

A New Class of Probability Distributions and its Application to Finance

CHRISTIAN MENN* AND SVETLOZAR T. RACHEV[†]

June 14, 2004

Christian Menn

Email: menn@statistik.uni-karlsruhe.de

Chair of Statistics, Econometrics and Mathematical Finance

Universität Karlsruhe, Geb. 20.12

D-76128 Karlsruhe, Germany

Fon: +49 (0) 721 608 8114

Fax: +49 (0) 721 608 3811

Svetlozar T. Rachev

Email: rachev@statistik.uni-karlsruhe.de

Chair of Statistics, Econometrics and Mathematical Finance

Universität Karlsruhe, Geb. 20.12

D-76128 Karlsruhe, Germany

and

Department of Statistics and Applied Probability

University of California, Santa Barbara, CA 93106, USA

*Corresponding author. The paper presents extensions and generalisations of my Ph.D. thesis (Menn 2004). The results have partly been presented in talks at University of California, Santa Barbara, University of Washington, Seattle, Hochschule für Banken, Frankfurt and at the EU-workshop on mathematical optimization models held in November 2003 in Cyprus. I gratefully acknowledge the helpful discussions and suggestions made by participants at these events.

[†]Rachev gratefully acknowledges research support by grants from Division of Mathematical, Life and Physical Sciences, College of Letters and Science, University of California, Santa Barbara, the Deutschen Forschungsgemeinschaft and the Deutscher Akademischer Austausch Dienst.

Stock market crashes like those in October '87 and October '97, the turbulent period around the Asian Crisis in 1998 through 1999 or the burst of the “dotcom bubble” together with the extreme volatile period after September 11, 2001 steadily remind financial engineers and risk managers how often extreme events actually happen in the reality of financial markets. These observations have led to increased effort to improve the flexibility and statistical reliability of existing models to describe the dynamics of economic variables. The history of probabilistic modelling of economic variables and especially price processes by means of stochastic processes goes back to Bachelier (1900) who suggested the Brownian Motion as a candidate to describe the evolution of stock markets. It took another 70 years until Black and Scholes (1973) and Merton (1973) used the Geometric Brownian Motion to describe the stock price movements in their famous solution of the option pricing problem. Their Nobel prize winning work inspired the foundation of the arbitrage pricing theory based on the martingale approach described in Harrison and Kreps (1979) and subsequently in Harrison and Pliska (1981). The key observation that pricing of derivatives has to be effected under a measure Q (usually denoted as risk neutral or equivalent martingale measure) which differs from the data generating “market measure” lead to an increasing literature on what we will call “implicit models”. Examples for implicit models include stochastic volatility models (see e.g. Heston (1993), Hull and White (1987)), local volatility models (see e.g. Derman and Kani (1994), Dupire (1994)), martingale models for the short rate and implicit volatility models. The joint characteristic inherent to implicit models is that the model parameters are determined through calibration on market prices of derivatives directly under the martingale measure and not through estimation from observations under the market measure. A direct consequence and simultaneously the main drawback of the calibration framework is that the market prices cannot be explained: Prices of liquid market instruments are used for the calibration procedure and consequently reproduced more or less perfectly but on the set of exotic derivatives the prices of the divers models differ substantially. Moreover from an objective viewpoint, there is no way to determine which pricing model is the most reliable one given that the statistical fit to historical realisations of the underlying isn't taken into account.

The opposite approach to model price processes is pursued by econometricians. The task is to provide the highest possible accuracy to the empirical observations or in other words to meet the statistical characteristics of financial data. This sentence already implies that the focus lays on the statistical properties of historical realisations and aspects of derivative pricing are neglected. Most econometric approaches neither present any risk neutral price processes nor the defined markets are checked for absence of arbitrage. Beside this criticism, econometric research led to a well founded knowledge of the statistical characteristics of financial data. Since the seminal work of Fama (1965) and Mandelbrot (1963a, b) it was frequently reported by various authors (see e.g. Mittnik and Rachev (2000) for an extensive overview) and is widely accepted among researchers that return distributions are left skewed, peaked around the mean and heavy tailed. Additionally, in longitudinal sections of return distributions one observes volatility clustering, i.e. calm periods are followed by highly volatile periods. Engle (1982) and Bollerslev (1986) introduced tractable time series models, denoted as ARCH and GARCH models, which gave raise to the explanation of the observed heteroscedasticity. In the subsequent years various generalisations and variants of the original models have been published. The overview of Duan (1997) introduces a general treatment of the different variants and examines their diffusion limits. Although members of the ARCH/GARCH-class generate stochastic processes with heavy tailed marginals, the results of applying such processes to the option pricing problem are disappointing in various aspects such as the defective statistical fit and the inability to explain prices of liquid derivatives. The main reason for the missing statistical reliability can be ascribed to the distribution of the innovation process. The underlying white noise process which can be seen as the driving risk factor is traditionally modelled with independent standard normal random variables. Various authors like e.g. Mittnik and Rachev (2000) and Carr and Wu (2003) have suggested to replace the normal by the stable distribution. The class of stable distributions forms an ideal alternative to the normal distributions which combines the stability property with modelling flexibility. The main drawback of stable non-Gaussian distributions is their infinite variance, a fact that is always mentioned by antagonists of stable models.

These observations motivated our research which tries to combine statistical reliability with market consistent derivative pricing. In other words,

we try to develop a model which leads to acceptable results in both worlds: Under the market measure P , on the one hand, prediction of future scenarios and value at risk considerations play an important role. On the other hand, in the risk neutral world, arbitrage free prices of derivatives are derived. In section I we introduce a new class of probability distributions denoted as smoothly truncated stable distributions (STS-distributions) which combine the modelling flexibility of stable distributions with the existence of arbitrary moments. The STS-distributions are generated by replacing the heavy tails of the stable density left of some lower truncation level and right of some upper truncation level by the densities of two appropriately chosen normal distributions. In subsection A we discuss the main properties and in subsection B we provide an efficient estimation procedure for the parameters of STS-distributions.

In section II we present a general time series model based on a GARCH-process with STS-distributed innovations. This time series model - denoted as STS-GARCH-model - combines properties of the process dynamic as volatility clustering and leverage effect with cross-sectional attributes as heavy tailed and skewed residuals. The model is applied to different data sets of the S&P 500 and Xerox Company. In all cases the model provides an excellent statistical fit which underpins our reasoning about the impact and importance of choosing an appropriate white noise distribution. The explaining power of the presented framework becomes evident in the comparison of the model dependent probabilities for the stock market crash in October '87: whereas this event is classified as impossible under models based on the normal assumption the STS-GARCH framework postulates an average time of occurrence of forty years for such an event.

Beside the positive results of our examinations under the market measure P we examine the results of applying the model to valuation problems, i.e. we examine the model properties under a risk neutral measure Q . In subsection A we show that the model generates market consistent smiles for European call options on the S&P 500. Subsection B reports the results of applying the STS-GARCH-model as firm value model for credit risk valuation. Simulated credit spreads and default probabilities show that the usual criticism against structural approaches is not valid. Section III concludes.

I. Smoothly Truncated α -Stable Distributions

A. Definition and Properties

In the sequel, we introduce a special class of truncated α -stable distributions. This new class of probability distributions is especially designed to meet the needs of economic applications and to provide the required properties for a reliable statistical description of financial time series. It combines the modelling power of stable distributions with the appealing property of finite moments of arbitrary order.

Definition I.1 (STS-distribution). *Let g_θ denote the density of some α -stable distribution with parameter-vector $\theta = (\alpha, \beta, \sigma, \mu)$ and $h_i, i = 1, 2$ denote the densities of two normal distributions with parameters (ν_i, τ_i) . Furthermore, let $a, b \in \mathbb{R}$ be two real numbers with $a \leq \mu \leq b$. The density of a “smoothly truncated stable distribution” (STS-distribution) is defined by:*

$$f(x) = \begin{cases} h_1(x) & \text{for } x < a \\ g_\theta(x) & \text{for } a \leq x \leq b \\ h_2(x) & \text{for } x > b \end{cases} . \quad (1)$$

In order to guarantee a well-defined continuous probability density, the following relations are imposed:

$$h_1(a) = g_\theta(a) \text{ and } h_2(b) = g_\theta(b) \quad (2)$$

and

$$p_1 := \int_{-\infty}^a h_1(x) dx = \int_{-\infty}^a g_\theta(x) dx \text{ and } \int_b^{\infty} h_2(x) dx = \int_b^{\infty} g_\theta(x) dx =: p_2 \quad (3)$$

Definition I.1 characterizes the 6-parameter family of STS-probability distributions, which will be denoted by \mathcal{S}^{tr} . The shortcut “tr” represents “truncated” and reflects the fact that the fat tails of the α -stable distribution were cut off and replaced by the thin tails of two appropriately chosen normal distributions. Elements of \mathcal{S}^{tr} are denoted by $S_\alpha^{[a,b]}(\sigma, \beta, \mu)$. Probability distributions used for modelling “white noise processes” e.g. the innovations of a time series model, are usually assumed to be standardized probability distributions with zero mean and unit variance. In the sequel, the notation

\mathcal{S}_0^{tr} will be used to denote the subclass of standardized STS-distributions, which will play a major role in this article.

The following proposition lists some generic properties of STS-distributions, which are most important for our applications.

Proposition I.2 (Properties of STS-distributions).

1. The parameters (ν_i, τ_i) of the two involved normal distributions are uniquely defined by the two equations (2) and (3) and given by the following expressions:

$$\tau_1 = \frac{\varphi(\Phi^{-1}(p_1))}{g_\theta(a)} \quad \text{and} \quad \nu_1 = a - \tau_1 \Phi^{-1}(p_1) \quad (4)$$

$$\tau_2 = \frac{\varphi(\Phi^{-1}(p_2))}{g_\theta(b)} \quad \text{and} \quad \nu_2 = b + \tau_2 \Phi^{-1}(p_2) \quad (5)$$

where φ denotes the density and Φ the distribution function of the standard normal distribution. p_1 and p_2 denote the ‘‘cut-off-probabilities’’ defined in equation (3) and G_θ is the distribution function of the α -stable distribution with parameter-vector $\theta = (\alpha, \beta, \sigma, \mu)$.

2. A useful property of α -stable distributions is the scale and translation invariance, which is transmitted to the class of STS-distributions: For $c, d \in \mathbb{R}$ and $X \sim S_\alpha^{[a,b]}(\sigma, \beta, \mu)$ we have:

$$Y := cX + d \sim S_{\tilde{\alpha}}^{[\tilde{a}, \tilde{b}]}(\tilde{\sigma}, \tilde{\beta}, \tilde{\mu}) \in \mathcal{S}^{tr} \quad \text{with}$$

$$\tilde{a} = ca + d, \tilde{b} = cb + d, \tilde{\alpha} = \alpha, \tilde{\sigma} = |c|\sigma, \tilde{\beta} = \text{sign}(c)\beta, \tilde{\mu} = c\mu + d$$

3. The mean EX and the second moment EX^2 of a STS-distributed random variable $X \sim S_\alpha^{[a,b]}(\sigma, \beta, \mu)$ is given by:

$$\begin{aligned} EX &= ap_1 - \tau_1 (\Phi^{-1}(p_1)p_1 + \varphi(\Phi^{-1}(p_1))) + \\ &\quad + \int_a^b x g_\theta(x) dx + \\ &\quad + bp_2 + \tau_2 (\Phi^{-1}(p_2)p_2 + \varphi(\Phi^{-1}(p_2))) \end{aligned} \quad (6)$$

$$\begin{aligned} EX^2 &= (\tau_1^2 + \nu_1^2)p_1 - \tau_1(a + \nu_1)\varphi(\Phi^{-1}(p_1)) + \\ &\quad + \int_a^b x^2 g_\theta(x) dx + \\ &\quad + p_2(\nu_2^2 + \tau_2^2) + \tau_2(\nu_2 + b) \cdot \varphi(\Phi^{-1}(p_2)) \end{aligned} \quad (7)$$

where, as above, φ denotes the density and Φ the distribution function of the standard normal distribution. p_1 and p_2 denote the “cut-off-probabilities” defined in equation (3) and G_θ is the distribution function of the α -stable distribution with parameter-vector $\theta = (\alpha, \beta, \sigma, \mu)$.

