Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Nonparametric Risk Management with Generalized Hyperbolic Distributions

Ying Chen[1],Wolfgang Härdle[2],Seok-Oh Jeong[3]

submitted: 02. November 2005

[1] Weierstrass Institute
     for Applied Analysis and Stochastics
     Mohrenstr. 39, 10117 Berlin, Germany
     CASE -
     Center for Applied Statistics and Economics
     Humboldt-Universität zu Berlin
     Wirtschaftswissenschaftliche Fakultät
     Spandauerstrasse 1, 10178 Berlin, Germany
     E-Mail: chen@wias-berlin.de

[2] CASE -
     Center for Applied Statistics and Economics
     Humboldt-Universität zu Berlin
     Wirtschaftswissenschaftliche Fakultät
     Spandauerstrasse 1, 10178 Berlin, Germany
     E-Mail: haerdle@wiwi.hu-berlin.de

[3] Institut de Statistique
     Université Catholique de Louvain
     Voie du Roman Pays, 20 1348 Louvain-la-Neuve, Belgium

No. 1063

Berlin 2005

W I A S

Abstract

In this paper we propose the GHADA risk management model that is based on the generalized hyperbolic (GH) distribution and on a nonparametric adaptive methodology. Compared to the normal distribution, the GH distribution possesses semi-heavy tails and represents the financial risk factors more appropriately. Nonparametric adaptive methodology has the desirable property of being able to estimate homogeneous volatility over a short time interval and reflects a sudden change in the volatility process. For DEM/USD exchange rate and German bank portfolio data, the proposed GHADA model provides more accurate Value at Risk calculations than the models with assumptions of the normal and t distributions. All calculations and simulations are done with XploRe.

1   INTRODUCTION

After the breakdown of the fixed exchange rate system of the Bretton Woods Agreement in 1971, a sudden increase of volatility was observed in financial markets. The following boom of financial derivatives accelerated the turbulence of the markets. The subsequent scale of losses astonished the world and pushed the development of sound risk management systems. One of the most challenging tasks in analyzing financial markets is to measure and manage risks properly. Financial risks have many sources and are typically mapped into a stochastic framework:

$$R_t = \sigma_t \varepsilon_t, \tag{1}$$

where R_t denotes the loss or the (log) negative return of financial instrument, i.e. $\log(S_{t-1}) - \log(S_t)$ with S_t the price of the financial instrument, σ_t is the time dependent volatility and ε_t is the white noise. Based on the distribution of the risk factor R_t, various kinds of risk measures such as Value at Risk (VaR), expected shortfall and lower partial moments can be calculated. Among them, VaR has become the standard measure of market risk since J.P. Morgan launched RiskMetrics with the distributional assumption of normality in 1994, making the analysis of VaR simple and standard, Jorion (2001).

The importance of VaR was reinforced after it was used by central banks to govern and supervise the capital adequacy of banks in the Group of Ten (G10) countries in 1995. For a given financial instrument, VaR indicates the possible loss at a certain risk level over a certain time horizon. Andersen, Bollerslev, Christoffersen and Diebold (2005) have pointed out, losses will converge to normality under temporal aggregation. This observation suggests that the principle of the RiskMetrics method is valid as longer time horizon such as two weeks or one month is considered. One the other hand, financial institutions are expected to report and control their daily VaRs as well and the daily VaR deviates from

the normal assumption. In this paper, we concentrate to propose a risk management model to improve the calculation of daily VaR. Based on (1), daily VaR at a risk level p is defined mathematically:

$$VaR_{p,t} = F_t^{-1}(p) = \sigma_t q_{\varepsilon_t}(p), \qquad (2)$$

where F_t^{-1} is the quantile function of R_t at time t, which equals the product of the volatility and the p-th quantile of the stochastic term ε_t, Franke, Härdle and Hafner (2004). It is clear that the accuracy of VaR and other risk measures heavily depends on the distributional assumption of the stochastic term and the volatility estimation. This observation motivates us to choose:

    a. a heavy tailed distribution family to mimic the empirical distribution of risk factors and

    b. an adaptive methodology to estimate and forecast volatility locally.

In literature, for reasons of stochastic and numerical simplicity, it is often assumed that the involved risk factors are normally distributed e.g. in the RiskMetrics framework. However this assumption contradicts the empirical fact observed in the market - daily financial time series are heavy tailed distributed. As more extreme risks happened in the market, VaRs with higher risk quantiles such as 99% quantile draw more attention of risk analysts and the difference of these VaRs with individual distributional assumption is evident. Although, with the normality assumption, the VaR at 95% confidence level is almost identical to that with a more realistic leptokurtic distribution, see Jaschke and Jiang (2002).

Figure 1 illustrates this empirical fact on the basis of the daily foreign exchange (FX) rates of the German Mark to the US Dollar (DEM/USD) from 1979-12-01 to 1994-04-01. Here we use the daily devolatilized returns, i.e. $\varepsilon_t = R_t/\hat{\sigma}_t$, to fit the previously assumed stochastic distribution. Compared to R_t, the devolatilized returns eliminate the influence of volatility clustering and are more stationary. The technique used to estimate the volatility σ_t will be discussed later. The nonparametrically estimated kernel density and log density are regarded as benchmarks, where the Quartic kernel function and Silverman's rule of thumb are applied to select the bandwidth h, Härdle, Müller, Sperlich and Werwatz (2004). In the figure, the estimated normal (log) densities (dotted line) obviously deviate from these benchmarks, which will lead to inaccurate VaR calculations. Due to the weak ability of normal distribution to capture the empirical distributional feature of financial risk factors, various heavy tailed distribution families such as the hyperbolic, Student-t distributions and the Lévy process have been introduced in finance by Eberlein and Keller (1995), Embrechts, McNeil and Straumann (1999) and Barndorff-Nielsen and Shephard (2001). Among them, the Generalized Hyperbolic (GH) distribution family has attracted the attention of researchers. With five parameters, it can match the distributional behavior of real data in a flexible way. Eberlein, Kallsen and Kristen (2003) have proposed a model

2

with the GH distribution that gives more accurate VaR values than the model with the normal distribution.

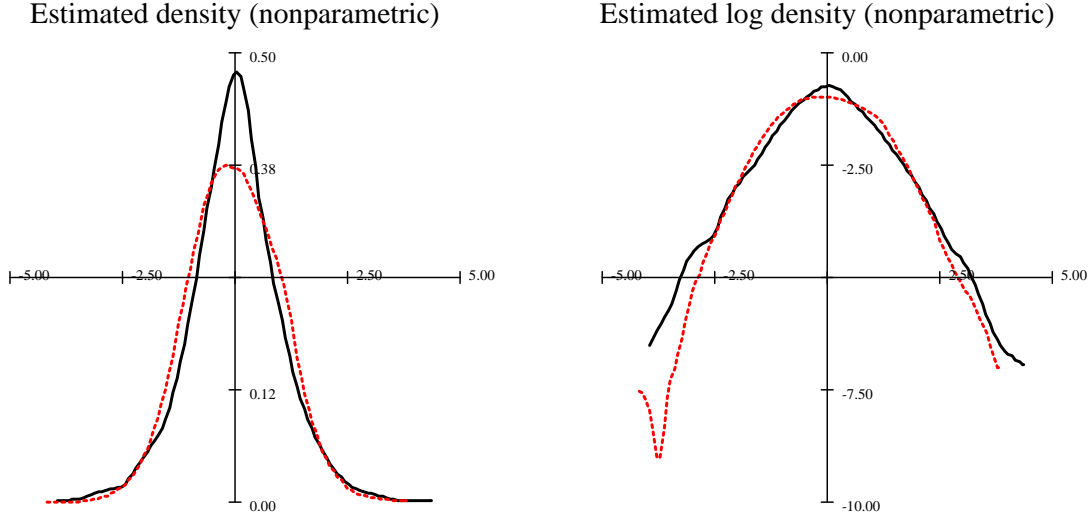Estimated density (nonparametric)   Estimated log density (nonparametric)

Figure 1: Graphical comparison of the density (left) and the log-density (right) of the daily DEM/USD devolatilized returns from 1979-12-01 to 1994-04-01 (3719 observations). The kernel density estimate is represented by a solid line and the normal density as a dotted line with $h \approx 0.54$. Data source: DEM/USD daily rates from 1979-12-01 to 1994-04-01 available at the FEDC (sfb649.wiwi.hu-berlin.de/fedc).

🔲 GHADAfx.xpl

In addition to the stochastic distributional assumption, the role of volatility models is of great significance in the VaR calculation as well. The most frequently used estimations are the ARCH (Engle, 1995), GARCH (Bollerslev, 1995) and stochastic volatility models (Harvey, Ruiz and Shephard, 1995). Although these models reflect the volatility clustering of financial time series, they are not uniformly applicable in risk management due to their theoretical drawbacks. The volatility process in these models is assumed to follow a time constant closed form, which is questionable especially in long time periods. To illustrate his problem, we estimated volatilities of the DEM/USD data by the GARCH(1,1) model: $\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$. The estimated parameters vary from each other as the time span increases from two years (1992-04-01 to 1994-04-01) to fourteen years (1979-12-01 to 1994-04-01), see Table 1. It is therefore plausible to use more flexible estimation methods by providing a data-driven "local" model, which can avoid this potential misspecification problem as much as possible. We follow the local constant model proposed by Mercurio and Spokoiny (2004) for its good performance. The philosophy of the local constant model, volatility changes little over a short interval, is numerically tractable and economic meaningful. Moreover, compared to most adaptive estimation methods such as the rectangular moving average, the local constant model can react very fast once a sudden jump happens.

3

Time period	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\beta}_1$
1979-12-01 to 1994-04-01	1.65e-06(3.90e-07)	0.07(0.01)	0.89(0.01)
1992-04-01 to 1994-04-01	1.09e-06(8.14e-07)	0.04(0.02)	0.93(0.03)

Table 1: Estimated parameters of the GARCH(1,1) process: $\sigma_t^2 = \omega + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_{t-1}^2$ based on different time periods. The standard deviations are in parentheses. Data source: DEM/USD daily rates from 1979-12-01 to 1994-04-01 available at the FEDC (sfb649.wiwi.hu-berlin.de/fedc).

An inconvenient assumption in Mercurio and Spokoiny (2004) is however that the stochastic term is normally distributed.

Motivated by the above two research lines, we estimate the local volatility adaptively and model the risk factors with heavy tails by the GH distribution. Here we name this new VaR technique as the **G**eneralized **H**yperbolic **Ada**ptive Volatility (GHADA) technique. The devolatilized return density plot in Figure 1 is in fact calculated with the GHADA technique. In addition, we check the validation of the GHADA technique by comparing with some other risk management models:

- volatility estimation with the GARCH(1,1) model and distributional fit with the normal (NGARCH), Student-t (tGARCH) or GH (GHGARCH) distribution,

- Local constant volatility estimation and distributional fit with the normal (NADA) or Student-t (tADA) distribution.

The paper is organized as follows: in Section 2, we will introduce the details of the GHADA technique. The validation of the GHADA model is illustrated through Monte Carlo simulation in Section 3. In Section 4, VaR calculations will be presented based on the DEM/USD and a German bank portfolio data. According to backtesting results, the GHADA technique provides more accurate forecasts than the models with assumptions of the normal and t distributions. Furthermore, the GHADA technique performs better than the models with GARCH(1,1) volatility processes in extreme events. Finally, we will briefly conclude our study in Section 5. All the figures may be recalculated and reproduced using the indicated links to the XploRe Quantlet Server.

2 GHADA TECHNIQUE AND OTHER VARIATIONS

In risk management modelling, a major task is to estimate the future loss distribution accurately. As illustrated in (2), a realistic distributional assumption of the stochastic term ε_t and an accurate volatility estimation play important roles in the modelling. In this

section, we describe two pillars of the proposed GHADA technique: the GH distribution and the local constant volatility estimation.

