$ observations are grouped into $m = 29$ classes with boundaries and frequencies given in Table 5.4. The obtained parameter values and goodness-of-fit of four distributions are summarized in Table 5.5 and Table 5.6. In this situation maximum likelihood estimation using the scoring method has been successful, and a further distinction as in Section 5.1 appears akward.

The fitted Burr, whose mean does not exist, is beaten by three distributions and has here the highest LE-value. A simple two parameter log-normal fits better than the Burr. The log-normal is beaten by both the lnDW and the lnNIG. While the lnNIG has the lowest LE-value, the lnDW is the preferred distribution. No formal test is undertaken.

TABLE 5.4

INDUSTRIAL FIRE LOSS DATA

| i | $\xi_i$ | $\lambda_i$ | i | $\xi_i$ | $\lambda_i$ |
|---|---------|-------------|---|---------|-------------|
| 0 | 0 | 0 | 15 | 6310 | 323 |
| 1 | 10 | 283 | 16 | 10000 | 179 |
| 2 | 16 | 280 | 17 | 15849 | 173 |
| 3 | 25 | 157 | 18 | 25119 | 112 |
| 4 | 40 | 464 | 19 | 39811 | 94 |
| 5 | 63 | 710 | 20 | 63096 | 57 |
| 6 | 100 | 781 | 21 | 100000 | 39 |
| 7 | 158 | 530 | 22 | 158489 | 22 |
| 8 | 251 | 446 | 23 | 251189 | 17 |
| 9 | 398 | 491 | 24 | 398107 | 12 |
| 10 | 631 | 673 | 25 | 630957 | 5 |
| 11 | 1000 | 779 | 26 | 1000000 | 5 |
| 12 | 1585 | 741 | 27 | 1584890 | 3 |
| 13 | 2512 | 520 | 28 | 2511890 | 1 |
| 14 | 3981 | 425 | 29 | 6309570 | 2 |

TABLE 5.5

PARAMETER VALUES OF FIRE LOSS DISTRIBUTIONS

| Overall rank | Distribution | Parameter values |
|---|---|---|
| 1 | $\ln DW(\mu, \sigma, a)$ | $\hat{a} = 1.41561$,   $\hat{\mu} = 5.79645$,   $\hat{\sigma} = 2.16935$ |
| 2 | $\ln NIG(\mu, c, a)$ | $\hat{a} = 25.7746$,   $\hat{\mu} = 5.89543$,   $\hat{c} = 15.23085$ |
| 3 | $\ln N(\mu, \sigma)$ | $\hat{\mu} = 5.90396$,   $\hat{\sigma} = 2.15982$ |
| 4 | $\mathrm{Burr}(a, \beta, \tau)$ | $\hat{\tau} = 0.80607$,   $\hat{a} = 0.98114$,   $\hat{\beta} = 110.35718$ |

TABLE 5.6

GOODNESS-OF-FIT OF FIRE LOSS DISTRIBUTIONS

| Overall rank | Distribution | $-\ln L$ | $\chi^2$ | K | LE | ME |
|---|---|---|---|---|---|---|
| 1 | lnDW | 24128 | 442 | 2.46 | 3.83 | 3.83 |
| 2 | lnNIG | 24213 | 637 | 1.86 | 2.53 | – |
| 3 | lnN | 24216 | 663 | 2.33 | 3.60 | 7.63 |
| 4 | Burr | 24281 | 758 | 1.71 | 9.09 | – |

## 6. FITTING FINANCIAL MARKET DATA

The distribution of the daily cumulative returns on a stock market index has been the subject of many past and current investigations. It is thus of great importance to look at the overall goodness-of-fit of the proposed models when compared with "good" competitors like the log-normal (justified by the model of Black and Scholes (1973)) and the symmetric $a$-stable distribution (justified by the work of Mandelbrot (1963), Fama (1965), and Peters (1994)). Our analysis is based on the SMI (Swiss Market Index) daily cumulative returns between September 29, 1998 and September 24, 1999.

The $n = 250$ observations are grouped into $m = 26$ classes with boundaries and frequencies given in Table 6.1. The parameter estimation is provided in Table 6.2 and the goodness-of-fit in Table 6.3. For the sake of comparisons, we distinguish between two NIG fits. The NIG minimizes approximately $\chi^2$ by varying $a$ and maximum likelihood estimation of $\mu, c$ while the NIG$_2$ uses maximum likelihood estimation of $\mu, c, a$.

The fitted log-normal, with high $\chi^2$- and K-values, seems unreasonable at first sight. The relative low LE- and ME-values in this example, which are quite smaller than the corresponding values of the two overall best fitted distributions, illustrate the apparent irrelevance of the LE- and ME-criteria (see however the comments to Figure 6.1). Also, the NIG$_2$, whose three parameters have been estimated with the maximum likelihood method, has high $\chi^2$- and K-values and takes only overall rank 5. By the way, it has the lowest LE-value and a quite small ME-value. The other four distributions seem to fit quite well.