As there exists no closed form expression for the density g_θ of a stable distribution, the mean and the variance given by equation (6) and (7) of an arbitrary STS-distribution $S_\alpha^{[a,b]}(\sigma, \beta, \mu)$ can only be evaluated with the help of numerical integration. We’d like to stress the fact that, due to modern computer power, the numerical integration needed for the “stable part” does not imply any limitation to our approach. Actually, we even use relations (6) and (7) to effectively calculate the truncation levels a and b of standardized STS-distributions. As extensive numerical experiments have shown that the subclass $\mathcal{S}_0^{tr} \subset \mathcal{S}^{tr}$ of standardized STS-distributions is uniquely defined by the vector of stable parameters $\theta = (\alpha, \beta, \sigma, \mu)$ due to moment matching conditions. In the sequel, we assume that the claimed relation holds true and we always calculate the truncation levels given the four stable parameters such that the resulting distribution is standardized.

It is important to see, that STS-distributions possess finite moments of arbitrary order and that they offer a large modelling-flexibility due to their six describing parameters. Even if STS-distribution possess thin tails, in the mathematical sense, they are powerful in describing the distribution of financial variables which typically admit fat-tails, left-skewness and excess kurtosis. Table I contains a comparison of the left tail probabilities of a standard normal and a standardized STS-distribution. The comparison of the tail-probabilities shows, that the standardized STS-distribution admits excess kurtosis and much heavier tails than the standard normal distribution.

Insert Table I somewhere around here

In Figure 1 and 2 we illustrate the results of a further examination of standardized STS-distributions. The graph in Figure 1 shows the influence of the distribution parameters $(\alpha, \sigma, \beta, \mu)$ on the truncation level a and b . We can see, that the left truncation level a decreases monotonically with increasing α , which follows mathematical intuition. In the two graphs of Figure 2 we have plotted the probability $P((-\infty, a - 5])$ that a standardized STS-distribution with predefined values for σ and μ admits a value less

or equal than the left truncation level a minus 5 in dependence of α and three different values of β . We can see that this probability, which can be interpreted as the probability for extreme events, increases monotonically with increasing α , decreasing β and decreasing σ .

Insert Figure 1 and 2 somewhere around here

B. Parameter Estimation

In practice, efficient and consistent estimation procedures for the parameters of probability distributions are needed. One popular method of estimating parameters even for very complex distributions is provided by the Maximum-Likelihood-framework.

Numerical algorithms provide density approximations of α -stable distributions at every needed accuracy level (for an overview see e.g. Racheva-Iotova and Stoyanov 2003). By the definition of STS-density in equation (1) together with the parameters of the normal tails given by equation (4) and (5) we are able to calculate the density of every STS-distribution for a given parameter-vector $(a, b, \alpha, \beta, \sigma, \mu)$. This enables us to effectuate a numerical Maximum-Likelihood-estimation (ML-estimation). The parameters to be estimated can be divided into two sets: The four stable parameters $(\alpha, \beta, \sigma, \mu)$ on the one hand and the two truncation levels a and b on the other. The stable parameters can be estimated very accurately by the suggested numerical ML-procedure as soon as the data sample becomes large enough¹. It is natural that the truncation levels a and b are difficult to estimate, because the resulting distribution is not very sensitive with respect to small changes in the truncation level. In the general case, for an arbitrary STS-distribution $S_\alpha^{[a,b]}(\sigma, \beta, \mu) \in \mathcal{S}^{tr}$ we have to accept this drawback, but for the special case of standardized STS-distribution we can circumvent this problem and get consistent estimates, also for the truncation levels. According to our remark above, a standardized STS-probability distribution $S_\alpha^{[a,b]}(\sigma, \beta, \mu) \in \mathcal{S}_0^{tr}$ is uniquely specified by the four stable parameters $(\alpha, \beta, \sigma, \mu)$. This means that the task of consistently estimating the distribution parameters $(a, b, \alpha, \beta, \sigma, \mu)$ of standardized STS-distributions

¹We show in the Appendix in the proof of Theorem 1 that the ML-estimates for the stable parameters are consistent.

$S_\alpha^{[a,b]}(\sigma, \beta, \mu) \in \mathcal{S}_0^{tr}$ actually reduces to an consistent estimation of the stable parameter vector, which is provided by the ML-framework. As we have already mentioned in the previous section, the assumption that the data sample is generated by a standardized probability distribution is often fulfilled for practical applications, especially when we are estimating the parameters of a “white noise”-distribution of a time series model.

II. Option pricing and credit risk valuation

The previous section has introduced a powerful class of probability distributions, designed to describe the behaviour of financial risk factors. For a modern approach to financial engineering and risk management respecting econometric considerations, not only is a flexible probability distribution for the risk factors necessary, but a time series model which is able to describe the connection between the unobservable shocks or innovations and the observable state variables such as stock returns, interest rates or exchange rates. Major issues, which the model must account for, are the possibility to describe the effect of volatility clustering, large unpredictable variations caused e.g. by crashes or the leverage effect observed in stock market data. In attempts to overcome the deficiencies of classical time series models but keeping their advantages, we have the following definition:

Definition II.1. *We consider a d -dimensional sequence $(X_t)_{t \in \mathbb{N}}$ with $X_t \in \mathbb{R}^d$ of predictable exogenous variables, parameter-vectors $a \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^{d+1}$, $\alpha \in \mathbb{R}_{\geq 0}^{p+1}$ and $\beta \in \mathbb{R}_{\geq 0}^q$ and a $\mathcal{B}(\mathbb{R}_{>0})$ -measurable function $f : \mathbb{R}_{>0} \rightarrow \mathbb{R}$.*

A generalised ARMAX-GARCH-process with STS-distributed innovations (shortly STS-GARCH-process) is a 1-dimensional stochastic process $(Y_t)_{t \in \mathbb{N}}$ in discrete time over some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{N}}, P)$ with:

$$Y_t = c_0 + \sum_{k=1}^d c_k X_t^{(k)} + \sum_{i=1}^n a_i Y_{t-i} + f(\sigma_t) + \sum_{j=1}^m b_j \sigma_{t-j} \epsilon_{t-j} + \sigma_t \epsilon_t, \quad t \in \mathbb{N}. \quad (8)$$

The innovations are assumed to follow a standardized STS-distribution:

$$\epsilon_t \stackrel{iid}{\sim} S_\alpha^{[a,b]}(\sigma, \beta, \mu) \in \mathcal{S}_0^{tr}. \quad (9)$$

The conditional variance σ_t^2 follows a GARCH(p, q) process of Bollerslev (1986):

$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^p \alpha_k \sigma_{t-k}^2 \epsilon_{t-k}^2 + \sum_{k=1}^q \beta_k \sigma_{t-k}^2, \quad t \in \mathbb{N}, \quad (10)$$

with $\sum_{k=1}^p \alpha_k + \sum_{k=1}^q \beta_k < 1$ in order to ensure covariance-stationarity.

The following proposition lists some of the generic properties of the STS-GARCH processes.

Proposition II.2.

1. The following relations holds for all $t \in \mathbb{N}$:

$$\text{sign}(C(\sigma_t \epsilon_t, \sigma_{t+1}^2)) = \text{sign}(E(\epsilon^3)), \quad (11)$$

Equation (11) will enable us to explain the leverage effect typically inherent to stock market or firm value data.

2. The following items will allow us to consider the process dynamic under a risk neutral measure, which is necessary for any pricing purpose.

- (a) For a predictable sequence $(\lambda_t)_{t \in \mathbb{N}}$ of random variables let Q^{λ_t} denote the probability measure on Ω such that for the distribution of $(\xi_t)_{t \in \mathbb{N}} := (\epsilon_t + \lambda_t)_{t \in \mathbb{N}}$ equals the one of $(\epsilon_t)_{t \in \mathbb{N}}$ under P , i.e. we have the relation:

$$\xi_t \stackrel{iid}{\sim} S_{\alpha}^{[a,b]}(\sigma, \beta, \mu). \quad (12)$$

Then the process $(Y_t)_{t \in \mathbb{N}}$ under Q possesses the following dynamic:

$$\begin{aligned} Y_t = & c_0 + \sum_{k=1}^d c_k X_t^{(k)} + \sum_{i=1}^n a_i Y_{t-i} + f(\sigma_t) + \dots \\ & \dots - \sum_{j=1}^m b_j \sigma_{t-j} \lambda_{t-j} - \lambda_t \sigma_t + \sum_{j=1}^m b_j \sigma_{t-j} \xi_{t-j} + \sigma_t \xi_t \end{aligned} \quad (13)$$

and for the conditional variance we have:

$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^p \alpha_k \sigma_{t-k}^2 (\xi_{t-k} - \lambda_{t-k})^2 + \sum_{k=1}^q \beta_k \sigma_{t-k}^2. \quad (14)$$

(b) Similarly to the results obtained by Duan (1995) in his definition of “Locally Risk Neutral Valuation Relationship”, the conditional variance remains unaffected by the change of measure:

$$V_Q(Y_t | \mathcal{F}_{t-1}) \stackrel{f.s.}{=} V_P(Y_t | \mathcal{F}_{t-1}) \quad (15)$$

(c) The unconditional variance will increase through the change of measure: If the sequence $(\lambda_t)_{1 \leq t \leq T}$ has a constant finite second moment $\lambda^2 := E_Q \lambda_t^2$ which fulfils the condition

$$\lambda^2 < \frac{1 - \sum_{k=1}^p \alpha_k - \sum_{k=1}^q \beta_k}{\sum_{k=1}^p \alpha_k} \quad (16)$$

and additionally the random variables λ_t^2 and σ_t^2 are uncorrelated for $t \in \mathbb{N}$, then the unconditional variance is given by:

$$V_Q(\sigma_t \xi_t) = E_Q \sigma_t^2 = \frac{\alpha_0}{1 - (1 + \lambda^2) \sum_{k=1}^p \alpha_k - \sum_{k=1}^q \beta_k}, \quad 0 \leq t \leq T. \quad (17)$$

3. As we want to estimate the process parameters from observations, we need the likelihood function which corresponds to the model (8). Assume that observations y_1, \dots, y_T of Y and x_1, \dots, x_T of X are given. Then the conditional Likelihood function is given by

$$\tilde{L}_{(x,y)}(a, b, c, \alpha, \beta) = \prod_{t=1}^T \frac{1}{\sigma_t} \varphi_\epsilon \left(\frac{y_t - c_0 - \sum_{k=1}^d c_k x_t^{(k)} - \sum_{i=1}^n a_i y_{t-i} - f(\sigma_t) - \sum_{j=1}^m b_j \sigma_{t-j} \epsilon_{t-j}}{\sigma_t} \right). \quad (18)$$

The conditional Likelihood function depends on the choice of the following starting values

$$Y_0 = y_0, Y_{-1} = y_{-1}, \dots, Y_{1-n} = y_{1-n},$$

$$\epsilon_0, \epsilon_{-1}, \dots, \epsilon_{1-\max\{p,m\}} \text{ and } \sigma_0, \sigma_{-1}, \dots, \sigma_{1-\max\{p,q,m\}}.$$

It's worth noting at this stage, that our approach can be seen as a new framework, which combines consistency requirements implied by the “no-arbitrage-principle” with statistical reliability offered by econometric models and needed by risk managers. The following two subsections will present possible applications for a special STS-GARCH-process.

