GHICA - Risk Analysis with GH Distributions and Independent Components

Ying Chen*
Wolfgang Härdle**
Vladimir Spokoiny*

* Center for Applied Statistics and Economics (CASE),
Humboldt-Universität zu Berlin
and Weierstraß-Institut für Analysis und Stochastik, Germany

** Center for Applied Statistics and Economics (CASE),
Humboldt-Universität zu Berlin, Germany

SFB 649 ECONOMIC RISK BERLIN

GHICA - Risk Analysis with GH Distributions and Independent Components

Ying Chen[1,2], Wolfgang Härdle[1] and Vladimir Spokoiny[1,2]

[1] CASE - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin
Wirtschaftswissenschaftliche Fakultät
Spandauerstrasse 1, 10178 Berlin, Germany
[2] Weierstraß - Institute für Angewandte Analysis und Stochastik
Mohrenstrasse 39, 10117 Berlin, Germany

Abstract

Over recent years, study on risk management has been prompted by the Basel committee for regular banking supervisory. There are however limitations of some widely-used risk management methods that either calculate risk measures under the Gaussian distributional assumption or involve numerical difficulty. The primary aim of this paper is to present a realistic and fast method, **GHICA**, which overcomes the limitations in multivariate risk analysis. The idea is to first retrieve independent components (ICs) out of the observed high-dimensional time series and then individually and adaptively fit the resulting ICs in the generalized hyperbolic (GH) distributional framework. For the volatility estimation of each IC, the local exponential smoothing technique is used to achieve the best possible accuracy of estimation. Finally, the fast Fourier transformation technique is used to approximate the density of the portfolio returns.

The proposed GHICA method is applicable to covariance estimation as well. It is compared with the dynamic conditional correlation (DCC) method based on the simulated data with $d = 50$ GH distributed components. We further implement the GHICA method to calculate risk measures given 20-dimensional German DAX portfolios and a dynamic exchange rate portfolio. Several alternative methods are considered as well to compare the accuracy of calculation with the GHICA one.

Keywords: multivariate risk management, independent component analysis, generalized hyperbolic distribution, local exponential estimation, value at risk, expected shortfall

1 Introduction

Over recent years, study on risk management has been prompted by the Basel committee for regular banking supervisory. Given a d-dimensional portfolio, the conditionally heteroscedastic model is widely used to describe the movement of the underlying series:

$$x(t) = \Sigma_x^{1/2}(t)\varepsilon_x(t), \tag{1}$$

where $x(t) \in \mathrm{IR}^d$ are risk factors of the portfolio, e.g. (log) returns of the financial instruments. The covariance Σ_x is assumed to be predictable with respect to (w.r.t.) the past information and $\varepsilon_x(t) \in \mathrm{IR}^d$ is a sequence of standardized innovations with $\mathsf{E}[\varepsilon_x(t)|\mathcal{F}_{t-1}] = 0$ and $\mathsf{E}[\varepsilon_x^2(t)|\mathcal{F}_{t-1}] = I_d$. There is a sizeable literature on risk management methods. Among others, we refer to Jorion (2001) for a systematic description.

In this paper, we focus on the calculation of two risk measures, value at risk (VaR) and expected shortfall (ES). These two risk measures are inherently related to the joint density of $x(t)$. The VaR is in fact the distributional quantile of loss, i.e. $-x(t)$, at a prescribed level over a target time horizon and the ES measures the size of loss once the loss exceeds the VaR value. Indicated by formula (1), the joint density estimation depends on the covariance estimation and the distributional assumption of the innovations.

The largest challenge of risk management is due to the high-dimensionality of real portfolios. Above all, the covariance estimation is really computationally demanding as high dimensional series, e.g. a dimension $d > 10$, is considered, see Härdle, Herwartz and Spokoiny (2003). For example, the dynamic conditional correlation (DCC) model proposed by Engle (2002), Engle and Sheppard (2001), which is one multivariate GARCH model, is recommended due to the good performance of its univariate version. In the estimation, the covariance matrix is approximated by the product of a diagonal matrix and a correlation matrix, which reduces the number of unknown parameters much relative to the BEKK specification proposed by Engle and Kroner (1995). In spite of the appealing dimensional reduction, the mentioned estimation method is time consuming and numerically difficult to handle given high-dimensional data.

Moreover, many widely-used risk management methods rely on the unrealistic Gaussian distributional assumption, e.g. the RiskMetrics product introduced by JP Morgan in 1994. In the Gaussian framework with an estimate $\hat{\Sigma}_x(t)$ of $\Sigma_x(t)$, the standardized returns $\hat{\varepsilon}_x(t) = \hat{\Sigma}_x^{(-1/2)}(t)x(t)$ are asymptotically independent and the joint distributional behavior can be easily measured by the marginal distributions. However the Gaussian distributional assumption is merely used for computational and numerical purposes and not for statistical reasons. The conditional Gaussian marginal distributions and the resulting joint Gaussian distribution are at odds with empirical facts, i.e. financial series are heavy tailed distributed.

The heavy tails are typically reduced but not eliminated as the series are standardized by the estimated volatility, see Anderson, Bollerslev, Diebold and Labys (2001).

We illustrate this effect based on two real data sets, the Allianz stock and a DAX portfolio from 1988/01/04 to 1996/12/30. The DAX is the leading index of Frankfurt stock exchange and a 20-dimensional hypothetic portfolio with a static trading strategy $b(t) = (1/20, \cdots, 1/20)^\top$ is considered. The portfolio returns $r(t) = b(t)^\top x(t)$ are analyzed in the univariate version of (1). This simplified calculation is used in practice, but it often suffers from low accuracy of calculation. Suppose now that the two return processes have been properly standardized, by using a local volatility estimation technique discussed later. The standardized returns are empirically heavy-tailed distributed, indicated by the sample kurtoses 12.07 for the Allianz and 22.38 for the portfolio respectively.

Figure 1 displays the estimated logarithmic density curves under several distributional assumptions. Among them, the estimate using the nonparametric kernel estimation is considered as benchmark. The comparison w.r.t. the Allianz stock shows that the GH estimate is most close to the benchmark among others. The Gaussian estimate presents lighter tails. To alleviate the limitation, the Student-$t(6)$ distribution with degrees of freedom of 6 has been recommended in practice. However this distribution is found to over-fit the heavy tails, namely the $t(6)$ estimate displays heavier tails relative to the benchmark. The similar result is observed w.r.t. the DAX portfolio. It is rational to surmise that the risk management methods under the Gaussian and $t(6)$ distributional assumptions generate low accurate results.

To overcome these limitations, Chen, Härdle and Spokoiny (2006) present a simple VaR calculation approach that achieves much better accuracy than the alternative RiskMetrics method. In their study, univariate approaches that involve more realistic but complex procedures can be easily extended for multivariate risk measurement. To be more specific, financial risk factors are first converted to independent components (ICs) using a linear filtering and the univariate method is applied to identify the distributional behavior of each IC. We name here two univariate approaches which measure the risk exposure in the realistic distributional framework. One is the univariate VaR calculation proposed by Chen, Härdle and Jeong (2005), which implements local constant model to estimate volatility and fit the standardized returns under the GH distributional assumption. The other is proposed by Chen and Spokoiny (2006), who apply the local exponential smoothing method to estimate volatility and calculate the risk measure in the GH distributional framework. The standardization of the Allianz and DAX returns in Figure 1 is in fact based on the local exponential smoothing technique.

The primary aim of this paper is to present an realistic and fast multivariate risk management method, **GHICA**, by implementing the IC analysis (ICA) to the high dimensional series and adaptively fitting the ICs in the GH distributional framework. The GHICA
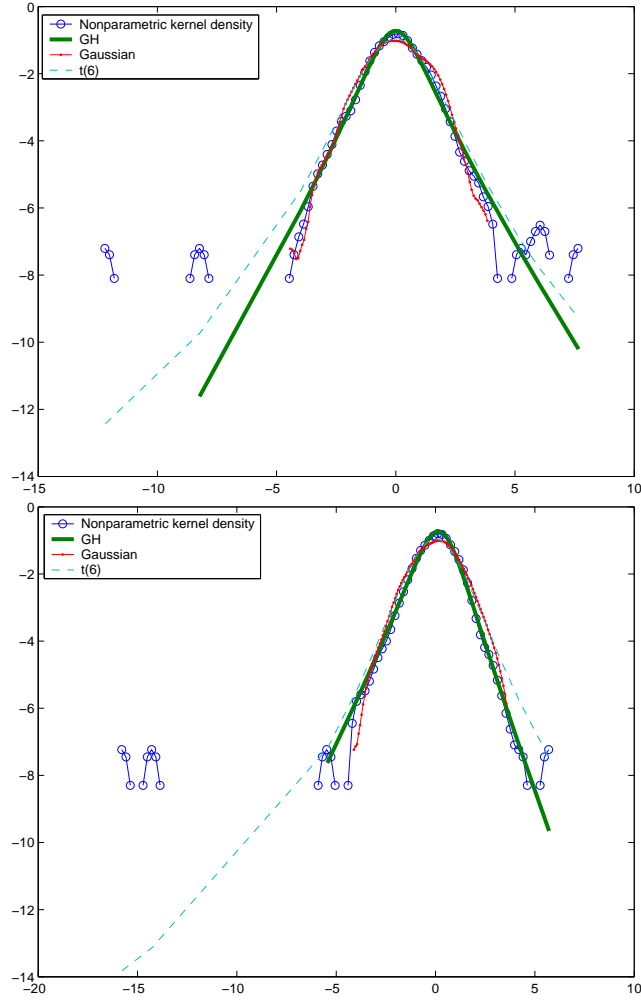
Fig. 1: Density comparisons of the standardized returns in log scale based on the Allianz stock (top) and the DAX portfolio (bottom) with static weights $b(t) = \text{unit}(1/20)$. Time interval: 1988/01/04 - 1996/12/30. The nonparametric kernel density is considered as benchmark. The GH distributional parameters are respectively $GH(-0.5, 1.01, 0.05, 1.11, -0.03)$ for the Allianz and $GH(-0.5, 1.21, -0.21, 1.21, 0.24)$ for the DAX portfolio. Data source: FEDC (http://sfb649.wiwi.hu-berlin.de).

method improves the work of Chen et al. (2006) from two aspects. The volatility estimation is driven by the local exponential smoothing technique to achieve the best possible accuracy of estimation. The fast Fourier transformation (FFT) technique is used to approximate the density of the portfolio returns. Compared to the Monte Carlo simulation technique used in the former study, it significantly speeds up the calculation.