It is remarkable that the two-parameter log-Laplace is not significantly beaten by the lnDW. The preferred SMI distributions are a NIG followed by a symmetric $a$-stable distribution, abbreviated *SaS*, whose characterisitcs are summarized in the Appendix. The ranked 3 lnDW and the log-Laplace have smaller LE- and ME-values.

A quick look at the graphs of the empirical and fitted mean excess functions in Figure 6.1 is very instructive (similar observations hold for the non-life insurance data sets, an analysis which can be left to the reader). The behavior of the empirical graph is quite erratic in the right tail. The simplest fit for this is anticipated by a parabola or more generally a convex curve. In contrast to this, in non-life insurance, an increasing and concave curve fitting has been proposed, at least in the right tail (see Benktander and Segerdahl (1960), and Benktander (1970) on this point). The two best fitting distributions distinguish themselves from the others by a considerable slope in the right tail (in accordance with extreme value theory). This is the reason for the high ME- and LE-values of these fitted distributions. A suggestion for future work might be the definition of more adequate *weighted* LE- and ME-statistics, which take this phenomenon into account. This could perhaps also allow these statistics to enter into an extended goodness-of-fit test.

TABLE 6.1

SMI DAILY CUMULATIVE RETURNS

| i | $\xi_i$ | $\lambda_i$ | i | $\xi_i$ | $\lambda_i$ |
|---|---|---|---|---|---|
| 0 | 0.950 | 0 | 14 | 1.020 | 22 |
| 1 | 0.955 | 1 | 15 | 1.025 | 8 |
| 2 | 0.960 | 4 | 16 | 1.030 | 3 |
| 3 | 0.965 | 0 | 17 | 1.035 | 2 |
| 4 | 0.970 | 1 | 18 | 1.040 | 1 |
| 5 | 0.975 | 1 | 19 | 1.045 | 0 |
| 6 | 0.980 | 11 | 20 | 1.050 | 0 |
| 7 | 0.985 | 14 | 21 | 1.055 | 0 |
| 8 | 0.990 | 15 | 22 | 1.060 | 1 |
| 9 | 0.995 | 43 | 23 | 1.065 | 0 |
| 10 | 1.000 | 31 | 24 | 1.070 | 0 |
| 11 | 1.005 | 36 | 25 | 1.075 | 0 |
| 12 | 1.010 | 25 | 26 | 1.080 | 1 |
| 13 | 1.015 | 30 | | | |

TABLE 6.2

PARAMETER VALUES OF SMI DISTRIBUTIONS

| Overall rank | Distribution | Parameter values |
|---|---|---|
| 1 | $NIG_1(\mu, c, a)$ | $a = 1.875,\quad \hat{\mu} = 1.00063587,\quad \hat{c} = 0.022579$ |
| 2 | $SaS(\mu, c, a)$ | $a = 1.8,\quad \hat{\mu} = 1.0006,\quad \hat{c} = 0.01012$ |
| 3 | $\ln DW(\mu, \sigma, a)$ | $a = 1.005,\quad \hat{\mu} = 0.00038779,\quad \hat{\sigma} = 0.016489$ |
| 4 | $\ln Laplace(\mu, \sigma)$ | $\hat{\mu} = 0.00038277,\quad \hat{\sigma} = 0.016547$ |
| 5 | $NIG_2(\mu, c, a)$ | $\hat{a} = 6.71936,\quad \hat{\mu} = 1.00063727,\quad \hat{c} = 0.049471$ |
| 6 | $\ln N(\mu, \sigma)$ | $\hat{\mu} = 0.00058706,\quad \hat{\sigma} = 0.0151181$ |