A. Option pricing and crash probabilities for the S&P 500

The following equation defines a special STS-GARCH-process designed to explain stock, firm value or index returns. Concretely, we assume the following dynamic for the logarithmic S&P 500 index return (for reasons of convenience we denote the S&P 500 from now on simply as stock):

$$\log S_t - \log S_{t-1} = r_t - d_t + \lambda_t \sigma_t - g(\sigma_t) + \sigma_t \epsilon_t, \quad t \in \mathbb{N}, \epsilon_t \stackrel{iid}{\sim} S_\alpha^{[a,b]}(\sigma, \beta, \mu). \quad (19)$$

S_t denotes the level of the S&P 500 ex dividend at date t and r_t (d_t) denotes the continuously compounded risk free rate of return (dividend rate) for the period $[t-1, t]$. The predictable parameter λ_t represents the market price of risk for the period $[t-1, t]$ and g the logarithmic moment generating function of the distribution $S_\alpha^{[a,b]}(\sigma, \beta, \mu)$. The conditional variance σ_t^2 is expected to follow a GARCH(p, q)-process:

$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^p \alpha_k \sigma_{t-k}^2 \epsilon_{t-k}^2 + \sum_{k=1}^q \beta_k \sigma_{t-k}^2, \quad t \in \mathbb{N}, \quad (20)$$

The process dynamic described by equation (19) combines essential needs of modern econometrics: Volatility clustering, leverage effect, skewed and heavy-tailed residuals in combination with finite moments of every order. If we choose the standard normal distribution instead of the STS-distribution, and assume further a constant dividend rate of $d = 0$ and a constant interest rate and market price of risk, then model (19) reduces to Duan's GARCH-model presented in (Duan 1995). We will see in the sequel that the choice of the distribution for the risk factor ϵ has a significant impact on the statistical fit as well as on the derivative pricing.

The process defined in equation (19) has some appealing properties which we list in the following proposition:

Proposition II.3.

1. If we denote by $S_t^{cd} = e^{d_t} \cdot S_t$ the index level “cum dividend” at the end of the trading period $[t-1, t]$ then the expected relative price increment for the stock price cum dividend equals:

$$E \left(\frac{S_t^{cd}}{S_{t-1}} \mid \mathcal{F}_{t-1} \right) = \exp(\underbrace{r_t + \lambda_t \sigma_t}_{=:\mu_t}). \quad (21)$$

This equation motivates the notion “market price of risk” for the parameter λ_t : The expected excess return rate for investing into the risky asset S with reinvesting the dividend, thus the difference between the expected total return rate and the risk free rate is proportional to the amount of risk inherent to the investment and measured by the volatility σ_t . The proportionality factor is given by λ_t .

2. If we change from the objective probability measure P to the equivalent measure Q^{λ_t} defined and examined in proposition II.2.2, the discounted total gain process $B_t^{-1}G_t$ will become a martingale, where the total gain process of investing into the stock with reinvesting the dividend is given by the equation:

$$\log G_t - \log G_{t-1} = r_t + \lambda_t \sigma_t - g(\sigma_t) + \sigma_t \epsilon_t, \quad t \in \mathbb{N}, \epsilon_t \stackrel{iid}{\sim} S_\alpha^{[a,b]}(\sigma, \beta, \mu). \quad (22)$$

As discount function B we use the money market account with $B_t = \exp(\sum_{k=0}^t r_k)$:

$$E_Q(B_t^{-1}G_t \mid \mathcal{F}_s) = B_s^{-1}G_s, \quad 0 \leq s \leq t. \quad (23)$$

The martingale property of G_t allows us to determine arbitrage free prices of derivatives depending on S .

In the following we will fit the time series model (19) to different historical data sets of the S&P 500. As input data we use historical realisations of the S&P 500, dividend yields and interest rates². In order to simplify the estimation procedure we assume a constant market price of risk λ . The set of unknown parameters can now be divided into two subsets: model parameters

Source: S&P 500/option data: <http://www.cboe.com>
² Dividend data: <http://www.economagic.com>
Interest rates: <http://www.federalreserve.gov>

$(\lambda, \alpha_k, k = 0, \dots, p$ and $\beta_k, k = 1, \dots, q)$ on the one hand and distribution parameters $(a, b, \alpha, \beta, \sigma, \mu)$ on the other. Assuming the STS-distribution $S_\alpha^{[a,b]}(\sigma, \beta, \mu)$ to be known we could effectuate a Maximum-Likelihood estimation for the model parameters using the likelihood-function given in equation (18). Having estimated the model parameters we can generate the series $(\hat{\epsilon}_t)_{t=1, \dots, n}$ of empirical residuals. From this series of empirical residuals we can estimate a new STS-distribution. This gives raise to the following iterative estimation procedure:

The iterative estimation procedure starts with an initial STS-distribution $S_\alpha^{[a,b]}(\sigma, \beta, \mu)_0$ and applying the numerical MLE according to equation (18) yields a series of empirical residuals $(\hat{\epsilon}_t)_{t=1, \dots, n}$. Fitting a STS-distribution to the empirical residuals gives us a standardized STS-distribution $S_\alpha^{[a,b]}(\sigma, \beta, \mu)_1$ which can be used for the next ML-estimation step. We stop the iterations as soon as the estimation sequence becomes constant in the sense $S_\alpha^{[a,b]}(\sigma, \beta, \mu)_{n-1} \approx S_\alpha^{[a,b]}(\sigma, \beta, \mu)_n$. The distance of $S_\alpha^{[a,b]}(\sigma, \beta, \mu)_{n-1}$ and $S_\alpha^{[a,b]}(\sigma, \beta, \mu)_n$ is measured by the Kolmogorov distance.

We start our examination with a time series beginning in 9/2/97 and ending in 5/13/04. This series has been chosen because it covers the extremely turbulent market period around and right after the millennium. Table II gives an overview of the estimated parameters and Figure 3 an overview of the different time series involved into the estimation procedure. In the estimation it turned out that $p = q = 2$ is a reasonable choice for the length of the GARCH process. Higher orders will lead to insignificant coefficient estimates and with GARCH(1,1) one loses significant statistical accuracy. The Box-Ljung- Q -statistics of the residuals (see Table III) indicate that the GARCH-process is able to explain the observed volatility clustering. To compare the STS-GARCH time series model (19) with STS-distributed residuals with Duans GARCH-model based on the normal distribution we compare the distribution of the empirical residuals under the two different model assumptions: The Kolmogorov-Smirnov-tests show a significant difference in statistical fit for the STS-distribution and the normal distribution: The value of the KS-statistic is 0.0158 (0.0372) in the STS (normal) case. The critical value at the 5% level equals 0.0330 and for the normal case this corresponds to a p -value of less than one percent. These result are exemplified by the QQ -plots in Figure 4. In summarising the results, we conclude that the presented stock price model based on the STS-distribution

outperforms its concurrent significantly. Especially the distribution of the risk factor, in this case the innovation ϵ , can be accurately modelled with the STS-distribution whereas the normal distribution fails completely in describing the left tail behaviour.

Insert Tables II and III somewhere around here

Insert Figures 3 and 4 somewhere around here

As already mentioned in the introduction, our goal is to combine statistical fit and reliability with the ability of consistent option pricing. Therefore we examine in a first step the qualitative properties of option prices generated with the presented STS-GARCH-model. It has frequently been reported by various authors (Rubinstein (1985) is the standard reference for the S&P 500) that market quotes of option prices admit the following anomalies which contradict the assumptions made by the Black and Scholes model:

- We observe high implicit volatility for options in the money.
- The implicit volatility is low for options out of the money.
- The graph of the implicit volatility versus strike prices (moneyness, delta) is strictly convex and sometimes it attains a strict minimum sometimes it decreases monotonically.
- Generally, the curve becomes flatter if time to maturity increases and very steep for the short term options, but the implicit volatility graphs for different maturities may intersect or may not.

Since Breeden and Litzenberger (1978) it was understood, that the set of option prices for a special maturity and therefore the shape of the smile implies a special shape for the risk neutral density of the underlying corresponding to this maturity. The implied density for stocks and stock indices is much heavier in the left tail and lighter in the right tail than any lognormal distribution. Especially the steepness of the smile which corresponds to the skewness of the distribution cannot be explained with symmetric distributions for the underlying risk factor. This explains, why the smiles obtained with Duan's GARCH model are almost symmetric and not adequate to price stock or index derivatives.

We calculate prices for European call options on the S&P 500 maturing in June '04 and November 04' for the last day of our estimation sample, i.e. May 13, 2004 by Monte Carlo simulation. Settlement of S&P 500 index options takes place on the third Friday of the month of maturity. This implies a time to maturity of 30 trading days for the June option and 135 trading days for the November option. The implicit volatility graphs generated with the model prices exhibiting exactly the same properties which were described by Rubinstein. The smile surface generated by the STS-GARCH model differs substantially from its counterpart based on the normal distribution and examined by Duan. In the Gaussian GARCH framework the smiles are rather symmetric which is not realistic in stock option markets. Our results are exemplarily illustrated in Figure 5. The different shapes of the two graphs can be explained by different initial volatilities for the Monte Carlo algorithm. There are known many thumb rules how to choose the initialising parameters for the recursion equation in Monte Carlo simulations. One could take the average over the last observations where the length of the averaging window equals the length of the simulation window or one could choose the last observation from the estimation sample or the stationary level of the volatility. We have chosen in the first case starting values which lay 20% above and in the second case 20% below the stationary level. This proceeding was already chosen by Duan (1995) and can be interpreted as imperfect information about the current state of the economy.

Insert Figure 5 somewhere around here

After this qualitative examination of model's option prices we examine whether it is possible to reproduce the prices of the liquid market instruments, i.e. the "at the money" and "nearby the money" options. Therefore, we divide the parameters determining our model into two parts: First, the fixed parameters consisting of the six distribution parameters of the estimated STS-distribution and the five volatility parameters of the GARCH(2,2) process. Second the variable market parameters representing the uncertainty in the actual market situation and including the market price of risk λ , the initial volatility σ_0 and the unknown future short rate r_t , $t = 0, \dots, T$. The model's option prices $C_T(K)$ of a European call option with strike K and maturity T (measured in trading days) do consequently depend on the concrete choice of these variable market parameters. By varying the market parameters we try to reproduce the market prices quoted on

May 13, 2004 for the two options maturing in June (30 trading days to maturity) and November (135 trading days to maturity) respectively. The results are presented in Figure 6 which shows the ratio of model prices and market prices. The relative error does not exceed 7 percent for the short term option and 5 percent for the long term option on the whole set of available strikes. These results are very encouraging as they are obtained without any automatic optimisation procedure and by keeping fixed all model parameters. Remember that concurrent models like stochastic volatility or local volatility models which try to reproduce the smile surface usually determine all model parameters through a fitting procedure. As a consequence, these models are not able to *explain* the smile through a reliable econometric model for the underlying time series as we have.

Insert Figure 6 somewhere around here

A second example for the modelling and explanatory power of the developed STS-GARCH-process addresses the most problematic event from a statistical viewpoint in the recent decades: The October crash in 1987. We fitted our stock price model to a data series of 1684 observation ending with October 18, 1987. On the 19th of October the S&P 500 dropped by 20%. Having estimated model parameters we can express this drop in terms of some implied realisation for the residual $\epsilon_{Oct.19}$. From that, we can derive the probability that such an event occurs under the model assumption. Additionally, we can assign the time period which we must wait in average for such an event to occur. In order to interpret the obtained results, we fit two other models to the same data set, namely the GARCH-model of Duan, which ignores the skewed and heavy-tailed shape of the residual distribution and the discrete geometric Brownian Motion which additionally ignores the effect of volatility clustering. The implicit probabilities and the mean time of occurrence for these three model assumptions are given by Table IV. We observe that the probability for such an event is practically zero in both models which are based on the normal distribution. Contrarily, the STS-GARCH-process leads to a probability of approximately 0.0001 which yields a mean time of occurrence of less than forty years. This example shows impressively that it is not sufficient to specify the dynamic of economic variables correctly way but that it is essential to capture the distribution of the risk factors accurately as well. Especially the ignorance

of the presence of heavy tails can lead to a fundamental misinterpretation of the market situation and possible future scenarios.