In addition, the proposed GHICA method is easily applicable for covariance estimation. Relative to the widely used DCC setup, the GHICA method is fast and delivers sensitive estimates. We demonstrate the comparison based on simulated data. Furthermore, the

5

GHICA method is implemented to risk management on the base of DAX stocks and foreign exchange rates. Several hypothetic portfolios are constructed by assigning static and dynamic trading strategies to the data sets. The results are compared with those calculated using alternative methods, i.e. the RiskMetrics method, the method using the exponential smoothing to estimate volatility and assuming the Student-$t(6)$ distribution, and the method using the DCC to estimate covariance in the Gaussian distributional framework. All the results are analyzed from the viewpoints of regulatory, investors and internal supervisory. The GHICA method, in general, produces better results than the others.

The paper is organized as follows. The GHICA method is described in Section 2, by which the ICA method, the local exponential smoothing technique and the FFT technique are detailed. Section 3 compares the covariance estimation using the GHICA and DCC methods based on the simulated data with $d = 50$ GH components. The real data analysis in Section 4 demonstrates the implementation of the GHICA method in risk management based on the 20-dimensional German DAX portfolios and a dynamic exchange rate portfolio. Several alternative methods are considered as well to compare the accuracy of calculation with the GHICA one.

2 GHICA Methodology

Given multidimensional time series, for example prices of financial assets, $s(t) \in \mathbb{R}^d$, the (log) returns are calculated as $x(t) = \log\{s(t)/s(t-1)\}$. Without loss of generality, the drift of the returns is set to be 0. Given the time homogeneous model, $x(t) = \Sigma_x^{1/2}\varepsilon_x(t)$ with standardized innovations $\varepsilon_x(t)$, the maximum Gaussian likelihood estimate of the time independent covariance Σ_x is the sample covariance based on the whole past information. Since the covariance is in fact time dependent, one considers the conditional heteroscedastic model:

$$x(t) = \Sigma_x^{1/2}(t)\varepsilon_x(t).$$

Many techniques have been used to approximate the local covariance by specifying a "local homogeneous" interval (e.g. one year or 250 trading days). Inside the homogeneous interval, the unknown covariance should be time-invariant and can be identified using the ML estimation. Among many others, the multivariate GARCH setup such as the DCC is successful in characterizing the clustering feature of covariance under the Gaussian distributional assumption. As the dimension d increases, it however needs to estimate many parameters and becomes numerically difficult. Moreover, the standardized returns $\hat{\varepsilon}_x(t) = \hat{\Sigma}_x^{-1/2}(t)x(t)$ are empirically not Gaussian distributed. Under a realistic distributional assumption, on the other hand, by which the distributional behaviors such as asymmetry and heavy tails are well matched, it is hard to identify the unknown distributional parameters due to complex density form.

The GHICA method proposes a solution to balance the numerical tractability and the realistic distributional assumption on the risk factors. It first converts the return series using a linear transformation and filters out ICs: $y(t) = Wx(t)$. The transformation matrix W is assumed to be time constant and nonsingular and $y(t)$ is the independent vector. The heteroscedastic model is now reformulated as:

$$x(t) = W^{-1}y(t) = W^{-1}\Sigma_y^{1/2}(t)\varepsilon_y(t) = W^{-1}D_y^{1/2}(t)\varepsilon_y(t).$$

Due to the statistical property of independence, the covariance of the ICs $\Sigma_y(t)$ is a diagonal matrix and is denoted as $D_y(t)$ to emphasize this feature. Its diagonal elements are the time varying variances of the ICs. The stochastic innovations $\varepsilon_y(t) = \{\varepsilon_{y_1}(t), \cdots, \varepsilon_{y_d}(t)\}^\top$ are cross independent and can be individually identified in the realistic and univariate distributional framework. By doing so, the GHICA method converts the high dimensional analysis to univariate study and significantly speeds up the calculation.

In this section, the building blocks of the GHICA method are detailed: The FastICA procedure is used to estimate the transformation matrix W; The resulting ICs are individually analyzed, by which the univariate volatility process is estimated using the local exponential smoothing technique and the innovations are assumed to be GH distributed; The quantile of the portfolio return is approximated using the FFT technique.

The GHICA algorithm is summarized as follows:

1. Do ICA to the given risk factors to get ICs.

2. Implement local exponential smoothing to estimate the variance of each IC

3. Identify the distribution of every IC's innovation in the GH distributional framework

4. Estimate the density of the portfolio return using the FFT technique

5. Calculate risk measures

In addition, the GHICA method can be used to estimate the covariance matrix $\Sigma_x(t)$. Given the matrix estimate $\hat{W}$ in the ICA and the variance estimates of the ICs, the covariance of the observed time series are: $\hat{\Sigma}_x(t) = \hat{W}^{-1}\hat{D}_y(t)\hat{W}^{-1\top}$. An alternative covariance estimation approach, the DCC, is briefly described as well. We will compare the GHICA-based covariance estimation with the DCC estimation in the later simulation study.

2.1 Independent component analysis (ICA) and FastICA approach

The aim of ICA is to retrieve, out of high dimensional time series, stochastically ICs through a linear transformation: $y(t) = Wx(t)$, where the transformation matrix $W = (w_1, \cdots, w_d)^\top$

is nonsingular. It is essential to use high order moments in the ICA. In the Gaussian framework, high order moments are however fixed such as skewness with value of 0 and kurtosis with value of 3. Therefore the ICs are assumed to be nongaussian distributed. Furthermore, the ICA transformation has scale identification problem, i.e. the equation holds true by simultaneously multiplying the same constants to the unknown terms $y(t)$ and W: $\{cy(t)\} = \{cW\}x(t)$. To avoid this problem, it is natural to standardize the dependent series and assume that every IC has unit variance $\mathsf{E}(y_j) = 1$ with $j = 1, \cdots, d$. The Mahalanobis transformation $\tilde{x}(t) = \tilde{\Sigma}_x^{-1/2} x(t)$ helps to standardize the return series and the resulting series are considered:

$$y(t) = \tilde{W}\tilde{x}(t),$$

where $\tilde{\Sigma}_x$ is the sample covariance based on the available data. It is easy to show that after the standardization the transformation matrix $\tilde{W}$ turns to be an orthogonal matrix with unit norm. The corresponding matrix w.r.t. the return series is $W = \tilde{W}\tilde{\Sigma}_x^{-1/2}$. For notational simplification, we eliminate the mark $\tilde{\ }$ in the following text in this section.

Various ideas have been proposed to estimate the transformation matrix W. Among others, one intuitive ICA estimation is motivated by the definition of mutual information. The mutual information is a natural measure of independence. It is defined as the difference of the sum of marginal entropy and the mutual entropy:

$$
\begin{aligned}
I(y) &= \sum_{j=1}^{d} H(y_j) - H(y) \\
\text{where } H(y_j) &= -\int f_{y_j}(u) \log f_{y_j}(u) du
\end{aligned}
\tag{2}
$$

The mutual information is nonnegative and goes to 0 if the vector y is cross independent, see Cover and Thomas (1991). Hence for a candidate transformation W, one can minimize the mutual information to achieve independence. Based on the linear transformation of the ICA, the mutual information in (2) can be reformulated as:

$$I(W, y) = \sum_{j=1}^{d} H(y_j) - H(x) - \log|\det(W)|.$$

Notice that the entropy of the return series $H(x)$ is a fixed value and does not depend on the ICs, and the last term in the equation is 0 due to the orthogonality of the transformation matrix W. The optimization problem is: $\min_W \sum_{j=1}^{d} H(y_j)$ and can be further simplified to d optimization problems according to the inequality:

$$\min_W \sum_{j=1}^{d} H(y_j) \geq \sum_{j=1}^{d} \min_{w_j} H(y_j)$$

8

This simplification leads to some loss in the W estimation but it extensively speeds up the estimation procedure by merely considering d elements of W every time. Equivalently, one can formulate the optimization problem concerning negentropy $J(y_j) = H(y_0) - H(y_j)$ since the entropy and the negentropy are in one-to-one correspondence, where $y_0 \sim \mathrm{N}(0,1)$ is a standard Gaussian vector and $H(y_0)$ is merely a constant. The negentropy is always nonnegative since the Gaussian random variable has the largest entropy given the same variance, see Hyvärinen (1998).

$$\hat{w}_j = \operatorname{argmin} H(y_j) = \operatorname{argmax} J(w_j, y_j).$$

In the estimation, the approximation of negentropy is used to construct the optimization object function w.r.t. the j-th row of the transformation matrix W:

$$
\begin{aligned}
\hat{w}_j &= \operatorname{argmin} H(y_j) = \operatorname{argmax} J(y_j) \\
J(y_j) &\approx const.\{\mathsf{E}[G(y)] - \mathsf{E}[G(y_0)]\}^2 \\
&= const.\{\mathsf{E}[G(w_j^\top x)] - \mathsf{E}[G(y_0)]\}^2 \\
G(y_j) &= \log \cosh(y_j)
\end{aligned}
\tag{3}
$$

This optimization problem is solved by using the symmetric FastICA algorithm, see Hyvärinen, Karhunen and Oja (2001):

1. Initialization: Choose initial vectors $\hat{w}_j^{(1)}$ for $W = \{w_1, \cdots, w_d\}^\top$ with $j = 1, \cdots, d$, each has a unit norm.

2. Loop:

   - At step n, Calculate $\hat{w}_j^{(n)} = \mathsf{E}\left[x^\top(t) g\left\{\hat{w}_j^{(n-1)\top} x(t)\right\}\right] - \mathsf{E}\left[g'\left\{\hat{w}_j^{(n-1)\top} x(t)\right\}\right] \hat{w}_j^{(n-1)}$, where g is the first derivative of $G(y)$ in form (3) and g' is the second derivative. The expectation $\mathsf{E}[\cdot]$ is approximated by the sample mean.

   - Do a symmetric orthogonalization of the estimated transformation matrix $\hat{W}^{(n)}$:

     $$\hat{W}^{(n)} = \{\hat{W}^{(n)} \hat{W}^{(n)\top}\}^{-1/2} \hat{W}^{(n)}$$

   - If not converged, i.e. $\det\{\hat{W}^{(n)} - \hat{W}^{(n-1)}\} \neq 0$, go back to 2. Otherwise, the algorithm stops.

3. Final result: the last (converged) estimate is the final estimate $\hat{W}$.

2.2 Local exponential smoothing and dynamically conditional correlation

Suppose that the ICs and the transformation matrix W are given. The covariance matrices of the ICs and the original return series are respectively:

$$
\begin{aligned}
D_y(t) &= \operatorname{diag}\{\sigma_{y_1}^2(t), \cdots, \sigma_{y_d}^2(t)\} \\
\Sigma_x(t) &= W^{-1} D_y(t) W^{-1\top}
\end{aligned}
\tag{4}
$$

where $\sigma_{y_j}(t)$ is the heteroscedastic volatility of the j-th IC with $j = 1, \cdots, d$. Recall that (4) has a similar decomposition structure as the often-used principal component analysis (PCA), by which the covariance is decomposed as: $\Sigma_x = \Gamma \Lambda \Gamma^\top$ with the eigenvector matrix Γ and the diagonal eigenvalue matrix Λ, see Flury (1998). Among other distinctions, the PCA method orders the resulting PCs whereas the ICs have equal importance. In the estimation of the unknown variance, the local exponential smoothing method is used.