TABLE 6.3

GOODNESS-OF-FIT OF SMI DISTRIBUTIONS

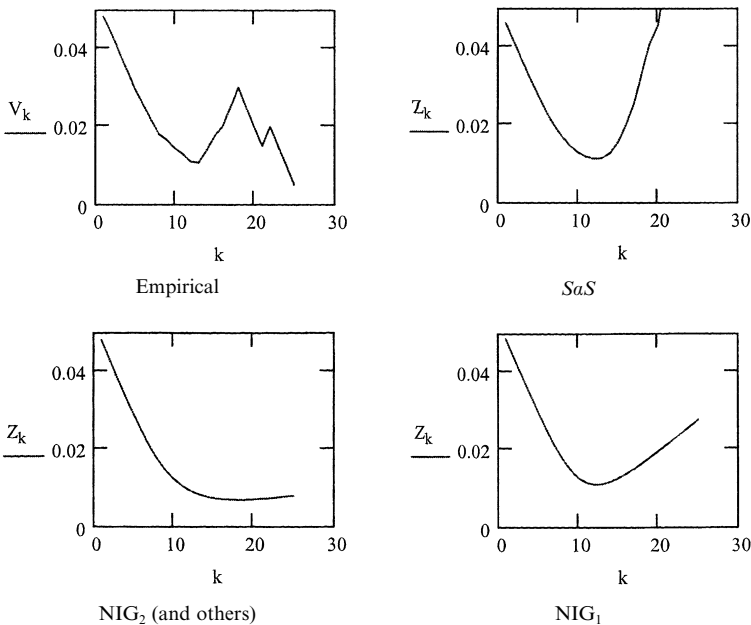| Overall rank | Distribution | $-\ln L$ | $\chi^2$ | K | $10^7 \cdot LE$ | ME |
|---|---|---|---|---|---|---|
| 1 | $NIG_1$ | 623.47 | 46.51 | 2.30 | 18.00 | 24.60 |
| 2 | $Sas$ | 623.14 | 49.53 | 2.73 | 12.77 | 303 |
| 3 | $\ln DW$ | 629.40 | 58.25 | 3.35 | 5.85 | 3.90 |
| 4 | $\ln Laplace$ | 629.64 | 58.26 | 3.36 | 6.11 | 3.97 |
| 5 | $NIG_2$ | 625.64 | 241.47 | 36.51 | 2.86 | 4.42 |
| 6 | $\ln N$ | 633.07 | 4490 | 485 | 5.96 | 5.87 |



FIGURE 6.1: graphs of mean excess functions

An ultimate validation and selection among the above models must necessarily be based on a formal statistical test and consider other alternative decision rules for selection. To perform a correct formal chi-square test, the raw data in Table 6.1 must be grouped in a different way. According to Moore (1978/86) a number of rules are to be fulfilled (e.g. Klugman et al. (1998), p. 121). Recommended is an expected frequency of at least 1% in each class and a 5% expected frequency in 80% of the classes. In view of this, the raw data is regrouped as in Table 6.4.

TABLE 6.4

SMI DAILY CUMULATIVE RETURNS FOR CHI-SQUARE TEST

| i | $\xi_i$ | $\lambda_i$ |
|---|---|---|
| 0 | 0.95 | 0 |
| 1 | 0.97 | 6 |
| 2 | 0.985 | 26 |
| 3 | 1.00 | 89 |
| 4 | 1.015 | 91 |
| 5 | 1.03 | 33 |
| 6 | 1.08 | 5 |

Based on the parameter values in Table 6.2, the up-dated goodness-of-fit statistics, together with the p-value of the test, are found in Table 6.5, which order the distributions according to the new overall rank.

TABLE 6.5

GOODNESS-OF-FIT UNDER CHI-SQUARE TEST

| Overall rank | Distribution | $-\ln L$ | $\chi^2$ | p-value | K | $10^7 \cdot LE$ | ME |
|---|---|---|---|---|---|---|---|
| 1 | $NIG_2$ | 351.87 | 0.75 | 0.69 | 0.014 | 7.76075 | 0.9382 |
| 2 | $SaS$ | 352.85 | 1.34 | 0.51 | 0.056 | 5.82119 | 0.5710 |
| 3 | $NIG_1$ | 353.28 | 2.63 | 0.27 | 0.055 | 11.14234 | 0.6414 |
| 4 | lnN | 353.11 | 3.08 | 0.38 | 0.061 | 2.67507 | 0.9077 |
| 5 | lnDW | 354.33 | 4.85 | 0.09 | 0.082 | 5.44844 | 0.6345 |
| 6 | lnLaplace | 354.42 | 5.00 | 0.17 | 0.084 | 5.45984 | 0.6275 |

With critical values of 5.99 (by 2 degrees of freedom for 3 parameters) and 7.82 (by 3 degrees of freedom for 2 parameters) for a 5% significance level, it is remarkable that all models are validated through this test. In view of the high $\chi^2$- and K-values in Table 6.3, this could not be expected a priori for the $NIG_2$ and the lnN. This shows once more that informal decision rules are to be applied very carefully. The dramatic change in the proposed overall ranking coincides exactly with the brute $\chi^2$-ranking and almost with the negative likelihood and K-ranking. The LE- and ME-values behave still quite erratically.

Since there are different number of parameters, the p-value ranking differs. This is a reason for considering further alternative selection rules.

