Smoothly Truncated Stable Distributions, GARCH-Models, and Option Pricing*

Christian Menn†‡
Cornell University

Svetlozar T. Rachev§
University of Karlsruhe and UCSB

June 20, 2005

*This paper subsumes the previous one under the title "A New Class of Probability Distributions and Its Application to Finance". The authors gratefully acknowledge comments made by seminar participants at University of California, Santa Barbara, University of Washington, Seattle, Hochschule für Banken, Frankfurt, Cornell University, Princeton University, American University , Washington DC, the Risk Management and Financial Engineering Conference held in Gainesville, FL in April 2005 and at the EU-Workshop on Mathematical Optimization Models held in November 2003 in Cyprus.

†**Correspondence Information:** Christian Menn, School of Operations Research and Industrial Engineering, Cornell University, 427 Rhodes Hall, Ithaca, NY 14853, tel: +1 (607) 254 4955, fax: +1 (607) 255 9129, email: menn@orie.cornell.edu.

‡Menn is grateful for research support provided by the German Academic Exchange Service (DAAD).

§Research support provided by grants from Division of Mathematical, Life and Physical Sciences, College of Letters and Science, University of California, Santa Barbara, the German Academic Exchange Service (DAAD), and the German Research Foundation (DFG) is gratefully acknowledged.

ABSTRACT

Although asset return distributions are known to be conditionally leptokurtic, this fact was rarely addressed in the recent GARCH model literature. For this reason, we introduce the class of smoothly truncated stable distributions (STS distributions) and derive a generalized GARCH option pricing framework based on non-Gaussian innovations. Our empirical results show that (1) the model's performance in the objective as well as the risk-neutral world is substantially improved by allowing for non-Gaussian innovations and (2) the model's best option pricing performance is achieved with a new estimation approach where all model parameters are obtained from time-series information whereas the market price of risk and the spot variance are inverted from market prices of options.

The failure to explain observed option prices is well-known for option pricing models which are solely based on time series information. For this reason, the option pricing research in the recent years has mainly focused martingale models: The stochastic dynamic of the underlying is directly specified under some pricing or martingale measure Q and all model parameters are estimated by calibrating the model to observed market prices of liquid options. Prominent examples include the stochastic volatility model of Heston (1993), more generally the affine jump diffusions proposed by Bakshi and Chen (1997) and Duffie, Pan, and Singleton (2000), and recently Carr, Geman, Madan, and Yor (2003) and Carr and Wu (2004) defined an option pricing framework with time-changed Lévy processes. In order to obtain the quasi-closed form solution for the option price in form of a generalized Fourier transform for the distribution of the terminal stock price, these models typically impose a Markovian structure for the stock returns contradicting empirical evidence.

Attempts to overcome this drawback were made in the GARCH option pricing literature. The GARCH-models of Duan (1995, 1999) and Heston and Nandi (2000) and more recently Duan, Ritchken, and Sun (2004), Christofferson and Jacobs (2004b) and Christofferson, Heston, and Jacobs (2004) are examples which take care of the non-Markovian structure of asset returns. To the best of our knowledge, Heston and Nandi were the first who promoted an estimation procedure for GARCH model parameters which is based on both sources of available information: Historical prices for the underlying and market quotes of options. The authors show that the gain of using the time series information to filter the spot variance from past returns is substantial with respect to the out-of-sample pricing performance. From a statistical viewpoint, the Heston-Nandi model and most of its extensions could be criticized for ignoring the conditional leptokurtosis and skewness of financial returns. Another drawback arises from the special form of the volatility filter which is used to retrieve the spot variance from past return observations given the set of risk-neutral parameters. As it will later be discussed in detail, the filter faces some identifiability problem and is only applicable by making an ad hoc choice for the market price of risk parameter governing the change of measure. This choice can create a bias in the estimate of the spot variance.

This is exactly the gap which we fill in: We present a time series model which combines statistical reliability with market consistent derivative pricing. To this goal, we need a probability distribution which is able to describe the major characteristics of the conditional return distribution. We address this issue by introducing the class of smoothly truncated stable distributions (STS-distributions). STS-distributions are obtained by smoothly replacing the upper and lower tail of an arbitrary stable cumulative distribution function by two appropriately chosen normal tails. Consequently the density of a smoothly truncated stable distribution consists of three parts. Left of some lower truncation level a and right of some upper truncation level b, it is described by two possibly different normal densities. In the center, the density equals the one of a stable distribution. As a result, STS distributions lay in the domain of attraction of the Gaussian distribution and possess even a finite moment generating function while offering at the same time a flexible tool to model extreme events. We will see that exactly this ability to assign a reasonable amount of probability to extreme events distinguishes the STS-distributions from concurrent probability distributions with heavier tails than a normal distribution but finite moment generating function. As a consequence, the speed of convergence to the normal distribution is extremely slow – a phenomenon which is well-known from empirical return distributions.

In a second step, we build a general NGARCH stock price model allowing for non-Gaussian innovation distributions. The model is mainly inspired by the Duan (1995) framework. The only conditions on the innovation distribution are that it is a continuous probability distribution with support on the whole real line which has zero mean, unit variance, and finite moment generating function. The dynamic of the log-returns is defined in a way such that the expected mean rate of return allows for the familiar decomposition into risk-free return minus dividends plus risk premium. The CAPM-like relation that the risk premium is proportional to the amount of risk measured in terms of volatility is maintained. This relation justifies the notion "market price of risk" for the model parameter describing the proportionality factor. In the most general case, the risk-free rate, the dividend rate as well as the market price of risk parameter are allowed to follow their own predictable stochastic processes. Supported by some recent evidence provided in the literature (see e.g. Christofferson and Jacobs 2004b) and our own findings in the empirical part, we propagate a parsimonious asymmetric GARCH(1,1) dynamic (NGARCH) as introduced by Engle and Ng (1993) for describing the evolution of the conditional variance. For the pricing of derivatives we will need a risk-neutral version of our time series model. We renounce to present an economical justification for the specific form of change of measure as the usual assumptions on investors' preferences seem to restrictive for practical purposes. Giving up these necessities, we are able to formulate a general form of change of measure which even allows for different residual distributions under the objective and the risk-neutral measure.

The STS-NGARCH model is tested along three different dimensions. We investigate the model's fit to a 14 years long history of S&P 500 log-returns. Not too surprisingly, we find that the model provides a better fit – judged on the log-likelihood and the Kolmogorov-Smirnov and Anderson-Darling distance between the empirical and theoretical residual distribution – than the NGARCH model with Gaussian or other non-Gaussian innovation distributions. In a second step we compare the model's ability to forecast the October '87 crash. NGARCH models with different innovation distributions are fitted to a data series ending on the trading day before the crash. Interestingly, a variant of the STS-NGARCH model with a predetermined STS distribution is the only model that assigns a reasonable probability to the event which took place on the next trading date, namely the Black Monday. However, the most interesting analysis is performed by comparing the impact of the model specification and the distributional assumption on the ability to explain market prices of S&P 500 index options. We examine three different estimation methodologies for the STS-NGARCH model. The first ("MLE") is only based on time series information, the second ("MLE/fitted") is mainly based on time series information but tries to eliminate its deficiencies. The market price of risk parameter λ – well-known to difficult to estimate and time varying – as well as the spot variance are backed out from a set of market prices of liquid options. The idea is to use the ML-estimates wherever they are known to be accurate and temporarily stable and to use the market data to make inferences about the time varying parameters. The third methodology ("NLS") is a slight variation of the Heston-Nandi approach. All parameters except from the spot variance are determined from the market prices of options through a non-linear least square optimization with respect to the sum of squared dollar pricing errors. The spot variance is filtered from the time series information. Our results which are also supported by other studies such as (Ritchken and Hsieh 2000), show that the resulting optimization problem is numerically difficult to handle and the resulting parameter estimates obtained by the non-linear least square methodology are unstable through time. Moreover, the resulting parameter estimates depend heavily on the choice of loss function which is applied in the optimization procedure (Christofferson and Jacobs 2004a). For these reasons, the suggested "MLE/fitted" procedure seems to be an interesting alternative and our in-sample and out-of-sample pricing performance comparison support this idea. Judged on the sum of squared relative and absolute pricing errors, the "MLE/fitted" methodology yields better results than the "NLS" approach. To assess the overall pricing performance of the STS-NGARCH variants, we adopt common practice and compare the different models with the ad hoc Black-Scholes model as introduced in (Dumas, Fleming, and Whaley 1998). The result show that in the out-of-sample study, the “MLE/fitted” outperforms the ad hoc Black-Scholes model. We renounce to present any results connected to the hedging performance of the different models. Even if this has been common practice in related publications, we think that the hedging performance is not an issue, an option pricing model should be judged on as long as liquid market prices are available. The reason is that any option pricing model which reproduces the actual observed price – especially the Black-Scholes model applied with the implied volatility of the specific option – will lead to the optimal hedge ratio. This fact follows directly from the homogeneity of option prices with respect to strike price and underlying spot price and the Euler formula.

The remainder of the article is organized as follows. In Section 1 we introduce the class of smoothly truncated stable distributions (STS-distributions) and we present a generalized NGARCH option pricing model with non-Gaussian innovations. The empirical results are discussed in Section 2. Firstly, we examine the statistical fit of the model to the S&P 500 under the objective measure. Secondly, we provide evidence that the model predicts reasonable probabilities for crash events such as in October 1987. Finally, we report the results of applying the STS-NGARCH-model to explain market prices of S&P 500 index options. Section 3 concludes and discusses possible directions of future research whereas the appendix contains the proofs of the main theoretical results.

1. The Model

1.1. Smoothly Truncated α-Stable Distributions

We introduce a special class of truncated α-stable distributions which we babtized as smoothly truncated stable distributions (henceforth denoted as STS distributions).[1] The name is due to the special form of tail truncation, which guarantees a continuously differentiable distribution function for the truncated α-stable distribution. Formally, we have:

Definition 1. *Let g_θ denote the density of an 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 mean ν_i and standard deviation τ_i, $i = 1, 2$. Furthermore, let $a, b \in \mathbb{R}$ be two real numbers with $a \leq m \leq b$, where m denotes the mode of g_θ . The density of a STS distribution is defined by:*

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

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

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

and

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

The family of STS distributions will be denoted by $\mathcal{S}$, the subclass of standardized STS distributions by $\mathcal{S}_0$. Elements of $\mathcal{S}$ are denoted by $S_\alpha^{[a,b]}(\sigma, \beta, \mu)$.