Local exponential smoothing: Given the univariate conditional heteroscedastic model: $y_j(t) = \sigma_{y_j}(t)\varepsilon_{y_j}(t)$ with $\mathsf{E}[\varepsilon_{y_j}(t)|\mathcal{F}_{t-1}] = 0$ and $\mathsf{E}[\varepsilon_{y_j}^2(t)|\mathcal{F}_{t-1}] = 1$, we now focus on the adaptive estimation of the volatility σ_{y_j} for $j = 1, \cdots, d$. For notational simplification, the subscripts y_j in σ_{y_j} and j in y_j are eliminated here.

Suppose that a finite set $\{\eta_k, k = 1, \cdots, K\}$ of values of smoothing parameter is given. Every value η_k leads to a localizing weighting scheme $\{\eta_k^{t-s}\}$ for $s \le t$ to the local Gaussian MLE $\tilde{\sigma}^{(k)}(t)$

$$
\tilde{\sigma}^{(k)}(t) = \left[\{\sum_{m=0}^{\infty} \eta_k^m y^2(t-m-1)\} / \{\sum_{m=0}^{\infty} \eta_k^m\} \right]^{1/2}
$$

In practice, one truncates the smoothing window at M_k such that $\eta_k^{M_k+1} \le c \to 0$:

$$
\tilde{\sigma}^{(k)}(t) = \left[\{\sum_{m=0}^{M_k} \eta_k^m y^2(t-m-1)\} / \{\sum_{m=0}^{M_k} \eta_k^m\} \right]^{1/2}
$$

where the Gaussian log-likelihood function given η_k is:

$$
\begin{aligned}
L(\eta_k, \tilde{\sigma}^{(k)}(t)) &= -\frac{N_k}{2} \log\left(2\pi\{\sigma^{(k)}(t)\}^2\right) - \frac{1}{2\{\sigma^{(k)}(t)\}^2} \sum_{m=0}^{M_k} \eta_k^m y^2(t-m-1) \\
\text{where } N_k &= \sum_{m=0}^{M_k} \eta_k^m
\end{aligned}
\tag{5}
$$

The fitted log-likelihood ratio $L\left(\eta_k, \tilde{\sigma}^{(k)}(t), \sigma(t)\right)$ reads as:

$$L\left(\eta_k, \tilde{\sigma}^{(k)}(t), \sigma(t)\right) = L\left(\eta_k, \tilde{\sigma}^{(k)}(t)\right) - L(\eta_k, \sigma(t))$$

The idea of local exponential smoothing is to aggregate all the local likelihood estimate to achieve the best possible accuracy of estimation. In this sense, the local MLEs $\tilde{\sigma}^{(k)}(t)$ are referred as "weak" estimates.

In our study, we concern the heavy-tailedness of financial time series and assume the normal inverse Gaussian (NIG) distribution, one subclass of the GH distribution, see Section 2.3 for more details. Since the NIG distributional parameters of the innovations are unknown at this stage, we use the quasi ML estimation instead of estimating the variance based on the NIG density form. The quasi ML estimation is applicable if the exponential moment of the squared innovations $\mathsf{E}[\exp\{\rho\varepsilon^2(t)\}]$ exists. A power transformation guarantees that:

$$
\begin{aligned}
y_p(t) &= \text{sign}\{y(t)\}|y(t)|^p \\
\theta(t) &= Var\{y_p(t)|\mathcal{F}_{t-1}\} = \mathsf{E}\{y_p^2(t)|\mathcal{F}_{t-1}\} = \mathsf{E}\{|y(t)|^{2p}|\mathcal{F}_{t-1}\} \\
&= \sigma^{2p}(t)\,\mathsf{E}\,|\varepsilon(t)|^{2p} = \sigma^{2p}(t)C_p
\end{aligned}
\tag{6}
$$

where $C_p = \mathsf{E}(|\varepsilon(t)|^{2p}|\mathcal{F}_{t-1})$ is a constant and only relies on $0 \leq p < 1/2$. Notice that the power transformed variable $\theta(t)$ is one-to-one correspondence to the variance $\sigma^2(t)$ and can be estimated on the base of the transformed observations $|y(t)|^{2p}$:

$$\tilde{\theta}^{(k)}(t) = \{\sum_{m=0}^{M_k} \eta_k^m |y(t-m-1)|^{2p}\}/N_k$$

Here the smoothing parameter η_k is designed to run over a wide range from values close to zero to one, so that the variability of the unknown process $\theta(t)$ reduces and at least one of the resulting MLEs is good in the sense of small estimation bias. Polzehl and Spokoiny (2006) show that the inverse of N_k in (5) is positively related to the variation of the MLEs. This result is used to construct the sequence of the smoothing parameter $\{\eta_k\}$:

$$\frac{N_{k+1}}{N_k} \approx \frac{1-\eta_k}{1-\eta_{k+1}} = a > 1, \tag{7}$$

where the coefficient a controls the decreasing speed of the variations.

The procedure is sequential and starts with the estimate $\tilde{\theta}^{(1)}(t)$ that has the largest variability but small bias, i.e. we set $\hat{\theta}^{(1)}(t) = \tilde{\theta}^{(1)}(t)$. At every step $k \geq 2$, the new estimate $\hat{\theta}^{(k)}(t)$ is constructed by aggregating the next "weak" estimate $\tilde{\theta}^{(k)}(t)$ and the previously constructed estimate $\hat{\theta}^{(k-1)}(t)$. Following to Belomestny and Spokoiny (2006), the aggregation is done in terms of the parameter $v = -1/(2\theta)$ so that the variable $y(t)$

belongs to the exponential distributional family with a density form: $p(y, v) = p(y) \exp\{yv - d(v)\}$:

$$\hat{v}^{(k)}(t) = \gamma_k \tilde{v}^{(k)}(t) + (1 - \gamma_k)\hat{v}^{(k-1)}(t)$$

$$\text{or equivalently, } \hat{\theta}^{(k)}(t) = \left(\frac{\gamma_k}{\tilde{\theta}^{(k)}(t)} + \frac{1 - \gamma_k}{\hat{\theta}^{(k-1)}(t)}\right)^{-1}$$

The mixing weights $\{\gamma_k\}$ are computed on the base of the fitted log-likelihood ratio by checking that the previously accepted estimate $\hat{\theta}^{(k-1)}(t)$ is in agreement with the next "weak" estimate $\tilde{\theta}^{(k)}(t)$, i.e. the difference between these two estimates is bounded by critical values $\mathfrak{z}_k$:

$$\gamma_k = K_{ag}\left\{L\left(\eta_k, \tilde{\theta}^{(k)}(t), \hat{\theta}^{(k-1)}(t)\right)/\mathfrak{z}_k\right\}$$

The aggregation kernel K_{ag} guarantees that the mixing coefficient γ_k is one if there is no essential difference between $\tilde{\theta}^{(k)}(t)$ and $\hat{\theta}^{(k-1)}(t)$, and zero if the difference is significant. The significance level is measured by the critical value ζ_k. In the intermediate case, the mixing coefficient γ_k is between zero and one. The procedure terminates after step k if $\gamma_k = 0$ and we define in this case $\hat{\theta}^{(m)}(t) = \hat{\theta}^{(k-1)}(t)$ for all $m \geq k$.

The critical values $\{\zeta_k\}$ are calculated by using Monte Carlo simulation. We briefly summarize the procedure here. Since the NIG distributional parameters of the innovations are unknown and the transformed variable is close to Gaussian variable, we start from the Gaussian assumption. To be more specific, we generate $y(t) = \sigma^* \varepsilon(t)$ with $\varepsilon(t) \sim N(0, 1)$ and $\sigma^* \stackrel{\text{def}}{=} 1$. The "weak" estimates are calculated given the sequence of $\{\eta_k\}$. For $k = 2, \ldots, K$ with $\zeta_1, \infty, \cdots, \infty$, the value ζ_1 is selected as the minimal one to fulfill

$$\mathsf{E}_{\theta^*}|L\left(\eta_k, \tilde{\theta}^{(k)}(t), \hat{\theta}_{\zeta_1}^{(k)}(t)\right)|^r \leq \frac{\alpha \tau_r}{K - 1}, \tag{8}$$

where $\tau_r = 2r \int_{\zeta \geq 0} \zeta^{r-1} e^{-\zeta} d\zeta = 2r\Gamma(r)$, and $r = 0.5$ and $\alpha = 1$ have been suggested in Chen and Spokoiny (2006). Consequently for $l = k + 1, \ldots, K$ with the parameters $\zeta_1, \ldots, \zeta_k, \infty, \ldots, \infty$, we select ζ_k as the minimal value which fulfills

$$\mathsf{E}_{\theta^*}|L\left(\eta_l, \tilde{\theta}^{(l)}(t), \hat{\theta}_{\zeta_1, \ldots, \zeta_k}^{(l)}(t)\right)|^r \leq \frac{k\alpha \tau_r}{K - 1}. \tag{9}$$

As said before, the transformed variable is close to Gaussian variable, we use the generated critical values under the Gaussian assumption to estimate the volatility. The constant C_p is calculated based on the estimates $\hat{\theta}(t)$ such that the innovation is standardized, i.e. $Var\{\hat{\varepsilon}(t)\} = Var\left[y(t)\{\hat{C}_p/\hat{\theta}(t)\}^{\frac{1}{2p}}\right] = 1$. One then estimates the NIG distributional parameters of $\hat{\varepsilon}(t) = y(t)/\hat{\sigma}(t)$ where $\hat{\sigma}(t) = \{\hat{\theta}(t)/\hat{C}_p\}^{\frac{1}{2p}}$. To get more accurate results, one

generates NIG innovations with the estimated distributional parameters and recalculates the critical values as in the Gaussian case.

The local exponential smoothing algorithm is described as follows:

1. Initialization: $\hat{\theta}^{(1)}(t) = \tilde{\theta}^{(1)}(t)$.

2. Loop: for $k \geq 2$,
$$\hat{\theta}^{(k)}(t) = \left(\frac{\gamma_k}{\tilde{\theta}^{(k)}(t)} + \frac{1 - \gamma_k}{\hat{\theta}^{(k-1)}(t)}\right)^{-1}$$

   where the aggregating parameter γ_k is computed as:
$$\gamma_k = K_{\text{ag}}(L(\eta_k, \tilde{\theta}^{(k)}(t), \hat{\theta}^{(k-1)}(t))/\zeta_{k-1}) \tag{10}$$

   If $\gamma_k = 0$ then terminate by letting $\hat{\theta}^{(k)}(t) = \ldots = \hat{\theta}^{(K)}(t) = \hat{\theta}^{(k-1)}(t)$.