If parsimony is a concern, the best three parameter $NIG_2$ needs to be compared with the log-normal and log-Laplace. For this it is usual to apply a likelihood ratio test (though this is here only an informal decision rule). The test statistics are $2 \cdot (353.11 - 351.87) = 1.24$ (lnN versus $NIG_2$) and $2 \cdot (354.42 - 351.87) = 2.55$ (lnLaplace versus $NIG_2$). With one degree of freedom, the critical value is 3.84, and the log-normal and log-Laplace are selected first. A perhaps more appropriate alternative to this informal hypothesis test is a penalized likelihood scoring method, called Schwartz Bayesian Criterion (SBC) and introduced by Schwartz (1978) (see Klugman et al. (1998)). To the negative likelihood one adds the penalty $p \cdot [\ln(n) - \ln(2\pi)]$, where $p$ is the number of estimated parameters and $n$ is the sample size, to obtain the SBC-score, which decides upon ranking. The result of this SBC selection is reported in Table 6.6.

TABLE 6.6

SBC RANKING OF SMI DISTRIBUTIONS

| SBC rank | Distribution | -lnL | penalty | SBC score |
|:---:|:---|:---:|:---:|:---:|
| 1 | lnN | 353.11 | 7.37 | 360.48 |
| 2 | lnLaplace | 354.42 | 7.37 | 361.79 |
| 3 | $NIG_2$ | 351.87 | 11.05 | 362.92 |
| 4 | $SaS$ | 352.85 | 11.05 | 363.90 |
| 5 | $NIG_1$ | 353.28 | 11.05 | 364.73 |
| 6 | lnDW | 354.33 | 11.05 | 365.38 |

To the knowledge of the author, the above study should be a unique first one, which justifies statistically several different and important models of financial returns motivated through financial economic and other principles. It places Black-Scholes (1973) model at the top rank, and justifies also the simple limiting log-Laplace model in Hürlimann (1995b). Furthermore, it does not reject other good alternative choices like the normal inverted gamma mixture by Praetz (1972) (used by J.P. Morgan Stanley), the prominent symmetric $a$-stable distribution by Mandelbrot (1963) and Fama (1965), and the more recent double Weibull as geometric-multiplication stable scheme in Mittnik and Rachev (1993).

## 7. On the prediction of one-year index returns from daily index returns

To conclude the present study with a practical illustration of the results in Section 6, it is interesting to compare the actual SMI index of 6966 at the end of the observation period with the SMI index resulting from a fitted distribution under a *strict white noise assumption* (independent and identically distributed

daily returns). The obtained 250 days return for the fitted lnN, lnLaplace, $NIG_2$, $SaS$, $NIG_1$ and lnDW are respectively 19.19%, 13.94%, 17.26%, 16.18%, 17.22% and 14.05%. The average between the two extremes is 16.565%, which is quite close to the observed $100 \cdot \left( \frac{6966}{6020} - 1 \right) = 15.71\%$. The range of variation for the SMI index at the end of the period is $[7017 \pm 158]$, where the midpoint is quite close to the actual index of 6966. The closest value to the actual index is 6994 for the $SaS$. Whether these extremely good fits are mere coincidence or of a deeper nature requires further investigations. Used as naive prediction value, if the index performs similarly in the next period, then the SMI index at the end of September 2000 should stay between 7945 and 8166. This corresponds approximately to the expected SMI index of 8200 at the end of year 2000 as predicted by a model of the Credit Suisse First Boston (see Tages-Anzeiger (1999)).

The above naive calculation is done under the strict white noise assumption, which is easy to test and refute from a pure statistical point of view (e.g. Taylor (1992), p. 19). However, it yields an acceptable value of the one-year return from a pure investment point of view. Must the independence hypothesis be rejected or can it be used for the present purpose? This well-known dilemma has been noted and studied in detail by Fama (1965), which states:

> "Dependence that is important from the trader's point of view need not be important from a statistical point of view, and conversely dependence which is important for statistical purposes need not be important for investment purposes."

Recall that Fama's tests did not reveal any evidence of important dependence from either an investment or a statistical point of view.

There exist some more formal mathematical calculations, which can justify the prediction of a one-year index return based on the distribution of the daily index returns. Under the made strict white noise assumption and for the distributions of Table 6.6 (except the $SaS$, for which more complex calculations are required), we have computed the four main characteristics of a distribution, namely the mean, standard deviation, skewness and kurtosis. For the $NIG(\mu, c, a)$ distributions, we have additionally calculated these characteristics for the Bayesian prediction models $NIG(\mu, c_n, a_n)$ with up-dated parameters $c_n = \sqrt{c^2 + \sum_{i=1}^{n} [x_1 - \mu]^2}$, $a_n = a + \frac{n}{2}$, as well as for the normal approximation to this prediction model as $a \to \infty$, where these models have been discussed in Section 3. The obtained results are summarized in Table 7.2. The required formulas for the first four moments $m_k, k = 1, 2, 3, 4$, are straightforward and listed below in Table 7.1 for the convenience of the reader. The one-year return corresponds here to $T = 250$ days. The mean of the one-year return is then $\mu_T = m_1$, the standard deviation is $\sigma_T = \sqrt{m_2 - m_1^2}$, while the skewness $\gamma_T$ and kurtosis $\gamma_{2,T}$ are calculated using the formulas