Definition 1 requires some discussion. Firstly, we mention that the special choice of parametrization of stable distributions we refer to, equals the one used in the book by Samorodnitsky and Taqqu (1994). This reference can be consulted for an introduction to stable distributions whose discussion is omitted in the present article. For a discussion of the different possible parameterizations and their special properties and advantages, the reader is referred to the standard reference (Zolotarev 1986). Secondly, the density f in equation (1) is indeed a continuous and bell-shaped probability density with a smooth distribution function and therefore the chosen name is justified. A further investigation reveals, that STS distributions actually form a six parameter distribution family. In other words, given an arbitrary STS distribution $S_\alpha^{[a,b]}(\sigma, \beta, \mu)$, then the parameters (ν_i, τ_i) of the two involved normal distributions are uniquely defined by the two equations (2) and (3). In order to provide the explicit formulas we need some notations. Let us denote by g_θ and G_θ the density and cumulative distribution function of the stable distribution with parameters $\theta = (\alpha, \beta, \sigma, \mu)$ which represents the center of the STS distribution $S_\alpha^{[a,b]}(\sigma, \beta, \mu)$. Let further $p_1 = G_\theta(a)$ and $p_2 = 1 - G_\theta(b)$ denote the the "cut-off-probabilities" defined in equation (3) and finally let the density and cumulative distribution function of the standard normal distribution be denoted as φ and Φ respectively. Then, the parameters (ν_i, τ_i) of the normal distributions describing the tails of the STS-distribution $S_\alpha^{[a,b]}(\sigma, \beta, \mu)$ can be obtained from the following two equations:

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

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

A derivation of formulas (4) and (5) is provided in Appendix A. The interpretation of the expression for the two standard deviations τ_1 and τ_2 is rather intuitive: τ_1 equals the ratio of two density values. In the numerator we have the value of the standard normal density evaluated at the p_1 quantile of a standard normal distribution. In the denominator we recognize the corresponding term for the stable distribution with parameter vector θ. The reader may notice that due to the relation $p_1 = G_\theta(a)$ the quantity a indeed equals the p_1-quantile of the stable distribution G_θ. The analogous argument works for τ_2.

A useful property of α-stable distributions – and normal distributions in particular – is their 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 that the random variable $Y = cX + d$ as an affine transform of the random variable X is again STS distributed, i.e. $Y \sim S_{\tilde{\alpha}}^{[\tilde{a},\tilde{b}]}(\tilde{\sigma}, \tilde{\beta}, \tilde{\mu}) \in \mathcal{S}$. The impact of the affine transformation on the distribution parameters is summarized in the following equation (the derivation is provided in Appendix A):

$$\tilde{a} = ca + d,\ \tilde{b} = cb + d,\ \tilde{\alpha} = \alpha,\ \tilde{\sigma} = |c|\sigma, \tag{6}$$

$$\tilde{\beta} = \operatorname{sign}(c)\beta,\ \ \tilde{\mu} = \begin{cases} c\mu + d & \alpha \neq 1 \\ c\mu - \frac{2}{\pi} c \log|c|\sigma\beta + d & \alpha = 1 \end{cases}. \tag{7}$$

Equations (6) and (7) identify the parameters a, b and μ as location parameters whereas σ serves a scale parameter in the case $\alpha \neq 1$.

Later in this paper, we will use standardized STS distributions to model the innovation process of a generalized GARCH stock price model. Therefore it is necessary, to have an efficient procedure to calculate the mean EX and the second moment EX^2 of a STS distributed random variable. The following two equations (8) and (9) provide expressions for the first two moments of an arbitrary STS distributed random variable X (the derivation is provided in Appendix A):

$$\begin{aligned} EX &= ap_1 - \tau_1\left(\Phi^{-1}(p_1)p_1 + \varphi(\Phi^{-1}(p_1))\right) + \dots \\ &\dots + \int_a^b x g_\theta(x)\, dx + \dots \\ &\dots + bp_2 + \tau_2\left(\Phi^{-1}(p_2)p_2 + \varphi(\Phi^{-1}(p_2))\right) \end{aligned} \tag{8}$$

$$\begin{aligned} EX^2 &= (\tau_1^2 + \nu_1^2)p_1 - \tau_1(a + \nu_1)\varphi(\Phi^{-1}(p_1)) + \dots \\ &\dots + \int_a^b x^2 g_\theta(x)\, dx + \dots \\ &\dots + p_2(\nu_2^2 + \tau_2^2) + \tau_2(\nu_2 + b)\cdot\varphi(\Phi^{-1}(p_2)) \end{aligned} \tag{9}$$

As before, φ denotes the density and Φ the distribution function of the standard normal distribution. $p_1 = G_\theta(a)$ and $p_2 = 1 - G_\theta(b)$ denote the "cut-off-probabilities" defined in equation (3) and g_θ (G_θ) is the density (cumulative distribution function) of the α-stable distribution with parameter-vector $\theta = (\alpha, \beta, \sigma, \mu)$. We will sometimes use relations (8) and (9) to effectively calculate the truncation levels a and b of standardized STS distributions. For practical applications, the subclass $\mathcal{S}_0 \subset \mathcal{S}$ of standardized STS distributions can be treated for all relevant parameter combinations as it was uniquely defined by the vector of stable parameters $\theta = (\alpha, \beta, \sigma, \mu)$ due to moment matching conditions. In the following, we casually calculate the truncation levels given the four stable parameters such that the resulting distribution is standardized. Figure 1 illustrates this procedure.

Insert Figure 1 somewhere around here

The graph in Figure 1 shows the influence of the distribution parameters $(\alpha, \sigma, \beta, \mu)$ on the truncation level a and b if they are determined by moment matching conditions such that the resulting STS distribution is standardized. Keeping the other stable parameters constant, we can see that the left truncation level a decreases and the right truncation level b increases monotonically with increasing α. This observation follows mathematical intuition. For small values of α, the α-stable distribution is extremely heavy-tailed. If we want the variance of the distribution to equal one, than the truncation of the heavy stable tails has to be carried out near to the mode of the distribution. If α increases, the truncation levels can move out into the tails. The effect of a change in the value of σ can easily be understood given the fact that σ represents the scale parameter of the stable distribution part. If σ increases and the other stable parameters are kept constant, the variation of the center distribution increases and therefore the truncation has to be accomplished near to the center to guarantee a variance of one.

Additionally to the first two moments discussed so far, STS distributions possess finite moments of arbitrary order and even exponential order. The latter property is important when log-returns of a financial asset are described by an STS distribution in order to guarantee a finite mean for the asset price and – as a consequence – finite prices for derivatives such as European options. Nevertheless, even if STS distributions possess thin tails in the mathematical sense, they are powerful in describing the distribution of financial variables which are typically leptokurtic and possibly admit skewness. Table 1 presents an ad hoc comparison of the left tail probabilities of a standard normal and an exemplary chosen standardized STS distribution. We emphasize the fact that both distribution have zero mean and unit variance. We observe that the STS distribution assigns significantly higher tail probabilities than the standard normal distribution (up to a factor of approximately 10^{17} for the probability of realizing a value smaller or equal to -10).

Insert Table 1 somewhere around here

One could argue that the standard normal distribution is not a challenging benchmark for comparing tail probabilities. Therefore, we continue the illustration by comparing the tail probabilities of standardized STS distributions with those of popular heavy-tailed distributions such as the standardized generalized extreme value distribution (GED), the standardized t distribution and the standardized skewed t distribution. Because an analytic comparison is not feasible, we have chosen the quantile-quantile plot as mean of illustration. The results are presented in Figure 2.

Insert Figure 2 somewhere around here

All distributions apart from the stable distribution which is included for pure comparison purposes possess zero mean and unit variance and the parameter values for the different distributions are chosen "reasonably", in the sense that they could arise from a time series estimation of financial data. The graph shows that the specific STS distribution assigns the highest tail probabilities within the examined range of 0.00001-quantiles to 0.1-quantiles. This result seems surprising as some of the distributions such as the standardized skewed t distribution possess much heavier tails in the mathematical sense than the STS distribution. Obviously the presented results depend on the specific chosen parameterization, and therefore we provide further insights by illustrating the influence of the distribution parameters α and β on the quantiles of the corresponding standardized STS distribution.

Insert Figure 3 somewhere around here

From Figure 3 we learn that the probability mass in the tails increases for increasing α – a result which seems to contradict the intuition. For the pure stable distribution the opposite relation holds true as in this case α measures the tail thickness. Less surprisingly, the left tail probabilities increase with increasing left-skewness.

So far we can state, that the family of STS distributions provides an impressive modeling flexibility and turns out to be a viable alternative against many popular heavy-tailed distributions. These observations raise the hope – which will be tested in the empirical part of this article – that a time series model based on STS distributed innovations may significantly improve the statistical fit to real financial data in comparison to models based on the normal or alternative non-normal distributions.

For practical applications the class of STS distributions needs to be generalized to the multivariate setting. This generalization is beyond the scope of the present article, but we outline one out of several plausible ways by mimicking the construction of the multivariate Gaussian distribution. In an initial step, let us consider a random vector $Y = (Y_1, Y_2, \ldots, Y_d)$ with independent components and each component follows an arbitrary standardized STS distribution, i.e. $Y_i \sim S_{\alpha_i}^{[a_i,b_i]}(\beta_i,\sigma_i,\mu_i) \in \mathcal{S}_0$. Next, consider a real matrix $A \in \mathbb{R}^{d\times d}$ and a real vector $b \in \mathbb{R}^d$ and define a general multivariate STS distributed random vector X by:

$$X = AY + b \tag{10}$$

The vector X has mean vector b and variance-covariance matrix $\Sigma = A \cdot A^*$, where A^* denotes the transpose of matrix A. Generally, the marginal distributions of X will no longer belong to the class of STS distributions. Nevertheless, a parameter estimation is possible on the basis of a two-step procedure where the generating vector Y is recovered from the observations of X with the help of a consistent estimate $\hat{\Sigma}$ for the variance-covariance matrix Σ by inverting the transformation in equation (10). This is always possible as long as matrix A possesses full rank or more precisely as long as the estimate for the variance-covariance matrix $\hat{\Sigma}$ is positive definite.