3. Aggregation estimate: $\hat{\theta}(t) = \hat{\theta}^{(K)}(t)$.

4. Final estimate: $\hat{\sigma}(t) = \{\hat{\theta}(t)/C_p\}^{\frac{1}{2p}}$, where the constant C_p is computed such that the residuals $\hat{\varepsilon}(t) = y(t)/\hat{\sigma}(t)$ have a unit variance as assumed in the heteroscedastic model.

Consequently, the covariance matrices $D_y(t)$ and $\Sigma_x(t)$ are calculated.

Dynamic conditional correlation (DCC) model: Alternatively, the covariance of the return series can be estimated by the DCC model:

$$\Sigma_x(t) = D_x(t)R_x(t)D_x(t)^\top.$$

This technique first identifies the elements of the diagonal matrix $D_x(t)$ in the GARCH(1,1) setup and adaptively specifies the correlation matrix as:

$$R_x(t) = \tilde{R}_x(1 - \theta_1 - \theta_2) + \theta_1\{\varepsilon_x(t-1)\varepsilon_x(t-1)^\top\} + \theta_2 R_x(t-1),$$

where $\tilde{R}_x$ is the sample correlation of the risk factors, $\varepsilon_x \in \mathbb{R}^d$ are the standardized returns, i.e. risk factors divided by the univariate GARCH(1,1) volatilities, or equivalently by the squared diagonal elements in $D_x(t)$. The standardized returns are assumed to be Gaussian distributed. The parameters θ_1 and θ_2 are identified by the ML estimation.

2.3 Normal inverse Gaussian (NIG) distribution and fast Fourier transformation (FFT)

The estimated ICs are assumed to be NIG distributed. The NIG is a subclass of the GH distribution with a fixed value of $\lambda = -1/2$, see Eberlein and Prause (2002). With 4

distributional parameters, the NIG distribution is flexible to well match the behavior of real data. Compared to many other subclasses of GH distribution, the NIG distribution has a desirable property, saying that the scaled NIG variable belongs to the NIG distribution as well. The density of NIG random variable has a form of:

$$f_{\text{NIG}}(y; \alpha, \beta, \delta, \mu) = \frac{\alpha \delta}{\pi} \frac{K_1 \left\{ \alpha \sqrt{\delta^2 + (y - \mu)^2} \right\}}{\sqrt{\delta^2 + (y - \mu)^2}} \exp\{\delta \sqrt{\alpha^2 - \beta^2} + \beta(y - \mu)\},$$

where the distributional parameters fulfill $\mu \in \mathbb{R}$, $\delta > 0$ and $|\beta| \leq \alpha$. The modified Bessel function of the third kind $K_\lambda(\cdot)$ with an index $\lambda = 1$ has a form of:

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

The characteristic function of the NIG variable is:

$$\varphi_y(z) = \exp\left[\mathbf{i}z\mu + \delta\{\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + \mathbf{i}z)^2}\} \right]$$

Proof: The characteristic function of the GH random variable has a form of:

$$\varphi_y(z) = \exp(\mathbf{i}z\mu) \left\{ \frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + \mathbf{i}z)^2} \right\}^{\lambda/2} \frac{K_\lambda\{\delta \sqrt{\alpha^2 - (\beta + \mathbf{i}z)^2}\}}{K_\lambda(\delta \sqrt{\alpha^2 - \beta^2})}$$

Using the representation of the modified Bessel function with a fixed index $\lambda = -1/2$ derived in Barndorff-Nielsen and Blæsild (1981):

$$K_\lambda(y) = \sqrt{\frac{2}{\pi}} y^{-1/2} e^{-y},$$

it is straightforwardly to show that the assertion holds. □

One desirable feature of the NIG distribution is its explicit scaling transformation. Multiplying the random variable by c, the resulting variable $y' = cy$ belongs to the NIG distribution as well:

$$f_{\text{NIG}}(y'; \alpha', \beta', \delta', \mu') = f_{\text{NIG}}(cy; \alpha/|c|, \beta/c, |c|\delta, c\mu). \tag{11}$$

Proof: It is easy to show the result by using the Jacobian transformation, see Härdle and Simar (2003). Given the density of y and let $\alpha' = \alpha/|c|$, $\beta' = \beta/c$, $\delta' = |c|\delta$ and $\mu' = c\mu$, the density of $y' = cy$ has a form of:

$$
\begin{aligned}
f(y') &= \frac{1}{|c|} f_y(\frac{y}{c}) = \frac{\alpha' \delta'}{\pi} \frac{K_1 \left\{ \alpha' \sqrt{\delta'^2 + (y' - \mu')^2} \right\}}{\sqrt{\delta'^2 + (y' - \mu')^2}} \exp\{\delta' \sqrt{\alpha'^2 - \beta'^2} + \beta'(y' - \mu')\} \\
&= f_{\text{NIG}}(y'; \alpha', \beta', \delta', \mu').
\end{aligned}
$$

14

$\square$

To calculate risk measures, it requires the identification of the portfolio returns' density. Based on the GHICA model, the portfolio returns are calculated as:

$$r(t) = b(t)^\top W^{-1} D_y(t)^{1/2} \varepsilon_y(t)$$

where $b(t)$ is the trading strategy. Notice that the linear transformation of the NIG variable is not necessarily NIG distributed. In other words, the density of the return is unknown although the marginal densities are clear. On the meanwhile its characteristic function is explicitly writable. This is the same case as approximating the α-stable distribution in Menn and Rachev (2004), by which the Fourier transformation is used to approximate the density of the variable based on its characteristic function. This motivates us to use the technique to approximate the density of the return in the GHICA procedure.

Set $a = (a_1, \cdots, a_d) = b(t)^\top W^{-1} D_y(t)^{1/2}$, the variable $\zeta_j = a_j \varepsilon_j$ is NIG distributed with $j = 1, \cdots, d$, according to (11):

$$\zeta_j \sim \mathrm{NIG}(\zeta_j, \breve{\alpha}_j, \breve{\beta}_j, \breve{\delta}_j, \breve{\mu}_j) = \mathrm{NIG}(\zeta_j, \alpha_j/|a_j|, \beta_j/a_j, |a_j|\delta_j, a_j\mu_j).$$

The characteristic function of the return $r = \sum_{j=1}^{d} \zeta_j$ at time t is:

$$\varphi_r(z) = \prod_{j=1}^{d} \varphi_{\zeta_j}(z) = \exp\left[\mathbf{i}z \sum_{j=1}^{d} \breve{\mu}_j + \sum_{j=1}^{d} \breve{\delta}_j\{\sqrt{\breve{\alpha}_j^2 - \breve{\beta}_j^2} - \sqrt{\breve{\alpha}_j^2 - (\breve{\beta}_j + \mathbf{i}z)^2}\}\right]$$

The density function is approximated by the Fourier transformation:

$$f(r) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-\mathbf{i}tr)\psi(z)dt \approx \frac{1}{2\pi} \int_{-s}^{s} \exp(-\mathbf{i}tr)\psi(z)dt$$

The procedure of quantile estimation is summarized as follows:

- Implement the discrete fast Fourier transformation (DFT) to approximate the density of r at every time point t:

  1. Let $N = 2^m$ with $m \in \mathbb{N}$ and define an equidistance grid over the integral interval $[-s, s]$ by setting $h = \frac{2s}{N}$ and the grid points $z_j = -s + j * h$ with $j = 0, \cdots, N$.
  2. Calculate the input of the DFT: $y_j = (-1)^j \psi(z_j^*)$ with $z_j^* = 0.5(z_j + z_{j+1})$ are the middle points. Notice that the characteristic function is time dependent.
  3. The density $f(r) = \frac{1}{2\pi} C_k \mathrm{DFT}(y)_k$ with $C_k = \frac{2s}{N}(-1)^k \exp(-\frac{\mathbf{i}k\pi}{N})\mathbf{i}$ with $k = 0, \cdots, N-1$. We refer to Borak, Detlefsen and Härdle (2005) and Menn and Rachev (2004) for more details. The corresponding values of $r = -\frac{N\pi}{2a} + \frac{\pi k}{a}$.

- The cumulative density function and the quantile are then approximated based on
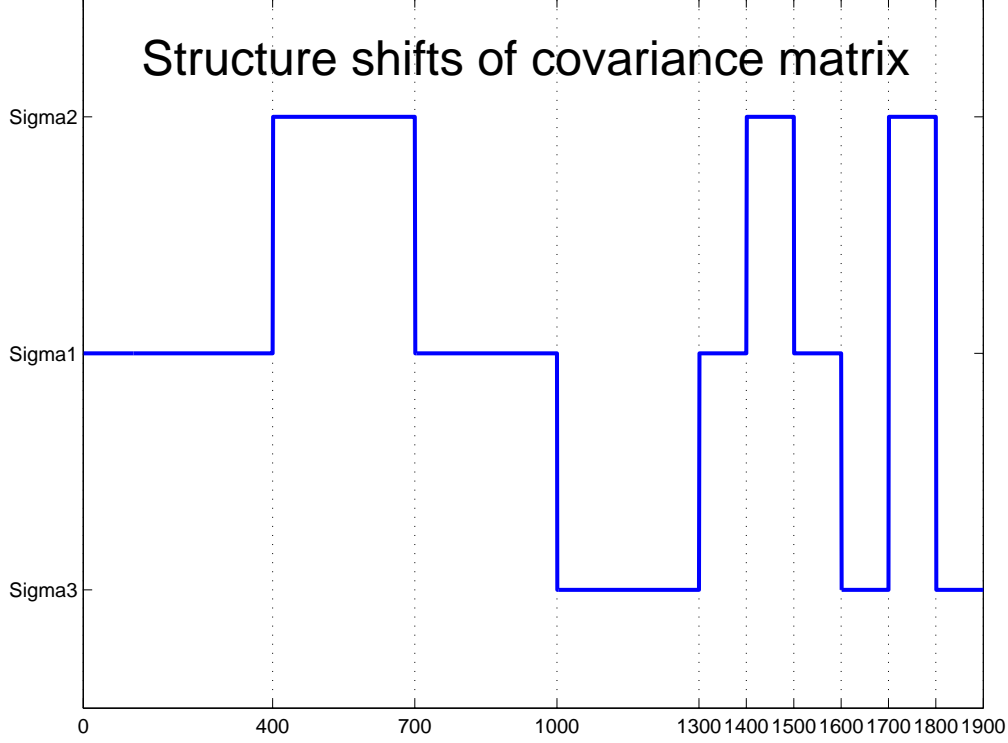
Fig. 2: Structure shifts of the generated covariance through time. Notice that there are shifts among matrices not up-and-down movements.

the resulting density.