$$\gamma_T = \frac{m_3 - 3m_1 m_2 + 2m_1^3}{\sigma_T^3} \tag{7.1}$$

$$\gamma_{2,T} = \frac{m_4 - 4m_1 m_3 + 6m_1^2 m_2 - 3m_1^4}{\sigma_T^4} \tag{7.2}$$

TABLE 7.1

MOMENT FORMULAS FOR ONE-YEAR INDEX RETURN PREDICTION MODELS

$\underline{\ln N(\mu, \sigma)}$

$m_k = \exp\left[k\mu T + \tfrac{1}{2}(k\sigma)^2 T\right], \quad k = 1, 2, 3, 4$

$\underline{\ln Laplace(\mu, \sigma)}$

$m_k = \dfrac{\exp(k\mu T)}{1 - \tfrac{1}{2}(k\sigma)^2 T}, \quad k = 1, 2, 3, 4$

$\underline{\ln DW(\mu, \sigma, a)}$

$m_k = \exp(k\mu T) \cdot \displaystyle\sum_{j=0}^{\infty} \frac{\Gamma\left(1 + \frac{2j}{a}\right)}{(2j)!} \cdot \left[(k\sigma)^2 \frac{T}{\Gamma\left(1 + \frac{2}{a}\right)}\right]^2, \quad k = 1, 2, 3, 4 \quad \text{(this is formula (4.11))}$

$\underline{NIG(\mu, c, a)}$

$m_1 = \mu^T, m_2 = \left(\mu^2 + \tfrac{1}{2}\frac{c^2}{a-1}\right)^T, \ m_3 = \left(\mu^3 + \tfrac{3}{2}\frac{\mu c^2}{a-1}\right)^T, \ m_4 = \left(\mu^4 + 3\frac{\mu^2 c^2}{a-1} + \tfrac{3}{4}\frac{c^4}{(a-1)(a-2)}\right)^T, \ a > 2$

$\underline{BNIG(\mu, c_n, a_n)}$ (Bayesian NIG prediction model)

The same formulas as for the NIG hold with $c$, $a$ replaced by the up-dated parameters $c_n$, $a_n$ calculated with the $n = 250$ daily observations.

$\underline{NNIG(\mu, \sigma_a)}$ (normal approximation to Bayesian NIG prediction model)

$\sigma_a = \dfrac{c_n}{\sqrt{2(a_n - 1)}}$

$m_1 = \mu^T, \ m_2 = (\mu^2 + \sigma_a^2)^T, \ m_3 = (\mu^3 + 3\mu\sigma_a^2)^T, \ m_4 = (\mu^4 + 6\mu^2\sigma_a^2 + 3\sigma_a^4)^T$

TABLE 7.2

ONE-YEAR RETURN PREDICTION FROM DAILY RETURN DISTRIBUTIONS

| Distribution | $\mu_T$ | $\sigma_T$ | $\gamma_T$ | $\gamma_{2,T}$ |
|---|---|---|---|---|
| LnN | 19.19% | 29.03% | 0.75 | 4.003 |
| LnLaplace | 13.94% | 32.36% | 2.61 | 30.717 |
| LnDW | 14.05% | 32.21% | 2.54 | 28.686 |
| $NIG_1$ | 17.22% | 32.20% | 0.84 | 4.144 |
| $BNIG_1$ | 17.22% | 28.70% | 0.75 | 4.004 |
| $NNIG_1$ | 17.22% | 28.70% | 0.75 | 4.004 |
| $NIG_2$ | 17.27% | 27.47% | 0.71 | 3.918 |
| $BNIG_2$ | 17.27% | 28.63% | 0.74 | 3.998 |
| $NNIG_2$ | 17.27% | 28.63% | 0.74 | 3.998 |

The comparison of the figures in Table 7.2 are quite instructive. The Bayesian NIG models and their normal approximations have skewness and kurtosis parameters very close to the best SBC ranked log-normal model in Section 6, and these values are also closest to empirical values obtained from long-term one-year returns. The skewness and kurtosis parameters are overestimated by the log-Laplace and log-double Weibull models. From this perspective, only the log-normal and normal inverted gamma mixtures are selected for practical purposes. For a cautious prediction, the Bayesian normal inverted gamma mixture and its normal approximation should be the preferred models for prediction.