1.2. A Generalized NGARCH Option Pricing Model

In this subsection we introduce a general option pricing model containing most of the GARCH-stock price models proposed in the literature as particular cases. We enhance the existing models in a way that we allow for alternative i.e. non-Gaussian distributions in the innovation process while keeping the intuitive decomposition of the expected rate of return into risk free rate plus risk proportional risk premium. Formally, the log-returns of the underlying are assumed to follow the following dynamic under the objective probability measure P:

$$\log S_t - \log S_{t-1} = r_t - d_t + \lambda_t\sigma_t - g(\sigma_t) + \sigma_t\epsilon_t, \quad t \in \mathbb{N},\ \epsilon_t \overset{iid}{\sim} F. \tag{11}$$

S_t denotes the price of the underlying *ex* dividend at date t and r_t and d_t denote the continuously compounded risk free rate of return and dividend rate respectively for the period $[t-1, t]$. Both quantities as well as λ_t are assumed to be predictable, but can in general be modeled by separate stochastic processes. F denotes the marginal distribution of the innovation process and we assume that F is a standardized continuous probability distribution whose support equals the whole real line $\mathbb{R}$ and whose moment generating function m is finite. g represents the logarithmic moment generating function of F, i.e. we have $g(u) = \log \int \exp(ux)dF$. The conditional variance σ_t^2 is assumed to follow an asymmetric NGARCH$(1,1)$-process:[2]

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 (\epsilon_{t-1} - \gamma)^2 + \beta_1 \sigma_{t-1}^2, \quad t \in \mathbb{N}. \tag{12}$$

where we assume $\alpha_1(1+\gamma^2) + \beta_1 < 1$ in order to guarantee the existence of a strong stationary solution with finite unconditional mean (this is a natural extension of the classical condition obtained by Nelson (1990) for the GARCH(1,1) and by Bougerol and Picard (1992) for the GARCH(p, q) model, see e.g. Duan (1997) for details). The choice of this particular GARCH specification is motivated by the empirical findings of Christofferson and Jacobs (2004b) and Ritchken and Hsieh (2000). The former showed that the NGARCH model possesses the best out-of-sample option pricing performance among many different GARCH specifications whereas the latter article reports the superior out-of-sample performance of the NGARCH model with respect to the popular Heston-Nandi GARCH option pricing model (Heston and Nandi 2000).

From the definition (11) we can deduce some characteristic properties of the process dynamic for the underlying. First we mention, that if the distribution of the innovations F equals the standard Gaussian distribution and if we assume constant r, d and λ then equation (11) reduces to Duan's option pricing model (For $\gamma = 0$ we obtain the model introduced in (Duan 1995) and in the general case the one treated in (Duan and Wei 1999)). For $\alpha_1 = \beta_1 = 0$ the model boils down to the discrete time Black-Scholes model.

Similarly as in Duan (1995), other subsequent publications, and well-known from the Black-Scholes world, can the parameter λ_t be interpreted as the market price of risk for the period $[t-1, t]$. Therefore, the model specification allows for an economically meaningful decomposition of the expected one period rate of return μ_t. More formally, we have that the expected excess return is proportional to the amount of risk taken by the investor - measured by the standard deviation:

$$E_{t-1}\left(\frac{S_t^{cd}}{S_{t-1}}\right) = \exp(\underbrace{r_t + \lambda_t \sigma_t}_{=\mu_t}). \tag{13}$$

where S_t^{cd} denotes the *cum* dividend price of the underlying at time t. Equation (13) follows directly from equation (11), the definition of the function g and the predictability of σ_t. Equation (13) already explains the occurrence of the logarithmic moment generating function g in equation (11). It compensates for the nonlinear transformation between asset price and return and $-g(\sigma_t)$ reduces to the familiar $-\frac{1}{2}\sigma_t^2$ in the case of standard normally distributed innovations.

The following equation is well-known for the Gaussian innovation case and provides the link between the skewness of the innovation distribution, the parameter γ, and the correlation between today's innovation $\sigma_t \epsilon_t$ and tomorrow's conditional return variance σ_t^2 (the equation is proved in Appendix A):

$$Cov(\sigma_t \epsilon_t, \sigma_{t+1}^2) = \alpha_1 \cdot E\left(\sigma_t^3\right) \cdot \left(E\left(\epsilon_t^3\right) - 2\gamma\right), \quad t \in \mathbb{N} \tag{14}$$

Equation (14) will enable us to explain the leverage effect typically inherent to stock markets as reported by Christie (1982) and Black (1976) in two different ways. On the one hand, it can be controlled by the skewness of the white noise distribution and on the other hand by the asymmetry parameter γ.

For the pricing of derivatives we will need a risk-neutral version of the model specification (11). A simple left-shift by the amount of the market price of risk applied to the distribution of the innovations at every time step enables us to formulate the dynamic of model (11) in one possible risk-neutral world. The following proposition presents the natural generalization of this basic idea:

Proposition 2. *For any standardized probability distribution G which is equivalent to the marginal distribution F of the innovations ϵ_t, the distribution of the following process is equivalent to the NGARCH stock price model described in equations* (11) *and* (12)*:*

$$\log S_t - \log S_{t-1} = r_t - d_t - h(\sigma_t) + \sigma_t \xi_t, \quad \xi_t \overset{iid}{\sim} G, \tag{15}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 (\xi_{t-1} - \lambda_t - \frac{h(\sigma_{t-1})}{\sigma_{t-1}} + \frac{g(\sigma_{t-1})}{\sigma_{t-1}} - \gamma)^2 + \beta_1 \sigma_{t-1}^2. \tag{16}$$

The corresponding discounted total return process $(\bar{S}_t^{tr})_{t \in \mathbb{N}} = (e^{-\sum_{k=1}^{t} r_k} S_t^{tr})_{t \in \mathbb{N}}$ *with dynamic*

$$\log \bar{S}_t^{tr} - \log \bar{S}_{t-1}^{tr} = -h(\sigma_t) + \sigma_t \xi_t \tag{17}$$

is a martingale. The total return process – obtained by reinvesting the dividends – is given by $S_t^{tr} = e^{\sum_{k=1}^{t} d_k} S_t$

The important issue about this proposition is that it allows for different innovation distributions under the objective and the risk-neutral measure. In the case where efficient calibration procedures and reliable market prices of derivatives are available, it is possible to estimate the model parameters and the objective distribution F from time-series data whereas the risk-neutral distribution can be retrieved from market prices. It would be interesting to see, whether the result of Chernov and Ghysels (2000), who mainly state that for option pricing purposes time series data possesses virtually no information content would likewise apply. In the remainder of the present note, we will however focus on the special case, where F and G coincide and the change of measure takes the following somewhat simpler form:

Proposition 3.

1. *For a predictable sequence $(\lambda_t)_{t\in\mathbb{N}}$ of random variables let Q^λ denote the probability measure such that the distribution of the random variables $(\xi_t)_{t\in\mathbb{N}} = (\epsilon_t + \lambda_t)_{t\in\mathbb{N}}$ equals the one of $(\epsilon_t)_{t\in\mathbb{N}}$ under P, i.e. $\xi_t \overset{Q^\lambda}{\sim} F$. Then, the dynamic of the log-return under Q^λ possesses the following form:*

$$\log S_t - \log S_{t-1} = r_t - d_t - g(\sigma_t) + \sigma_t \xi_t, \quad t \in \mathbb{N},\ \xi_t \overset{iid}{\sim} F. \tag{18}$$

   and for the conditional variance we obtain:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 (\xi_t - \lambda_t - \gamma)^2 + \beta_1 \sigma_{t-1}^2, \quad t \in \mathbb{N}, \tag{19}$$

2. *The conditional variance remains unaffected by the change of measure:*

$$V_{t-1}^{Q^\lambda}(\log S_t - \log S_{t-1}) \overset{a.s.}{=} V_{t-1}^{P}(\log S_t - \log S_{t-1}) \tag{20}$$

3. *Under suitable regularity conditions, the unconditional variance under Q^λ will increase but still be finite: If the sequence $(\tilde{\lambda}_t)_{t\in\mathbb{N}} := (\lambda_t + \gamma)_{t\in\mathbb{N}}$ has a constant finite second moment $\tilde{\lambda}^2 := E^{Q^\lambda} \tilde{\lambda}_t^2$ which fulfills the condition*

$$\tilde{\lambda}^2 < \frac{1 - \alpha_1 - \beta_1}{\alpha_1} \tag{21}$$

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

$$V^{Q^\lambda}(\sigma_t \xi_t) = E^{Q^\lambda} \sigma_t^2 = \frac{\alpha_0}{1 - (1 + \tilde{\lambda}^2)\alpha_1 - \beta_1} \tag{22}$$

The proof of proposition 3 is similar to equivalent results stated by various other authors ((Duan 1995) being the first) and is provided in the appendix. We would like to emphasize that the measure Q^λ is indeed equivalent to P, which follows from the condition that the support of F equals whole $\mathbb{R}$. Similar arguments to those of Duan (1995) in the derivation of the locally risk-neutral valuation relationship could be found in order to "prove" that the risk-neutral dynamic in equation (18) is actually the "right" one for pricing derivatives. As it is discussed in Bates (2003), it is at least questionable whether the conditions imposed on the existence and the properties of the representative investor are realistic and can be applied to index option pricing. Another possibility which is open to the same criticism is to simply assume, that the market option prices follow some kind of Rubinstein-formula as done in Heston and Nandi (2000). We restrict our self to an empirical investigation of the usefulness of the presented change of measure: We will check the quality of the model under the objective probability measure and then verify whether the option prices obtained with the help of the risk-neutral dynamic as specified in equation (18) are able to explain the observed prices – at least partially.

2. Empirical Analysis

Similar to previous studies, we test the option pricing model presented in the previous section on the S&P 500 index market using S&P 500 index options. We are particularly interested in the following two questions, namely (1) how does the use of the STS distribution affect the statistical fit and the forecasting properties of the asymmetric NGARCH stock price model and (2) can the STS distribution provide a step to improve model's internal consistency? Therefore,we divide our statistical analysis into two parts: In the first part, we investigate the statistical properties under the objective probability measure P whereas the second part examines the model's ability to explain observed market prices of liquid options.

2.1. Data Description

The time series data consists of all 3784 S&P 500 closing prices from January 2, 1990 through May 4, 2005. Dividend data on a daily basis is extracted from the corresponding S&P 500 total return series available from the Chicago Board Options Exchange (CBOE). As approximation for the risk free rate we use appropriately interpolated T-Bill rates. Our set of market quotes for S&P 500 index calls consists of 4942 daily closing prices sampled on all 53 Wednesdays between May 5, 2004 and May 4, 2005 and including all traded options whose time to maturity on the specific Wednesday lays between 6 and 100 trading days (cf. Dumas, Fleming, and Whaley 1998). The option price sample contains all maturities between May 2004 and July 2005. As a consequence it is guaranteed that on each trading day under consideration there are at least three options "alive". As in Dumas, Fleming, and Whaley (1998), we take into account only options with a forward moneyness $K/F - 1$ between -0.1 and 0.1 where K denotes the dollar strike and F the forward price for the maturity of the option. Furthermore, all option prices have to obey the no-arbitrage condition derived from the put-call-parity (Merton 1973):

$$C_t(T, K) \geq e^{-d\cdot(T-t)} \cdot S_t - e^{-r(T-t)} \cdot K,$$

where $C_t(T, K)$ denotes the price at time t for a European call option with strike K maturing at time T and r denotes continuously compounded risk free rate for the lifetime of the option and d the corresponding dividend rate. Additionally, option prices violating the "convex in strike" condition are eliminated. Starting initially with 4942 price quotes we end up with 1920 remaining option prices after applying these filters. The average price in the remaining sample is $28.23. The average number of contracts taken into consideration on each trading day is 36 with a minimum number of 11 and a maximum number of 58.