2.1 GHADA Technique

2.1.1 Generalized Hyperbolic Distribution

The GH distribution introduced by Barndorff-Nielsen (1977) is a heavy tailed distribution that can well replicate the empirical distribution of financial risk factors. The density of the GH distribution for $x \in \mathbb{R}$ is:

$$f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu) = \frac{(\iota/\delta)^{\lambda}}{\sqrt{2\pi}K_{\lambda}(\delta\iota)} \frac{K_{\lambda-1/2}\left\{\alpha\sqrt{\delta^2 + (x-\mu)^2}\right\}}{\left\{\sqrt{\delta^2 + (x-\mu)^2}/\alpha\right\}^{1/2-\lambda}} \cdot e^{\beta(x-\mu)} \tag{3}$$

under the conditions:

$$\begin{aligned}
\delta \geq 0, &\quad |\beta| < \alpha \quad \text{if } \lambda > 0 \\
\delta > 0, &\quad |\beta| < \alpha \quad \text{if } \lambda = 0 \\
\delta > 0, &\quad |\beta| \leq \alpha \quad \text{if } \lambda < 0
\end{aligned}$$

where λ, α, β, δ and $\mu \in \mathbb{R}$ are the GH parameters with $\iota^2 = \alpha^2 - \beta^2$. The density's location and scale are mainly controlled by μ and δ respectively:

$$\begin{aligned}
\mathsf{E}[X] &= \mu + \frac{\delta^2 \beta}{\delta\iota} \frac{K_{\lambda+1}(\delta\iota)}{K_{\lambda}(\delta\iota)} \\
\mathrm{Var}[X] &= \delta^2 \left\{ \frac{K_{\lambda+1}(\delta\iota)}{\delta\iota K_{\lambda}(\delta\iota)} + (\frac{\beta}{\iota})^2 [\frac{K_{\lambda+2}(\delta\iota)}{K_{\lambda}(\delta\iota)} - \{\frac{K_{\lambda+1}(\delta\iota)}{K_{\lambda}(\delta\iota)}\}^2] \right\},
\end{aligned}$$

whereas β and α play roles in the skewness and kurtosis of the distribution. For more details of the parameters' domains, we refer to Bibby and Sørensen (2001). $K_{\lambda}(\cdot)$ is the modified Bessel function of the third kind with index λ, Barndorff-Nielsen and Blæsild (1981):

$$K_{\lambda}(x) = \frac{1}{2} \int_0^{\infty} y^{\lambda-1} exp\{-\frac{x}{2}(y + y^{-1})\} \, dy$$

Furthermore, the GH distribution has a tail behavior:

$$f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu = 0) \sim x^{\lambda-1} e^{(\mp\alpha+\beta)x} \text{ as } x \to \pm\infty, \tag{4}$$

where $a(x) \sim b(x)$ as $x \to \infty$ means that both $a(x)/b(x)$ and $b(x)/a(x)$ are bounded as $x \to \infty$. Recall that the negative return, $R_t = \log(S_{t-1}) - \log(S_t)$, is regarded as a risk factor in risk management, indicating that the right tail of losses with large values is more important. The left tail, on the other hand, concerns the values of profits that are less interesting in

risk management. Compared to three popular heavy tailed distributions, the Student-t, Laplace and Cauchy distributions, the exponentially decaying speed of the GH distribution is faster and better matches the empirical tail behavior of real data. Comparisons of the four heavy tailed and normal distributions, and especially their tail behaviors, are displayed in Figure 2.

In order to retain the comparability of these distributions, we normalized the variables to have means equal to 0 and variances equal to 1. Here we used one important subclass of the GH distribution: the normal-inverse Gaussian (NIG) distribution with $\lambda = -\frac{1}{2}$, which is introduced more precisely in the following text. In the left panel, the complete
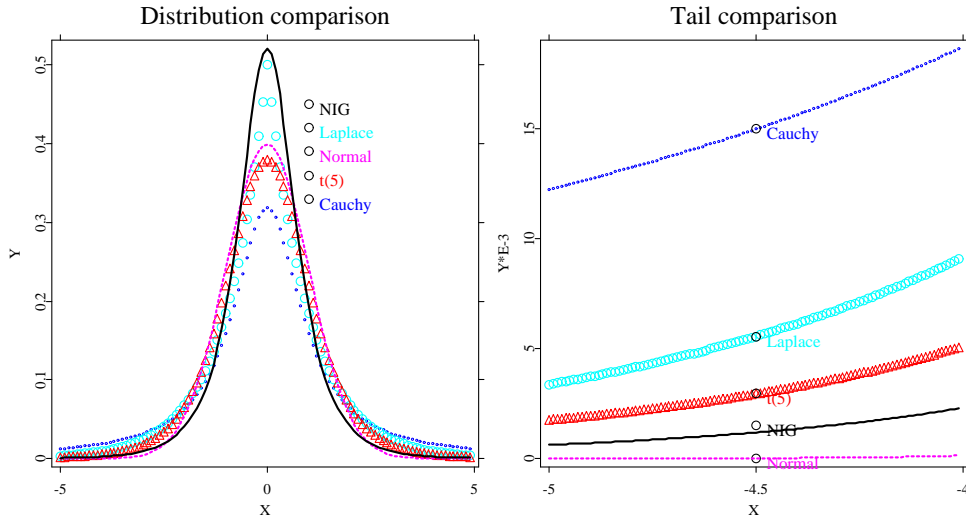

Figure 2: Graphical comparison of the tail-behavior of five standardized distributions: NIG distribution (solid line), standard normal distribution (dashed line), Student-t distribution with 5 degrees of freedom (triangles), Laplace distribution (circles) and Cauchy distribution (dotted line).

Q GHADAtail.xpl

shapes of these distributions are shown. Among them, the Cauchy distribution has the lowest peak and the fattest tails; briefly to say, it has the flattest distribution. The NIG distribution decays second fastest in the tails but with the highest peak, which is similar to the distributional shape of real data, see e.g. Figure 1.

The moment generating function of the GH distribution is:

$$m_f(z) = e^{\mu z} \cdot \frac{\iota^{\lambda}}{\iota_z^{\lambda}} \cdot \frac{K_{\lambda}(\delta \iota_z)}{K_{\lambda}(\delta \iota)}, \quad |\beta + z| < \alpha, \ \iota_z^2 = \alpha^2 - (\beta + z)^2 \tag{5}$$

indicating that m_f is differentiable infinitely many times near 0. As a result, every moment of a GH variable exists. In Section 2.1.2, this feature as well as the tail behavior (4) of the GH distribution will help to extend the local constant volatility methodology from the

normal distribution to the GH distribution, see Appendix.

Given the distributional form of the GH distribution in (3), maximum likelihood (ML) estimation can be applied straightforwardly. Such a direct estimation based on the GH distribution is however rarely applied since it is computationally cumbersome and numerically unstable due to the estimation of λ in the modified Bessel function. Instead, subclasses of the GH distribution such as the hyperbolic (HYP) and normal-inverse Gaussian (NIG) distributions are frequently used. These subclasses fix the value of λ to avoid the numerical problem. Eberlein and Keller (1995), Barndorff-Nielsen (1997) have shown that these subclasses are rich enough to model financial time series in an efficient way. In addition, the popularity of the subclasses is also motivated by the observation that the four parameters $(\mu, \delta, \beta, \alpha)^\top$ can simultaneously control the four moment functions of the distribution, i.e. the trend, scale, asymmetry and likeliness of extreme events. In our study we concentrate on the two subclasses of the GH distribution: HYP with $\lambda = 1$ and NIG distribution with $\lambda = -1/2$. In the following simulation and empirical studies, one can see that each subclass performs better than the other depending on the case. The corresponding density functions are given as:

- Hyperbolic (HYP) distribution: $\lambda = 1$,

$$f_{HYP}(x; \alpha, \beta, \delta, \mu) = \frac{\iota}{2\alpha\delta K_1(\delta\iota)} e^{\{-\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu)\}}, \tag{6}$$

  where $x, \mu \in \mathbb{R}$, $0 \le \delta$ and $|\beta| < \alpha$,

- Normal-inverse Gaussian (NIG) distribution: $\lambda = -1/2$,

$$f_{NIG}(x; \alpha, \beta, \delta, \mu) = \frac{\alpha\delta}{\pi} \frac{K_1\left\{\alpha\sqrt{\delta^2 + (x-\mu)^2}\right\}}{\sqrt{\delta^2 + (x-\mu)^2}} e^{\{\delta\iota + \beta(x-\mu)\}}. \tag{7}$$

  where $x, \mu \in \mathbb{R}$, $\delta > 0$ and $|\beta| \le \alpha$.

In order to estimate the unknown parameters $(\alpha, \beta, \delta, \mu)^\top$, ML and numerical optimization methods such as the Powell method (Press, Teukolsky, Vetterling and Flannery, 1992) are used. For an independently and identically distributed (i.i.d.) HYP respectively NIG distributed variable X, the log-likelihood functions are:

$$
\begin{aligned}
L_{HYP} &= T\log\iota - T\log 2 - T\log\alpha - T\log\delta - T\log K_1(\delta\iota) \\
&\quad + \sum_{t=1}^{T}\{-\alpha\sqrt{\delta^2 + (x_t - \mu)^2} + \beta(x_t - \mu)\} \\
L_{NIG} &= T\log\alpha + T\log\delta - T\log\pi + T\delta\iota \\
&\quad + \sum_{t=1}^{T}\left[\log K_1\left\{\alpha\sqrt{\delta^2 + (x_t - \mu)^2}\right\} - \frac{1}{2}\log\{\delta^2 + (x_t - \mu)^2\} + \beta(x_t - \mu)\right]
\end{aligned}
$$

$$\tag{8}$$
$$\tag{9}$$

7

Figure 3 shows the estimated HYP and NIG densities with the corresponding ML estimators of the DEM/USD devolatilized returns. The estimated densities graphically coincide with the kernel and log densities of the financial risk factor.
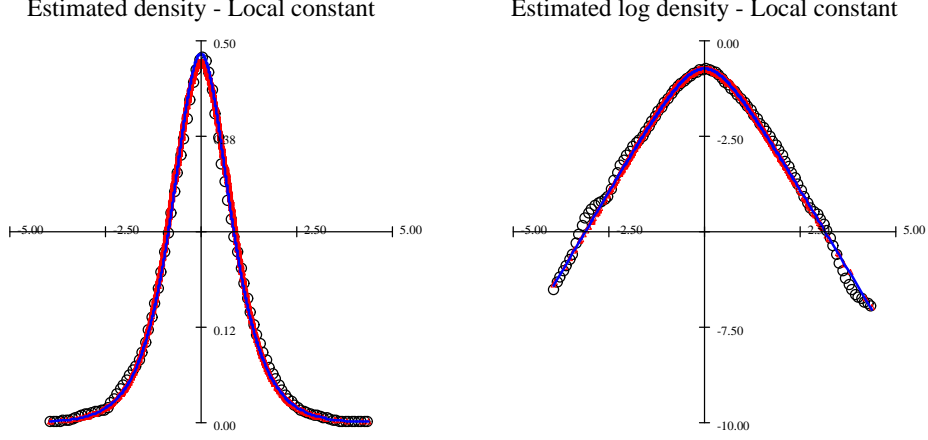

Figure 3: The kernel estimated density (left) and log density (right) of the devolatilized return of FX rates (circles) ($h \approx 0.55$). The HYP (solid line) parameters are $\hat{\alpha} = 1.744$, $\hat{\beta} = -0.017$, $\hat{\delta} = 0.782$, $\hat{\mu} = 0.012$, and the NIG (triangles) parameters are $\hat{\alpha} = 1.340$, $\hat{\beta} = -0.015$, $\hat{\delta} = 1.337$, $\hat{\mu} = 0.010$.

Q GHADAfx.xpl

2.1.2 Adaptive Volatility Estimation

Now we describe the adaptive estimation procedure for the volatility coefficients when risk factors follow the subclasses of GH distributions. Originally the concept of adaptive volatility estimation was proposed by Mercurio and Spokoiny (2004): There exists an interval of local homogeneity of the volatility process σ_t, which means that for τ fixed there exists a time interval $I = [\tau - m, \tau)$ such that σ_t varies little over I. Once an interval of homogeneity I is specified, the volatility at time τ is simply estimated by averaging the squared returns over the time interval I:

$$\hat{\sigma}_\tau^2 = \frac{1}{|I|} \sum_{t \in I} R_t^2, \tag{10}$$

where $|I|$ denotes the cardinality of I.