**Appendix:** comparative distributions

The formulas for the scoring method and the stop-loss transforms for the evaluation of the LE- and ME-statistics are listed.

### Normal inverted gamma mixture

$$S_X(x) = S_Z\left[\frac{x-\mu}{c}\right], \text{ with defined in } (3.4)$$

$$\frac{\partial}{\partial\mu} S_X(x) = f_X(x), \quad \frac{\partial}{\partial c} S_X(x) = \left[\frac{x-\mu}{c}\right] \cdot f_X(x)$$

$$\frac{\partial}{\partial a} S_X(x) = \begin{cases} \left[\psi\left(a+\frac{1}{2}\right) - \psi(a)\right] \cdot S_X(x) + \frac{1}{4}(2a-1)\cdot\beta\left(a-1,\frac{1}{2};\frac{c^2}{c^2+(x-\mu)^2}\right), & x \geq \mu, \\ -\left[\psi\left(a+\frac{1}{2}\right) - \psi(a)\right] \cdot F_X(x) - \frac{1}{4}(2a-1)\cdot\beta\left(a-1,\frac{1}{2};\frac{c^2}{c^2+(x-\mu)^2}\right), & x \leq \mu, \end{cases}$$

where $\psi(x) = \frac{d}{dx}\ln\Gamma(x)$ is the digamma function or psi function.

$$\pi_X(x) = \frac{c^2+(x-\mu)^2}{2a-1} \cdot f_X(x) - (x-\mu)\cdot S_X(x)$$

### Double Weibull distribution

$$S_X(x) = S_Z\left[\frac{x-\mu}{c}\right], \text{ with } S_Z(z) \text{ defined in } (4.2)$$

$$\frac{\partial}{\partial\mu} S_X(x) = f_X(x), \quad \frac{\partial}{\partial\sigma} S_X(x) = \left[\frac{x-\mu}{\sigma}\right] \cdot f_X(x)$$

$$\frac{\partial}{\partial a} S_X(x) = -\text{sgn}\left[\frac{x-\mu}{\sigma}\right] \cdot \frac{1}{2}\left[\lambda_a' + \lambda_a \ln\left|\frac{x-\mu}{\sigma}\right|\right] \cdot \left|\frac{x-\mu}{\sigma}\right|^a \cdot \exp\left[-\lambda_a\left|\frac{x-\mu}{\sigma}\right|^a\right],$$

$$\lambda_a' = -\frac{1}{a}\Gamma\left(1+\frac{2}{a}\right)^{\frac{a}{2}} \cdot \psi\left(1+\frac{2}{a}\right)$$

$$\pi_X(x) = \sigma \cdot \pi_Z\left(\frac{x-\mu}{\sigma}\right), \text{ with } \pi_Z(d) = \frac{1}{2}\lambda_a^{-\frac{1}{a}}\Gamma\left(1+\frac{1}{a}\right)\cdot\left[1 - \Gamma\left(1+\frac{1}{a};\lambda_a|d|^a\right)\right] - d\cdot S_Z(d),$$

where $\Gamma(\beta; x)$ is a gamma distribution with shape parameter $\beta$.

### BenktanderI

$$S(x) = [1 + 2a\beta \ln(1+x)] \cdot \exp\left\{-\left(\tfrac{\beta+1}{\beta}\right)\ln(1+x) - a \cdot [\ln(1+x)]^2\right\}$$

$$\frac{\partial S(x)}{\partial a} = \frac{2\beta \cdot \ln(1+x)}{1 + 2a\beta \ln(1+x)} \cdot S(x) - [\ln(1+x)]^2 \cdot S(x)$$

$$\frac{\partial S(x)}{\partial \beta} = \frac{2a \cdot \ln(1+x)}{1 + 2a\beta \ln(1+x)} \cdot S(x) + \frac{1}{\beta^2} \cdot \ln(1+x) \cdot S(x)$$

$$\pi(x) = S(x) \cdot e(x), \ e(x) = \frac{\beta \cdot (1+x)}{1 + 2a\beta \cdot \ln(1+x)}$$

### BenktanderII

$$S(x) = (1+x)^{a-1} \cdot \exp\left\{-\frac{1}{a\beta}\left[(1+x)^a - 1\right]\right\}$$

$$\frac{\partial S(x)}{\partial a} = \left[\frac{1}{a^2\beta}\left[(1+x)^a - 1\right] - \frac{1}{a\beta}\ln(1+x)(1+x)^a + \ln(1+x)\right] \cdot S(x)$$

$$\frac{\partial S(x)}{\partial \beta} = \frac{1}{a\beta^2}\left[(1+x)^a - 1\right] \cdot S(x)$$