2.2. Results under the objective probability measure

We are interested in answering the following question: To what extent can the use of an appropriate probability distribution for the innovation process improve the statistical fit and the forecasting properties of popular GARCH models for stock-returns? We run a comparison study, where the model candidates differ by their innovation distribution as well as by the specific GARCH form. Due to the general structure of the generalized NGARCH model (cf. equation (11) and (12) in section 1.), all models can be expressed as a special subtype of the generalized NGARCH model. To simplify the estimation procedure we impose a constant market price of risk λ. As candidates for the innovation distribution F, we will consider the normal distribution, the generalized error distribution (GED) and the class of STS distribution. From our analysis in subsection 1.1.1 it became clear that the skewed t distribution is the only standardized distribution which admits comparable tail probabilities as the STS distribution. Therefore, we take an additional model into consideration which is based on the skewed t distributed innovations. Due to the infinite moment generating function of skewed t distributions, the logarithmic moment generating function g in equation (11) will be replaced by σ_t^2. Even if this approach is internally inconsistent, it enables us to evaluate the relative performance of the STS distribution with respect to a more challenging benchmark. For all models, the conditional variance of the log-returns is assumed to follow a subtype of the asymmetric NGARCH model as introduced in equation (12). To emphasize the importance of the GARCH components (α_1, β_1) and especially the asymmetric GARCH component γ on the statistical fit, we will consider the constant conditional variance case $\alpha_1 = \beta_1 = 0$ and the symmetric GARCH model with $\gamma = 0$. In summary, we will examine and compare the following models:

DGBM : The model as given by equations (11),(12) with constant market price of risk λ, standard normally distributed innovations, and constant conditional variance ($\alpha_1 = \beta_1 = 0$).

Gaussian-GARCH : The model as given by equations (11),(12) with constant market price of risk λ, with standard normally distributed innovations and symmetric GARCH conditional variance ($\gamma = 0$).

Gaussian-NGARCH : The model as given by equations (11),(12) with constant market price of risk λ, with standard normally distributed innovations, and asymmetric NGARCH conditional variance.

GED-NGARCH : The model as given by equations (11),(12) with constant market price of risk λ, with standardized GED distributed innovations, and asymmetric NGARCH conditional variance.

Skewed-t-NGARCH : The model as given by equations (11),(12) with constant market price of risk λ, with standardized skewed t distributed innovations, and asymmetric NGARCH conditional variance. The logarithmic moment generating function in (11) is replaced by the one of a standard normal distribution.

STS-NGARCH : The model as given by equations (11),(12) with constant market price of risk λ, with standard STS distributed innovations, and asymmetric NGARCH conditional variance. This case is subdivided into two variants: First, we estimate the model with a fixed STS distribution (STS-NGARCH fixed), where the predetermined STS distribution has reasonably chosen parameter values. Second, we estimated a standardized STS distribution (STS-NGARCH estimated).

All models are estimated by a numerical maximum likelihood routine. We emphasize the fact that Gaussian quasi maximum likelihood methods are not applicable as the distributional assumption enters the dynamic of the conditional mean in form of its logarithmic moment generating function. For this reason, our estimation methodology is based on the following intuitive iteration procedure. Let us assume for a moment, that the parameters of the standardized distribution F governing the innovation process are known (this is only the case when we assume standard normally distributed innovations). If we denote by l the logarithm of the corresponding density function, then the standard argument leads to the following conditional log-likelihood function for the NGACRH stock price model with innovation distribution F:

$$\tilde{L}_{(x,y)}(\lambda, \alpha_0, \alpha_1, \beta_1, \gamma) = \sum_{t=1}^{T} l\left(\frac{y_t - r_t + d_t - \lambda\sigma_t + g(\sigma_t)}{\sigma_t}\right) - \frac{1}{2}\log(\sigma_t). \tag{23}$$

The conditional variance is recursively obtained from:

$$\sigma_t^2 = \alpha_0 + \alpha_1\sigma_{t-1}^2(\epsilon_{t-1} - \gamma)^2 + \beta_1\sigma_{t-1}^2. \tag{24}$$

$y_t = \log S_t - \log S_{t-1}$ denotes a series of log-returns r_t a corresponding series of risk free returns and d_t the corresponding dividend yields. As usual, the conditional likelihood function depends on the choice of the starting values ϵ_0 and σ_0, but for increasing sample size the impact of the starting values on the estimation results is negligible. Maximizing the conditional likelihood function leads to estimates for the unknown model parameters $(\lambda, \alpha_0, \alpha_1, \beta_1, \gamma)$. From these estimates we can recursively recover the time series of empirical residuals $(\epsilon_t)_{t=1,\ldots,T}$. The empirical distribution $\hat{F}$ of these residuals can now be used to update the distributional assumption F by estimating new distribution parameters. The estimation of the model parameters can now be repeated with the updated distribution and this gives raise to an iterative procedure. After each iteration step we calculate the Kolmogorov-Smirnov distance $d(F, \hat{F})$ between the distributional assumption and the empirical distribution of the residuals. We finish the iteration procedure and the estimation procedure ends as soon as this distance stops decreasing.

Insert Table 2 somewhere around here

Table 2 summarizes the estimation results for the seven different variants of the generic generalized NGARCH model. Judged purely on the log-likelihood, the second variant of the STS-NGARCH-model performs best. The second best performance is achieved by the inconsistent model which is based on the skewed t distribution followed by GED-NGARCH model. The first variant of the STS-NGARCH model where the innovation process is modeled by using a predetermined STS distribution provides a comparable fit as the GED-NGARCH model. A completely unsatisfactory fit is provided by the three models which are based on standard normally distributed innovations (DGBM, Gaussian GARCH and NGARCH) and these models are rejected by a likelihood ratio test. At the same time, the results show that the NGARCH model is superior to its classical counterpart - a result which has previously been recognized by other authors. We state without presenting all the estimation results that this fact generalizes to the alternative distributional assumptions. The difference in performance between GARCH and NGRACH reduces for the cases where the innovation distribution admits skewness. The reason is that in these cases the residual distribution is able to explain part of the leverage effect which can be explained by the parameter γ in the NGARCH model.

Judged on the distance between the empirical and the theoretical distribution of the residuals the picture changes slightly. The Kolmogorov-Smirnov test allows us to reject all Gaussian models and interestingly also the skewed t-NGARCH which provided the second best value for the log-likelihood. With p-values of 20% and higher, the GED-NGARCH model as well as the three variants of the STS-NGARCH model cannot be rejected. As the Kolmogorov-Smirnov statistic measures the uniform distance between two distribution functions, and as it might be of interest to test the model's ability to appropriately model extreme events, we also provide the corresponding values for the Anderson-Darling statistic which has its focus on the tails of the distribution. We use the following discrete version of the Anderson-Darling statistic AD, measuring the distance between the theoretical distribution function F and the empirical distribution $\hat{F}$:

$$AD(F,\hat{F}) = \sup_{x\in\mathbb{R}} \frac{|F(x) - \hat{F}(x)}{\sqrt{F(x)(1-F(x))}} \tag{25}$$

The results coincide with our intuition in the following sense: Models based on the STS distribution – which was designed to provide a flexible description for empirically observed tail probabilities – outperform all other distributions. Summarizing the results and repeating findings of various other authors, we can state the following: Albeit GARCH-models are well-known to be one of the best-performing models to describe the evolution of volatility, a satisfactory statistical fit can only be provided when the distribution of the innovation is non-Gaussian. In a recent study by Barone-Adesi, Engle, and Mancini (2004) the authors try to circumvent alternative distributional assumptions by applying a methodology called "filtering historical simulation", i.e. using the empirical distribution of the filtered historical residuals for simulation purposes. Christofferson and Jacobs (2004b) emphasize that the NGARCH model with Gaussian innovations works has the best option pricing performance but the filtered residuals differ strongly from the underlying normal assumption. As we have shown, the STS distributions are able to capture all important properties of the empirical residual distribution and can therefore help building a consistent model where the theoretical assumption coincides with the empirical distribution of the filtered residuals. Due to its modeling flexibility, the class of STS distributions turns out to be a viable alternative to other popular heavy-tailed distributions.

The following example will emphasize this fact. It deals with one of the most problematic events from a statistical viewpoint in the recent decades: The October crash in 1987. This event has the reputation to be “unexplainable” by every reasonable time series model. To prove the exceptional power of STS distributions in modeling extreme events, we compare the different NGARCH stock price models in their ability to “forecast the crash”. The setup for the comparison study is similar to the previous one: All six models (DGBM, Gaussian GARCH and NGARCH, GED-NGARCH, skewed t-NGARCH and STS-NGARCH) are fitted to the same data series consisting of 1000 S&P 500 log-returns preceding the crash and ending with the observation on Friday, October 16, 1987. On the next trading day, namely on the 19th of October, the S&P 500 dropped by more than 20%. Having estimated a specific time series model, we can express the drop which occured on Black Monday in terms of some implied realisation for the residual $\hat{\epsilon}_{Oct.19}$. Given that value of the “crash residual”, we can derive the model dependent implicit probability $\hat{p} = P(\epsilon \leq \hat{\epsilon}_{Oct.19})$ for such an event. More intuitively, we can express the information content of the implicit probability $\hat{p}$ by the quantity $\hat{n} = 1/(\hat{p} \cdot 252)$ which is the average time in years we have to wait for observing such an event under the specific model assumption. A summary of the results is reported in Table 3.

Insert Table 3 somewhere around here

First, we observe that from a statistical viewpoint, only the models with non-Gaussian innovations are competitive. The highest log-likelihood values are achieved by the GED-NGARCH and skewed t-NGARCH models directly followed by the two variants of the STS-NGARCH. None of these models can be rejected by the Kolmogorov-Smirnov test and the values of the Anderson-Darling statistic (0.06-0.08) are similar apart from the GED-NGARCH model whose value is slightly greater (0.15). Given the fact, that the skewed t-NGARCH model is inconsistent and only taken into account for comparison purposes, we can state that the STS-NGARCH models perform best. Now we focus on the ability to "forecast the crash". We observe that the model implicit probability for the Black Monday crash is practically zero in all Gaussian models and also for the GED-NGARCH model. The skewed t-NGARCH and the estimated STS-NGARCH models assign a probability which is by magnitudes better but still the average time of occurrence (800 and 5000 years) is by far bigger than the lifetime of the typical investor. Interestingly, the situation changes when we focus on the NGARCH model with a predetermined STS distribution. The model implicit probability equals approximately 0.0001 which in turn implies a mean time of occurrence of approximately 25 years.