Insert Table IV somewhere around here

B. Credit risk valuation

This section presents the results of applying the STS-distribution framework to the problem of credit risk valuation. Inspired by the ideas of Merton Merton (1974) we can try to explain the evolution of firm values with the STS-GARCH-process defined in equation (19). Similar to the previous subsection we assume the following dynamic for the logarithmic firm value changes:

$$\log V_t - \log V_{t-1} = r_t - d_t + \lambda_t \sigma_t - g(\sigma_t) + \sigma_t \epsilon_t, \quad t \in \mathbb{N}, \epsilon_t \stackrel{iid}{\sim} S_\alpha^{[a,b]}(\sigma, \beta, \mu). \quad (24)$$

V_t denotes the firm value ex dividend at date t and r_t (d_t) denotes the continuously compounded risk free rate of return (dividend rate) for the period $[t-1, t]$. The predictable parameter λ_t represents the market price of risk for the period $[t-1, t]$ and g the logarithmic moment generating function of the distribution $S_\alpha^{[a,b]}(\sigma, \beta, \mu)$. Again, the conditional variance σ_t^2 is expected to follow a GARCH(p, q)-process:

$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^p \alpha_k \sigma_{t-k}^2 \epsilon_{t-k}^2 + \sum_{k=1}^q \beta_k \sigma_{t-k}^2, \quad t \in \mathbb{N}, \quad (25)$$

The total market value of the firm V_t at time t is given by the sum of equity's market value S_t and liability's market value B_t . The unobservability of the firm value process imposes problems on the parameter estimation. For exchange traded companies the equity value is quoted on a daily basis and at least the book value of the liabilities can be obtained quarterly from balance sheet data³. Consequently, we use the daily stock market data for S_t and interpolate the liabilities B_t for the missing days linearly. The example we present is based on the Xerox company and the estimation sample contains 1453 observations between 3/31/94 and 12/31/99 (Figure 7 illustrates the evolution of Xerox company for this time period).

Insert Figure 7 somewhere around here

The results of our parameter estimation are reported in Table V

Insert Table V somewhere around here

We forbear from presenting the detailed estimation results and the analysis of goodness of fit at this point. The results are very similar to the ones presented in the last subsection. No correlation can be detected in the empirical residuals and squared residuals and the distributional assumption cannot be rejected at the 10% level.

The problem of credit risk valuation is equivalent to the determination of risk neutral default probabilities and recovery values in case of default. Default occurs at time t if the value of the liabilities exceeds the value of the firm $V_t < B_t$. For the ease of computation we assume a constant exogenously given recovery rate of $\mu = 0.5$. The credit spread s_T is the constant extra-interest the company has to pay for borrowing money for the the period $[0, T]$. By definition s_T fulfils the equation:

$$e^{-(r+s_T)T} \stackrel{!}{=} \underbrace{q_T \cdot \mu + (1 - q_T) \cdot 1}_{\text{expected payoff}}, \quad (26)$$

where q_T denotes the cumulative risk neutral probability that default occurs until time T . These default probabilities can be obtained through Monte Carlo simulation.

On the last day of the estimation period the equity of Xerox had a value of $S_0 = 16\,867$ million US \$ with a book value of liabilities of 23 546 million US \$. The average annualised volatility of Xerox's firm value over the estimation period was $\bar{\sigma}_V = 18.6\%$, which implies a distance to default of 2.25 standard deviations. As in our analysis of S&P 500 index options we can reflect the imperfect information about the current state of the economy through a variation of some parameters, in the discussion above denoted as "market parameters". The set of undetermined variables consists of the market price of risk λ , the future short rate r_t and the initial volatility σ_0 . Figures 8, 9 and 10 report our results for a simulation period of 500 trading days which corresponds to approximately two years. For comparison purposes we have added in each case the corresponding value obtained with the Merton approach which is based on a geometric Brownian motion.

³Flannery and Sorescu (1996) proceeded in a similar way and set the book value of assets equal to the market value.

Insert Figures 8, 9 and 10 somewhere around here

The Merton approach and all its refinements (see e.g. Uhrig-Homburg (2002) for a detailed overview) are mainly criticised for the fact that the implied term structure of credit spreads differs significantly in shape and level from its market counterparts. Typically short term market spreads are many times higher than implied by any firm value process. The shape of plots of credit spreads versus time is typically concave and sometimes the spreads do even decrease for longer maturities. This is what lead practitioners and academics to the development of intensity or reduced form models where the default is exogenously modelled by the first jump of some counting process. Nevertheless, from an economic viewpoint, the structural approach to credit risk valuation seems to be the more intuitive one: It is more convincing that the default occurs when the liabilities exceed the value of the assets than that it occurs when some process admits a jump. Our results show, that with an accurately chosen distribution for the driving risk factor together with an adequate process dynamic, the results may change significantly. The credit spreads obtained with the STS-GARCH-model reflect exactly the properties of their market counterparts. Through a variation of the free input variables, various different shapes may be obtained and simultaneously the statistical fit to the underlying time series is reliable.

III. Conclusion

This article introduces the class of STS-probability distribution designed to meet the needs of modelling financial data. The STS-GARCH model, which captures the major phenomena of financial data like volatility clustering, leverage effect and heavy tailed and skewed residuals, is defined and theoretically examined. The pricing measure is constructed in a canonical way and is driven by only one preference parameter which can be used to calibrate the model to market prices. The empirical investigation of the model leads to encouraging results: Market prices of liquid derivatives can be reproduced, qualitative properties of volatility smiles and default intensities match the characteristics of their market counterparts and extreme events like stock market crashes can be explained.

References

- 1 Bachelier, L., 1900, Théorie de la Spéculation, *Annales d'Ecole Normale Supérieure* 3, 21–86.
- 2 Black, F., and M. Scholes, 1973, The Pricing of Options and Corporate Liabilities, *Journal of Political Economy* 81, 637–345.
- 3 Bollerslev, T., 1986, Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics* 52, 115–127.
- 4 Breeden, D., and R. Litzenberger, 1978, Prices of State Contingent Claims Implicit in Option Prices, *Journal of Business* 51, 621–651.
- 5 Carr, P., and L. Wu, 2003, The Finite Moment Log Stable Process and Option Pricing, *Journal of Finance* 58, 753–778.
- 6 Derman, E., and I. Kani, 1994, Riding on a Smile, *Risk* 7, 32–39.
- 7 Duan, J.-C., 1995, The GARCH Option Pricing Model, *Mathematical Finance* 5, 13–32.
- 8 Duan, J.-C., 1997, Augmented GARCH(p, q) Processes and its Diffusion Limit, *Journal of Econometrics* 79, 97–127.
- 9 Dupire, B., 1994, Pricing with a Smile, *Risk* 7, 18–20.
- 10 Engle, R. F., 1982, Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50, 987–1007.
- 11 Fama, E., 1965, The Behaviour of Stock Market Prices, *Journal of Business* 38, 34–105.
- 12 Flannery, M.J., and S. Sorescu, 1996, Evidence of Bank Market Discipline in Subordinated Debenture Yields: 1983-1991., *Journal of Finance* 51, 1347–1377.
- 13 Harrison, M., and D. Kreps, 1979, Martingales and Arbitrage in Multiperiod Securities Markets, *Journal of Economic Theory* 20, 381–408.
- 14 Harrison, M., and S. Pliska, 1981, Martingales and Stochastic Integrals in the Theory of Continuous Trading, *Stochastic Processes and their Applications* 11, 215–260.

- 15 Heston, S. L., 1993, A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies* 6, 327–343.
- 16 Hull, J. C., and A. White, 1987, The Pricing of Options on Assets with Stochastic Volatilities, *Journal of Finance* 42, 281–300.
- 17 Mandelbrot, B. B., 1963a, New Methods in Statistical Economics, *Journal of Political Economy* 71, 421–440.
- 18 Mandelbrot, B. B., 1963b, The Variation of Certain Speculative Prices, *Journal of Business* 36, 394–419.
- 19 Menn, C., 2004, Optionspreistheorie: Ein ökonometrischer Ansatz, Ph.D. thesis Universität Karlsruhe.
- 20 Merton, R., 1973, The Theory of Rational Option Pricing, *Bell Journal of Economics and Management Sciences* 4, 141–183.
- 21 Merton, R. C., 1974, On the Pricing of Corporate Debt: The Risk Structure of Interest Rates, *Journal of Finance* 29, 449–470.
- 22 Mittnik, S., and S. T. Rachev, 2000, *Stable Paretian Models in Finance*. (John Wiley & Sons, Inc. New York, London).
- 23 Pfanzagl, J., 1994, *Parametric Statistical Theory*. (Walter de Gruyter Berlin, New York).
- 24 Racheva-Iotova, B., and S. Stoyanov, 2003, Univariate stable laws in the field of finance-approximation of density and distribution functions, .
- 25 Rubinstein, M., 1985, Nonparametric Tests of Alternative Option Pricing Models Using All Reported Trades and Quotes on the 30 Most Active CBOE Option Classes from August 23, 1976 through August 31, 1978, *Journal of Finance* 40, 455–480.
- 26 Uhrig-Homburg, Marliese, 2002, Valuation of Defaultable Claims - A Survey, *Schmalenbach Business Review* 54, 24–57.

Appendix A. Proofs

Proof of Proposition I.2:

1. Let p_1 and p_2 denote the “cut-off-probabilities”, i.e.

$$p_1 := \int_{-\infty}^a g_\theta(x) dx \quad \text{and} \quad p_2 := \int_b^{\infty} g_\theta(x) dx$$

and Φ denotes the distribution function of the standard normal distribution. From equation (3) we have:

$$\begin{aligned} \int_{-\infty}^a h_1(x) dx &= \int_{-\infty}^a \frac{1}{\sqrt{2\pi\tau_1^2}} e^{-\frac{(x-\nu_1)^2}{2\tau_1^2}} dx \\ &= \Phi\left(\frac{a-\nu_1}{\tau_1}\right) \stackrel{!}{=} p_1 \\ \Rightarrow \nu_1 &= a - \tau_1\Phi^{-1}(p_1). \end{aligned}$$

Plugging this relation into equation (2) yields:

$$\begin{aligned} g_\theta(a) &= h_1(a) = \frac{1}{\sqrt{2\pi\tau_1^2}} e^{-\frac{(a-\nu_1)^2}{2\tau_1^2}} \\ &= \frac{1}{\sqrt{2\pi\tau_1^2}} e^{-\frac{(a-(a-\tau_1\Phi^{-1}(p_1)))^2}{2\tau_1^2}} \\ &= \frac{1}{\sqrt{2\pi\tau_1^2}} e^{-\frac{(\Phi^{-1}(p_1))^2}{2}} \\ \Rightarrow \tau_1 &= \frac{e^{-\frac{1}{2}(\Phi^{-1}(p_1))^2}}{\sqrt{2\pi} g_\theta(a)} \end{aligned}$$

An analogue calculation for b yields finally:

$$\tau_1 = \frac{e^{-\frac{1}{2}(\Phi^{-1}(p_1))^2}}{\sqrt{2\pi} g_\theta(a)} \quad \text{and} \quad \nu_1 = a - \tau_1\Phi^{-1}(p_1) \quad (\text{A1})$$

$$\tau_2 = \frac{e^{-\frac{1}{2}(\Phi^{-1}(p_2))^2}}{\sqrt{2\pi} g_\theta(b)} \quad \text{and} \quad \nu_2 = b + \tau_2\Phi^{-1}(p_2) \quad (\text{A2})$$

2. Obviously the following relations hold true:

$$\begin{aligned}
f_Y(y) &= \frac{1}{c} f_X\left(\frac{y-d}{c}\right) \\
&= \begin{cases} \frac{1}{c} h_1\left(\frac{y-d}{c}\right) & \text{for } y < ca + d \\ \frac{1}{c} g_\theta\left(\frac{y-d}{c}\right) & \text{for } ca + d \leq y \leq cb + d \\ \frac{1}{c} \left(\frac{y-d}{c}\right) & \text{for } y > cb + d \end{cases} \\
&= \begin{cases} \tilde{h}_1(y) & \text{for } y < \tilde{a} \\ \tilde{g}_\theta(y) & \text{for } \tilde{a} \leq y \leq \tilde{b} \\ \tilde{h}_2(y) & \text{for } y > \tilde{b} \end{cases}
\end{aligned}$$

Because of the scale and translation invariance of the normal and α -stable distribution, the function \tilde{h}_1 and \tilde{h}_2 are equal to the densities of the normal distributions $\mathcal{N}(c\nu_1 + d, (c\tau_1)^2)$, $\mathcal{N}(c\nu_2 + d, (c\tau_2)^2)$ and \tilde{g}_θ corresponds to the density of the α -stable distribution with parameter $\tilde{\theta} = (\alpha, \text{sign}(c)\beta, |c|\sigma, c\mu + d)$. The consistency conditions (2) and (3) remain fulfilled.