$$\pi(x) = S(x) \cdot e(x), \ e(x) = \beta \cdot (1+x)^{1-a}$$

### Burr

$$S(x) = \left[\frac{\beta}{\beta + x^\tau}\right]^a$$

$$\frac{\partial S(x)}{\partial a} = -\ln\left[\frac{\beta}{\beta + x^\tau}\right] \cdot S(x), \ \frac{\partial S(x)}{\partial \beta} = -\frac{ax^\tau}{\beta^2} \cdot S(x)^{\frac{a+1}{a}}, \ \frac{\partial S(x)}{\partial \tau} = -\frac{ax^\tau}{\beta} \cdot \ln(x) \cdot S(x)^{\frac{a+1}{a}}$$

$$\pi(x) = \mu \cdot \left[1 - \beta\left(1+c, a-c; \tfrac{x^\tau}{\beta + x^\tau}\right)\right] - x \cdot S(x), \text{ with the mean}$$

$$\mu = \frac{\beta^c \Gamma(a-c) \cdot \Gamma(1+c)}{\Gamma(a)}, \ a > c = \frac{1}{\tau}$$

$$LE(x) = \xi_0 + \frac{\beta^c}{S(\xi_0) \cdot \tau} \cdot \int_{(1+\beta^{-1}x^\tau)^{-1}}^{(1+\beta^{-1}\xi_0^\tau)^{-1}} y^{a-c-1}(1-y)^{c-1}\,dy, \ a \le c \text{ (case of infinite mean)}$$

### Log-normal

$$S(x) = 1 - N\left(\frac{\ln x - \mu}{\sigma}\right)$$

$$\frac{\partial S(x)}{\partial \mu} = \frac{1}{\sigma} \cdot \Phi\left[\frac{\ln x - \mu}{\sigma}\right], \quad \frac{\partial S(x)}{\partial \sigma} = \frac{\ln x - \mu}{\sigma^2} \cdot \Phi\left[\frac{\ln x - \mu}{\sigma}\right]$$

$$\pi(x) = \exp(\mu + \tfrac{1}{2}\sigma^2) \cdot \left[1 - N\left(\frac{\ln x - \mu}{\sigma}\right) - \sigma\right] - x \cdot S(x)$$

### Symmetric α-stable distribution

Explicit expressions exist only in the special cases $a = 1$ (Cauchy) and $a = 2$ (normal). Bergström (1952) developed series expansions that Fama and Roll (1968/71) and other authors applied in case $a > 1$. The density and distribution of the normalized case $\mu = 0$, $c = 1$ is first stated.

$$f_Z(z) = \begin{cases} \dfrac{1}{\pi a} \cdot \displaystyle\sum_{k=0}^{\infty} (-1)^k \dfrac{\Gamma\left(\frac{2k+1}{a}\right)}{(2k)!} z^{2k}, & |z| \leq 5a - 4, \\[3mm] -\dfrac{1}{\pi} \cdot \displaystyle\sum_{k=0}^{n} (-1)^k \dfrac{\Gamma(ak+1)}{k! z^{ak+1}} \sin\left[\frac{ak\pi}{2}\right] + R(z), & |z| > 5a - 4. \end{cases}$$

The remainder satisfies $|R(z)| < C \cdot z^{-a(n+1)-1}$, $C$ a constant, and becomes smaller than the previous term in the summation as $z$ gets larger.

$$F_Z(z) = \begin{cases} \dfrac{1}{2} + \dfrac{1}{\pi a} \cdot \displaystyle\sum_{k=0}^{\infty} (-1)^{k-1} \dfrac{\Gamma\left(\frac{2k-1}{a}\right)}{(2k-1)!} z^{2k-1}, & |z| \leq 5a - 4, \\[3mm] 1 + \dfrac{1}{\pi} \cdot \displaystyle\sum_{k=0}^{n} (-1)^k \dfrac{\Gamma(ak+1)}{k! z^{a\kappa}} \sin\left[\frac{ak\pi}{2}\right] - \displaystyle\int_z^{\infty} R(t)\,dt, & |z| > 5a - 4. \end{cases}$$

## REFERENCES

BEARD, R.E., PENTIKÄINEN, T. and PESONEN E. (1984) *Risk Theory*. Chapman and Hall.

BENKTANDER, G. (1970) Schadenverteilung nach Grösse in der Nicht-Lebensversicherung. *Bulletin of the Swiss Association of Actuaries*, 263-84.

BENKTANDER, G. and SEGERDAHL C.-O. (1960) On the analytical representation of claim distribution with special reference to excess of loss reinsurance. *Transactions of the International Congress of Actuaries*, 626-46.

BERGSTRÖM, H. (1952) On some expansions of stable distribution functions. *Arkiv för Matematik* **2**, 375-78.

BLACK, F. and SCHOLES M. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy* **81**, 637-59.