3 Covariance estimation with simulated data

In this section, the GHICA versus the DCC, are implemented to estimate covariance of simulated data. The dimension is set to be $d = 50$. The simulation study is designed to include structure shifts of covariance. To be more specific, the designed covariance changes among three matrices over time, one is an identity matrix denoted as Σ_1, meaning uncorrelatedness, and two symmetric and semi-positive defined matrices Σ_2 and Σ_3. (Here we first generate $d * d$ matrix U_1 whose elements are uniform random variables for Σ_2 and standard Gaussian variables for Σ_3, then calculate a new matrix $U_2 = U_1 * U_1'$ to guarantee the semi-positiveness. The elements $\Sigma(i,j)$ of the target matrix are calculated as $\Sigma(i,j) = U_2(i,j)/\sqrt{U_2(i,i)U_2(j,j)}$.) The eigenvalues of these two matrices are distributed in $[5.92e-004, 3.779]$ (Σ_2) and $[0.002, 3.573]$ (Σ_3) respectively. The off-diagonal values span over $[-0.433, 0.468]$ in the first self-correlated matrix (Σ_2) and $[-0.447, 0.464]$ in the second one (Σ_3). Temporal stationarity is assumed to be long for 400 time units and short for 100 units. The structure shifts of the generated covariance are illustrated in Figure 2. The level of the shifts is either small with a shift from one self-correlated matrix (Σ_2 or Σ_3) to the identity matrix or contrariwise, e.g. at the point 700, or large with a shift between the two

16

self-correlated matrices, e.g. at the point 1800.

Furthermore, two distributional parameters μ and β of the standardized NIG innovations $\varepsilon_x(t)$ are set to be 0, meaning that the innovations are centered around 0 and symmetric distributed, see Barndorff-Nielsen and Blæsild (1981). By doing so, the mean and variance of the NIG innovations only depend on α and δ:

$$
\begin{aligned}
\mathsf{E}(\varepsilon_x) &= \mu + \frac{\beta\delta}{\sqrt{\alpha^2 - \beta^2}} = 0 \\
\mathrm{Var}(\varepsilon_x) &= \frac{\delta}{\sqrt{\alpha^2 - \beta^2}} + \frac{\beta^2}{\delta^3\sqrt{\alpha^2 - \beta^2}} = \frac{\delta}{\alpha} = 1
\end{aligned}
$$

This result is used to generate the standardized innovations, by which $\alpha \sim U[1, 2]$ is suggested by our experience on real data analysis and $\delta = \alpha$.

In the Monte Carlo simulation, we generate $d = 50$ NIG variables with the designed covariance and distributional parameters:

$$
x(t) = \Sigma_x^{1/2}(t)\varepsilon_x(t).
$$

The sample size is $T = 1900$ and the scenarios are repeated $N = 100$ times. The covariance matrix is estimated using the GHICA procedure and the DCC method respectively.

The GHICA method first converts the underlying series to ICs by a linear transformation:

$$
x(t) = W^{-1}y(t) = W^{-1}D_y^{1/2}(t)\varepsilon_y(t),
$$

by which the elements of $D_y(t)$ on the diagonal are estimated using the local exponential smoothing method. In the local exponential smoothing estimation, we set the involved parameters $c = 0.01$, $a = 1.25$ and $p = 0.25$. The sequence of the smoothing parameters $\{\eta_k\}$ are $0.600, \cdots, 0.982$ with $K = 15$, based on the condition $(1-\eta_k)/(1-\eta_{k+1}) = a$ in (7). The first 300 observations are reserved as training set for the very beginning estimations, since the largest smoothing parameter used in this study corresponds to a window with 259 observations.

The covariance of $x(t)$ is calculated by the basic statistical property:

$$
\Sigma_x(t) = W^{-1}D_y(t)W^{-1\top}
$$

The DCC method assumes that the underlying series are Gaussian distributed. It decomposes the covariance matrix to a product of diagonal variance matrix and correlation matrix:
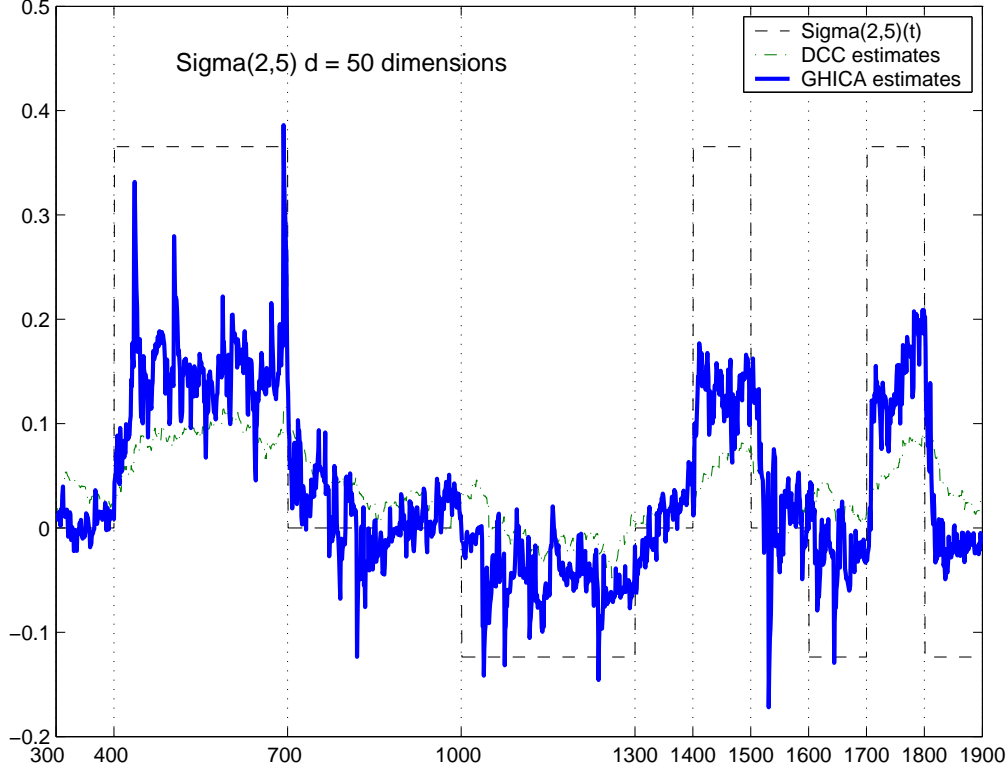
$$
\Sigma_x(t) = D_x(t)R_x(t)D_x(t)^\top.
$$

Fig. 3: Realized estimates of $\Sigma(2,5)$ based on the GHICA and DCC methods. The generated data consists of 50 NIG distributed components.

where $D_x(t)$ consists of the variances of $x(t)$ on the diagonal that are estimated in the GARCH(1,1) setup.

Figure 3 displays one realization of $\Sigma(2,5)$, i.e. the covariance of the second and fifth risk factors $x_2(t)$ and $x_5(t)$, based on one simulation data. The true values are 0.365 in Σ_2 and -0.124 in Σ_3. As expected, the GHICA estimates are sensitive to structure shifts through time. The DCC estimates, on the contrary, are over-smooth and slowly follow the shifts. Given more often shifts around the last hundreds of time points, the DCC estimates deliver less information on the movements. Recall that 100 points correspond to 4 months observations of daily returns. It is rational to surmise that structure shifts happen so often in the active financial markets, see Merton (1973). The similar estimation results are observed in the other elements of the covariance, which are eliminated here.

To measure the accuracy of estimation, ratio of absolute estimation error (RAE) of the estimates w.r.t. the true covariance are calculated pointwise.

$$\text{RAE}(i,j) \;=\; \frac{\sum_{t=301}^{T} |\hat{\Sigma}_{(i,j)}^{\text{GHICA}}(t) - \Sigma_{(i,j)}(t)|}{\sum_{t=301}^{T} |\hat{\Sigma}_{(i,j)}^{\text{DCC}}(t) - \Sigma_{(i,j)}(t)|}$$

If $\text{RAE}(i,j) \leq 1$, it means that the GHICA method reaches higher accuracy in the estima-
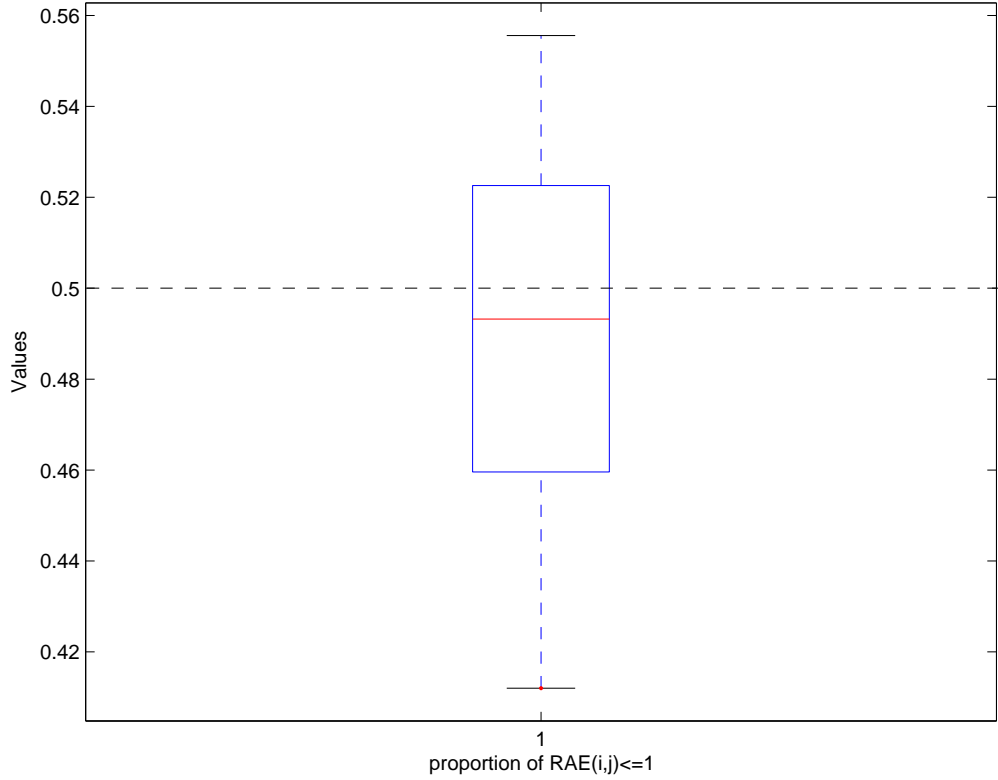
18

Fig. 4: Boxplot of the proportion $\frac{\sum_i \sum_j \mathbf{1}(\text{RAE}(i,j) \leq 1)}{d \times d}$ for $i, j = 1, \cdots, d$. Here $d = 50$ and the proportions on the base of 100 simulations are considered.

tion of $\Sigma(i,j)$ than the DCC. To compare the general performance of these two methods in covariance estimation, we check the proportion of the RAEs among the 2500 ($d*d$) elements that are smaller or equal to one, i.e. $\frac{\sum_i \sum_j \mathbf{1}(\text{RAE}(i,j) \leq 1)}{d \times d}$ for $i, j = 1, \cdots, d$. Notice that the proportion with value of 0.5 indicates that half elements are better estimated by using the GHICA and the other half are better done by the DCC. In other words, the considered methods have a comparable accuracy of estimation. Figure 4 displays the boxplot of the 100 proportions. The mean of the proportion is 0.4904 among the 100 simulations. It states that the DCC method performs a little bit better than the GHICA in the sense of accuracy. On the meanwhile, the GHICA method is much fast and sensitive to structure shifts.