3. Denoting the density of X for $x < a$ ($x > b$) by h_1 (h_2) we have:

$$\begin{aligned}
EX &= \int_{-\infty}^a x h_1(x) dx + \int_a^b x g_\theta(x) dx + \int_b^{\infty} x h_2(x) dx \\
EX^2 &= \int_{-\infty}^a x^2 h_1(x) dx + \int_a^b x^2 g_\theta(x) dx + \int_b^{\infty} x^2 h_2(x) dx
\end{aligned}$$

The statement will be proved for $x < a$. The case $x > b$ can be verified by analogue calculations.

$$\begin{aligned}
\int_{-\infty}^a x h_1(x) dx &= \int_{-\infty}^{\frac{a-\nu_1}{\tau_1}} (\tau_1 y + \nu_1) \varphi(y) dy \\
&= \tau_1 \int_{-\infty}^{\frac{a-\nu_1}{\tau_1}} y \varphi(y) dy + \nu_1 \Phi\left(\frac{a-\nu_1}{\tau_1}\right) \\
&= -\tau_1 \varphi(y) \Big|_{-\infty}^{\frac{a-\nu_1}{\tau_1}} + \nu_1 \Phi\left(\frac{a-\nu_1}{\tau_1}\right) \\
&= \nu_1 \Phi\left(\frac{a-\nu_1}{\tau_1}\right) - \tau_1 \varphi\left(\frac{a-\nu_1}{\tau_1}\right)
\end{aligned}$$

and with the definition of $\nu_1 = a - \tau_1 \Phi^{-1}(p_1)$ we get

$$\begin{aligned} &= \nu_1 p_1 - \tau_1 \varphi(\Phi^{-1}(p_1)) \quad \text{and} \\ &= a p_1 - \tau_1 (\Phi^{-1}(p_1) p_1 + \varphi(\Phi^{-1}(p_1))). \end{aligned}$$

With partial integration and substitution we get for the second integral:

$$\begin{aligned} \int_{-\infty}^a x^2 h_1(x) dx &= \int_{-\infty}^{\frac{a-\nu_1}{\tau_1}} (\tau_1 y + \nu_1)^2 \varphi(y) dy \\ &= \tau_1^2 \int_{-\infty}^{\frac{a-\nu_1}{\tau_1}} y^2 \varphi(y) dy + 2\tau_1 \nu_1 \int_{-\infty}^{\frac{a-\nu_1}{\tau_1}} y \varphi(y) dy + \nu_1^2 \int_{-\infty}^{\frac{a-\nu_1}{\tau_1}} \varphi(y) dy \\ &= \tau_1^2 \left[-y \varphi(y) \Big|_{-\infty}^{\frac{a-\nu_1}{\tau_1}} + \int_{-\infty}^{\frac{a-\nu_1}{\tau_1}} \varphi(y) dy \right] + \dots \\ &\quad \dots - 2\tau_1 \nu_1 \varphi(y) \Big|_{-\infty}^{\frac{a-\nu_1}{\tau_1}} + \nu_1^2 \Phi \left(\frac{a - \nu_1}{\tau_1} \right) \end{aligned}$$

and again with the definition of $\nu_1 = a - \tau_1 \Phi^{-1}(p_1)$ we have finally

$$\begin{aligned} &= \tau_1^2 (-\Phi^{-1}(p_1) \varphi(\Phi^{-1}(p_1)) + p_1) - 2\tau_1 \nu_1 \varphi(\Phi^{-1}(p_1)) + \nu_1^2 p_1 \\ &= (\tau_1^2 + \nu_1^2) p_1 - \tau_1 \underbrace{(\tau_1 \Phi^{-1}(p_1) + 2\nu_1)}_{a+\nu_1} \varphi(\Phi^{-1}(p_1)) \\ &= (\tau_1^2 + \nu_1^2) p_1 - \tau_1 (a + \nu_1) \varphi(\Phi^{-1}(p_1)). \end{aligned}$$

q. e. d.

Consistency of ML-estimates for STS-distributions:

First we introduce some notations: A STS-distribution is defined by its parameter-vector $(a, b, \theta) \in \mathbb{R}^6$, where $\theta = (\alpha, \beta, \sigma, \mu)$ denotes the four stable parameters. For a STS-distribution P_γ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ we denote by $P_\gamma^n := P_\gamma \otimes \dots \otimes P_\gamma$ the corresponding product measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. For the following we assume that the parameter space $\Gamma := \Gamma_0 \times \Theta \subset \mathbb{R}^6$ is of

the form:

$$\Gamma_0 = \{(a, b) \in \mathbb{R}^2 : a \leq b\} \quad (\text{A3})$$

$$\Theta = [\alpha_0, 2] \times [-1, 1] \times (0, \infty) \times \mathbb{R} \quad \text{for some } \alpha_0 > 1. \quad (\text{A4})$$

Further we set:

$$\begin{aligned} L : \Gamma \times \mathbb{R} &\rightarrow \mathbb{R} \\ (\gamma, x) &\mapsto L(x, \gamma) := \log f_\gamma(x) \end{aligned} \quad (\text{A5})$$

and for arbitrary $n \in \mathbb{N}$

$$\begin{aligned} L_n : \Gamma \times \mathbb{R}^n &\rightarrow \mathbb{R} \\ (\gamma, x) &\mapsto L_n(x, \gamma) := \frac{1}{n} \sum_{i=1}^n \log f_\gamma(x_i), \end{aligned} \quad (\text{A6})$$

where f_γ symbolizes the density of a STS-distribution with parameter vector γ . The function L fulfils the following requirements:

$$\begin{aligned} x &\mapsto L(x, \gamma) \text{ is measurable for all } \gamma \in \Gamma \\ \gamma &\mapsto L(x, \gamma) \text{ is continuous for all } x \in \mathbb{R} \\ E_\gamma(L(x, \tilde{\gamma}) - L(x, \gamma)) &< 0 \text{ for all } \tilde{\gamma} \in \Gamma \text{ with } \tilde{\gamma} \neq \gamma. \end{aligned}$$

The last equation follows from Jensen's inequality:

$$\begin{aligned} E_\gamma(L(x, \tilde{\gamma}) - L(x, \gamma)) &= E_\gamma\left(\log \frac{f_{\tilde{\gamma}}(x)}{f_\gamma(x)}\right) \\ &< \log\left(\int_{\mathbb{R}} f_{\tilde{\gamma}}(x) dx\right) = 0. \end{aligned}$$

For realizations (x_1, \dots, x_n) , $n \in \mathbb{N}$ we are defining the ML-estimator sequences $(a^{(n)}, b^{(n)}, \theta^{(n)})$ for $\gamma = (a, b, \theta)$ with the help of the function L_n by:

$$\frac{1}{n} \sum_{i=1}^n L(x_i, (a^{(n)}, b^{(n)}, \theta^{(n)})) = \sup_{\tilde{\gamma} \in \Gamma} \frac{1}{n} \sum_{i=1}^n L(x_i, \tilde{\gamma}). \quad (\text{A7})$$

We will prove the following result:

Theorem 1. *For every compact $K \subset \Theta$ and every $\epsilon > 0$ the following holds true:*

$$\lim_{n \rightarrow \infty} \sup_{\theta \in K} P_{(a,b,\theta)}^n \{x \in \mathbb{R}^n : \|\theta^{(n)}(x) - \theta\| > \epsilon\} = 0. \quad (\text{A8})$$

Theorem 1 states, that not the entire parameter vector γ but the stable part θ will consistently be estimated by the ML-procedure.

Proof. The proof of 1 is divided into several steps and lemmas.

Lemma A.1. *It exists $n_0 \in \mathbb{N}$, such that for all functions L_k with $k \geq n_0$ we have:*

1. $E_{(a,b,\theta)}(\bar{L}_k(x, \Gamma) - L_k(x, (a, b, \theta))) < \infty$

2. *It exists a compact neighbourhood C of θ with:*

$$E_{(a,b,\theta)}(\bar{L}_k(x, \Gamma_0 \times C^c) - L_k(x, (a, b, \theta))) < 0.$$

where C^c denotes the complement of C in Θ .

For a subset $M \subset \Gamma$ we have used the following notation:

$$\bar{L}_k(x, M) = \sup_{\gamma \in M} \frac{1}{k} \sum_{i=1}^k L(x_i, \gamma)$$

Proof. The following proof is an adoption of the classical proof of the consistency of ML-estimators for location scale families as it can be found e.g. in (Pfanzagl 1994, pp329). For simplicity we use the same notations as in Pfanzagl (1994).

Let p be the density of the standardized STS-distribution with parameter vector $\gamma' := ((a - \mu)/\sigma, (b - \mu)/\sigma, \alpha, \beta, 1, 0)$. We have the following relation between f and p :

$$f_\gamma(x) = \frac{1}{\sigma} \cdot p\left(\frac{x - \mu}{\sigma}\right) \quad (\text{A9})$$

For arbitrary $0 < \epsilon < \alpha_0$, the function p fulfils the following relation due to the condition $\alpha \in [\alpha_0, 2]$:

$$\lim_{|x| \rightarrow \infty} |x|^{1+\epsilon} \sup_{a,b,\alpha,\beta} p(x) = 0 \quad (\text{A10})$$

From the continuity of p on \mathbb{R} we deduce the existence of a constant $M > 0$ with

$$\sup_{a,b,\alpha,\beta} p(x) \leq \frac{M}{1 + |x|^{1+\epsilon}} \quad (\text{A11})$$

Furthermore, it directly follows from equation (A10):

$$\int \log(p(x))p(x) dx > -\infty$$

and consequently

$$E_{(a,b,\theta)}(L_k(x, (a, b, \theta))) > -\infty \quad (\text{A12})$$

For the proof of relation (1) in lemma A.1 it remains to show:

$$E_{(a,b,\theta)}(\bar{L}_k(x, \Gamma)) < \infty \quad (\text{A13})$$

With direct calculations, we get:

$$\begin{aligned} \bar{L}_k(x, \Gamma) &= \sup_{\gamma \in \Gamma} \frac{1}{k} \sum_{i=1}^k f_{\gamma}(x_i) \\ &\leq \sup_{\mu, \sigma} \sup_{a, b, \alpha, \beta} \log \prod_{i=1}^k f_{(a, b, \alpha, \beta, \sigma, \mu)}(x_i) \\ &= \sup_{\mu, \sigma} \sup_{a, b, \alpha, \beta} \log \prod_{i=1}^k \frac{1}{\sigma} p\left(\frac{x - \mu}{\sigma}\right) \end{aligned}$$

and due to relation (A11), we have

$$\leq \log \sup_{\mu, \sigma} \prod_{i=1}^k \frac{1}{\sigma} \cdot \frac{M}{1 + \left|\frac{x_i - \mu}{\sigma}\right|^{1+\epsilon}}$$

For $x = (x_1, \dots, x_k)$ let $m(x) := \min\{0.5 \cdot |x_i - x_j| : i \neq j\}$. Thus, for all $\mu \in \mathbb{R}$ we have $|x_i - \mu| \geq m(x)$ for all $i = 1, \dots, k$ with at most one exception. We deduce:

$$\log \sup_{\mu, \sigma} \prod_{i=1}^k \frac{1}{\sigma} \cdot \frac{M}{1 + \left|\frac{x_i - \mu}{\sigma}\right|^{1+\epsilon}} \leq \log \sup_{\sigma} \frac{M^k}{\sigma^k} \left(1 + \left(\frac{m(x)}{\sigma}\right)^{1+\epsilon}\right)^{-(k-1)}$$

Choose $k > 1 + 1/\epsilon$. We have $1 + \epsilon > k/(k-1)$ and consequently for all $u > 0$, we get $1 + u^{1+\epsilon} > u^{k/(k-1)}$. Applied to $u := m(x)/\sigma$ this yields:

$$\begin{aligned} \bar{L}_k(x, \Gamma) &\leq \log \sup_{\sigma} \frac{M^k}{\sigma^k} \left(1 + \left(\frac{m(x)}{\sigma}\right)^{1+\epsilon}\right)^{-(k-1)} \leq \log(M^k m(x)^{-k}) \\ &= k \log M - k \log m(x) \end{aligned}$$

Condition (A13) is proved and as a direct consequence also relation (1) in lemma A.1, if we have additionally $\int \log m(x) P^k(dx) > -\infty$. The proof of the latter is similar to the first part and not shown in detail.