The squared returns R_t^2 are always nonnegative and have a skewed distribution with the stochastic errors ε_t. Therefore, the problem of estimating σ_t is transformed into an additive regression problem by a power transformation:

$$
\begin{aligned}
|R_t|^\gamma &= C_\gamma \sigma_t^\gamma &+& D_\gamma \sigma_t^\gamma \zeta_{\gamma,t} \\
&= \theta_t &+& s_\gamma \theta_t \zeta_{\gamma,t}
\end{aligned} \tag{11}
$$

8

with γ the power transformation parameter. In the following Lemma 1, one can see that γ is a constant bounded in $[0, 1]$. Furthermore, $\zeta_{\gamma,t} = (|\varepsilon_t|^\gamma - C_\gamma)/D_\gamma$, $C_\gamma = \mathsf{E}(|\varepsilon_t|^\gamma | \mathcal{F}_{t-1})$, $D_\gamma^2 = \mathsf{E}[(|\varepsilon_t|^\gamma - C_\gamma)^2 | \mathcal{F}_{t-1}]$ and $s_\gamma = D_\gamma/C_\gamma$. This equation (11) can be considered as a regression model to estimate θ_t with heteroscedastic additive errors $s_\gamma \theta_t \zeta_{\gamma,t}$. Since ε_t are assumed to be i.i.d., C_γ, D_γ and s_γ are merely nonstochastic constants. For example, when ε_t are $N(0, 1)$ distributed, one may easily compute the exact values of C_γ, D_γ and hence s_γ. However, the calculations of these constants are not necessary in order to specify an interval of homogeneity, as illustrated in the following.

Note that σ_t has one-to-one correspondence with θ_t. Therefore, if I is an interval of homogeneity, then θ_t is nearly constant for $t \in I$ and can be estimated by:

$$\hat{\theta}_I = \frac{1}{|I|} \sum_{t \in I} |R_t|^\gamma. \tag{12}$$

By (11), we have

$$\hat{\theta}_I = \frac{1}{|I|} \sum_{t \in I} \theta_t + \frac{s_\gamma}{|I|} \sum_{t \in I} \theta_t \zeta_{\gamma,t}.$$

The conditional expectation and variance of $\hat{\theta}_I$ are as follows:

$$\mathsf{E}[\hat{\theta}_I | \mathcal{F}_{\tau-1}] = \mathsf{E} \frac{1}{|I|} \sum_{t \in I} \theta_t,$$

$$v_I^2 = \operatorname{Var}[\hat{\theta}_I | \mathcal{F}_{\tau-1}] = \frac{s_\gamma^2}{|I|^2} \mathsf{E}(\sum_{t \in I} \theta_t \zeta_{\gamma,t})^2 = \frac{s_\gamma^2}{|I|^2} \mathsf{E} \sum_{t \in I} \theta_t^2,$$

when I is an interval of homogeneity, v_I can be estimated by:

$$\hat{v}_I = s_\gamma \hat{\theta}_I |I|^{-1/2}.$$

For testing whether an interval I is of homogeneity, Mercurio and Spokoiny (2004) suggested to investigate the homogeneity of θ_t in I instead of σ_t. They proposed a homogeneity test on θ_t based on Martingale deviation probability bound, see section 3 in Mercurio and Spokoiny (2004), but with $\varepsilon_t \sim N(0, 1)$. Hence we made a theoretical justification before we adopt their procedure for determining the interval of homogeneity with ε_t from the GH distribution. Details are given in Appendix. We precisely address this issue in Theorem 1 below. The proof is given in Mercurio and Spokoiny (2004).

THEOREM 1 *If the volatility coefficient σ_t satisfies the condition $b \leq \sigma_t^2 \leq bB$ with some positive constants b and B, then it holds that:*

$$P\left\{ |\hat{\theta}_I - \theta_\tau| > \Delta_I(1 + \eta s_\gamma |I|^{-1/2}) + \eta \hat{v}_I \right\}$$

9

$$\leq 4\sqrt{e}\eta(1+\log B)\exp\left\{-\frac{\eta^2}{2a_\gamma(1+\eta s_\gamma|I|^{-1/2})^2}\right\}$$

where Δ_I is the squared bias defined as $\Delta_I^2 = |I|^{-1}\sum_{t\in I}(\theta_t - \theta_\tau)^2$.

This theorem indicates that, if I is a time homogeneous interval, the squared bias Δ_I is negligible and the estimation error $|\hat{\theta}_I - \theta_\tau|$ is small relative to $\eta\hat{v}_I$ for $\tau \in I$ with a high probability. Moreover it holds that for any subinterval J of a homogeneity interval I

$$\left|\hat{\theta}_{I\setminus J} - \hat{\theta}_J\right| \leq \eta\left(\hat{v}_J + \hat{v}_{I\setminus J}\right) = \eta'\left(\hat{\theta}_J|J|^{-1/2} + \hat{\theta}_{I\setminus J}|I\setminus J|^{-1/2}\right) \tag{13}$$

with high probability when $\eta' = \eta s_\gamma$ is large enough. Therefore, if there exists a subinterval $J \subset I$ which makes $\left|\hat{\theta}_{I\setminus J} - \hat{\theta}_J\right|$ have significantly large positive value, the homogeneity of the interval I should be denied.

There are still two parameters to be specified: γ in the power transformation and the thresholding parameter η' in (13). According to Lemma 1, the parameter γ should be bounded in $[0,1]$. In our study, we chose $\gamma = 0.5$ like Mercurio and Spokoiny (2004). In fact the choice of γ does not affect much during the procedure for estimating the interval of homogeneity. On the contrary, the thresholding parameter η' is crucial in the homogeneity test. We pursue a selection procedure, similar to the choice of smoothing amount in non-parametric estimations. Take t_0 such that there are enough past observations to estimate $\hat{\theta}_{(t_0,\eta')}$ properly. Then select η' that gives minimal forecast error:

$$\eta' = \operatorname{argmin}\sum_{t=t_0}^{\tau-1}\left\{|R_t|^\gamma - \hat{\theta}_{(t,\eta')}\right\}^2 \tag{14}$$

where $\hat{\theta}_{(t,\eta')}$ is the estimate (12) of θ_t from observations $R_1,\ldots,R_{t-1}$ with the thresholding parameter η'. Note that such a selection is distributional free and avoids the calculation of these constants based on the GH distribution in (11).

Now we are ready to describe the precise procedure for estimating the interval of homogeneity with a constant η'. We start with an initial small interval $I = [\tau - m_0, \tau)$ that satisfies the homogeneity and suitably choose a step-increasing parameter m_0 and an integer $k(\geq 1)$. The choice of the step-increasing parameter will influence the sensitivity of the estimation to a change point. A smaller value increases the sensitivity but slows down the estimation speed, which will be illustrated in the simulation study later. For convenience m_0 is suggested to be a multiple of 5.

Step 1 Increase k to $k+1$, and enlarge the interval I to $[\tau - m, \tau)$ with $m = k \times m_0$.

Step 2 Reject I if there exists $J(\ell) = [\tau - \frac{2m}{3} + \ell, \tau), \ell = 1, 2, \cdots, \frac{m}{3}$ such that

$$\left| \hat{\theta}_{I \setminus J(\ell)} - \hat{\theta}_{J(\ell)} \right| > \eta' \left(\hat{\theta}_{J(\ell)} |J(\ell)|^{-1/2} + \hat{\theta}_{I \setminus J(\ell)} |I \setminus J(\ell)|^{-1/2} \right) \tag{15}$$

Step 3 If I is rejected, set $I = [\tau - m_0, \tau)$ as the interval of homogeneity and stop. If I is not rejected, then set $m_0 = m$ and go to **Step 1**.

2.2 Some Risk Management Models

The research on VaR models has been ignited and prompted by the rule of Basel Committee on Banking Supervision in 1995: financial institutions may use their internal VaR models to calculate VaRs. Recall that VaR is formulated as: $\text{VaR}_{p,t} = F_t^{-1}(p) = \sigma_t q_{\varepsilon_t}(p)$ with the notations defined before. In Table 2, eight models are listed according to two factors: the estimation of volatility and the distributional assumption of stochastic term. The GARCH(1,1) and local constant models are implemented with four distributional assumptions: the HYP, NIG, standard normal and t with degrees of freedom (df) distributions. Among these models, the tGARCH combining the GARCH(1,1) technique and the Student-t distribution, to the best of our knowledge, is the most frequently used model in practice. Compared to all other models mentioned here, the GHADA (HYPADA and NIGADA) models are expected to perform superiorly due to their desirable statistical characteristics discussed before. In the simulation and empirical studies, we will compare the accuracy of these various combinations and check the expectation.

Model	Volatility Estimation	Distributional Assumption
HYPGARCH	GARCH(1,1)	$\text{HYP}(\alpha, \beta, \delta, \mu)$
NIGGARCH	GARCH(1,1)	$\text{NIG}(\alpha, \beta, \delta, \mu)$
NGARCH	GARCH(1,1)	$N(0,1)$
tGARCH	GARCH(1,1)	$t(df)$
HYPADA	Local Constant	$\text{HYP}(\alpha, \beta, \delta, \mu)$
NIGADA	Local Constant	$\text{NIG}(\alpha, \beta, \delta, \mu)$
NADA	Local Constant	$N(0,1)$
tADA	Local Constant	$t(\text{df})$

Table 2: Risk management models based on the heteroscedastic model: $R_t = \sigma_t \varepsilon_t$.

3 MONTE CARLO SIMULATION

As discussed before, a good risk management modelling relies on two factors: estimating the volatility and fitting the distribution of risk factors. The previous distribution estimation

based on the DEM/USD data together with the comparison to the nonparametric density estimation in Section 2.1.1, provide evidence that with four parameters, the HYP and NIG distributions can represent the empirical distribution of the stochastic term very well. In this section, we illustrate the reliability of the local constant volatility model with four different distributional assumptions. Two simple volatility processes with jumps and a GARCH(1,1) process are considered here:

$$\sigma_{1,t} = \begin{cases} |0.02t - 5|/100 & , \quad 1 \le t \le 300 \\ |0.02t - 20|/100 & , \quad 300 < t \le 600 \\ |0.12t - 30|/100 & , \quad 600 < t \le 1000 \end{cases} \tag{16}$$

$$\sigma_{2,t} = \begin{cases} 0.01 & , \quad 1 \le t \le 400 \\ 0.03 & , \quad 400 < t \le 750 \\ 0.015 & , \quad 750 < t \le 1000 \end{cases} \tag{17}$$

$$\sigma_{3,t} = 1.65e - 06 + 0.07\varepsilon_{t-1}^2 + 0.89\sigma_{t-1}^2 \tag{18}$$

where the parameters of the GARCH(1,1) process (18) are the estimates of the DEM/USD negative returns, see Table 1.

In each scenario, we generate $n = 1000$ observations with the HYP$(2, 0, 1, 0)$, NIG$(2, 0, 1, 0)$, N$(0, 1)$ and $t(6)$ distributions. The risk factors are generated based on the heteroscedastic model:

$$R_{ij,t} = \sigma_{it}\varepsilon_{jt}, \quad i = 1, 2, 3 \text{ and } j = \text{HYP}, \text{NIG}, \text{N}, t.$$

In the local constant (LC) model, the first 200 observations of $R_{ij,t}$ are considered as a training set. The transformation parameter γ is fixed at 0.5 and a global η' that minimizes the mean of forecast error for $t \in [201, 1000]$ is selected to perform the homogeneity test. Last but not least, two values of the step-increasing parameter m_0 in the homogeneous interval are used since the value of m_0 will influence the detecting speed of the LC model as a jump appears: the recommended value $m_0 = 5$ and a more sensitive value $m_0 = 2$. Note that a smaller value of m_0 in general can find jumps faster than a larger one. All scenarios are repeated $M = 200$ times.

Three examples of the estimated volatility series of σ_{1t} (top), σ_{2t} (middle) and σ_{3t} (bottom) with the HYP or NIG distributional assumption are displayed in Figure 4, where the LC (solid line) and GARCH(1,1) estimations (dotted line) are compared with the generated volatilities (circles). In the first two estimations, the LC and GARCH(1,1) models display comparable results, whereas the GARCH(1,1) setup performs better in the estimation of σ_{3t}.