2.3. Explaining S&P 500 Option Prices

In this section we examine the pricing performance of the STS-NGARCH option pricing model. For the reader's convenience, the following two equations recapitulate the specific dynamic under the objective probability measure P which we will assume throughout this section.

$$\begin{aligned} \log S_t - \log S_{t-1} &= r_t - d_t + \lambda\sigma_t - g(\sigma_t) + \sigma_t\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} S_\alpha^{[a,b]}(\sigma, \beta, \mu) \\ \sigma_t^2 &= \alpha_0 + \alpha_1\sigma_{t-1}^2(\epsilon_{t-1} - \gamma)^2 + \beta_1\sigma_{t-1}^2. \end{aligned}$$

This specific version differs from the general model by assuming a constant market price of risk λ and the special choice for the innovation distribution F. The meaning of the different parameters is the same as in section 1. Throughout the remainder of this article, we impose the especially simple form of the risk-neutral measure Q as discussed in proposition 3 which leads to the following process dynamic under Q:

$$\begin{aligned} \log S_t - \log S_{t-1} &= r_t - d_t - g(\sigma_t) + \sigma_t\xi_t, \quad \xi_t \overset{iid}{\sim} S_\alpha^{[a,b]}(\sigma, \beta, \mu) \\ \sigma_t^2 &= \alpha_0 + \alpha_1\sigma_{t-1}^2(\xi_{t-1} - \tilde{\gamma})^2 + \beta_1\sigma_{t-1}^2, \end{aligned}$$

where $\tilde{\gamma} = \gamma + \lambda$ is the asymmetry parameter under the risk neutral measure. This dynamic can be used to price any derivative security with underlying S by means of Monte Carlo simulation. The set of parameters which is needed to simulate paths of the STS-NGARCH model under Q consists of the six STS distribution parameters, the 4 risk-neutral model parameters and the spot variance h_0. We will analyze the following three estimation methodologies:

MLE : The STS distribution and the model parameters under the objective probability measure are estimated from a history of log-returns, dividend and interest rates by the same iterative maximum-likelihood procedure which we already described and applied in the previous subsection. The spot variance h_0 is obtained as a by-product of the estimation and the risk-neutral asymmetry parameter $\tilde{\gamma}$ is obtained as $\tilde{\gamma} = \gamma + \lambda$.

MLE/fitted : As in "MLE" all distribution and model parameters are estimated from past information. In a second step, we determine new values of $\tilde{\gamma}$ and h_0 by fitting the model prices to the available market prices of liquid options. This is done by minimizing the sum of squared dollar pricing errors over these two parameters.

NLS : This approach can be seen as complementary to the previous one. All model parameter values are obtained by minimizing the sum of squared dollar pricing errors between model and market prices. For every choice for the values of the model parameters, the value of the spot variance is filtered from the time series information. The STS distribution parameters are exogenously set to the average values which were obtained by approach "MLE".

As it is only based on time series information, the advantage of the "MLE" methodology is that it can even be applied in situations where no reliable market data of derivatives is available. The approach "MLE/fitted" is inspired by two empirical findings. First, the market price of risk parameter λ – which is known to be time varying in reality – is difficult to estimate and is very likely to be different for the life time of some derivative contract than for the estimation window. This creates reduced confidence in the estimated value based on past information and therefore it seems plausible to rely on an estimate which is based on the forward-looking information provided by option prices. The reader may notice at this point that determining $\tilde{\gamma} = \gamma + \lambda$ from option prices is actually equivalent to determining λ, if the value of γ is considered to be given. Second, the option prices generated with the NGARCH model are very sensitive to the value of spot volatility. The question is whether the "right value" can be better revealed from the time series information or from the option market? Filtering of the spot variance needs a long data history in order to minimize the starting value bias. Therefore, one could argue that current option prices contain more information on the current level of volatility than a return from five years ago. The third methodology "NLS" is equivalent to the approach of Heston and Nandi (2000). These authors enhanced the pure NLS-methodologies as e.g. applied in (Bakshi, Cao, and Chen 1997) by estimating all model parameters from the market prices but filtering the spot variance from a series of past log-returns. The underlying idea for this approach is that GARCH models are well-known as being excellent in describing the dynamic of volatility. Therefore, one hopes that a spot volatility estimate based on time series information by applying a GARCH filter could result in increased estimation accuracy. The disadvantage about this approach is that from the option prices only risk-neutral parameters can be determined whereas the available data is supposed to be generated from the objective distribution. As the values of the objective parameters cannot be retrieved from the risk-neutral ones (in the risk-neutral world, γ and λ are only observed through their sum as $\tilde{\gamma}$), the filtering of volatility will lead to a biased estimate for the spot variance. In our application we solve the identifiability problem by setting the market price of risk parameter equal to its time series estimate.

The three different estimation techniques are compared with respect to their in-sample and out-of-sample pricing performance. In the in-sample analysis, the estimation is carried out for each of the 53 Wednesdays in our data sample. For every Wednesday, the preceding 2500 data points of the daily log-return, dividend and interest rate series and the closing prices of all liquid options on this particular Wednesday are available for the estimation. The model option prices are determined by Monte Carlo simulation with 10'000 paths. We apply the empirical martingale simulation as introduced in (Duan and Simonato 1998), (Duan, Gauthier, and Simonato 2001) as variance reduction technique. The increased number of paths per simulation with respect to similar studies is to account for the STS distributed innovations. As empirical experiments have shown, the convergence of the option prices is slower than for the classical Gaussian GARCH model.

For the out-of-sample comparison, the models are only allowed to use the estimation result of the previous week plus the observations of the log-returns, dividend and interest rate series between the previous and the current Wednesday. Therefore, the model parameter estimates of last week are used to price the option contracts of this week. Only the spot variance can be updated from last week's spot variance by applying the GARCH filter to the five returns between the last and the current Wednesday. The out-of-sample analysis can only be carried out for the last 52 Wednesdays in our sample.

The results of the in-sample estimation are reported in Table 4. As the estimation is carried out on 53 different trading days, the Table contains the average parameter estimates. For the "MLE" approach, the individual estimate fluctuate very little around the reported average estimate. The "MLE/fitted" approach does only differ from "MLE" in the estimates for h_0 and $\tilde{\gamma}$. Even if the estimates for these two parameters do not possess the same temporal stability as their ML counterparts – which is certainly expected due to the heteroskedasticity prevalent in the return series – the optimization problem is still well-behaved. The situation changes for the "NLS" approach. The empirical distribution of the NGARCH-parameters turns out to be more or less bimodal. This phenomenon, which is connected to the difficult structure of the underlying optimization problem and the fact that our option sample for every single trading day is rather small, has already been reported by other authors (see e.g. Ritchken and Hsieh 2000). Apart from the numerical problems, the estimation results of the NLS approach differ strongly from their time series counterparts – a phenomenon which is well-known and usually denoted as internal inconsistency (cf. Bates 1996). Therefore, it will be interesting to see how the different approaches behave, when we start comparing their ability to explain observed option prices.

Insert Table 4 somewhere around here

So far we have only talked about different estimation techniques for one and the same model. Obviously, we also need an objective benchmark which allows us to judge the absolute performance of the STS-NGARCH option pricing model. As other authors before, we choose the Ad-Hoc-Black-Scholes (AHBS). Although this model is purely descriptive and does not try to explain any aspect of the observed market prices, it is known since Dumas, Fleming, and Whaley (1998) that this model is a challenging benchmark. The AHBS model is basically a standard Black-Scholes model, but the variance which is used to price a specific option may depend on strike and time to maturity of the option. Specifically, we assume the following parametric form for the implied volatility surface:

$$\sigma^2(m,\tau) = \delta_0 + \delta_1 \cdot m + \delta_2 \cdot \tau + \delta_3 \cdot m^2 + \delta_4 \cdot m \cdot \tau + \delta_5 \cdot \tau^2, \tag{26}$$

where $m = K/F$ denotes the forward moneyness and τ the time to maturity expressed as a fraction of a year. For the in-sample evaluation the six parameters are determined by minimizing the sum of square dollar pricing errors using all available market prices on the specific trading day. For the out-of-sample pricing comparison the parameter estimates from the last week are used for pricing.

The resulting root mean squared dollar pricing errors (absolute RMSE) and root mean squared relative pricing errors (relative RMSE) for the different models in the in-sample and out-of-sample comparison are provided by Table 5. The greater values of the absolute RMSE of all models compared to similar studies can be explained by our specific data set: Due to the poor quality of the out-of-the-money option prices, most of our market quotes are at-the-money or in-the-money. The absolute valuation errors for options with greater prices tend to be greater and this explains the increased values of the RSME. However, the numbers reported for the relative RSME are in accordance with similar studies. It can be seen that in-sample, the AHBS model outperforms all three variants of the STS-NGARCH model. These result are similar to other studies where the AHBS model performs always at least as good as its GARCH competitor. However, out-of-sample the MLE/fitted model performs best. Interestingly, the NLS variant performs worse than the MLE/fitted in-sample as well as out-of-sample. This can be seen as an indicator for our new estimation approach. It seems that the value for the market price of risk as well as the spot variance can be retrieved from option prices in a more reliable way than by filtering it from the return series. This might partially be due to the identification problem which arises in the NLS approach. As mentioned before, the values of λ and γ cannot be separately identified from option prices, but only their sum in from of $\tilde{\gamma}$. Another interesting issue about the pricing performance results is that the ordering of the different models in the out-of-sample comparison, depends on the performance criteria. Whereas for the absolute RMSE the pure "MLE" approach performs second worst, it it the second best with respect to the relative RSME. Given the fact that the parameters of the AHBS and "NLS" model are determined by minimizing the absolute RSME, these results are not too surprising. A recent study of (Christofferson and Jacobs 2004a), where the authors report about the impact of the loss function used for the estimation of the model parameters and its impact on the out-of-sample performance, supports these findings. The already mentioned problem with the NLS approach is that the estimated parameter values don not only vary through the different trading dates but also depend significantly on the loss function used for the minimization procedure. A similar effect has been reported by Ritchken and Hsieh (2000) who also mentioned the complex structure of the minimization problem. Given the competitive out-of-sample performance of the "MLE/fitted" and even the pure "MLE" approach together with the temporal stability of their model parameter estimates, we believe that these approaches should be revisited in future research.