For the proof of relation (2) in lemma A.1 it suffices, due to equation A12, to prove that for each $K > 0$ it exists a compact set $C_K \subset \Theta$ with

$$E_{(a,b,\theta)}(\bar{L}_k(x, \Gamma_0 \times C_K^c)) < -K \quad (\text{A14})$$

We show first that the set C_r with

$$C_r(x_1, \dots, x_k) := \{\theta \in \Theta : \sup_{a,b} \prod_{i=1}^k \log f_\gamma(x_i) \geq r\} \quad (\text{A15})$$

is bounded:

$$\begin{aligned} C_r(x_1, \dots, x_k) &\subset [\alpha_0, 2] \times [-1, 1] \times \dots \\ &\dots \times \{(\sigma, \mu) \in (0, \infty) \times \mathbb{R} : \prod_{i=1}^k \log(\sup_{a,b,\alpha,\beta} \frac{1}{\sigma} p(\frac{x-\mu}{\sigma})) \geq r\} \\ &\subset [\alpha_0, 2] \times [-1, 1] \times \dots \\ &\dots \times \{(\sigma, \mu) \in (0, \infty) \times \mathbb{R} : \prod_{i=1}^k \log(\frac{1}{\sigma} \cdot \frac{M}{1 + (\frac{x_i - \mu}{\sigma})^{1+\epsilon}}) \geq r\} \end{aligned}$$

and it is easy to deduce the existence of $0 < \sigma_0 < \sigma_1$ and $\mu_0 < \mu_1$ with:

$$C_r(x_1, \dots, x_k) \subset [\alpha_0, 2] \times [-1, 1] \times [\sigma_0, \sigma_1] \times [\mu_0, \mu_1].$$

Now, equation A14 is a direct consequence of the following lemma A.2 (see e.g. Pfanzagl 1994, p331) applied to the function $\tilde{L}(x, \theta) := \bar{L}(x, \Gamma_0 \times \{\theta\})$.

Lemma A.2. *Let Γ be σ -locally compact but not compact. Furthermore, the function $L : \mathbb{R}^n \times \Gamma \rightarrow \mathbb{R}$ fulfils the following conditions:*

1. $x \mapsto L(x, \gamma)$ is measurable for all $\gamma \in \Gamma$.
2. $\gamma \mapsto L(x, \gamma)$ is continuous for all $x \in \mathbb{R}^n$.
3. $E_\gamma(\bar{L}(x, \Gamma)) < \infty$.
4. $\{\gamma \in \Gamma : L(x, \gamma) \geq r\}$ is compact or empty for all $x \in \mathbb{R}^n, r \in \mathbb{R}$.

Then it exists for every $K > 0$ a compact set C_K with:

$$E_\theta(\bar{L}(x, C_K^c)) < -K \quad (\text{A16})$$

The proof of lemma A.2 can be found in Pfanzagl (1994) on page 331. *q.e.d.*

The rest of the proof of theorem 1 consists of the standard machinery for uniform consistency proofs: With the results shown above it is easy to deduce that the function L_k of lemma A.1 fulfils the locally uniform covering condition for the set $\Gamma_0 \times V^c$ in γ , which means that it exists a finite covering of $\Gamma_0 \times V^c$ with sets V_1, \dots, V_k such that for every $i = 1, \dots, k$ we have:

1. The function $L(x, V_i) - L(x, \gamma)$ is uniformly $P_{\tilde{\gamma}}$ -integrable in some neighbourhood U_i of γ
2. $E_{\gamma}(\bar{L}(x, V_i) - L(x, \gamma)) < 0$

From the locally uniform covering condition it can be deduced the validity of the following condition:

Lemma A.3. *For each neighbourhood $V \subset \Theta$ of θ it exists a neighbourhood $U \subset \Theta$ of θ and some $\delta > 0$ with:*

$$\lim_{n \rightarrow \infty} \sup_{\tau \in U} P_{(a,b,\tau)}^n \{x \in \mathbb{R}^n : \bar{L}_n(x, \Gamma_0 \times V^c) > L_n(x, (a, b, \theta)) - \delta\} = 0, \quad (\text{A17})$$

This completes the proof. *q.e.d.*

Proof of Proposition II.2:

1. From the \mathcal{F}_{t-1} -measurability of σ_k , $k = 1, \dots, t$ and ϵ_k , $k = 1, \dots, t-1$ we deduce with the law of iterated expectations:

$$\begin{aligned} C(\sigma_t \epsilon_t, \sigma_{t+1}^2) &= E(\sigma_t \epsilon_t \sigma_{t+1}^2) - \underbrace{E(\sigma_t \epsilon_t)}_{=0} \cdot E(\sigma_{t+1}^2) \\ &= E\left(\sigma_t \epsilon_t \left(\alpha_0 + \sum_{k=1}^p \alpha_k \sigma_{t-k+1}^2 \epsilon_{t-k+1}^2 + \sum_{k=1}^q \beta_k \sigma_{t-k+1}^2\right)\right) \\ &= E\left(\alpha_0 \sigma_t \epsilon_t + \sum_{k=1}^p \alpha_k \sigma_t \epsilon_t \sigma_{t-k+1}^2 \epsilon_{t-k+1}^2 + \sum_{k=1}^q \beta_k \sigma_t \epsilon_t \sigma_{t-k+1}^2\right) \\ &= \alpha_0 E(E(\sigma_t \epsilon_t | \mathcal{F}_{t-1})) + \alpha_1 E(E(\sigma_t^3 \epsilon_t^3 | \mathcal{F}_{t-1})) + \dots \\ &\quad \dots + \sum_{k=2}^p \alpha_k E(E(\sigma_t \epsilon_t \sigma_{t-k+1}^2 \epsilon_{t-k+1}^2 | \mathcal{F}_{t-1})) + \dots \\ &\quad \dots + \sum_{k=1}^q \beta_k E(E(\sigma_t \epsilon_t \sigma_{t-k+1}^2 | \mathcal{F}_{t-1})) \end{aligned}$$

$$\begin{aligned}
&= \alpha_0 E \left(\underbrace{\sigma_t E(\epsilon_t | \mathcal{F}_{t-1})}_{=0} \right) + \alpha_1 E \left(\underbrace{\sigma_t^3 E(\epsilon_t^3 | \mathcal{F}_{t-1})}_{=\kappa} \right) + \dots \\
&\dots + \sum_{k=2}^p \alpha_k E \left(\underbrace{\sigma_{t-k+1}^2 \epsilon_{t-k+1}^2 \sigma_t E(\epsilon_t | \mathcal{F}_{t-1})}_{=0} \right) + \dots \\
&\dots + \sum_{k=1}^q \beta_k E \left(\underbrace{\sigma_{t-k+1}^2 \sigma_t E(\epsilon_t | \mathcal{F}_{t-1})}_{=0} \right) \\
&= \underbrace{\alpha_1}_{>0} \underbrace{E(\sigma_t^3)}_{>0} \cdot \kappa.
\end{aligned}$$

2.

- (a) The dynamic under the new probability measure Q is simply obtained by replacing ϵ_t through $\xi_t - \lambda$.
- (b) Similarly, the equality of the conditional variances follows directly from the fact

$$V_Q(\epsilon_t) = 1 = V_P(\epsilon_t).$$

- (c) From the fact that λ_t^2 and σ_t^2 are uncorrelated and the independence of ξ_t and \mathcal{F}_{t-1} we deduce:

$$\begin{aligned}
E_Q \sigma_t^2 &= \alpha_0 + \sum_{k=1}^p \alpha_k E_Q (\sigma_{t-k}^2 (\xi_{t-k} - \lambda_{t-k})^2) + \sum_{k=1}^q \beta_k E_Q \sigma_{t-k}^2 \\
&= \alpha_0 + \sum_{k=1}^p \alpha_k E_Q (\sigma_{t-k}^2 E_Q (\xi_{t-k}^2 - 2\xi_{t-k} \lambda_{t-k} + \lambda_{t-k}^2 | \mathcal{F}_{t-k-1})) \\
&\quad + \sum_{k=1}^q \beta_k E_Q \sigma_{t-k}^2 \\
&= \alpha_0 + \sum_{k=1}^p \alpha_k E_Q (\sigma_{t-k}^2 (1 + \lambda_{t-k}^2)) + \sum_{k=1}^q \beta_k E_Q \sigma_{t-k}^2 \\
&= \alpha_0 + \sum_{k=1}^p \underbrace{\alpha_k (1 + \lambda^2)}_{=: \alpha_k^*} E_Q \sigma_{t-k}^2 + \sum_{k=1}^q \beta_k E_Q \sigma_{t-k}^2
\end{aligned}$$

Under the assumptions made on the parameters λ , α_k and β_k the stationary solution has the property:

$$E_Q \sigma_t^2 = \frac{\alpha_0}{1 - \sum_{k=1}^p \alpha_k^* - \sum_{k=1}^q \beta_k},$$

which yields the result.

3. The result is obtained through straight calculation.

q.e.d.

Proof of Proposition II.3:

1. Due to the predictability of the processes $(r_t), (d_t), (\lambda_t)$ and (σ_t) we have

$$\begin{aligned} E\left(\frac{S_t^{cd}}{S_{t-1}} \mid \mathcal{F}_{t-1}\right) &= E\left(e^{r_t + \lambda_t \sigma_t - g(\sigma_t) + \sigma_t \epsilon_t} \mid \mathcal{F}_{t-1}\right) \\ &= e^{r_t + \lambda_t \sigma_t - g(\sigma_t)} \cdot E(e^{\sigma_t \epsilon_t} \mid \mathcal{F}_{t-1}) \\ &= e^{r_t + \lambda_t \sigma_t}, \end{aligned}$$

where the last equation follows from the relation

$$\log(E(\exp(\sigma_t \epsilon_t) \mid \mathcal{F}_{t-1})) = g(x) \Big|_{x=\sigma_t}$$

2. From equation (13) in Proposition II.2 and by the definition of the discounted total gain process $\bar{G}_t := B_t^{-1} G_t$ we obtain the following dynamic under Q :

$$\log \bar{G}_t - \log \bar{G}_{t-1} = -g(\sigma_t) + \sigma_t \xi_t, \quad t \in \mathbb{N}, \quad \xi_t \stackrel{iid}{\sim} S_\alpha^{[a,b]}(\sigma, \beta, \mu).$$

If we use the \mathcal{F}_{t-1} -measurability of (σ_t) and the definition of g as logarithmic moment generating function corresponding to ξ_t we get:

$$\begin{aligned} E_Q(\bar{G}_t \mid \mathcal{F}_{t-1}) &= E_Q(\bar{G}_{t-1} \exp(-g(\sigma_t) + \sigma_t \xi_t) \mid \mathcal{F}_{t-1}) \\ &= \bar{G}_{t-1} \cdot \underbrace{\exp(-g(\sigma_t))}_{=1/m(\sigma_t)} \cdot \underbrace{E_Q(\exp(\sigma_t \xi_t) \mid \mathcal{F}_{t-1})}_{=m(\sigma_t)} \\ &= \bar{G}_{t-1}, \end{aligned}$$

where m denotes the moment generating function of the distribution $S_\alpha^{[a,b]}(\sigma, \beta, \mu)$. As $(G_t)_{t \in \mathbb{N}_0}$ is a discrete stochastic process the martingale property follows from the law of iterated expectations.

q.e.d.

Table I
Comparison of $P(X \leq x)$ for the normal and a standardized
STS-distribution.