The quality of these two volatility estimation techniques is further measured by two ratios, the ratio of mean absolute error (**RMAE**) and the ratio of mean squared error
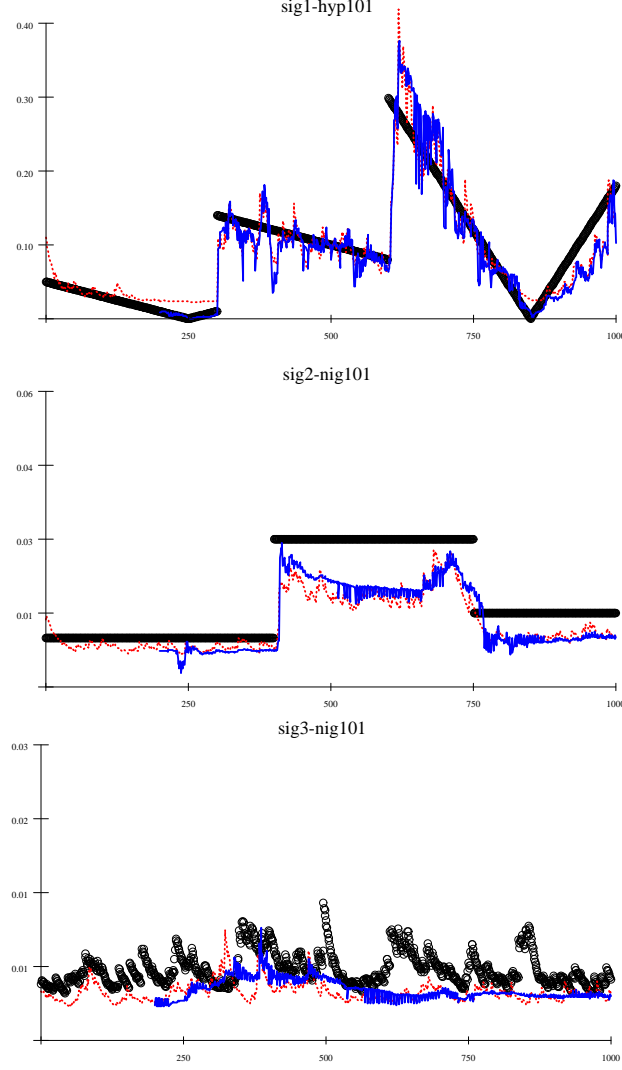
Figure 4: The estimated volatility processes on the basis of the three simulated examples (circles) with the HYP variables for σ_{1t}, the NIG for σ_{2t} and the NIG for σ_{3t}. The power transformation parameter is $\gamma = 0.5$, $m_0 = 5$ and the starting point $t_0 = 201$. The GARCH(1,1) process is displayed as a dotted line whereas the local constant is a solid line.

Q GHADAsim.xpl

(**RMSE**):

$$\text{RMAE} = \frac{\sum_{t=201}^{1000} |\hat{\sigma}_{it}^{LC} - \sigma_{it}|}{\sum_{t=201}^{1000} |\hat{\sigma}_{it}^{GARCH} - \sigma_{it}|},$$

$$\text{RMSE} = \frac{\sum_{t=201}^{1000} (\hat{\sigma}_{it}^{LC} - \sigma_{it})^2}{\sum_{t=201}^{1000} (\hat{\sigma}_{it}^{GARCH} - \sigma_{it})^2}, \quad i = 1, 2, 3$$

If the value of RMAE or RMSE is smaller than 1, it means that the LC model has a smaller estimation error on average than the GARCH and vice versa. Based on 200 repetitions,

13

the mean, standard deviation (sd), maximum and minimum of the two criteria are reported in Table 3. In the simulation of σ_{1t}, the LC model with $m_0 = 5$ gives more accurate volatility estimations on average than the GARCH(1,1) technique. Concerning σ_{2t}, the GARCH model performs better than the LC when the stochastic terms are HYP and NIG distributed, but has lower accuracy when ε_t are normal and t distributed. As supposed, the GARCH model can match the generated GARCH process σ_{3t} better. Based on the simulation results, the LC model is comparable to the GARCH(1,1) technique. However, the assumption of the GARCH technique, i.e. the estimation form is time constant, is in dispute as the nonstationary volatility processes σ_{1t} and σ_{2t} are given. In this sense, the LC model is considered better since it not only represents the empirical characteristics of volatility movements, but also provides a reasonable theory for the estimation.

	RMAE				RMSE			
$\sigma_{1t}\varepsilon_t$	mean	sd	max	min	mean	sd	max	min
HYP	0.77	1.05	1.04	0.06	0.78	1.23	1.08	0.00
NIG	0.87	1.00	1.04	0.06	0.83	1.15	1.07	0.00
N	0.76	1.18	1.04	0.05	0.87	1.22	1.09	0.00
t	0.93	1.28	1.06	0.05	0.95	1.14	1.11	0.00
$\sigma_{2t}\varepsilon_t$	mean	sd	max	min	mean	sd	max	min
HYP	1.31	0.76	1.27	1.42	1.47	0.89	1.39	1.73
NIG	1.25	0.70	1.24	1.29	1.44	0.85	1.43	1.55
N	0.61	0.57	0.61	0.62	0.50	0.00	0.33	0.50
t	0.69	0.78	0.70	0.69	0.51	0.88	0.60	0.50
$\sigma_{3t}\varepsilon_t$	mean	sd	max	min	mean	sd	max	min
HYP	1.21	1.07	1.16	1.51	1.31	1.20	1.29	1.89
NIG	1.07	1.03	1.12	1.53	1.11	1.15	1.24	2.25
N	1.32	1.19	1.21	1.25	1.58	1.31	1.40	1.40
t	1.22	1.39	1.23	1.50	1.49	1.67	1.38	2.12

Table 3: Descriptive statistics of the RMAE and RMSE with respect to volatility estimations. Two volatility models: local constant (LC) ($\gamma = 0.5$ and $m_0 = 5$) and GARCH(1,1) models are applied to estimate three volatility processes based on 4 variables: HYP$(2, 0, 1, 0)$, NIG$(2, 0, 1, 0)$, N$(0, 1)$ and $t(6)$.

Moreover, the sensitivity of the LC and GARCH(1,1) models to jumps in volatility is compared. We introduce a percentage rule to study the sensitivity of the two volatility estimation techniques. The detection speed of the estimated volatility to a sudden jump is measured at a 40%, 50% or 60% level of the jump size. The 40% rule, for example, refers to the number of time steps to reach 40% of the jump size as it happens. Table 4 gives some examples of the detection steps. The GARCH(1,1) process has a naturally fast reaction to jumps in a short interval since it is actually an exponential smoothing process. In general, the LC model needs more time to detect a jump than the GARCH, but the difference is very small. Sometimes the LC model reacts even faster than the GARCH(1,1) by adjusting the

		$\sigma_{1,t=300}$			$\sigma_{1,t=600}$		
Model	m_0	40% rule	50% rule	60% rule	40% rule	50% rule	60% rule
HYPADA	2	3.72(1.7)	4.66(2.9)	6.10(4.7)	5.01(3.2)	6.81(4.7)	8.82(6.5)
HYPADA	5	4.47(2.2)	5.85(3.3)	7.87(4.8)	5.52(3.0)	7.28(3.9)	9.66(5.6)
HYPGARCH		3.63(2.3)	5.17(3.5)	7.27(5.3)	3.42(2.6)	5.28(3.8)	7.69(5.5)
NIGADA	2	4.35(2.5)	5.64(4.3)	8.88(8.3)	5.81(3.6)	8.09(6.2)	13.22(13.4)
NIGADA	5	5.92(2.9)	7.94(4.5)	14.69(20.5)	9.25(4.2)	11.98(7.2)	17.89(13.1)
NIGGARCH		4.73(3.3)	6.74(4.9)	10.90(8.1)	4.19(2.9)	6.95(5.2)	10.76(9.7)
NADA	2	3.05(1.6)	3.62(1.9)	4.23(2.4)	3.77(2.1)	5.00(2.7)	6.14(3.5)
NADA	5	4.29(1.8)	5.17(2.0)	6.19(2.4)	7.24(3.6)	9.49(4.1)	11.19(5.4)
NGARCH		2.56(1.7)	3.27(2.2)	4.27(2.9)	2.27(1.5)	3.11(2.0)	4.00(2.5)
tADA	2	2.86(1.4)	3.43(1.8)	3.82(2.0)	1.97(1.9)	3.76(2.5)	4.83(2.9)
tADA	5	3.78(1.7)	4.50(1.9)	5.14(2.2)	1.39(2.1)	6.69(3.8)	8.61(3.9)
tGARCH		2.42(1.6)	3.03(2.1)	3.76(2.5)	1.64(1.4)	2.47(2.0)	3.25(2.4)
		$\sigma_{2,t=400}$			$\sigma_{2,t=750}$		
Model	m_0	40% rule	50% rule	60% rule	40% rule	50% rule	60% rule
HYPADA	2	5.24(3.7)	7.58(4.9)	10.79(7.5)	48.51(33.5)	28.81(19.2)	19.22(12.6)
HYPADA	5	6.90(3.9)	9.44(5.2)	12.74(9.5)	60.67(45.5)	30.60(21.8)	20.20(13.8)
HYPGARCH		4.09(3.0)	7.65(5.0)	12.27(9.2)	73.28(35.4)	30.96(11.4)	17.10(7.7)
NIGADA	2	6.84(4.3)	10.09(6.9)	15.70(12.9)	29.77(21.7)	18.84(13.6)	9.92(8.4)
NIGADA	5	8.93(4.6)	12.04(6.7)	18.03(13.5)	39.27(30.1)	22.56(13.8)	13.38(9.0)
NIGGARCH		6.63(4.4)	12.71(7.8)	20.49(12.4)	39.49(18.1)	18.17(8.9)	7.94(6.1)

Table 4: Mean of the detection step for several sudden jumps based on the LC model with $m_0 = 2$ and $m_0 = 5$ and the GARCH(1,1) model. The standard deviations of the detection steps are in parenthesis. Two jumps w.r.t. σ_{1t} at $t = 300$ and $t = 600$ and two jumps w.r.t. σ_{2t} at $t = 400$ and $t = 750$ are considered.

value of m_0. For example, concerning the jump of σ_{1t} at $t = 300$, the HYPADA needs 4.66 steps on average to detect the 50% jump sizes and 6.10 steps to detect 60% jump sizes while the GARCH(1,1) requires 5.17 and 7.27 steps, respectively. In addition, the deviations of these two detections based on the LC method with values of 2.9 and 1.7, are smaller than those of the GARCH technique. On the meanwhile, we find that the detection speed is slow for a deceased jump. For σ_{2t}, a downward jump from 5% to 1% happens at $t = 750$. The LC model with $m_0 = 2$ needs 19.22 steps on average to detect the 60% jump sizes. This number is three times more than that of detection steps for an increased jump with 40% sizes at $t = 400$. This phenomenon results from a low test power in the homogeneity test (13), where the squared conditional variance v_I depends on θ_t and a larger value of θ_t will lead to a low test power.

4 EMPIRICAL STUDY

4.1 Data Set

Two data sets, the DEM/USD exchange rate and a German bank portfolio, are used in the empirical analysis. They are available at FEDC (sfb649.wiwi.hu-berlin.de/fedc).

The loss of the exchange rate is calculated daily from 1979-12-01 to 1994-04-01. There are 3719 observations. The bank portfolio data reports the market value of the portfolio held by a German bank (anonymous due to the privacy protection law in Germany). There are 5603 daily observations.

The mean, standard deviation, skewness, kurtosis and the first two autocorrelations ρ_1 and ρ_2 of these two data sets are listed in Table 5. All time series are centered around 0 and have leptokurtic distributions as indicated by their kurtoses. Two processes of devolatilized returns $\varepsilon_t^{LC} = R_t/\hat{\sigma}_t^{LC}$ and $\varepsilon_t^{GARCH} = R_t/\hat{\sigma}_t^{GARCH}$ are analyzed according to the two volatility estimation techniques. As discussed before, the devolatilized returns ε_t^{LC} and ε_t^{GARCH}, compared to the loss series R_t, are expected to be more stationary. Note that these devolatilized returns still have the heavy tailed distributional property even after eliminating the influence of the time varying volatility.

Data	mean	sd	skewness	kurtosis	ρ_1	ρ_2
exchange rate: $t \in [501, 3719]$						
R_t	8.30e-05	7.00e-03	0.07	4.94	0.02	0.01
ε_t^{LC}	5.24e-03	0.99	0.01	4.03	0.03	0.02
ε_t^{GARCH}	7.13e-03	0.99	0.04	4.38	0.03	0.02
bank portfolio: $t \in [501, 5602]$						
R_t	-9.51e-05	1.59e-02	0.28	8.08	-0.04	-0.03
ε_t^{LC}	1.13e-02	0.96	-0.08	5.18	-0.04	-0.02
ε_t^{GARCH}	-1.31e-02	0.99	0.08	7.38	-0.03	-0.02

Table 5: Descriptive statistics for the daily devolatilized residuals of the exchange rate data and bank portfolio data.