Insert Table 4 somewhere around here

3. Conclusion

This article introduces the class of smoothly truncated stable distribution (STS distribution) designed to meet the needs of modeling financial data. STS distributions combine the modeling power of stable distributions with the appealing property of finite moments of arbitrary order. In short, a STS distribution is obtained by replacing the fat tails of the stable density by the thins tails of two appropriately chosen normal distributions. Furthermore, we enhance the classical NGARCH option pricing model in a way that it allows for non-normal innovation distributions. In this sense we create an time series model, which combines the three major characteristics of financial data: volatility clustering, leverage effect and conditionally leptokurtic returns. The pricing measure is constructed in a canonical way and is driven by only by dynamic of the market price of risk. In its most general case, the model allows for a different distribution of the innovations under the risk-neutral and the objective probability measure, a property which could be helpful on the way towards an internally consistent option pricing model.

The empirical investigation of the model leads to encouraging results. First, we show that the use of non-Gaussian innovation distributions significantly improves the statistical fit of the NGARCH model under the objective probability measure. The result are exemplified by means of the October crash 1987: A special variant of the STS-NGARCH model which uses a predetermined STS distribution for the innovation process is the only process in our study which assigns a realistic probability to the event which took place on October 19, 1987 and made the S&P 500 drop by more than 20%. The second part of the empirical part is dedicated to the analysis of the model's ability to explain market prices of S&P 500 index options. We compare three varaints of the STS-NGARCH option pricing model with an *ad hoc* Black-Scholes model where the variance used to price the option is determined from a parameteric fit to the implied volatility surface. Interestingely, we find that the STS-NGARCH model where the model parameters are estimated from time series information and only the market price of risk and the spot variance are inverted from option prices performs best on an out-of-sample basis. These findings could be a basis for future research to investigate the information content provided by time series information and option prices with respect to the different parameters of GARCH models.

4. Figure Legends

The Figure 1 shows how the parameters α, β and σ influence the truncation levels of standardized STS distributions. The truncation levels a and b are calculated by moment matching conditions from the stable parameters $(\alpha, \beta, \sigma, \mu)$ such that the resulting distribution is standardized. The graph shows in the upper (lower) part the variation of the right truncation level b (the left truncation level a) depending on α. The different branches of the plot correspond to different combinations of the skewness parameter β ($\beta = 0, -0.1, -0.2$) and the scale parameter σ ($\sigma = 0.55, 0.6$). The magnitude $|a|$ and $|b|$ of the truncation levels is increasing in α and decreasing in σ. We have $|a| > |b|$ when $\beta < 0$, i.e. when the center distribution admits left-skewness.

Figure 2 contains a plot of the 0.00001−0.1 quantiles of a standard normal distribution against the corresponding quantiles of various standardized probability distributions. In detail, we consider a standard normal distribution, a generalized extreme value distribution with parameter $\nu = 1.3$, a standardized t distribution with $n = 6$ degrees of freedom, a standardized skewed t distribution with $n = 6$ degrees of freedom and a skewness parameter of $\lambda = -0.5$ and the previously used standardized STS distribution $S_{1.85}^{[-5.92,3.33]}(0.6, -0.1, 0)$. For comparison purposes we have also included the corresponding stable distribution with parameter $(1.85, -0.1, 0.6, 0)$, which is not standardized but possesses an infinite variance. The exhibit aims to exemplify the fact that STS distributions are light-tailed distributions in the mathematical sense which nevertheless can possess extreme quantile functions. The illustration provides only one example and the general behavior depends on the parameterization of the STS as well as the other distributions.

Figure 3 contains a plot of the 0.00001−0.1 quantiles of a standard normal distribution against the corresponding quantiles of different standardized STS distributions. In Panel (a) we consider the distribution $S_\alpha^{[a,b]}(0.6, -0.1, 0)$ with different values for α while fixing the parameters σ, β and μ. Given the specific value for α, the truncation levels a and b are determined such that the resulting distribution is standardized. In Panel (b) we consider the distribution $S_{1.85}^{[a,b]}(0.6, \beta, 0)$ with different values for β. Again, for given β, the truncation levels a and b are determined such that the resulting distribution is standardized.

Appendix A. Proofs

Proof of some statements in section 1:

1. **Equations** (4) **and** (5) **for the parameters of the normal distributions describing the tails:**

   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 cumulative distribution function and φ the density of the standard normal distribution. From equation (3) we have:

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

   Inserting this relation into equation (2) yields:

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

   Finally, we have:

$$\tau_1 = \frac{\varphi\left(\Phi^{-1}(p_1)\right)}{g_\theta(a)} \quad \text{and} \quad \nu_1 = a - \tau_1\Phi^{-1}(p_1) \tag{A1}$$

and similar calculations prove equation (5):

$$\tau_2 \;=\; \frac{\varphi\left(\Phi^{-1}(p_2)\right)}{g_\theta(b)} \quad \text{and} \quad \nu_2 \;=\; b + \tau_2\Phi^{-1}(p_2) \tag{A2}$$

2. **Translation and scale invariance for STS distributions:**

Let $X \sim S_\alpha^{[a,b]}(\sigma, \beta, \mu)$ and for $d \in \mathbb{R}$ and $\sigma > 0$ define $Y = cX + d$. Then we have:

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

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 $\tilde{g}_\theta$ 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, which completes the proof.

3. **Formulae** (8) **and** (9) **for the first moments:**

Denoting as previously the density of a STS distributed random variable 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

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

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

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

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

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

q.e.d.

Proof of equation (14)**:** From the predictability of $\sigma_k,\ k = 1,\dots,t$ and $\epsilon_k,\ k = 1,\dots,t-1$ we deduce with the help of the law of iterated expectations:

$$
\begin{aligned}
Cov\left(\sigma_t\epsilon_t, \sigma_{t+1}^2\right) &= E\left(\sigma_t\epsilon_t\sigma_{t+1}^2\right) - \underbrace{E(\sigma_t\epsilon_t)}_{=0} \cdot E(\sigma_{t+1}^2) \\
&= E\left(\sigma_t\epsilon_t(\alpha_0 + \alpha_1\sigma_t^2(\epsilon_t - \gamma)^2 + \beta_1\sigma_t^2)\right) \\
&= E\left(\alpha_0\sigma_t\epsilon_t + \alpha_1\sigma_t^3(\epsilon_t^3 - 2\gamma\epsilon_t^2 + \gamma^2\epsilon_t) + \beta_1\sigma_t^3\epsilon_t\right) \\
&= \alpha_0 E(\sigma_t \underbrace{E_{t-1}(\epsilon_t)}_{=0}) + \alpha_1 E(\sigma_t^3\,(E_{t-1}\epsilon_t^3 - 2\gamma \underbrace{E_{t-1}\epsilon_t^2}_{=1} + \gamma^2 \underbrace{E_{t-1}\epsilon_t}_{=0})) + \dots \\
&\quad \dots + \beta_1 E(\sigma_t^3 \underbrace{E_{t-1}(\epsilon_t)}_{=0}) \\
&= \alpha_1 \cdot E(\sigma_t^3) \cdot \left(E(\epsilon_t^3) - 2\gamma\right).
\end{aligned}
$$

q.e.d.

Proof of Proposition 2 in section 1: In the discrete time framework, it is enough to specify the change of measure for the conditional distribution one step ahead. The new dynamic can be obtained by replacing the random variable ϵ_t by $\xi_t - \frac{h(\sigma_t)}{\sigma_t} + \lambda_t - \frac{g(\sigma_t)}{\sigma_t}$. Under the assumption made in the proposition, this algebraic rearrangement is indeed expressible by an equivalent change of measure, which follows from the fact that the distribution F of ϵ_t is equivalent to the distribution G of ξ_t and from the fact that the support of both F and G equals the whole real line.

The martingale property of the discounted total return process with risk neutral dynamic

$$\log \bar{S}_t^{tr} - \log \bar{S}_{t-1}^{tr} = -h(\sigma_t) + \sigma_t \xi_t, \quad t \in \mathbb{N}, \ \xi_t \overset{iid}{\sim} G.$$

is obtained as follows. If we use the predictability of σ_t and the definition of h as logarithmic moment generating function corresponding to ξ_t we get:

$$\begin{aligned} E_{t-1}^Q(\bar{S}_t^{tr}) &= E_{t-1}^Q(\bar{S}_{t-1}^{tr} \ \exp(-h(\sigma_t) + \sigma_t \xi_t)) \\ &= \bar{S}_{t-1}^{tr} \cdot \underbrace{\exp(-h(\sigma_t))}_{=1/m(\sigma_t)} \cdot \underbrace{E_{t-1}^Q\left(\exp(\sigma_t \xi_t)\right)}_{=m(\sigma_t)} = \bar{S}_{t-1}^{tr}, \end{aligned}$$

where m denotes the moment generating function of the distribution G. As $(\bar{S}_t^{tr})_{t \in \mathbb{N}_0}$ is a discrete stochastic process the martingale property follows from the law of iterated expectations.

q.e.d.

Proof of Proposition 3 in section 1:

1. The dynamic under the new probability measure Q^λ is simply obtained by replacing ϵ_t through $\xi_t - \lambda$.

2. Similarly, the equality of the conditional variances follows directly from the fact

$$V^{Q^\lambda}(\epsilon_t) = 1 = V^P(\epsilon_t).$$

3. From the fact that $\tilde{\lambda}_t^2 = (\lambda_t + \gamma)^2$ and σ_t^2 are uncorrelated and independent of ξ_t we deduce:

$$
\begin{aligned}
E^{Q^\lambda}\sigma_t^2 &= \alpha_0 + \alpha_1 E^{Q^\lambda}(\sigma_{t-1}^2(\xi_{t-1} - \tilde{\lambda}_{t-1})^2) + \beta_1 E^{Q^\lambda}\sigma_{t-1}^2 \\
&= \alpha_0 + \alpha_1 E^{Q^\lambda}(\sigma_{t-1}^2 E_{t-1}^{Q^\lambda}(\xi_{t-1}^2 - 2\xi_{t-1}\tilde{\lambda}_{t-1} + \tilde{\lambda}_{t-1}^2)) + \ldots \\
&\quad \ldots + \beta_1 E^{Q^\lambda}\sigma_{t-1}^2 \\
&= \alpha_0 + \alpha_1 E^{Q^\lambda}(\sigma_{t-1}^2(1 + \tilde{\lambda}_{t-1}^2)) + \beta_1 E^{Q^\lambda}\sigma_{t-1}^2 \\
&= \alpha_0 + \underbrace{\alpha_1(1 + \tilde{\lambda}^2)}_{=:\alpha^*} E^{Q^\lambda}\sigma_{t-1}^2 + \beta_1 E^{Q^\lambda}\sigma_{t-1}^2
\end{aligned}
$$

Under the assumptions made on the parameters $\tilde{\lambda}$, α_1 and β_1 the stationary solution is given by:

$$
E^{Q^\lambda}\sigma_t^2 = \frac{\alpha_0}{1 - \alpha_1^* - \beta_1},
$$

q.e.d.