4   Risk management with real data

In this section, we implement the proposed GHICA method to calculate risk measures using real data sets: 20-dimensional German DAX portfolio and 7-dimensional exchange rate portfolio. The results are compared with those based on alternative risk management models. The data sets have been kindly provided by the financial and economic data center (FEDC) of the Collaborative Research Center 649 on Economic Risk of the Humboldt-

19

Universität zu Berlin (http://sfb649.wiwi.hu-berlin.de). Before giving detailed description of the data sets, we analyze the risk measures from the viewpoints of regulatory, investors and internal supervisory.

Regulatory requirement: Financial institutions generally face market risk that arises from the uncertainty due to changes in market prices and rates such as share prices, foreign exchange rates and interest rates, the correlations among them and their levels of volatility, see Jorion (2001). The market risk is the main risk source and has a great negative influence on the development of economic. The famous example is the stock crashes in the autumn 1929 and 1987 which caused a violent depression in the United States and some other countries, with the collapse of financial markets and the contraction of production and employment. To alleviate the down influence of market risks, regulation on banking and other financial institutions has been strengthened since the mid-1990s. The goals of the regulation are to restrict the happening of extremely large losses and require banks to reserve adequate capital. In 1998 the Basel accord officially allowed financial institutions to use their internal models to measure market risks. Among others, Value at Risk (VaR) has been considered as industry standard risk measure:

$$\text{VaR}_{t,\text{pr}} = -\text{quantile}_{\text{pr}}\{r(t)\}.$$

where pr is the $h = 1$-day or $h = 5$-day forecasted probability of the portfolio returns. Internal models for risk management are verified in accordance with the "traffic light" rule that counts the number of exceptions over VaR at 1% probability spanning the last 250 days and identifies the multiplicative factor M_f in the market risk charge calculation, see Franke, Härdle and Hafner (2004):

$$\text{Risk charge}_t = \max\left(M_f \frac{1}{60} \sum_{i=1}^{60} \text{VaR}_{t-i,1\%}, \text{VaR}_{t,1\%}\right)$$

The multiplicative factor M_f has a floor value 3. It increases corresponding to the number of exceptions, see Table 1. For example, if an internal model generates 7 exceptions at 1% probability over the last 250 days, the model is in the yellow zone and its multiplicative factor is $M_f = 3.65$. Financial institutions whose internal model is located in the yellow or red zone, with a very high probability, are required to reserve more risk capital than their internal-model-based VaRs. Notice that the increase of risk charge will reduce the ratio of profit since the reserved capital can not be invested. On the meanwhile, an internal model is automatically accepted if the number of exceptions does not exceed 4. This regulatory rule in fact suggests banks to control VaR at 1.6% (i.e. 4/250) instead of 1% probability. It is clear that 1.6%-VaR is smaller than 1%-VaR. Therefore an internal model is particularly desirable by financial institutions if its empirical probability is smaller or equal to 1.6%, and simultaneously requires risk charge as small as possible. Here a simplified calculation

No. exceptions	Increase of M_f	Zone
0 bis 4	0	**green**
5	0.4	yellow
6	0.5	yellow
7	0.65	yellow
8	0.75	yellow
9	0.85	yellow
More than 9	1	**red**

Tab. 1: Traffic light as a factor of the exceeding amount, cited from Franke, Härdle and Hafner (2004).

on the average value of VaRs is used as risk charge for comparison:

$$\text{Risk charge (RC)} = \text{mean}\left(\text{VaR}_{t,\text{pr}}\right)$$

Investor: It is known that VaR is inappropriate for the measurement of capital adequacy, since it controls only the probability of default, i.e. the frequency of losses, but not the size of losses in the case of default. For this reason, investors concern expected shortfall (ES) more than VaR to measure and control their risks.

$$\text{ES} = \mathsf{E}\{-r(t)| - r(t) > \text{VaR}_{t,\text{pr}}\}$$

Investors suffer loss once bankruptcy happens. Even in the "best" situation, their loss equals to the difference between the total loss and the reserved risk capital, i.e. the value of ES. Generally risk-averse investors care the amount of loss and thus prefer an internal model with small value of ES. Risk-seeking investors, on the other hand, care profit and hence the small value of risk charge favors their requirement.

Internal supervisory: It is important for internal supervisory to exactly measure the market risk exposures before risk controlling. For this reason, internal supervisory prefers the model delivering accurate probability prediction, i.e. the empirical probability p̂r is as close to the expected values as possible:

$$\hat{\text{pr}} = \frac{\text{No. exceptions}}{\text{No. total observations}}$$

Given two models with the same empirical probability, the model has a smaller value of ES is considered better than the other. Here two extreme probabilities are considered, i.e. pr = 1% for regulatory reason and pr = 0.5% used by financial institutions with AAA rating.

4.1 Data analysis 1: DAX portfolio

The primary target of the real data analysis is to compare the forecasting ability of the GHICA method with two alternatives, the RiskMetrics method under the Gaussian distributional assumption and a modification with the Student-$t(6)$ distributional assumption (abbreviated as $t(6)$ method) in the market. The comparison is demonstrated based on 20 DAX stocks over a long time period, starting on 1974/01/02 and ending on 1996/12/30 (5748 observations). The return series are all centered around 0 and have heavy tails (kurtosis> 3), the smallest correlation coefficient is 0.3654. Hypothetical German DAX portfolios are constructed with two static trading strategies $b(t) = b^{(1)} = (1/d, \cdots, 1/d)^{\top}$ and $b(t) = b^{(2)} \sim U[0, 1]^d$. Such a simple portfolio construction eliminates the influence of strategy adjustments on the calculation. The portfolio returns are analyzed using the RiskMetrics or the $t(6)$ method. Here the unknown volatility process of the portfolio is estimated using the exponential smoothing method with $\eta = 0.94$:

$$
\begin{aligned}
r(t) &= b^{\top}x(t) = \sigma_r(t)\varepsilon_r(t) \\
\sigma_r^2(t) &= \{\sum_{m=0}^{M} \eta^m r^2(t - m - 1)\}/(\sum_{m=0}^{M} \eta^m)
\end{aligned}
$$

where the truncated value M fulfills the condition $\eta^{(M+1)} \leq 0.01$. Notice that given a dynamic trading strategy, this simplification needs to repeatedly estimate the density of the time varying hypothetical portfolio returns, and it often suffers from a low accuracy of estimation.

Figure 5 depicts the one day log-returns of the DAX portfolio with the static trading strategy $b(t) = b^{(1)}$. The VaRs from 1975/03/17 to 1996/12/30 at pr $= 0.5\%$ are displayed w.r.t. three methods, the GHICA, the RiskMetrics and the $t(6)$. The most volatile time period over $t \in [3300, 4300]$ is detailed in the bottom diagram. Recall that on the Monday, 19 October 1987, the worldwide downward jump of stocks happened. Dow Jones Industrial Average for example dropped by over 500 points. At this market quiver around $t = 3446$, the GHICA method exactly achieves the locations of extreme losses whereas the RiskMetrics and $t(6)$ methods over-react to them. Such over reactions induce large risk charges unnecessarily. On the other hand, it is observed that these two alternative methods give close forecasts to some extreme losses, e.g. around time points 4000 and 4500. As a result, the associating values of ES are small and satisfy the requirement of risk-averse investors.

Table 2 reports the risk measures based on the three methods. In general, the Risk-Metrics is successful in fulfilling the minimal requirement of regulatory. The $t(6)$ method is preferred by investors who consider risk happened with 1% probability. The GHICA method performs better than the other two for internal supervisory and requirement of
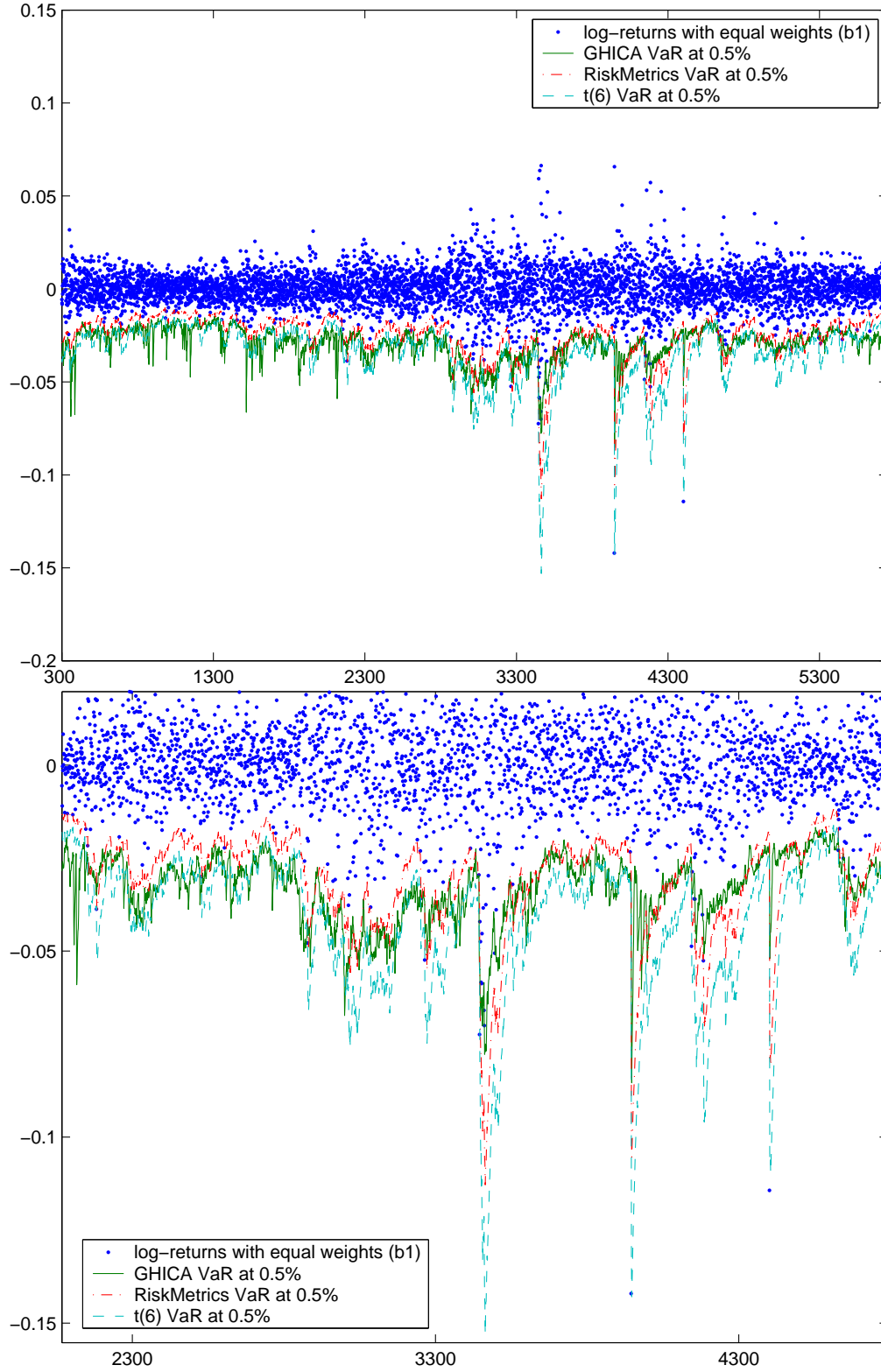
Fig. 5: One day log-returns of the DAX portfolio with the static trading strategy $b(t) = b^{(1)}$. The VaRs are from 1975/03/17 to 1996/12/30 at $pr = 0.5\%$ w.r.t. three methods, the GHICA, the RiskMetrics and the $t(6)$. Part of the VaR time plot is enlarged and displayed on the bottom.