The time plots of the daily losses, the local constant and GARCH(1,1) volatilities of the two data sets are displayed in Figure 5. The plots of volatility estimations show the volatility clustering, which reflects the movement pattern of risks graphically. For example, the local constant volatility displays a jump at $t = 3044$ when a large loss is observed in the time plot of the exchange rate series. In the German bank portfolio data, the simultaneous movements of loss and volatility are more evident. High risk, i.e. large value of volatility, is observed over the turbulent period as $t \in (3000, 4000)$ and small volatility appears in comparably

quiet periods. The GARCH(1,1) technique gives comparable estimations. Compared to the
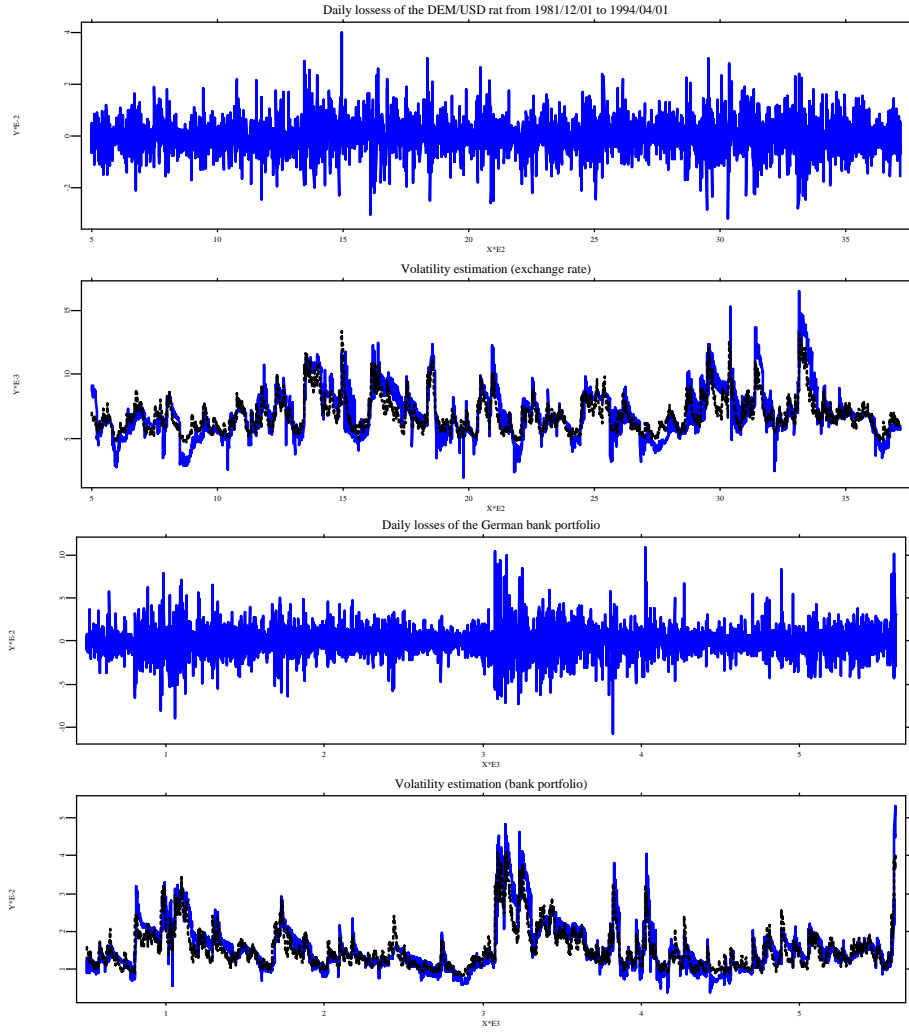

Figure 5: The time series of the daily losses, the local constant volatility (solid line) and the GARCH(1,1) volatility (dashed line) w.r.t. the DEM/USD rate (top) and the German bank portfolio (bottom). The parameters in the local constant models are $t_0 = 501$, $m_0 = 5$ and $\eta' = 1.06$ in the DEM/USD data and $\eta' = 1.23$ for the German bank portfolio data.

Q GHADAfx.xpl **Q** GHADAkupfer.xpl

German bank portfolio, the loss series of the DEM/USD displays a more regular fluctuation, since exchange rate market is liquid. In the German bank data, long quiet periods are observed with two extremely turbulent periods, which suggests that large homogeneous intervals will be specified on average in the German bank data. Boxplots in Figure 6 provide evidence of this suppose. The means of homogeneous intervals w.r.t. the two data sets are 51.37 (DEM/USD) and 76.42 (German bank), and further, many outliers with large value of interval length (circles or stars with 1.5 or 3 times box length distance from the upper level) are observed in the German bank data.
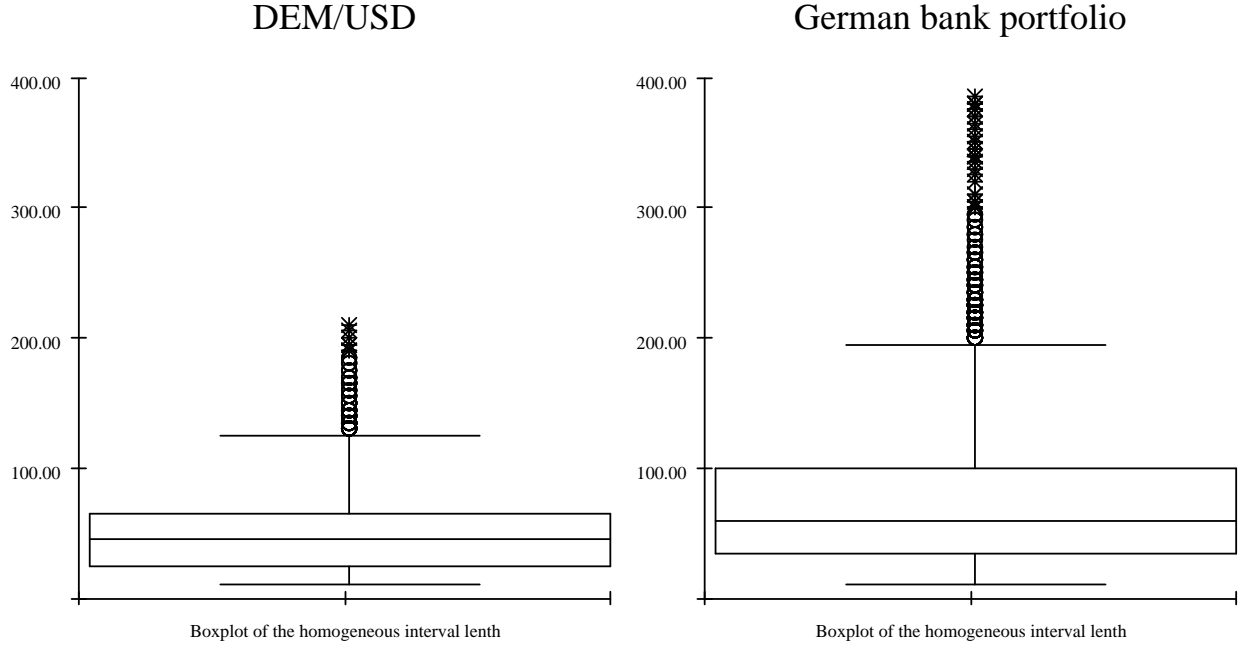
DEM/USD   German bank portfolio

Figure 6: Boxplots of the homogeneous interval length w.r.t. the DEM/USD exchange rates (left) and the German bank portfolio data (right).

Q GHADAfx.xpl **Q** GHADAkupfer.xpl

Given the estimated volatilities, the devolatilized returns are calculated and used to estimate the distributional parameters of four different assumptions: the HYP, NIG, standard normal and t with 6 degrees of freedom distributions. The HYP and NIG distributional parameters w.r.t. the local constant and GARCH(1,1) volatilities are listed in Table 6. The density estimations with the four assumptions are compared graphically. Once again, the nonparametric kernel densities estimation implemented in Section 1 are used as benchmarks. Considering the influence of volatility estimation techniques on the devolatilized returns, the probability densities w.r.t. the local constant (top) and GARCH(1,1) (bottom) techniques are graphed individually. Based on the DEM/USD data, the HYP (solid line) and NIG (triangle) models can better describe the distributions of the devolatilized returns ε_t^{LC} and ε_t^{GARCH} than the normal (dashed line) and $t(6)$ (dotted line), see Figure 7. The normal density underestimates the right tail of the devolatilized returns whereas the $t(6)$ displays a heavier right tail than the benchmark. This misspecification is enlarged in the log-density comparison (right) in Figure 7. Additionally, the devolatilized returns ε_t^{LC} with assumptions of the HYP and NIG distributions match the shape of benchmark better than ε_t^{GARCH}. Based on the GARCH(1,1) process, all the density estimations deviate from these benchmarks. This weak performance of the GARCH devolatilization is more obvious in the German bank data, see Figure 8. The benchmarks with the GARCH devolatilization is weakly matched by all the four distributional assumptions. It is possible that the German bank portfolio data is less liquid and therefore weaklier stationary than the exchange rate

18

Model	exchange rate: $t \in [501, 3719]$				bank portfolio: $t \in [501, 5602]$			
	α	β	δ	μ	α	β	δ	μ
HYPADA	1.74	0.01	0.78	-0.01	1.44	0.01	0.00	0.00
NIGADA	1.34	0.01	1.33	-0.00	0.99	-0.01	0.94	0.02
HYPGARCH	1.65	0.02	0.63	-0.02	1.41	-0.01	0.00	0.00
NIGGARCH	1.20	0.02	1.21	-0.01	0.87	0.00	0.88	-0.02

Table 6: Distributional parameters of the devolatilized residuals w.r.t. the local constant (LC) volatility and the GARCH(1,1) volatility of the DEM/USD data and the German bank portfolio data.

Q GHADAfx.xpl **Q** GHADAkupfer.xpl

data, where the local constant model is more suitable, at least theoretically, to capture the movement of the volatility process.

A latent problem of density estimation is whether the distribution of stochastic term is really stationary. A small experiment indicates that the distributional parameters, like volatility, could be time-variant as well. Figure 9 shows the HYP-quantile forecasts based on 500 historical devolatilized returns of the exchange rate for each point in time. It provides evidence that quantiles vary as time passes, especially for extreme probability levels such as $p = 0.995$. The same phenomenon holds for the NIG distribution, which is omitted here. If the sample size is small, we could not stick to the assumption that the innovations are identically distributed, although it assumes that the historical observations are i.i.d. as well. Instead, one should update the distributional parameters daily. For example, on may estimate the local distribution based on previous 500 data points. However as sample size increases to infinity, the distribution will converge to the unconditional distribution asymptotically. Given the two data sets with large sample size, we assume that all the observations have an identical distribution.