References

Bakshi, G., C. Cao, and Z. Chen, 1997, "Empirical Performance of Alternative Option Pricing Models," *Journal of Finance*, 52, 2003–2049.

Bakshi, G., and Z. Chen, 1997, "An Alternative Valuation Model for Contingent Claims," *Journal of Financial Economics*, 44, 123–165.

Barone-Adesi, G., R. F. Engle, and L. Mancini, 2004, "GARCH Options in Incomplete Markets," NCCR-FinRisk Paper, University of Zürich, Paper No. 155.

Bates, D., 1996, "Testing Option Pricing Models," in *North-Holland Handbooks of Statistics*, ed. by G. S. Maddala, and C. R. Rao. Elsevier, Amsterdam, New York, vol. 9, pp. 567–611.

——— , 2003, "Empirical Option Pricing: A Retrospection," *Journal of Econometrics*, 116, 387–404.

Black, F., 1976, "Studies of Stock Price Volatility Changes," in *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.

Bougerol, P., and N. Picard, 1992, "Stationarity of GARCH Processes and of some Nonnegative Time Series," *Journal of Econometrics*, 52, 115–127.

Carr, P., H. Geman, D. B. Madan, and M. Yor, 2003, "Stochastic Volatility for Lévy Processes," *Mathematical Finance*, 13, 345–382.

Carr, P., and L. Wu, 2004, "Time-Changed Lévy Processes and Option Pricing," *Journal of Financial Economics*, 71, 113–141.

Chernov, M., and E. Ghysels, 2000, "A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Option Valuation," *Journal of Financial Economics*, 56, 407–458.

Christie, A. A., 1982, "The Stochastic Behaviour of Common Stock Variances: Value, Leverage and Interest Rate Effects," *Journal of Financial Economics*, 10, 407–432.

Christofferson, P., S. L. Heston, and K. Jacobs, 2004, "Option Valuation with Conditional Skewness," working paper, EFA 2004 Maastricht Meetings Paper No. 2964, forthcoming in: *Journal of Econometrics*.

Christofferson, P., and K. Jacobs, 2004a, "The Importance of the Loss Function in Option Valuation," *Journal of Financial Economics*, 72, 291–318.

——— , 2004b, "Which GARCH Model for Option Valuation," *Management Science*, 50, 1204–1221.

Duan, J.-C., 1995, "The GARCH Option Pricing Model," *Mathematical Finance*, 5, 13–32.

——— , 1997, "Augmented GARCH(p, q) Processes and its Diffusion Limit," *Journal of Econometrics*, 79, 97–127.

Duan, J.-C., G. Gauthier, and J.-G. Simonato, 2001, "Asymptotic Distribution of the EMS Option Price Estimator," *Management Science*, 47, 1122–1132.

Duan, J.-C., P. Ritchken, and Z. Sun, 2004, "Jump Starting GARCH: Pricing and Hedging Options with Jumps in Returns and Volatilities,," working paper, University of Toronto and Case Western Reserve University.

Duan, J.-C., and J.-G. Simonato, 1998, "Empirical Martingale Simulation for Asset Prices," *Management Science*, 44, 1218–1233.

Duan, J.-C., and J. Wei, 1999, "Pricing Foreign Currency and Cross-Currency Options under GARCH," *Journal of Derivatives*, 7, 51–63.

Duffie, D., J. Pan, and K. J. Singleton, 2000, "Transform Analysis and Asset Pricing for Affine Jump Diffusions," *Econometrica*, 68, 1343–1376.

Dumas, B., J. Fleming, and R. E. Whaley, 1998, "Implied Volatility Functions: Empirical Tests," *Journal of Finance*, 53, 2059–2106.

Engle, R. F., and V. K. Ng, 1993, "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1749–1778.

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.

Heston, S. L., and S. Nandi, 2000, "A Closed-Form GARCH Option Valuation Model," *The Review of Financial Studies*, 13, 585–625.

Menn, C., 2004, *Optionspreistheorie: Ein ökonometrischer Ansatz*. Dr. Kovač Verlag, Hamburg, Doctoral Thesis at the University of Karlsruhe, Germany.

Menn, C., and S. T. Rachev, 2005, "A GARCH Option Pricing Model with α-Stable Innovations," *European Journal of Operations Research*, 163, 201–209.

Merton, R. C., 1973, "The Theory of Rational Option Pricing," *Bell Journal of Economics and Management Sciences*, 4, 141–183.

Nelson, D. B., 1990, "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6, 318–334.

Ritchken, P., and K. C. Hsieh, 2000, "An Empirical Comparison of GARCH Option Pricing Models," Technical Memorandum 734, Case Western Reserve University.

Samorodnitsky, G., and M. S. Taqqu, 1994, *Stable Non-Gaussian Random Processes.* Chapmann & Hall/CRC, Boca Raton.

Zolotarev, V. M., 1986, *One-dimensional Stable Distributions.* American Mathematical Society, Providence.

Notes

[1]The present definition of STS distributions extends earlier versions which were presented in Menn and Rachev (2005) and Menn (2004). The reader may notice that the suggested methodology of "smooth truncation" can be generalized in various obvious ways. Either the tail distribution, as well as the center distribution could be chosen differently.

[2]The restriction to the case $p = q = 1$ has mainly been made for practical reasons: First, it seems sufficient for our purposes and second it simplifies the exposition. The generalization to the NGARCH(p, q) case or different GARCH variants is straightforward.

Table 1

Comparison of $P(X \leq x)$ for the Standard Normal and a Standardized STS Distribution.

Argument x	Probability $P(X \leq x)$	
	$X \sim \mathcal{N}(0,1)$	$X \sim S_{1.85}^{[-5.92,3.33]}(0.6,-0.1,0)$
-10	$7.620e-24$	0.0002840
-9	$1.129e-19$	0.0004099
-8	$6.221e-16$	0.0005860
-7	$1.280e-12$	0.0008299
-6	$9.866e-10$	0.001164
-5	$2.867e-07$	0.001679
-4	$3.167e-05$	0.002684
-3	0.001350	0.005307
-2	0.02275	0.01889
-1	0.1587	0.1236

The Table exemplifies the difference in the left tail probabilities $P(X \leq x)$ of a specific standardized STS distribution and the standard normal law. The STS distribution exhibits significantly higher tail probabilities than the standard normal law.

Table 2

Maximum Likelihood Estimation and Model Comparison under P.

Model	Model Parameters					Characteristics/Statistics		
	λ	α_0	α_1	β_1	γ	LL	KS	AD
DGBM	0.029	0.00011	-	-	-	11930	1.6e-10	283
$(N(0,1))$	(0.016)	(1.5e-06)						
Gaussian-GARCH	0.065	5.3e-07	0.057	0.94	-	12423	4.3e-05	198
$(N(0,1))$	(0.016)	(1.2e-07)	(4.5e-03)	(4.8e-03)				
Gaussian-NGARCH	0.029	1.1e-006	0.054	0.89	0.94	12479	4.7e-04	51.4
$(N(0,1))$	(0.016)	(1.3e-06)	(4.7e-03)	(7.1e-03)	(0.10)			
GED-NGARCH	0.041	9.0e-07	0.053	0.89	0.93	12530	0.7304	0.636
$(GED(1.45))$	(0.015)	(1.7e-07)	(6.9e-03)	(0.010)	(0.14)			
Skewed t-NGARCH	0.033	8.9e-07	0.054	0.89	0.92	12537	0.0063	0.136
(skewed-$t(8.56, -0.05)$)	(0.016)	(1.7e-07)	(7.1e-03)	(0.010)	(0.13)			
STS-NGARCH (fixed)	0.033	1.1e-06	0.067	0.89	0.85	12525	0.3387	0.0632
$(S_{1.85}^{[-5.94,3.33]}(0.6, -0.1, 0))$	(0.014)	(2.0e-07)	(8.5e-03)	(0.010)	(0.12)			
STS-NGARCH (estimated)	0.035	9.8e-07	0.063	0.90	0.84	12541	0.2253	0.0364
$(S_{1.84}^{[-3.30,1.16]}(0.65, 0, -0.1))$	(0.014)	(1.8e-07)	(7.5e-03)	(9.3e-03)	(0.11)			

The Table reports the estimation results for seven different generalized NGARCH models. The parameter λ represents the estimated market price of risk and $\alpha_0, \alpha_1, \beta_1$ are the parameters of the GARCH(1,1) process. For the NGARCH variants, we have additionally the asymmetry parameter γ. The asymptotic standard errors for the model parameter estimates are provided in parenthesis. The prefix of every model identifies the distribution of the innovation process and DGBM stands for discrete geometric Brownian motion. The residual distribution – and where necessary the estimated distribution parameters – are provided in parenthesis below the model's name. The column "LL" reports the log-likelihood value which is achieved for the given parameter estimates and the column "KS" reports the p-value for the Kolmogorov-Smirnov test with the zero hypothesis that the empirical distribution of the model's residuals equals the distributional assumption for the innovations. The column "AD" reports the values of the Anderson-Darling distance between the assumed and the empirical residual distribution.

Table 3

Model Comparison: Crash Probabilities.

Model	Distribution	Estimation Statistics			Crash Forecasting		
		LL	KS	AD	residual $\hat{\epsilon}_{Oct,19}$	probability $\hat{p}$	mean time $\hat{n}$
DGBM	$N(0,1)$	3336	0.01371	63.11	-26.70	2.523e-157	1.57e+154
Gaussian-GARCH	$N(0,1)$	3365	0.004088	4.386	-13.71	4.35e-43	9.128e+39
Gaussian-NGARCH	$N(0,1)$	3367	0.01116	3.671	-12.22	1.19e-34	3.347e+31
GED-NGARCH	$GED(1.31)$	3391	0.9837	0.157	-12.00	1.93e-12	2.05e+09
Skewed t-NGARCH	skewed-$t(5.7, 0.046)$	3391	0.8908	0.06745	-11.45	4.59e-06	863.9
STS-NGARCH (fixed)	$S_{1.85}^{[-5.94,3.33]}(0.6, -0.1, 0)$	3383	0.2421	0.08161	-11.52	1.60e-04	24.82
STS-NGARCH(estimated)	$S_{1.70}^{[-2.88,2.00]}(0.57, 0, 0)$	3387	0.6203	0.07944	-11.70	7.89e-07	5032

We compares the ability of seven different generalized NGARCH models to forecast the stock market crash in October '87. The different models are fitted to a time series of 1000 S&P 500 log-returns ending the day before the crash. The estimated model parameters are omitted for the ease of exposition. The column "Distribution" reports the assumed distribution for the innovation process together with the estimated distribution parameters. The column "LL" reports the log-likelihood value which is achieved with the specific model and the column "KS" reports the p-value for the Kolmogorov-Smirnov test with the zero hypothesis that the assumed distribution for the innovations equals the empirical distribution of the model's residuals. The column "AD" reports the values of the Anderson-Darling distance between the assumed and the empirical residual distribution. Given the estimated model parameters it is possible to back out the model implicit residual $\hat{\epsilon}_{Oct,19}$ for October 19 from the actual observation. The result is reported in the column "$\hat{\epsilon}_{Oct,19}$", the column entitled $\hat{p} = P(\epsilon \leq \hat{\epsilon}_{Oct,19})$ represents the corresponding crash probability under the distributional assumption of the specific model and $\hat{n} = 1/(\hat{p} \cdot 252)$ represents the mean time of occurrence in years for an event at least as bad as what happened on Black Monday.