23

h	$b(t)$	pr	GHICA			RiskMetrics $N(\mu, \sigma^2)$			Exponential smoothing $t(6)$		
			$\hat{pr}$	RC	ES	$\hat{pr}$	RC	ES	$\hat{pr}$	RC	ES
1	$b^{(1)}$	1%	0.55%	0.0264	0.0456	1.18%^{s}	**0.0229**^{r}	0.0279	0.40%	0.0292	**0.0269**^{i}
	$b^{(1)}$	0.5%	0.44%^{s}	0.0297	**0.0472**^{i}	0.75%	0.0254	0.0317	0.23%	0.0345	0.0506
	$b^{(2)}$	1%	0.59%	0.0265	0.0448	1.03%^{s}	**0.0231**^{r}	0.0288	0.38%	0.0294	**0.0406**^{i}
	$b^{(2)}$	0.5%	0.42%^{s}	0.0298	**0.0476**^{i}	0.71%	0.0256	0.0315	0.21%	0.0347	0.0514
5	$b^{(1)}$	1%	0.83%	0.0550	0.0841	1.15%^{s}	**0.0481**^{r}	0.0602	0.19%	0.0665	**0.0833**^{i}
	$b^{(1)}$	0.5%	0.51%^{s}	0.0612	**0.0939**^{i}	0.64%	0.0536	0.0683	0.09%	0.0784	0.1067
	$b^{(2)}$	1%	0.83%^{s}	0.0554	**0.0828**^{i}	1.18%	**0.0488**^{r}	0.0613	0.16%	0.0673	0.0852
	$b^{(2)}$	0.5%	0.50%^{s}	0.0617	**0.0943**^{i}	0.63%	0.0543	0.0676	0.07%	0.0794	0.1218

Tab. 2: Risk analysis of the DAX portfolios with two static trading strategies. The concerned forecasting interval is $h = 1$ or $h = 5$ days. The best results to fulfill the regulatory requirement are marked by ^{r}. The method preferred by investor is marked by ^{i}. For the internal supervisory, the method marked by ^{s} is recommended.

risk-averse investors who care the extreme risk happened with 0.5% probability.

4.2 Data analysis 2: Foreign exchange rate portfolio

In financial markets, traders adjust trading strategy according to information obtained. The GHICA is easily applicable to dynamic portfolios. We consider here 7 actively traded exchange rates, Euro (EUR), the US dollar (USD), the British pounds (GBP), the Japanese yen (JPY) and the Singapore dollar (SGD) from 1997/01/02 to 2006/01/05 (2332 observations). The foreign exchange rate (FX) market is the most active and liquid financial market in the world. It is realistic to analyze a dynamic portfolio with daily time varying trading strategy $b^{(3)}(t)$. The strategy at time point t relies on the realized returns at $t-1$, the proportions of which w.r.t the sum of returns:

$$b^{(3)}(t) = \frac{x(t-1)}{\sum_{j=1}^{d} x_j(t-1)}$$

where $x(t) = \{x_1(t), \cdots, x_d(t)\}^\top$. Among these data sets, the returns of the EUR/SGD and USD/JPY rates are least correlated with the correlation coefficient 0.0071 whereas the returns of the EUR/USD and EUR/SGD rates are most correlated with the coefficient 0.6745. The resulting portfolio returns span over $[-0.7962, 0.7074]$.

The GHICA method is compared with an alternative method, abbreviated as DCCN, that applies the DCC covariance estimation under the Gaussian distributional assumption.

$$r(t) = b(t)^\top x(t) = b(t)^\top \Sigma_x^{(1/2)}(t)\varepsilon_x(t)$$

where $\varepsilon_x \sim N(\mu, \Sigma_\varepsilon)$ with the diagonal covariance matrix Σ_ε. Notice that the quantile

			GHICA			DCCN		
h	$b(t)$	pr	$\hat{pr}$	RC	ES	$\hat{pr}$	RC	ES
1	$b^{(3)}(t)$	1%	$1.28\%^s$	$\mathbf{0.0453}^r$	0.0778	1.59%	0.0494	$\mathbf{0.0254}^i$
	$b^{(3)}(t)$	0.5%	$0.59\%^s$	0.0493	$\mathbf{0.1944}^i$	0.94%	0.0547	0.0289
5	$b^{(3)}(t)$	1%	$1.53\%^s$	$\mathbf{0.0806}^r$	$\mathbf{0.2630}^i$	4.17%	0.0993	0.1735
	$b^{(3)}(t)$	0.5%	$0.79\%^s$	0.1092	$\mathbf{0.2801}^i$	3.44%	0.1100	0.1389

Tab. 3: Risk analysis of the dynamic exchange rate portfolio. The best results to fulfill the regulatory requirement are marked by ^{r}. The recommended method to the investor is marked by ^{i}. For the internal supervisory, we recommend the method marked by ^{s}.

vector with pr-quantiles of individual innovations does not necessarily correspond to the pr-quantile of the portfolio return. Under the Gaussian distributional assumption, the standardized DCCN returns are theoretically cross independent and the Gaussian quantiles of the portfolio can be easily calculated. The dynamic mean, variance of the portfolio's returns have values of:

$$\mathsf{E}\{r(t)\} = b(t)^\top \Sigma_x^{(1/2)}(t)\,\mathsf{E}\{\varepsilon_x(t)\}$$
$$Var\{r(t)\} = b(t)^\top \Sigma_x^{(1/2)}(t)\,Var\{\varepsilon_x(t)\}\Sigma_x^{(1/2)\top}(t)b(t)$$

The GHICA method in general presents better results than the DCCN. Except the value of ES at 1% level, the GHICA fulfills the requirements of regulatory, internal supervisory and investors, see Table 3. For $h = 1$ day forecasts, the DCCN gives although a closer VaR value to 1.6%, i.e. the ideal probability for regulatory, its risk charge with a value of 0.0494 is larger than that based on the GHICA, 0.0453. Therefore the GHICA is more favored in fulfilling the minimal regulatory requirement.

The two real data studies show that the GHICA method fulfills the minimal regulatory requirement by controlling the risk inside 1.6% level and requiring small risk charge, in particular satisfies the internal supervisory requirement by precisely measuring risk level as expected and favors the investors' requirement by delivering small size of loss. In summary, the GHICA method is not only a realistic and fast procedure given either static or dynamic portfolios but also produces better results than several alternative risk management methods.

References

Anderson, T., Bollerslev, T., Diebold, F. and Labys, P. (2001). The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* pp. 42–55.

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

Belomestny, D. and Spokoiny, V. (2006). Spatial aggregation of local likelihood estimates with applications to classification, *WIAS Preprint*.

Borak, S., Detlefsen, K. and Härdle, W. (2005). FFT-based option pricing, *in* P. Cizek, W. Härdle and R. Weron (eds), *Statistical Tools for Finance and Insurance*, Springer Verlag.

Chen, Y. and Spokoiny, V. (2006). Local exponential smoothing with applications to volatility estimation and risk management, *working paper*.

Chen, Y., Härdle, W. and Jeong, S. (2005). Nonparametric risk management with generalized hyperbolic distributions, *SFB 649, Discussion paper 2005-001, http://sfb649.wiwi.hu-berlin.de*.

Chen, Y., Härdle, W. and Spokoiny, V. (2006). Portfolio value at risk based on independent components analysis, *Journal of Computational and Applied Mathematics, forthcoming*.

Cover, T. and Thomas, J. (1991). *Elements of information theory*, Wiley.

Eberlein, E. and Prause, K. (2002). The generalized hyperbolic model: financial derivatives and risk measures, *in* H. Geman, D. Madan, S. Pliska and T. Vorst (eds), *Mathematical Finance-Bachelier Congress 2000*, Springer Verlag.

Engle, R. (2002). Dynamic conditional correlation - a simple class of multivariate garch models, *Journal of Business and Economic Statistics, 20(3)* pp. 339–350.

Engle, R. and Kroner, F. (1995). Multivariate simultaneous generalized arch, *Econometric Theory 11* pp. 122–150.

Engle, R. and Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate garch, *NBER Working Paper 8554*.

Flury, B. (1998). *Common Principal Components and Related Multivariate Models*, John Wiley & Sons, Inc.

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

Härdle, W. and Simar, L. (2003). *Applied Multivariate Statistical Analysis*, 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.

Hyvärinen, A. (1998). *New Approximations of Differential Entropy for Independent Component Analysis and Projection Pursuit*, MIT Press, pp. 273–279.

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

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

Menn, C. and Rachev, S. (2004). Calibrated FFT-based density approximations for α-stable distributions.

Merton, R. (1973). Theory of rational option pricing, *The Bell Journal of Economics and Management Science* **4**: 141–183.

Polzehl, J. and Spokoiny, V. (2006). Propagation-separation approach for local likelihood estimation, *Probability Theory and Related Fields* pp. 335–362.

SFB 649 Discussion Paper Series 2006

001   "Calibration Risk for Exotic Options" by Kai Detlefsen and Wolfgang K. Härdle, January 2006.

002   "Calibration Design of Implied Volatility Surfaces" by Kai Detlefsen and Wolfgang K. Härdle, January 2006.

003   "On the Appropriateness of Inappropriate VaR Models" by Wolfgang Härdle, Zdeněk Hlávka and Gerhard Stahl, January 2006.

004   "Regional Labor Markets, Network Externalities and Migration: The Case of German Reunification" by Harald Uhlig, January/February 2006.

005   "British Interest Rate Convergence between the US and Europe: A Recursive Cointegration Analysis" by Enzo Weber, January 2006.

006   "A Combined Approach for Segment-Specific Analysis of Market Basket Data" by Yasemin Boztuğ and Thomas Reutterer, January 2006.