4.2 VaR

In this section, we focus on the model selection from the proposals in Table 2. Above all, the selected model should be theoretically reasonable and practically tractable to implement. Based on this criterion, we prefer the GHADA model due to its desirable properties discussed before. Furthermore, another important criterion of model selection is to compare the accuracy of VaR forecasts. The VaR at the probability level p is forecasted as:

$$\tilde{\text{VaR}}_{p,t+1} = \tilde{\sigma}_{t+1}\hat{q}_{\varepsilon_t}(p). \tag{19}$$
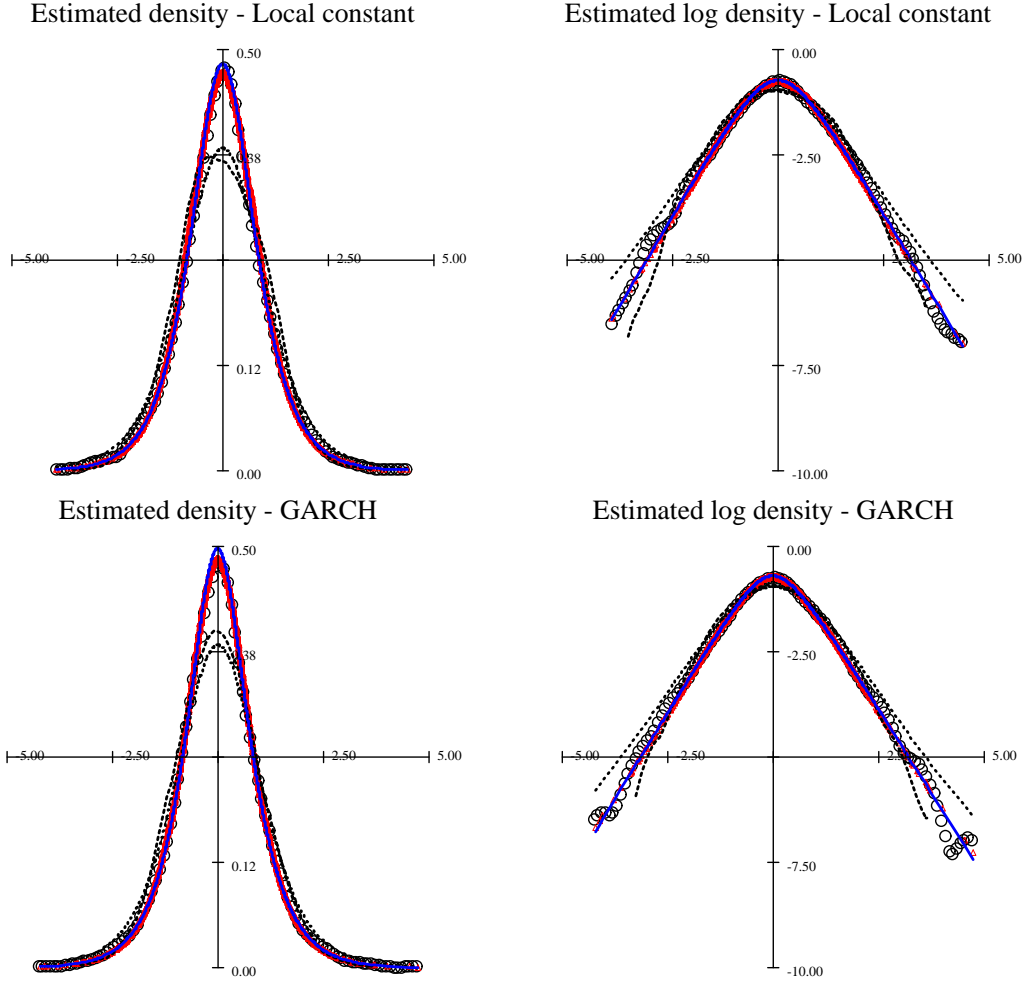
Figure 7: The estimated densities (top left) based on the HYP (solid line), NIG (triangle), standard normal (dashed line) and $t(6)$ (dotted line) compare with the nonparametric kernel density (circles) in the left panel. The local constant model is applied to devolatilize R_t of the DEM/USD data on the top and the GARCH(1,1) on the bottom. The corresponding log densities are displayed in the right panel. The distributional parameters are listed in Table 6.

◘ GHADAfx.xpl

The quantile $q_{\varepsilon_t}(p)$ is computed according to the distributional estimation. The volatility in the future is unavailable, but empirical studies find volatility time series appear to have a unit root, see Poon and Granger (2003). Therefore, the estimated volatility $\hat{\sigma}_t$ today is naturally used as the forecast of tomorrow.

Observation exceeding the forecasted VaR is called exceedance. A validate VaR model should neither underestimate nor overestimate the market risk. To evaluate the validation of VaR calculation, backtesting presented in Christoffersen (1998) is implemented. The empirical risk level, the ratio of exceedances in the time interval under consideration, is
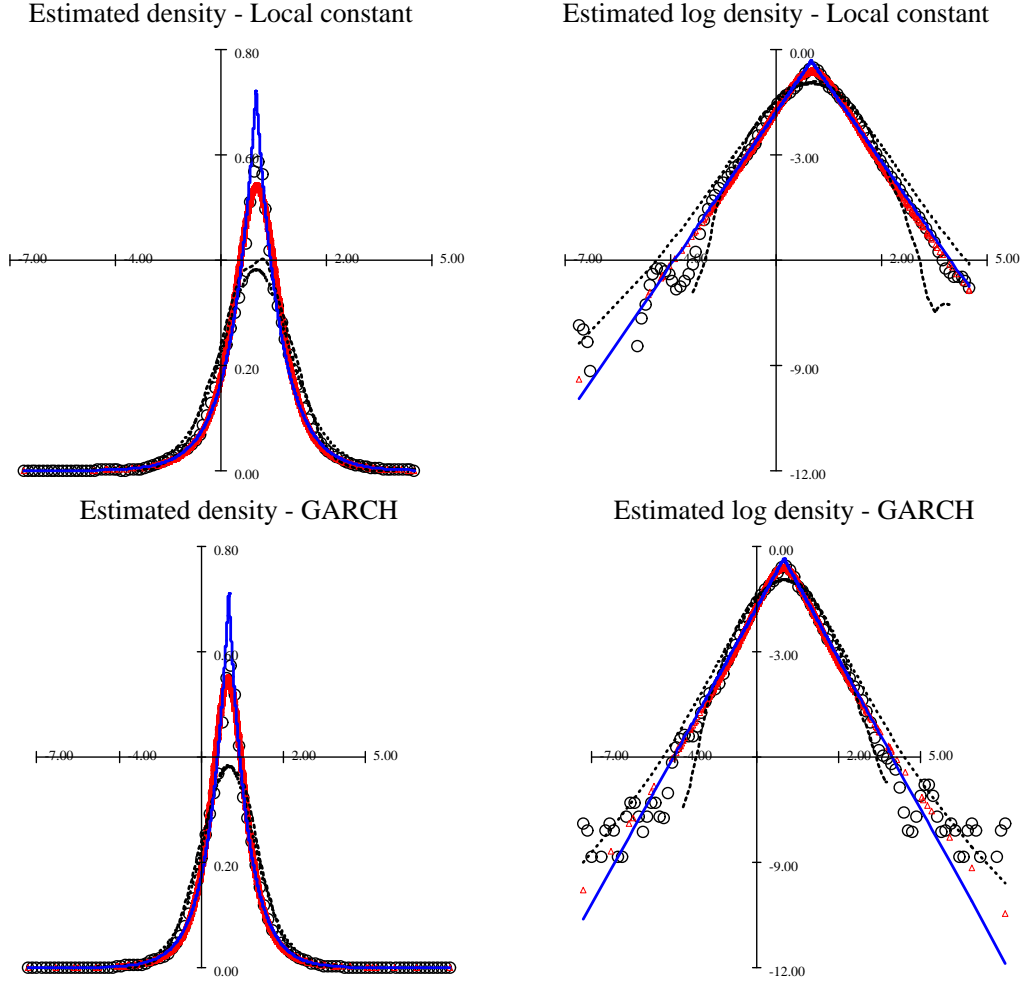
Figure 8: The estimated densities (top left) based on the HYP (solid line), NIG (triangle), standard normal (dashed line) and $t(6)$ (dotted line) compare with the nonparametric kernel density (circles) in the left panel. The local constant model is applied to devolatilize R_t of the German bank data on the top and the GARCH(1,1) on the bottom. The corresponding log densities are displayed in the right panel. The distributional parameters are listed in Table 6.

🔍 GHADAkupfer.xpl

compared to the expected risk level p:

$$N/T = T^{-1} \sum_{t=1}^{T} 1_t, \tag{20}$$

where 1_t denotes the indicator of exceedances at time point t. If the empirical risk level is larger than p, it indicates an overestimation of the selected model. In this case, additional capital requirements than necessary need to be located in central bank and prospective business opportunities are lost. On the contrary, if the empirical risk level is much smaller
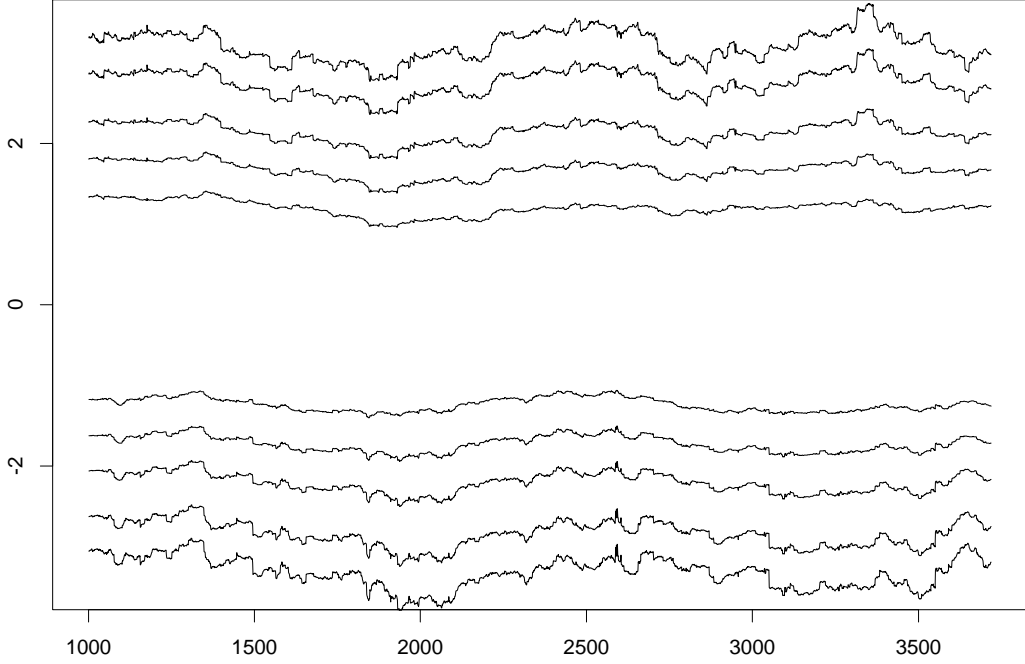
21

Figure 9: Quantiles estimated based on the past 500 devolatilized returns of the exchange rate. From the top the evolving HYP quantiles for $p = 0.995$, $p = 0.99$, $p = 0.975$, $p = 0.95$, $p = 0.90$, $p = 0.10$, $p = 0.05$, $p = 0.025$, $p = 0.01$, $p = 0.005$.

than p, a punishment due to risk underestimation will occur. The null hypothesis of the backtesting is formulated as:

$$H_0 : \mathsf{E}[N] = Tp \tag{21}$$

Under H_0, N is a Binomial random variable with parameters T and p, the likelihood ratio test statistic can be derived as:

$$\text{LR} = -2 \log \left\{ (1 - p)^{T-N} p^N \right\} + 2 \log \left\{ (1 - N/T)^{T-N} (N/T)^N \right\}, \tag{22}$$

which is asymptotically $\chi^2(1)$ distributed with critical values 3.84 (95%) and 6.63 (99%), Jorion (2001).

Table 7 summarizes the backtesting results of eight different models of the two data. As illustrated in Figure 7, the HYP and NIG distributional assumptions reflect the empirical distribution of the DEM/USD stochastic term better than the normal and t distributions, therefore models based on these two distributions should give more accurate VaRs. The

backtesting provides evidence for it, see Table 7. In general, the normal distribution underestimates while the $t(6)$ overestimates the risk. Moreover given the DEM/USD data, there is no large difference between models with the HYP and NIG distributions and the two volatility estimation techniques. Concerning the German bank data, the HYP model gives more accurate VaR forecasts than the NIG, and further the local constant technique performs better than the GARCH(1,1) setup.

Exemplary time plots of the forecasted VaR are displayed in Figure 10 and Figure 11. The exceedances are represented by crosses marked for the HYPADA, NIGADA, NADA and tADA.
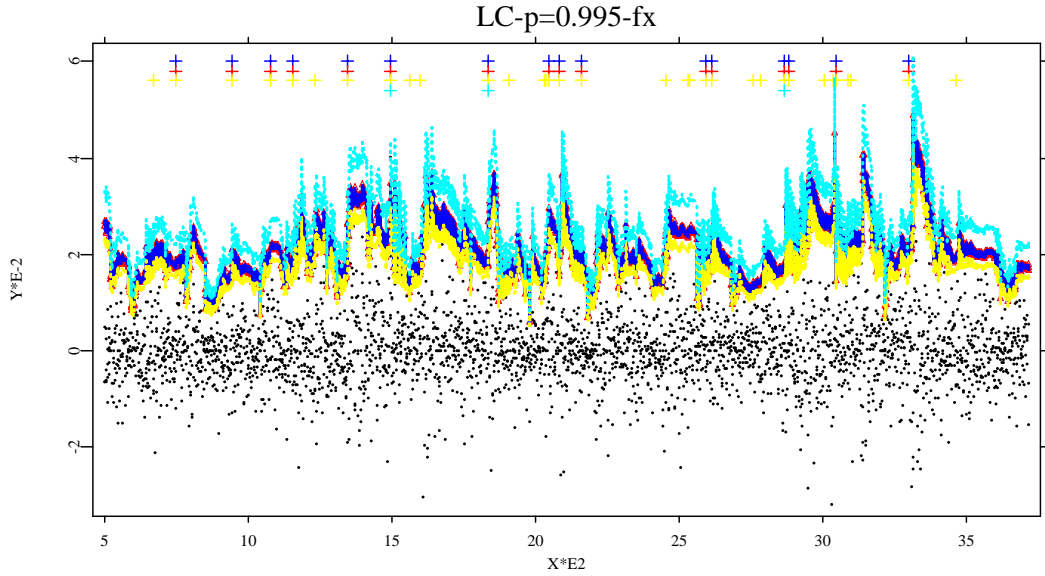

Figure 10: Time plot of the VaR forecasts for the DEM/USD data at $p = 0.995$. The returns are displayed as dots whereas the exceedances are marked as crosses w.r.t. the HYPADA (solid line), the NIGADA (triangles), the NADA (dashed line) and the tADA (dotted line) from the top down.