The Table shows the difference in the tail probabilities of an exemplarily standardized STS-distribution and the standard normal law. The STS-distribution exhibits a significantly higher tail probability than the standard normal law.

x	$X \sim \mathcal{N}(0, 1)$	$X \sim S_{1.8}^{[-4.0, 3.3]}(0.58, -0.1, 0)$
-10	$7.6198530242 \cdot 10^{-24}$	0.00009861132775
-9	$1.1285884060 \cdot 10^{-19}$	0.00019210547239
-8	$6.2209605743 \cdot 10^{-16}$	0.00036395064728
-7	0.00000000000128	0.00067063777602
-6	0.00000000098659	0.00120208371192
-5	0.00000028665157	0.00209626995052
-4	0.00003167124183	0.00355718712680
-3	0.00134989803163	0.00669781592407
-2	0.02275013194818	0.02013650454786
-1	0.15865525393146	0.11793584416637

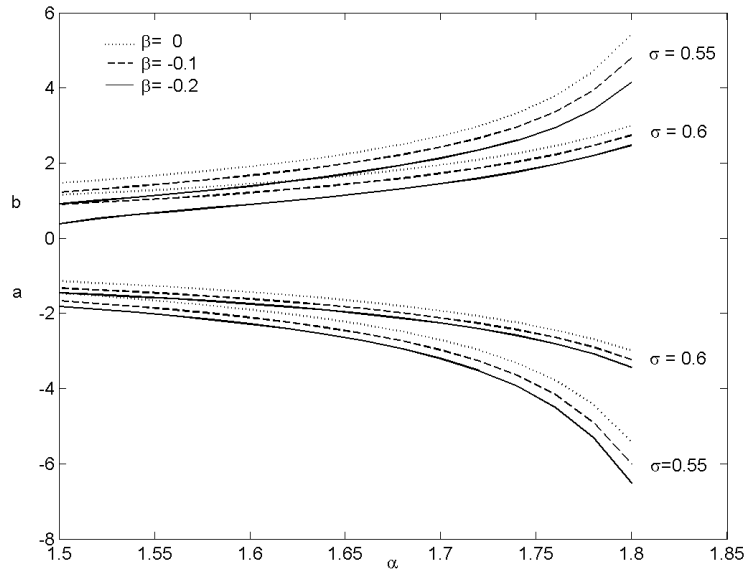
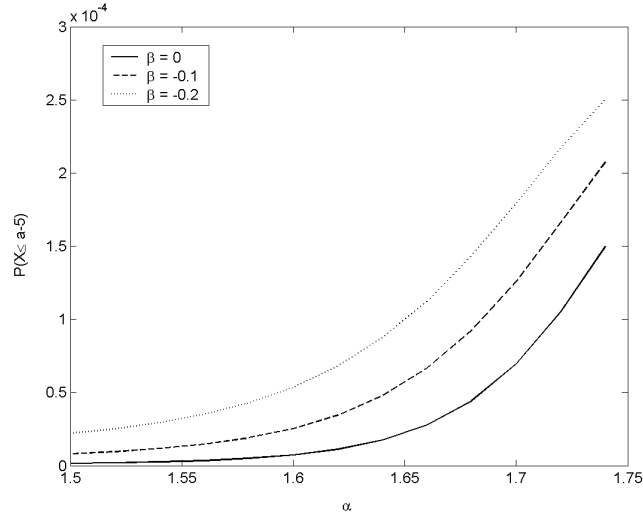
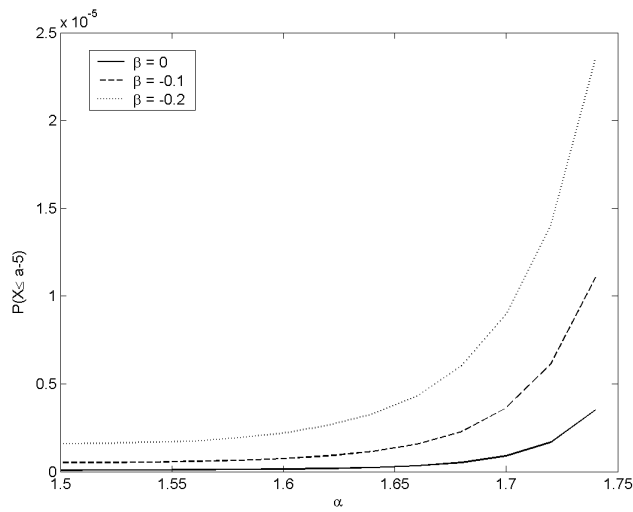


Figure 1. Truncation levels of standardized STS-distributions. The Figure shows how the parameters α , β and σ influence the truncation levels for standardized STS-distributions. The graph shows in the upper (lower) part the right truncation level b (the left truncation level a) in dependence of α for all combinations of three different values of β ($\beta = 0, -0.1, -0.2$) and two different values of σ ($\sigma = 0.55, 0.6$).



(a) $\mu = 0, \sigma = 0.55$



(b) $\mu = 0, \sigma = 0.6$

Figure 2. Left tail probabilities of standardized STS-distributons.

This figure reports the dependence of the probability $P(-\infty, a - 5)$ from the distribution parameters α and β . The first (second) graph contains this probability for varying α and three different values of β ($\beta = 0, -0.1, -0.2$). The values of $\mu = 0$ and $\sigma = 0.55$ ($\sigma = 0.6$) are predefined.

Table II
Estimation results for the model and distribution parameters

Fitting the presented STS-GARCH-model and Duan's stock price model to a S&P 500 time series yields the following parameter estimates. We distinguish between distribution parameters which are only needed in the STS-case and model parameters for both approaches.

Parameters of STS-distribution					
a	b	α	β	σ	μ
-4.6714	1.7581	1.9203	-0.1023	0.6683	0.0064

Model parameters						
Model	λ	α_0	α_1	α_2	β_1	β_2
STS-GARCH	0.0479	5.557e-006	0.0389	0.0976	0.5606	0.2771
Duan	0.0500	9.826e-006	0.0467	0.1411	0.3498	0.4108

Table III**Box-Ljung- Q -statistic for autocorrelation in the residuals.**

The Table shows the values of the Box-Ljung- Q -statistic and the corresponding p -value for the first 10 lags in the empirical residuals obtained by fitting the STS-GARCH-model to a time series of S&P 500 observations. The test does not detect any significant autocorrelation neither in the residuals nor in the squared residuals.

lag	ϵ_t		ϵ_t^2	
	statistic	p -value	statistic	p -value
1	0.3649	0.5458	2.3968	0.1216
2	0.4596	0.7947	5.4927	0.0642
3	2.8433	0.4164	5.5009	0.1386
4	2.8484	0.5835	5.9000	0.2067
5	7.6404	0.1772	5.9013	0.3159
6	7.6922	0.2615	7.2148	0.3014
7	9.1587	0.2415	7.3913	0.3893
8	9.6562	0.2900	7.3995	0.4942
9	9.6722	0.3777	9.5473	0.3884
10	9.8625	0.4526	10.1044	0.4314

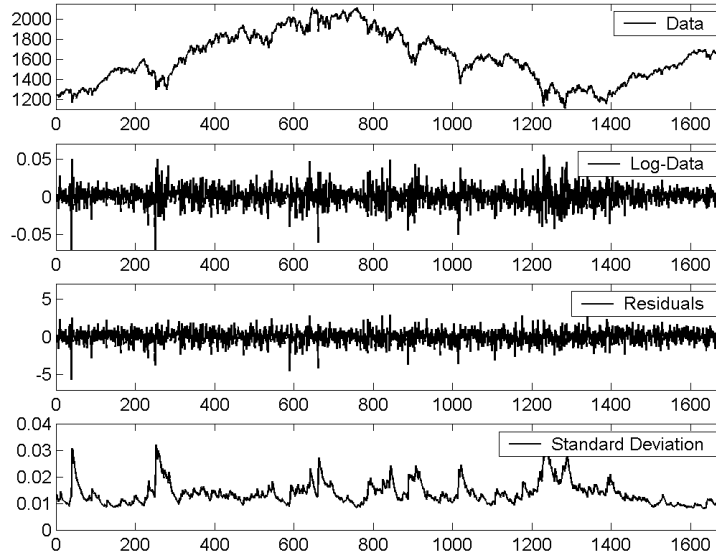
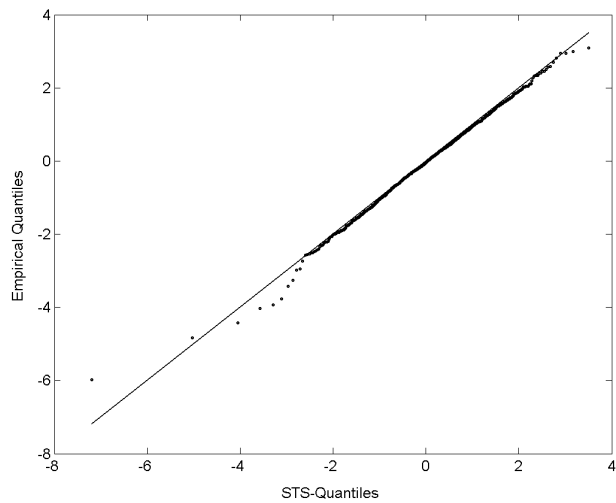
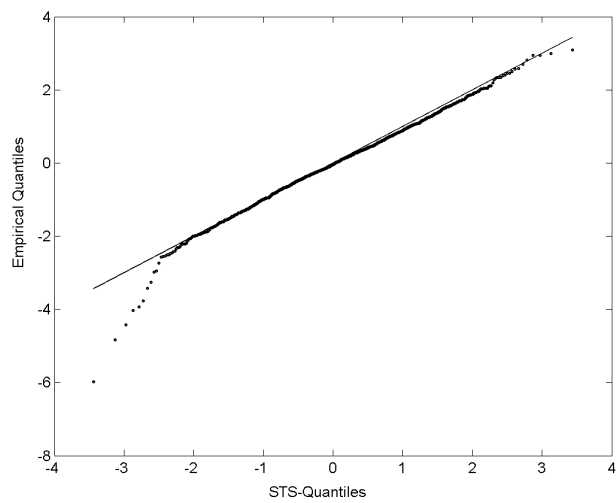


Figure 3. Time series involved in the estimation procedure. The Figure shows the time series for the S&P 500 with reinvested dividends, the corresponding Logdifferences, which serve as dependent variable in the regression model and which exhibit volatility clustering, the standardized residuals which can be recognized as left skewed and not normal and the time series of estimated volatilities.

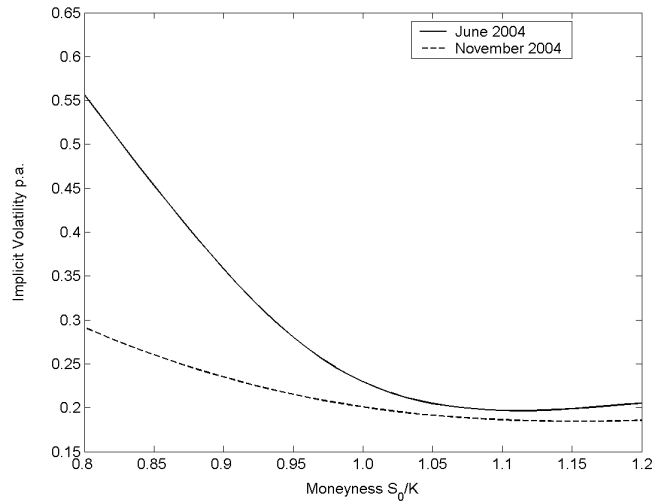


(a) STS-distribution

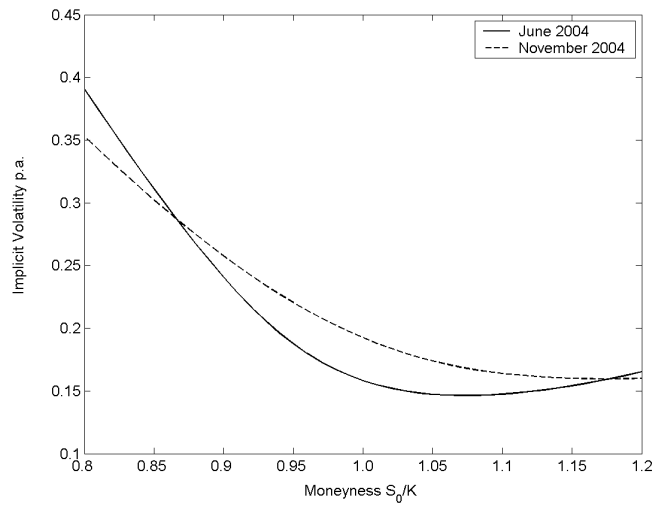


(b) Normal distribution

Figure 4. Q-Q-plots for the empirical residuals. This Figure shows in panel (a) the ability of the STS-distributions to describe the distribution of the innovations accurately. The normal distribution fails completely in modelling the left tail behaviour of the innovation process (panel (b)), which is essential for explaining and forecasting large downward-movements in the market.