Table 4

In-Sample Parameter Estimates.

Model	Distribution	Model Parameters				Volatility		
		α_0	α_1	β_1	$\tilde{\gamma}$	h_0	σ^P	σ^Q
STS-NGARCH (MLE)	$S_{1.87}^{[-2.05,2.08]}(0.67, 0, 0)$	1.903e-006	0.06501	0.8447	1.141	0.1270	0.2091	0.2906
STS-NGARCH (MLE/fitted)	$S_{1.87}^{[-2.05,2.08]}(0.67, 0, 0)$	1.903e-006	0.06501	0.8447	1.120	0.1159	0.2091	0.2352
STS-NGARCH (NLS)	$S_{1.87}^{[-2.05,2.08]}(0.67, 0, 0)$	1.413e-005	0.1078	0.5092	1.8601	0.1091	–*	0.1231

The Table reports the estimation results for the STS-NGARCH model with different estimation techniques. $\alpha_0, \alpha_1, \beta_1$ are the parameters of the NGARCH(1,1) process. The asymmetry parameter under the risk neutral measure Q is denoted by $\tilde{\gamma}$ and consists of the sum of the market price of risk parameter λ and the asymmetry parameter γ under the objective probability measure P. h_0 denotes the estimated spot variance, $\sigma^P := \sqrt{251 \cdot \alpha_0/(1 - (1+\gamma^2)\alpha_1 - \beta_1)}$ is the annualized volatility under the objective probability measure derived from the stationary level of local variance and σ^Q the corresponding risk-neutral annualized volatility. "MLE" refers to a pure maximum-likelihood approach where all parameters are estimated from a time series of length 2500. "MLE/fitted" is a hybrid approach where only the spot variance h_0 and the market price of risk λ are estimated from market option prices. "NLS" determines all model parameters – except of the spot variance which is filtered from past log-returns – by a non-linear least square optimization procedure from market option prices. All models are estimated for all 53 Wednesdays in our data sample. The Table reports the average values of these 53 estimates.

*: The stationary level of variance under the objective probability measure can not be uniquely identified from the set of risk-neutral parameters.

Table 5

Model Comparison: In- and Out-of-Sample Pricing Performance

Model	**In-Sample**		**Out-of-Sample**	
	abs. RMSE	rel. RMSE	abs. RMSE	rel. RMSE
Black-Scholes	9.10	8.5%	9.11	8.2%
STS-NGARCH (MLE)	4.13	3.1%	4.10	3.1%
STS-NGARCH (MLE/fitted)	2.29	1.4%	3.17	2.9%
STS-NGARCH (NLS)	2.40	1.7%	3.58	3.7%
Ad-Hoc-BS	1.54	1.1%	3.93	3.2%

The Table reports the root mean absolute and relative pricing error over all 1920 option quotes in the sample. "In-sample" means that every model is estimated/fitted on every of the 53 Wednesdays in the sample and the market prices of the specific trading day enter in the parameter estimation. "Out-of-sample" means that the parameter estimates from the last week are used for pricing. "Black-Scholes" stands for the classical Black-Scholes model, the STS-NGARCH models differ by the estimation methodology: "MLE" is only based on time series information, "MLE/fitted" uses mainly the time series estimates but determines the market price of risk λ and the spot variance h_0 from option prices whereas "NLS" inverts the risk neutral model parameters from observed market prices but filters the spot variance from the historic time series. "Ad-Hoc-BS" is a benchmark model which uses a parametric approximation to the implied volatility surface depending on six parameters.

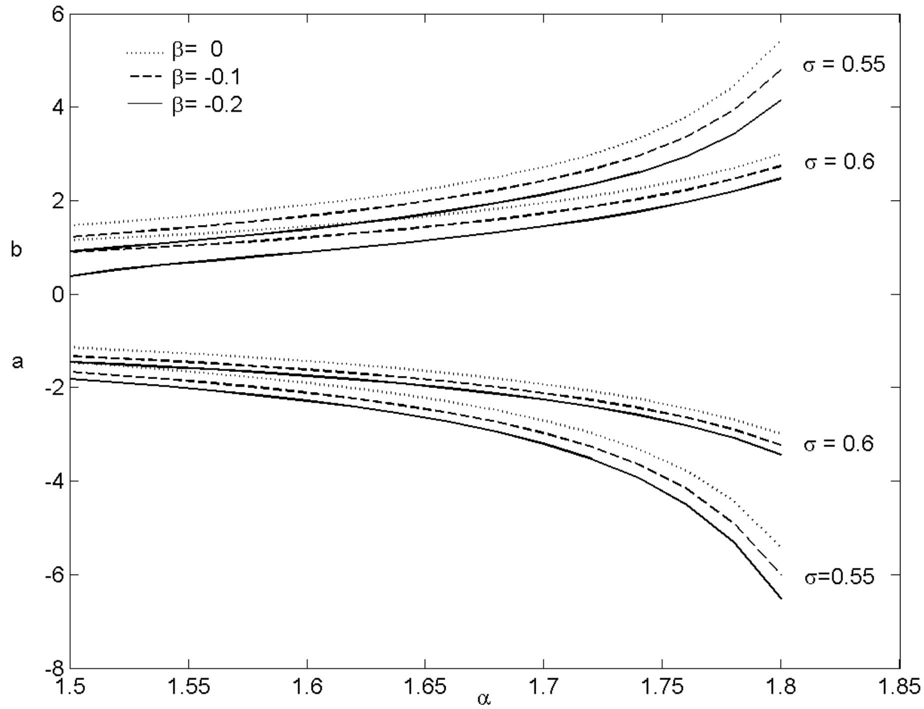


Figure 1. Truncation Levels of Standardized STS Distributions.

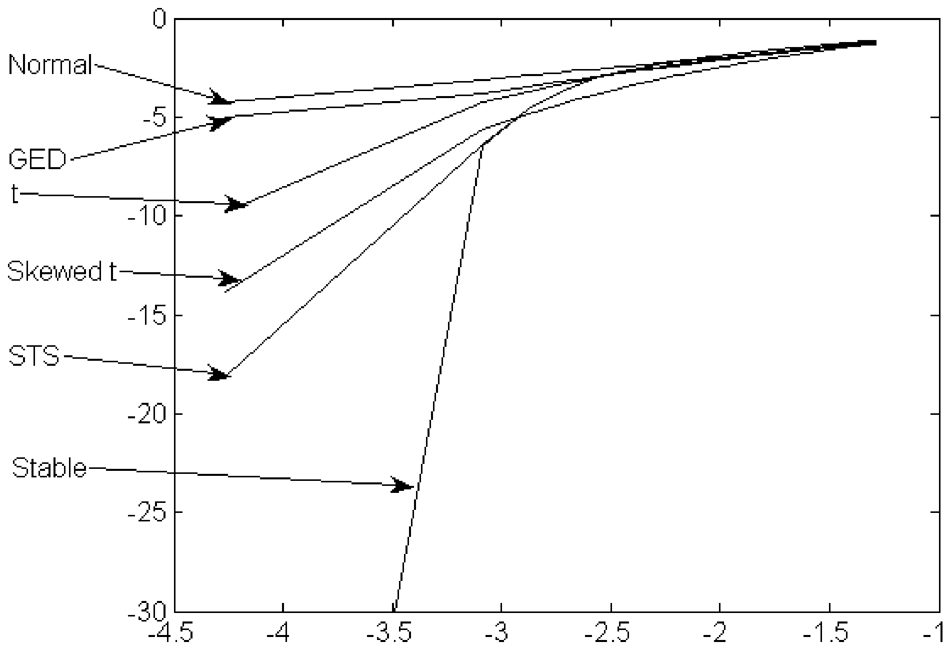
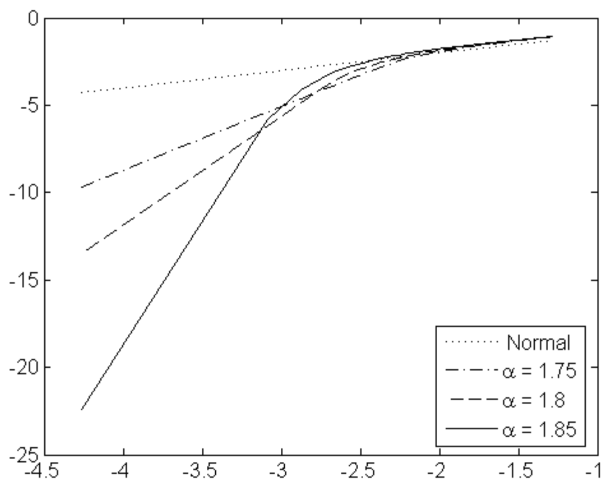
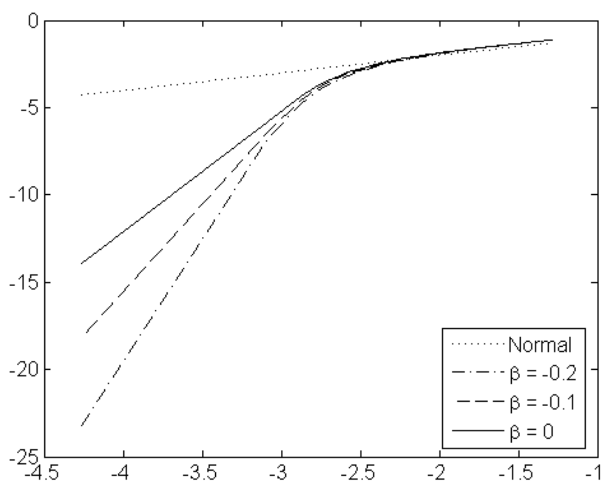


Figure 2. Quantile Comparison for Different Standardized Distributions.



(a) Influence of α



(b) Influence of β

Figure 3. Impact of the Distribution Parameters α and β on the Quantiles.