GHADAfxvar.xpl

5 CONCLUSION

In this paper, we have proposed a risk management (GHADA) model based on the adaptive volatility estimation and the GH distribution. Compared to some other proposed risk management models in Table 2, the GHADA technique gives more accurate VaR forecasts in real data analysis. Some interesting points are summarized:

- Above all, the two subclasses, HYP and NIG, of the GH distribution can better

23

DEM/USD data				German bank data			
$p = 0.95$	N/T	LR1	p-value	$p = 0.95$	N/T	LR1	p-value
HYPADA	0.9443	2.05	0.15	HYPADA	0.9515	-NAN	-NAN
NIGADA	0.9443	2.05	0.15	NIGADA	0.9482	-NAN	-NAN
NADA	0.9475	0.41	0.51	NADA	0.9547	-NAN	-NAN
tADA	0.9711	*35.37	0.00	tADA	0.9680	-NAN	-NAN
HYPGARCH	0.9521	0.32	0.57	HYPGARCH	<u>0.9512</u>	-NAN	-NAN
NIGGARCH	<u>0.9503</u>	0.00	0.93	NIGGARCH	0.9486	-NAN	-NAN
NGARCH	0.9521	0.32	0.57	NGARCH	0.9535	-NAN	-NAN
tGARCH	0.9692	*28.92	0.00	tGARCH	0.9696	+INF	0.00
$p = 0.975$	N/T	LR1	p-value	$p = 0.975$	N/T	LR1	p-value
HYPADA	0.9742	0.08	0.77	HYPADA	<u>0.9737</u>	0.32	0.56
NIGADA	0.9732	0.38	0.53	NIGADA	0.9698	5.28	0.02
NADA	0.9714	1.61	0.20	NADA	0.9682	-NAN	-NAN
tADA	0.9888	*31.65	0.00	tADA	0.9843	*20.91	0.00
HYPGARCH	<u>0.9754</u>	0.02	0.86	HYPGARCH	0.9772	1.10	0.29
NIGGARCH	0.9745	0.02	0.86	NIGGARCH	0.9764	0.46	0.49
NGARCH	0.9708	2.21	0.13	NGARCH	0.9706	3.83	0.05
tGARCH	0.9844	*13.65	0.00	tGARCH	0.9856	*28.22	0.00
$p = 0.99$	N/T	LR1	p-value	$p = 0.99$	N/T	LR1	p-value
HYPADA	<u>0.9897</u>	0.02	0.88	HYPADA	<u>0.9894</u>	0.17	0.67
NIGADA	<u>0.9897</u>	0.02	0.88	NIGADA	0.9884	1.20	0.27
NADA	0.9854	6.02	0.01	NADA	0.9811	*31.81	0.00
tADA	0.9972	*23.60	0.00	tADA	0.9941	*10.26	0.00
HYPGARCH	0.9906	0.15	0.69	HYPGARCH	0.9909	0.51	0.47
NIGGARCH	0.9906	0.15	0.69	NIGGARCH	0.9919	2.13	0.14
NGARCH	0.9822	*15.71	0.00	NGARCH	0.9835	*18.02	0.00
tGARCH	0.9956	*13.17	0.00	tGARCH	0.9954	*19.54	0.00
$p = 0.995$	N/T	LR1	p-value	$p = 0.995$	N/T	LR1	p-value
HYPADA	<u>0.9950</u>	0.00	0.98	HYPADA	<u>0.9945</u>	0.23	0.62
NIGADA	<u>0.9950</u>	0.00	0.98	NIGADA	0.9939	1.11	0.29
NADA	0.9897	*13.66	0.00	NADA	0.9880	*35.62	0.00
tADA	0.9990	*16.16	0.00	tADA	0.9968	4.11	0.04
HYPGARCH	<u>0.9950</u>	0.00	0.98	HYPGARCH	0.9960	1.29	0.25
NIGGARCH	<u>0.9950</u>	0.00	0.98	NIGGARCH	0.9962	1.83	0.17
NGARCH	0.9891	*16.68	0.00	NGARCH	0.9892	*25.70	0.00
tGARCH	0.9987	*13.09	0.00	tGARCH	0.9972	6.24	0.01

Table 7: Backtesting results for the DEM/USD data and the German bank potfolio data. * indicates that the model is rejected at 99% confidence level.
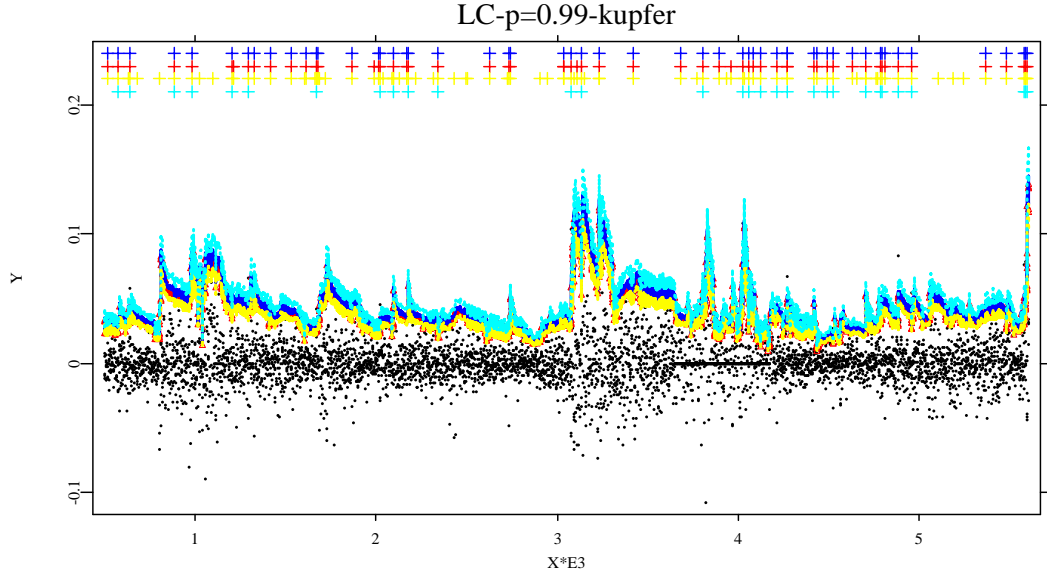
Figure 11: Time plot of the VaR forecasts for the German bank portfolio data at $p = 0.99$. The returns are displayed as dots whereas the exceedances are marked as crosses w.r.t. the HYPADA (solid line), the NIGADA (triangles), the NADA (dashed line) and the tADA (dotted line) from the top down.

describe the distributional features of the risk factors than the normal and t distributions. Both distributions well match the empirical density, especially the right tail behavior of losses, that in turn leads to precise quantile estimations at extreme risk levels. However we don't have enough evidence to say that one subclass is better than the other, although in the simulation study as well as the empirical study based on German bank portfolio, the HYP performs better. We consider that the performance of these two subclasses relies on the data set considered. A subjective suggestion is however that the HYP is more suitable for less liquid portfolio.

- The local constant technique gives at least comparable volatility estimation as the GARCH(1,1). The simulation study shows that sometimes the former provides more accurate volatility estimation on average. Furthermore, the difference between the detection speeds of two techniques is trivial. We want to emphasize here that although the GARCH(1,1) performs well in the simulation study, it is weakened by its theoretical assumption, i.e. the volatility estimation follows a time constant closed form. This influence is illustrated in the German bank portfolio analysis where even the HYP and NIG weakly represent the empirical density of the devolatilized return based on the GARCH(1,1) technique. On the contrary, the devolatilized return based on the local constant technique displays a nice fit without problem.

- The proposed approach can be easily applied to calculate and forecast other risk

25

measures such as expected shortfall.

In financial markets, it is more interesting and challenging to measure the risk levels of multiple time series. Härdle, Herwartz and Spokoiny (2003) have proposed an idea of the adaptive volatility estimation method for multiple time series. Prause (1999) has discussed the estimation of the high dimensional GH distribution. Both studies are however computationally cumbersome and infeasible as the dimension of data set increases. Instead, we expect to apply a different research thought, independent component analysis (ICA) to convert the high dimensional problem to univariate study via a linear transformation, see Hyvärinen, Karhunen and Oja (2001). The whole procedure can be formulated as:

$$
\begin{aligned}
R_t &= b_t^\top X_t \\
&= b_t^\top W Y_t \\
&= b_t^\top W \mathrm{diag}(\sigma_{y1,t}, \cdots, \sigma_{yd,t}) \varepsilon_t
\end{aligned}
$$

where R_t denotes the portfolio loss with d-dimensional individual instruments that have losses $X_t \in \mathbb{R}^d$. The weights of the portfolio is expresses as b_t. After the linear transformation with the matrix W, one obtains (approximately) independent components Y_t. For each IC, the GHADA technique helps to specify the volatility and marginal distribution. Since the components are independent, then joint density is simply the product of these marginals. A detailed study on this approach is available at Chen, Härdle and Spokoiny (2005).

6 APPENDIX

For a GH distributed random variable, we have the following lemma:

LEMMA 1 *For every $0 \leq \gamma \leq 1$ there exists a constant $a_\gamma > 0$ such that*

$$
\log \mathsf{E}[e^{u\zeta_\gamma}] \leq \frac{a_\gamma u^2}{2},
$$

where $\zeta_\gamma = (|\varepsilon|^\gamma - C_\gamma)/D_\gamma$ with ε from a GH distribution.

Proof of Lemma 1.

Proof:
Firstly we show that the moment generating function $\mathsf{E}[e^{u\zeta_\gamma}]$ exists for all $u \in \mathbb{R}$.