BLATTBERG, R.C. and GONEDES N.J. (1974) A comparison of the stable and Student distributions as statistical models for stock prices. *Journal of Business* **47**, 244-80.

CARRIERE, J. (1992) Limited expected value comparison tests. *Statistics and Probability Letters* **15**, 321-27.

CRAMÉR, H. (1946) *Mathematical Methods of Statistics*. Princeton University Press.

EFRON, B. (1982) Transformation theory: how normal is a family of distributions? (The 1981 Wald Memorial Lecture). *The Annals of Statistics* **10**, 323-339.

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH TH. (1997) Modelling Extremal Events for Insurance and Finance. *Applications of Mathematics – Stochastic Modelling and Applied Probability*, vol. 33. Springer.

FAMA, E. (1963) Mandelbrot and the stable Paretian hypothesis. *J. of Business* **36**, 420-29.

FAMA, E. (1965) The behavior of stock market prices. *Journal of Business* **38**, 34-105.

FAMA, E. and ROLL R. (1968) Some properties of symmetric stable distributions. *Journal of the American Statistical Association* **63**, 817-36.

FAMA, E. and ROLL R. (1971) Parameter estimates for symmetric stable distributions. *Journal of the American Statistical Association* **66**, 331-38.

FISZ, M. (1973) *Wahrscheinlichkeitsrechnung und Mathematische Statistik*. VEB Deutscher Verlag der Wissenschaften, Berlin.

GERBER, H.-U. (1979) *An Introduction to Mathematical Risk Theory*. Hübner Foundation. University of Pennsylvania.

HEILMANN, W.-R. (1989) Decision theoretic foundations of credibility theory. *Insurance: Mathematics and Economics* **8**, 77-95.

HOGG, R. and KLUGMAN S. (1984) *Loss Distributions*. John Wiley. New York.

HÜRLIMANN, W. (1993) Solvabilité et réassurance. *Bulletin of the Swiss Association of Actuaries*, 229-49.

HÜRLIMANN, W. (1995a) Predictive stop-loss premiums and Student's t-distribution. *Insurance: Mathematics and Economics* **16**, 151-59.

HÜRLIMANN, W. (1995b) *Is there a rational evidence for an infinite variance asset pricing model?* Proceedings of the 5th International AFIR Colloquium.

HÜRLIMANN, W. (1997) Fonctions extrémales et gain financier. *Elemente der Mathematik* **52**, 152-68.

HÜRLIMANN, W. (1998a) On the characterization of maximum likelihood estimators for location-scale families. *Communications in Statistics – Theory and Methods* **27(2)**, 495-508.

HÜRLIMANN, W. (1998b) *Extremal Moment Methods and Stochastic Orders*. Application in Actuarial Science. Monograph manuscript (available from the author).

HÜRLIMANN, W. (2000) Higher-degree stop-loss transforms and stochastic orders (I) Theory. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik* **XXIV(3)**, 449-63.

JOHNSON, N.L., KOTZ, A. and BALAKRISHNAN N. (1995) *Continuous Univariate Distributions*. (2nd ed.). John Wiley, New York.

KLUGMAN, S., PANJER, H. and WILLMOT G. (1998) *Loss Models. From Data to Decisions*. John Wiley, New York.

KON, S.J. (1984) Models of stock returns – a comparison. *Journal of Finance* **39**, 147-65.

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

MITTNIK, S. and RACHEV S.T. (1993) Modeling asset returns with alternative stable distributions. *Econometric Reviews* **12(3)**, 261-330.

MOORE, D. (1978) Chi-square tests. In Hogg, R. (Ed.). Studies in Statistics, vol. 19. *Mathematical Association of America*, 453-63.

MOORE, D. (1986) *Tests of chi-squared type*. In D'Agostino, R. and M. Stephens (Eds.). Goodness-of-Fit Techniques. Marcel Dekker, New York, 63-95.

MÜLLER, A. (1996) Ordering of risks: a comparative study via stop-loss transforms. *Insurance: Mathematics and Economics* **17**, 215-22.

PETERS, E. (1994) *Fractal Market Analysis*. Wiley Finance Editions.

PRAETZ, P.D. (1972) The distribution of share price changes. *Journal of Business* **45**, 49-65.

SCHWARTZ, G. (1978) Estimating the dimension of a model. *Annals of Statistics* **6**, 461-64.

TAGES-ANZEIGER (1999) *CSFB: Günstiger Ausblick auf 2000*. October 6, 1999, p. 43.

TAYLOR, S.J. (1992) *Modeling Financial Time Series* (3rd reprint). John Wiley.

WERNER HÜRLIMANN
*Value and Risk Management*
*Winterthur Life and Pensions*
*Postfach 300*
*CH-8401 Winterthur*