007   "Robust utility maximization in a stochastic factor model" by Daniel Hernández–Hernández and Alexander Schied, January 2006.

008   "Economic Growth of Agglomerations and Geographic Concentration of Industries - Evidence for Germany" by Kurt Geppert, Martin Gornig and Axel Werwatz, January 2006.

009   "Institutions, Bargaining Power and Labor Shares" by Benjamin Bental and Dominique Demougin, January 2006.

010   "Common Functional Principal Components" by Michal Benko, Wolfgang Härdle and Alois Kneip, Jauary 2006.

011   "VAR Modeling for Dynamic Semiparametric Factors of Volatility Strings" by Ralf Brüggemann, Wolfgang Härdle, Julius Mungo and Carsten Trenkler, February 2006.

012   "Bootstrapping Systems Cointegration Tests with a Prior Adjustment for Deterministic Terms" by Carsten Trenkler, February 2006.

013   "Penalties and Optimality in Financial Contracts: Taking Stock" by Michel A. Robe, Eva-Maria Steiger and Pierre-Armand Michel, February 2006.

014   "Core Labour Standards and FDI: Friends or Foes? The Case of Child Labour" by Sebastian Braun, February 2006.

015   "Graphical Data Representation in Bankruptcy Analysis" by Wolfgang Härdle, Rouslan Moro and Dorothea Schäfer, February 2006.

016   "Fiscal Policy Effects in the European Union" by Andreas Thams, February 2006.

017   "Estimation with the Nested Logit Model: Specifications and Software Particularities" by Nadja Silberhorn, Yasemin Boztuğ and Lutz Hildebrandt, March 2006.

018   "The Bologna Process: How student mobility affects multi-cultural skills and educational quality" by Lydia Mechtenberg and Roland Strausz, March 2006.

019   "Cheap Talk in the Classroom" by Lydia Mechtenberg, March 2006.

020   "Time Dependent Relative Risk Aversion" by Enzo Giacomini, Michael Handel and Wolfgang Härdle, March 2006.

021   "Finite Sample Properties of Impulse Response Intervals in SVECMs with Long-Run Identifying Restrictions" by Ralf Brüggemann, March 2006.

022   "Barrier Option Hedging under Constraints: A Viscosity Approach" by Imen Bentahar and Bruno Bouchard, March 2006.
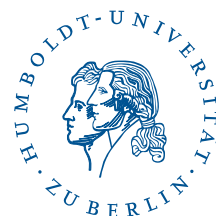
023 "How Far Are We From The Slippery Slope? The Laffer Curve Revisited" by Mathias Trabandt and Harald Uhlig, April 2006.

024 "e-Learning Statistics – A Selective Review" by Wolfgang Härdle, Sigbert Klinke and Uwe Ziegenhagen, April 2006.

025 "Macroeconomic Regime Switches and Speculative Attacks" by Bartosz Maćkowiak, April 2006.

026 "External Shocks, U.S. Monetary Policy and Macroeconomic Fluctuations in Emerging Markets" by Bartosz Maćkowiak, April 2006.

027 "Institutional Competition, Political Process and Holdup" by Bruno Deffains and Dominique Demougin, April 2006.

028 "Technological Choice under Organizational Diseconomies of Scale" by Dominique Demougin and Anja Schöttner, April 2006.

029 "Tail Conditional Expectation for vector-valued Risks" by Imen Bentahar, April 2006.

030 "Approximate Solutions to Dynamic Models – Linear Methods" by Harald Uhlig, April 2006.

031 "Exploratory Graphics of a Financial Dataset" by Antony Unwin, Martin Theus and Wolfgang Härdle, April 2006.

032 "When did the 2001 recession *really* start?" by Jörg Polzehl, Vladimir Spokoiny and Cătălin Stărică, April 2006.

033 "Varying coefficient GARCH versus local constant volatility modeling. Comparison of the predictive power" by Jörg Polzehl and Vladimir Spokoiny, April 2006.

034 "Spectral calibration of exponential Lévy Models [1]" by Denis Belomestny and Markus Reiß, April 2006.

035 "Spectral calibration of exponential Lévy Models [2]" by Denis Belomestny and Markus Reiß, April 2006.

036 "Spatial aggregation of local likelihood estimates with applications to classification" by Denis Belomestny and Vladimir Spokoiny, April 2006.

037 "A jump-diffusion Libor model and its robust calibration" by Denis Belomestny and John Schoenmakers, April 2006.

038 "Adaptive Simulation Algorithms for Pricing American and Bermudan Options by Local Analysis of Financial Market" by Denis Belomestny and Grigori N. Milstein, April 2006.

039 "Macroeconomic Integration in Asia Pacific: Common Stochastic Trends and Business Cycle Coherence" by Enzo Weber, May 2006.

040 "In Search of Non-Gaussian Components of a High-Dimensional Distribution" by Gilles Blanchard, Motoaki Kawanabe, Masashi Sugiyama, Vladimir Spokoiny and Klaus-Robert Müller, May 2006.

041 "Forward and reverse representations for Markov chains" by Grigori N. Milstein, John G. M. Schoenmakers and Vladimir Spokoiny, May 2006.

042 "Discussion of 'The Source of Historical Economic Fluctuations: An Analysis using Long-Run Restrictions' by Neville Francis and Valerie A. Ramey" by Harald Uhlig, May 2006.

043 "An Iteration Procedure for Solving Integral Equations Related to Optimal Stopping Problems" by Denis Belomestny and Pavel V. Gapeev, May 2006.

044 "East Germany's Wage Gap: A non-parametric decomposition based on establishment characteristics" by Bernd Görzig, Martin Gornig and Axel Werwatz, May 2006.

045 "Firm Specific Wage Spread in Germany - Decomposition of regional differences in inter firm wage dispersion" by Bernd Görzig, Martin Gornig and Axel Werwatz, May 2006.
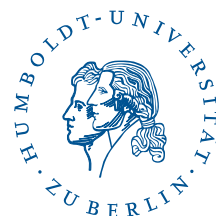
046 "Produktdiversifizierung: Haben die ostdeutschen Unternehmen den Anschluss an den Westen geschafft? – Eine vergleichende Analyse mit Mikrodaten der amtlichen Statistik" by Bernd Görzig, Martin Gornig and Axel Werwatz, May 2006.

047 "The Division of Ownership in New Ventures" by Dominique Demougin and Oliver Fabel, June 2006.

048 "The Anglo-German Industrial Productivity Paradox, 1895-1938: A Restatement and a Possible Resolution" by Albrecht Ritschl, May 2006.

049 "The Influence of Information Costs on the Integration of Financial Markets: Northern Europe, 1350-1560" by Oliver Volckart, May 2006.

050 "Robust Econometrics" by Pavel Čížek and Wolfgang Härdle, June 2006.

051 "Regression methods in pricing American and Bermudan options using consumption processes" by Denis Belomestny, Grigori N. Milstein and Vladimir Spokoiny, July 2006.

052 "Forecasting the Term Structure of Variance Swaps" by Kai Detlefsen and Wolfgang Härdle, July 2006.

053 "Governance: Who Controls Matters" by Bruno Deffains and Dominique Demougin, July 2006.

054 "On the Coexistence of Banks and Markets" by Hans Gersbach and Harald Uhlig, August 2006.

055 "Reassessing Intergenerational Mobility in Germany and the United States: The Impact of Differences in Lifecycle Earnings Patterns" by Thorsten Vogel, September 2006.

056 "The Euro and the Transatlantic Capital Market Leadership: A Recursive Cointegration Analysis" by Enzo Weber, September 2006.

057 "Discounted Optimal Stopping for Maxima in Diffusion Models with Finite Horizon" by Pavel V. Gapeev, September 2006.

058 "Perpetual Barrier Options in Jump-Diffusion Models" by Pavel V. Gapeev, September 2006.

059 "Discounted Optimal Stopping for Maxima of some Jump-Diffusion Processes" by Pavel V. Gapeev, September 2006.

060 "On Maximal Inequalities for some Jump Processes" by Pavel V. Gapeev, September 2006.

061 "A Control Approach to Robust Utility Maximization with Logarithmic Utility and Time-Consistent Penalties" by Daniel Hernández–Hernández and Alexander Schied, September 2006.

062 "On the Difficulty to Design Arabic E-learning System in Statistics" by Taleb Ahmad, Wolfgang Härdle and Julius Mungo, September 2006.

063 "Robust Optimization of Consumption with Random Endowment" by Wiebke Wittmüß, September 2006.

064 "Common and Uncommon Sources of Growth in Asia Pacific" by Enzo Weber, September 2006.

065 "Forecasting Euro-Area Variables with German Pre-EMU Data" by Ralf Brüggemann, Helmut Lütkepohl and Massimiliano Marcellino, September 2006.

066 "Pension Systems and the Allocation of Macroeconomic Risk" by Lans Bovenberg and Harald Uhlig, September 2006.

067 "Testing for the Cointegrating Rank of a VAR Process with Level Shift and Trend Break" by Carsten Trenkler, Pentti Saikkonen and Helmut Lütkepohl, September 2006.

068 "Integral Options in Models with Jumps" by Pavel V. Gapeev, September 2006.

069 "Constrained General Regression in Pseudo-Sobolev Spaces with Application to Option Pricing" by Zdeněk Hlávka and Michal Pešta, September 2006.
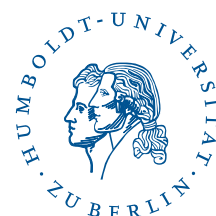
070 "The Welfare Enhancing Effects of a Selfish Government in the Presence of Uninsurable, Idiosyncratic Risk" by R. Anton Braun and Harald Uhlig, September 2006.

071 "Color Harmonization in Car Manufacturing Process" by Anton Andriyashin, Michal Benko, Wolfgang Härdle, Roman Timofeev and Uwe Ziegenhagen, October 2006.

072 "Optimal Interest Rate Stabilization in a Basic Sticky-Price Model" by Matthias Paustian and Christian Stoltenberg, October 2006.

073 "Real Balance Effects, Timing and Equilibrium Determination" by Christian Stoltenberg, October 2006.

074 "Multiple Disorder Problems for Wiener and Compound Poisson Processes With Exponential Jumps" by Pavel V. Gapeev, October 2006.

075 "Inhomogeneous Dependency Modelling with Time Varying Copulae" by Enzo Giacomini, Wolfgang K. Härdle, Ekaterina Ignatieva and Vladimir Spokoiny, November 2006.

076 "Convenience Yields for CO_2 Emission Allowance Futures Contracts" by Szymon Borak, Wolfgang Härdle, Stefan Trück and Rafal Weron, November 2006.

077 "Estimation of Default Probabilities with Support Vector Machines" by Shiyi Chen, Wolfgang Härdle and Rouslan Moro, November 2006.

078 "GHICA - Risk Analysis with GH Distributions and Independent Components" by Ying Chen, Wolfgang Härdle and Vladimir Spokoiny, November 2006.