Suppose that $\mathcal{L}(x) = GH(\lambda, \alpha, \beta, \delta, \mu)$ with the density function f for the transformed variable $y \stackrel{\text{def}}{=} |x|^\gamma$, we have

$$P(y \leq z) = P(-z^{\frac{1}{\gamma}} \leq x \leq z^{\frac{1}{\gamma}}) = \int_{-\infty}^{z^{\frac{1}{\gamma}}} f(x)dx - \int_{-\infty}^{-z^{\frac{1}{\gamma}}} f(x)dx, \quad z > 0$$

Then the density of $y \in (0, \infty)$ is:

$$\begin{aligned}
g(z) = \frac{d}{dz}P(y \leq z) &= \gamma^{-1}\{f(z^{\frac{1}{\gamma}})z^{\frac{1}{\gamma}-1} + f(-z^{\frac{1}{\gamma}})z^{\frac{1}{\gamma}-1}\} \\
&= \gamma^{-1}z^{\frac{1}{\gamma}-1}\{f(z^{\frac{1}{\gamma}}) + f(-z^{\frac{1}{\gamma}})\}, \quad z > 0.
\end{aligned}$$

Since $f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu = 0) \sim x^{\lambda-1}e^{-(\alpha-\beta)x}$ as $x \to \pm\infty$, it follows

$$\begin{aligned}
g(z) &\sim \frac{z^{\frac{1}{\gamma}-1}}{\gamma}\{z^{\frac{\lambda-1}{\gamma}}e^{(\beta-\alpha)z^{\frac{1}{\gamma}}} + z^{\frac{\lambda-1}{\gamma}}e^{-(\beta+\alpha)z^{\frac{1}{\gamma}}}\} \\
&= \frac{z^{\frac{\lambda}{\gamma}-1}}{\gamma}\{e^{(\beta-\alpha)z^{\frac{1}{\gamma}}} + e^{-(\beta-\alpha)z^{\frac{1}{\gamma}}}\}, \quad z \longrightarrow \infty
\end{aligned}$$

For $\gamma < 1$, it holds that $\int_0^\infty e^{uz}g(z)dz < \infty \ \forall u \in \mathbb{R}$, since

$$\begin{aligned}
\lim_{z\to\infty}(\beta - \alpha)z^{\frac{1}{\gamma}} + uz &\to -\infty \quad \forall u \in \mathbb{R} \\
\lim_{z\to\infty} -(\beta + \alpha)z^{\frac{1}{\gamma}} + uz &\to -\infty \quad \forall u \in \mathbb{R}
\end{aligned}$$

Since the integration depends only on the exponential part, it holds also that

$$\int_0^\infty z^n e^{uz}g(z)dz = \int_0^\infty \frac{\partial^n}{\partial u^n}(e^{uz})g(z)dz = \frac{\partial^n}{\partial u^n}\mathsf{E}[e^{uy}] < \infty,$$

then it can be shown that the moment generating function and $\log(\mathsf{E}[e^{uy}])$ are smooth. It holds for every $t > 0$,

$$\begin{aligned}
\mathsf{E}[e^{uy}] = \mathsf{E}[e^{u|x|^\gamma}] &= \mathsf{E}[e^{u|x|^\gamma}\mathbb{1}(|x| \leq t)] + \mathsf{E}[e^{u|x|^\gamma}\mathbb{1}(|x| > t)] \\
&\leq e^{ut^\gamma} + \mathsf{E}[e^{|x|ut^{\gamma-1}}I(|x| > t)],
\end{aligned} \quad (23)$$

Without loss of generality, we assume $\mu = 0$. Further

$$f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu = 0) \sim x^{\lambda-1}e^{-(\alpha-\beta)x} \text{ as } x \to \infty,$$

and $\int_y^\infty x^{\lambda-1}e^{-x}dx \sim y^{\lambda-1}e^{-y}$ as $y \to \infty$, Press et al. (1992).

For an arbitrary but fixed $u \in \mathbb{R}_+$ and $t_0 > 1$ so that $ut^{\gamma-1} < \alpha - \beta$, it holds for all $t \geq t_0$

$$f(t) \leq C_1 t^{\lambda-1}e^{(\beta-\alpha)t}$$

27

$$\int_{(\alpha-\beta-ut^{\gamma-1})t}^{\infty} x^{\lambda-1}e^{-x}dx \leq C_2[(\alpha-\beta-ut^{\gamma-1})t]^{\lambda-1}e^{-(\alpha-\beta-ut^{\gamma-1})t}$$

where $C_1, C_2 > 1$.

Consequently for $t \geq t_0$,

$$\begin{aligned}
\mathsf{E}[e^{u|t|^{\gamma-1}x}1(|x| > t)] &= \int_t^{\infty} e^{ut^{\gamma-1}x}f(x)dx \leq C_1 \int_t^{\infty} e^{ut^{\gamma-1}x}x^{\lambda-1}e^{-(\alpha-\beta)x}dx \\
&= C_1 \int_t^{\infty} x^{\lambda-1}e^{-(\alpha-\beta-ut^{\gamma-1})x}dx \\
&= C_1(\alpha-\beta-ut^{\gamma-1})^{-\lambda} \int_{(\alpha-\beta-ut^{\gamma-1})t}^{\infty} x^{\lambda-1}e^{-x}dx \\
&\leq C_1 C_2 t^{\lambda-1}e^{-(\alpha-\beta-ut^{\gamma-1})t}(\alpha-\beta-ut^{\gamma-1}t)^{-1} \qquad (24)
\end{aligned}$$

If u is so large that $t \overset{\text{def}}{=} (\frac{\alpha-\beta}{2})^{\frac{1}{\gamma-1}}u^c \geq t_0$ with $\frac{1}{1-\gamma} \leq c$, then (24) holds true since $ut^{\gamma-1} = (\frac{\alpha-\beta}{2})uu^{c(\gamma-1)} \leq \frac{\alpha-\beta}{2} < \alpha-\beta$.

Given $t = (\frac{\alpha-\beta}{2}u)^{\frac{1}{1-\gamma}}$, we get

$$\mathsf{E}[e^{ut^{\gamma-1}x}1(|x| > t)] \leq \frac{2C_1 C_2}{\alpha-\beta}(\frac{\alpha-\beta}{2}u)^{\frac{\lambda-1}{1-\gamma}}e^{-\frac{\alpha-\beta}{2}(\frac{\alpha-\beta}{2}u)^{\frac{1}{1-\gamma}}}.$$

From which we get

$$\log(\mathsf{E}[e^{ut^{\gamma-1}}1(x > t)]) \leq C_3 + \frac{\lambda-1}{1-\gamma}\log(u) - (\frac{\alpha-\beta}{2})^{\frac{2-\gamma}{1-\gamma}}u^{\frac{1}{1-\gamma}}$$

Further $\log(\mathsf{E}[e^{ut^{\gamma-1}}1(x > t)])u^{-\frac{1}{1-\gamma}}$ is also bounded for $u \to \infty$. Analogously we can show the bounding of $\log(\mathsf{E}[e^{ut^{\gamma-1}}1(x < -t)])u^{-\frac{1}{1-\gamma}}$. Therefore for $\gamma < 1$ the whole term $\mathsf{E}[e^{u|x|^{\gamma}}1(|x| > t)]u^{-\frac{1}{1-\gamma}}$ is bounded as $u \to \infty$.

Given $t = (\frac{\alpha-\beta}{2}u)^{\frac{1}{1-\gamma}}$, we have

$$\begin{aligned}
e^{ut^{\gamma}} &= e^{(\frac{\alpha-\beta}{2})^{\frac{\gamma}{1-\gamma}}u^{\frac{1}{1-\gamma}}} \\
u^{-\frac{1}{1-\gamma}}\log(e^{ut^{\gamma}}) &= (\frac{\alpha-\beta}{2})^{\frac{\gamma}{1-\gamma}} = \text{constant}
\end{aligned}$$

Thus $u^{-\frac{1}{1-\gamma}}\log(\mathsf{E}[e^{u|x|^{\gamma}}]) \leq u^{-\frac{1}{1-\gamma}}[\log(e^{ut^{\gamma}}) + \log\{E[e^{ut^{\gamma-1}|x|}1(|x| > t)]\}]$ is bounded for $u \to \infty$, i.e. for a sufficient large u_0 there exist a constant $C_u > 0$ such that

$$\mathsf{E}[e^{u|x|^{\gamma}}] \leq C_u u^{\frac{1}{1-\gamma}}, \quad u \geq u_0.$$

$\square$

Proof of the Martingale property of θ_t Consider a predictable process p_t (such as the volatility σ_t or the local parameter θ_t) w.r.t. the information set $\mathcal{F}_{t-1}$:

$$\Upsilon_t = \exp\left(\sum_{s=1}^{t} p_s \zeta_s - (a_\gamma/2) \sum_{s=1}^{t} p_s^2\right)$$

Υ_t is a supermartingale, since

$$
\begin{aligned}
\mathsf{E}(\Upsilon_t|\mathcal{F}_{t-1}) - \Upsilon_{t-1} &= \mathsf{E}(\Upsilon_t|\mathcal{F}_{t-1}) - \mathsf{E}(\Upsilon_{t-1}|\mathcal{F}_{t-1}) \\
&= \mathsf{E}[\exp\left(\sum_{s=1}^{t} p_s \zeta_s - (a_\gamma/2) \sum_{s=1}^{t} p_s^2\right) \\
&\quad - \exp\left(\sum_{s=1}^{t-1} p_s \zeta_s - (a_\gamma/2) \sum_{s=1}^{t-1} p_s^2\right) |\mathcal{F}_{t-1}] \\
&= \mathsf{E}[\exp\left(\sum_{s=1}^{t-1} p_s \zeta_s - (a_\gamma/2) \sum_{s=1}^{t-1} p_s^2\right) \{\exp(p_t \zeta_t - a_\gamma/2 p_t^2) - 1\}|\mathcal{F}_{t-1}] \\
&= \underbrace{\frac{\exp(p_1 \zeta_1)}{\exp(a_\gamma/2 p_1)}}_{\leq 1, Lemma 1} \cdots \underbrace{\frac{\exp(p_{t-1}\zeta_{t-1})}{\exp(a_\gamma/2 p_{t-1})}}_{\leq 1} \cdot \underbrace{\mathsf{E}[\frac{\exp(p_t \zeta_t)}{\exp(a_\gamma/2 p_t)} - 1|\mathcal{F}_{t-1}]}_{\leq 1} \\
&\leq 0
\end{aligned}
$$

i.e. $\mathsf{E}(\Upsilon_t|\mathcal{F}_{t-1}) \leq \Upsilon_{t-1}$. By this lemma, we obtain a generalized version of Theorem 3.1 in Mercurio and Spokoiny (2004) to the case when ε are from a GH distribution. The statistical properties of $\hat{\theta}_I$ are given in Theorem 1.

References

Andersen, T., Bollerslev, T., Christoffersen, P. and Diebold, F. (2005). Volatility and correlation forecasting, *in* G. Elliott, C. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Amsterdam: North-Holland.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size, *Proceedings of the Royal Society of London* **A 353**: 401–419.

Barndorff-Nielsen, O. (1997). Normal inverse gaussian distributions and stochastic volatility modelling, *Scandinavian Journal of Statistics* **24**: 1–13.

Barndorff-Nielsen, O. and Shephard, N. (2001). Modelling by lévy processes for financial econometrics, *in* O. Barndorff-Nielsen, T. Mikosch and S. Resnik (eds), *Lévy Processes : Theory and Applications*, Birkhauser Boston.

Barndorff-Nielsen, O. E. and Blæsild, P. (1981). Hyperbolic distribution and ramifications: Contributions to theory and applications, *in* C. Taillie, P. G. Patil and A. Baldessari (eds), *Statistical Distributions in Scientific Work*, Vol. 4, D. Reidel, pp. 19–44.

Bibby, B. M. and Sørensen, M. (2001). *Hyperbolic Processes in Finance, Technical Report 88*, University of Aarhus, Aarhus School of Business.

Bollerslev, T. (1995). Generalied autoregressive conditional heteroskedasticity, *in* R. Engle (ed.), *ARCH, selected readings*, Oxford University Press, pp. 42–60.

Chen, Y., Härdle, W. and Spokoiny, V. (2005). Ghica - risk analysis with gh distributions and independent components.

Christoffersen, P. F. (1998). Evaluating interval forecast, *International Economic Review* **39**: 841–862.

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

Eberlein, E., Kallsen, J. and Kristen, J. (2003). Risk management based on stochastic volatility, *Journal of Risk* **5**: 19–44.

Embrechts, P., McNeil, A. and Straumann, D. (1999). Correlation: pitfalls and alternatives, *Risk* **12**: 69–71.

Engle, R. F. (1995). *Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, ARCH*, Oxford University Press.

Franke, J., Härdle, W. and Hafner, C. (2004). *Statistics of Financial Markets*, Springer-Verlag Berlin Heidelberg New York.

Härdle, W., Herwartz, H. and Spokoiny, V. (2003). Time inhomogeneous multiple volatility modelling, *Journal of Financial Econometrics* **1**: 55–95.

Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and Semi-parametric Models*, Springer-Verlag Berlin Heidelberg New York.

Harvey, A., Ruiz, E. and Shephard, N. (1995). Multivariate stochastic variance models, *in* R. Engle (ed.), *ARCH, selected readings*, Oxford University Press, pp. 256–276.

Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis*, John Wiley & Sons, Inc.

Jaschke, S. and Jiang, Y. (2002). Approximating value at risk in conditional gaussian models, *in* W. Härdle, T. Kleinow and G. Stahl (eds), *Applied Quantitative Finance*, Springer-Verlag Berlin Heidelberg New York.

Jorion, P. (2001). *Value at Risk*, McGraw-Hill.

Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time inhomogeneous volatility models, *Annals of Statistics* **32**: 577–602.

Poon, S. and Granger, C. (2003). Forecasting volatility in financial markets: A review, *Journal of Economic Literature* **XLI**: 478–539.

Prause, K. (1999). The generalized hyperbolic model: Estimation, financial derivatives and risk measures, *dissertation*.

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. (1992). *Numerical Recipes in C*, Cambridge University Press.