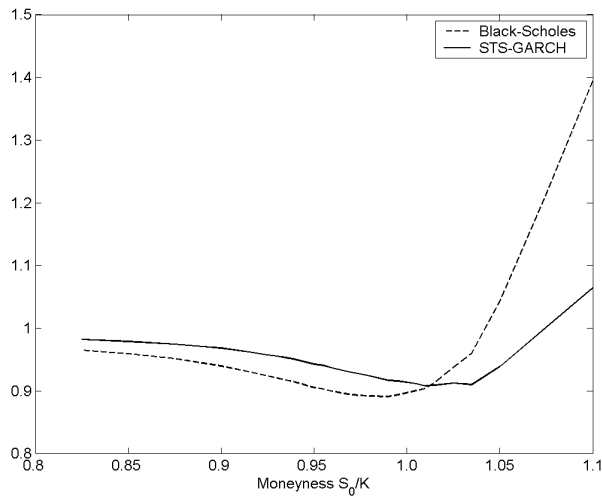


(a) High initial volatility

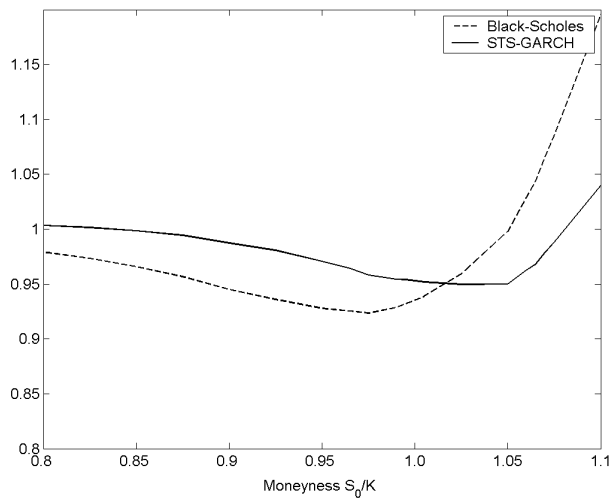


(b) Low initial volatility

Figure 5. Implicit volatilities of S&P 500 call options calculated with model prices. The Figure reports the qualitative properties of volatility smiles calculated with STS-GARCH-option prices. The prices are calculated on May 13, 2004. Different smiles can be generated by varying the unknown initial volatility for the GARCH-process in the Monte Carlo simulation. Panel (a) (panel (b)) is based on a starting value laying 20% above (below) the stationary level.



(a) June 2004



(b) November 2004

Figure 6. Ratio of model prices and market prices Figure 6 shows the ratio of model prices and market for two different call options on the S&P 500: One matures in June '04 and the other in November 2004. The prices are calculated on May 13, 2004. For illustration purposes the ratio of Black and Scholes prices with market prices is included.

Table IV
Comparison of crash probabilities.

The Table lists the probability of observing a crash like it happened on October 19, 1987 in the framework of a discrete geometric Brownian motion (DGBM), Duans GARCH-model (Duan) and the STS-GARCH-process. The model parameters are estimated with the same data sample of 1684 observations ending on October 18, 1987.

Model	$\hat{\epsilon}_{oct,19}$	$P(\epsilon_t \leq \hat{\epsilon}_{oct,19})$	Average time of occurrence
DGBM	-20	$4.27 \cdot 10^{-91}$	$\approx 10^{87}$ years
Duan	-9	$5.76 \cdot 10^{-21}$	$\approx 10^{22}$ years
STS-GARCH	-11	0.0001	≈ 38 years

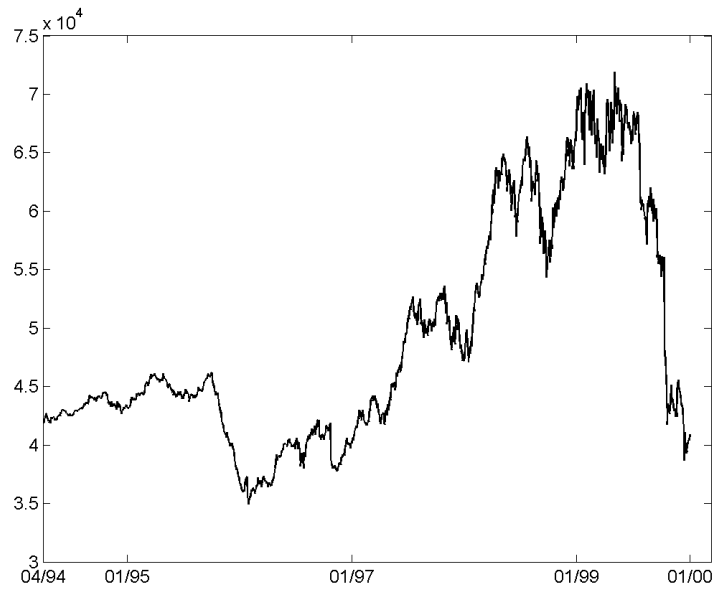


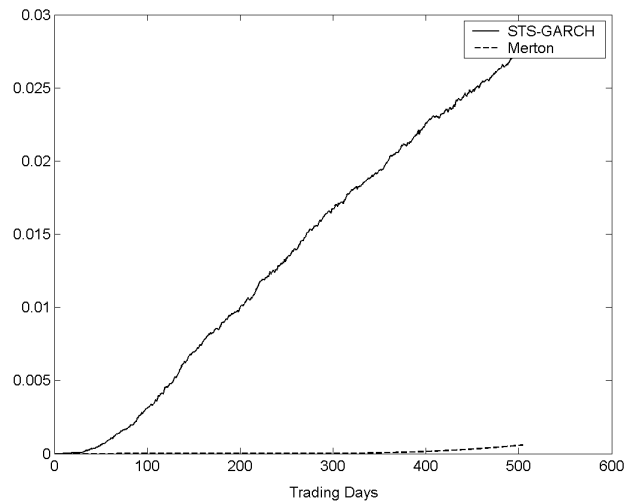
Figure 7. Evaluation of Xerox's firm value between 1994/03/31 and 1999/12/31. The firm value is measured in million US\$.

Table V
Estimation results for the model and distribution parameters

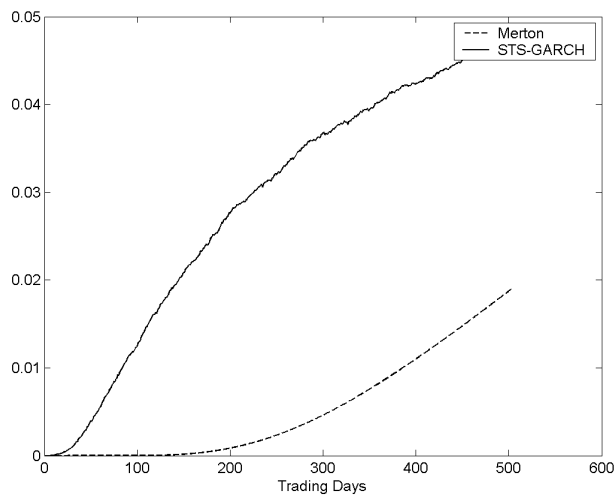
Fitting the presented STS-GARCH-model to 1453 observations of Xerox's firm value between 04/94 and 12/99 yields the following parameter estimates.

Parameters of STS-distribution					
a	b	α	β	σ	μ
-3.1024	1.7879	1.7151	-0.104	0.5781	-0.0012
Model parameters					
λ	α_0	α_1	α_2	β_1	β_2
0.0106	6.823e-007	0.0759	0.0896	-(*)	0.8341

(*): the value of β_1 is not significant at the 5% level.

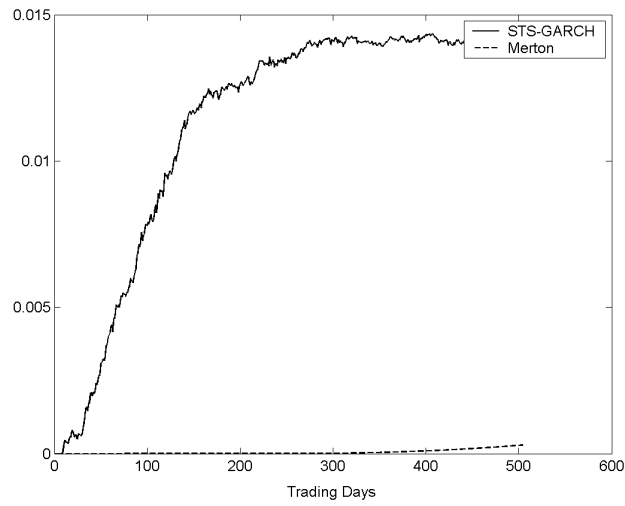


(a) Low initial volatility

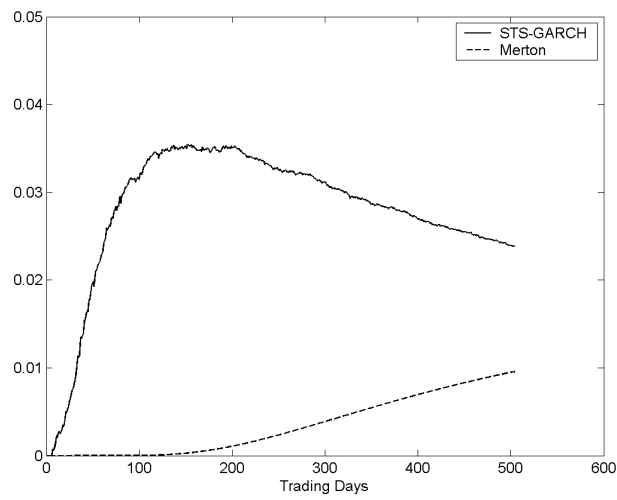


(b) Medium initial volatility

Figure 8. Cumulative default probabilities for Xerox The graphs show the cumulative risk neutral default probabilities for the Xerox company viewed from December 31, 1999 and obtained with the STS-GARCH-model and the Merton model. Time scale is measured in trading days.



(a) Low initial volatility



(b) Medium initial volait

Figure 9. Implied credit spreads The graphs show the implied credit spreads for different maturities calculated with the risk neutral default probabilities and under the assumption of 50% recovery value in case of default. Time scale is measured in trading days.

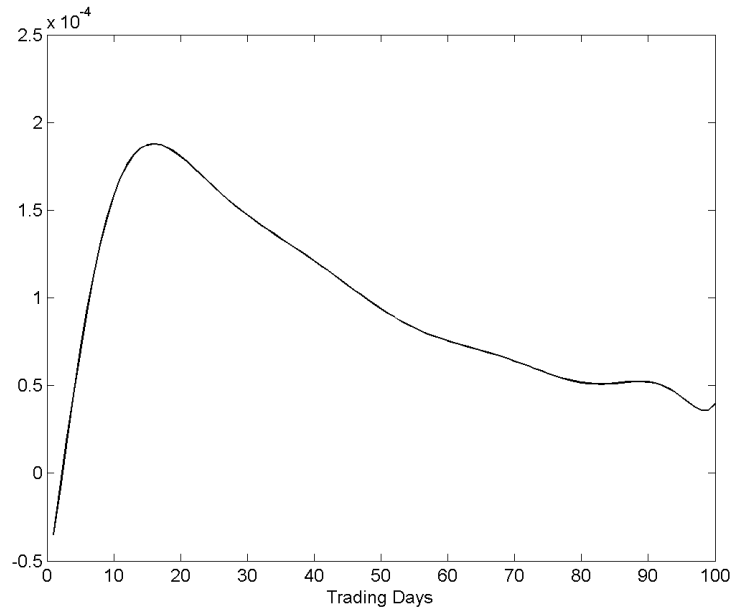


Figure 10. Implied default intensity. The figure illustrates the short term implied default intensity which corresponds to the low initial volatility case. The negative values for the first days and the oscillations at the right end are due to numerical instabilities.