Abstract
Over recent years, study on risk management has been prompted by the Basel committee for regular banking supervisory. There are however limitations of many risk management methods applied to high dimensional portfolios: These methods either calculate risk measures under the Gaussian distributional assumption or involve high computational demand and numerical difficulty. The primary aim of this paper is to introduce a simple and fast multivariate risk management method, GHICA, by implementing the independent component analysis (ICA) to the high dimensional series and fitting the resulting ICs in the generalized hyperbolic (GH) distributional framework. The volatility process of every IC is estimated using the local exponential smoothing technique to achieve the best possible accuracy of estimation, and the fast Fourier transformation (FFT) technique is used to approximate the density of the portfolio returns. Furthermore, the GHICA method is applicable to covariance estimation. The results are compared with the dynamic conditional correlation (DCC) method based on simulated data with $d = 50$ GH distributed components. The GHICA method is faster and delivers more sensitive estimates than the DCC. In addition, the real data analysis demonstrates the quality of the GHICA method in risk management based on 20-dimensional German DAX portfolios and a dynamic exchange rate portfolio. Several alternative methods are implemented 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

JEL Codes: C14, C16, C32, C61, G20

Acknowledgement: This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk". In the numerical analysis, the Matlab DCC function developed by Kevin Sheppard and the Matlab FastICA function developed by Aapo Hyvärinen are used.
1 Introduction

For regulatory purposes, study on risk management is important for precise banking supervisory. Recent research provides detailed methodologies to calculate various risk measures. Among others, we refer to Jorion (2001) for a systematic description. Given a $d$-dimensional portfolio, the conditionally heteroscedastic model is widely used to describe the movement of the underlying series:

\[
x(t) = \Sigma_x^{1/2}(t)x(t),
\]

where $x(t) \in \mathbb{R}^d$ are risk factors of the portfolio, e.g. (log) returns of the involved components. The covariance $\Sigma_x$ is time dependent and needs to be assessed in a meaningful way. The stochastic innovations $\varepsilon_x(t) \in \mathbb{R}^d$ are assumed to be standardized with $\mathbb{E}[\varepsilon_x(t)|\mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[\varepsilon^2_x(t)|\mathcal{F}_{t-1}] = I_d$. The popular risk measures are calculated based on the estimated joint density of the risk factors. For example, value at risk (VaR) is in fact the distributional quantile at a prescribed level over a target time interval and expected shortfall (ES) measures the size of losses once the realized losses exceed the VaR values. Indicated by formula (1), the joint density mainly relies on the covariance estimation and innovations’ distributional assumption.

There are however some pitfalls and limitations when many risk management methods are applied to high dimensional portfolios. First risk measures are often calculated under the 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 therefore 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). To alleviate the limitation, the Student-$t(6)$ distribution with degrees of freedom of 6 has been recommended. However this distribution often over-fits the heavy tails.

Figure 1 demonstrates the effect of the distributional assumptions for 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
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. Three density estimations under the generalized hyperbolic (GH), Gaussian and $t(6)$ distributional assumptions are depicted in the figure. In the density comparison, the logarithmic density 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 $t(6)$ estimate displays heavier tails relative to the benchmark, and the Gaussian estimate, on the contrary, presents lighter tails. The similar result is observed w.r.t. the DAX portfolio. It is therefore rational to surmise that the risk management methods under the Gaussian and $t(6)$ distributional assumptions generate low accurate results.

The second limitation of the popular multivariate risk management methods is due to high computational demand. Above all, these methods concern the covariance estimation to recover the empirical behavior of high dimensional series. Among many others, the constant conditional correlation (CCC) model proposed by Bollerslev (1990) and the subsequent dynamic conditional correlation (DCC) model proposed by Engle (2002), Engle and Sheppard (2001) are very successful. In the estimation, the covariance matrix is approximated by the product of a diagonal matrix and a correlation matrix: $\Sigma_x(t) = D_x(t)R_x(t)D_x(t)^\top$, which reduces the number of unknown parameters relative to the BEKK specification proposed by Engle and Kroner (1995). In spite of the appealing dimensional reduction, the mentioned estimation methods are still time consuming and numerically difficult to handle as really high dimensional series, e.g. a dimension $d > 10$, is considered, see Härdle, Herwartz and Spokoiny (2003). Moreover, these methods rely on the questionable Gaussian distributional assumption to ensure the independence of the resulting standardized returns. Otherwise, the distributional identification under a realistic assumption, such as the multivariate GH distribution with at least $4d$ parameters, involves once again numerical problem.

In accordance with the discussed limitations, Chen, Härdle and Spokoiny (2006) present a simple VaR calculation approach that achieves much better accuracy of calculation than the alternative RiskMetrics method. In their study, financial risk factors are first converted to independent components (ICs) using a linear filtering. The covariance matrix of the resulting ICs is a diagonal matrix and the standardized ICs are independent and individually identifiable. By doing so, many univariate methods, that involve more realistic but complex procedures for local volatility estimation and distributional identification, can be easily applied to high dimensional series. Chen, Härdle and Jeong (2005), for example, propose a univariate VaR calculation approach by locally estimating volatility and well imitating the empirical distributional behavior of the underlying series. Chen and Spokoiny (2006) present the local exponential smoothing method to estimate volatility and implement it in univariate risk management, by which the volatility estimates are sensitive to structure shifts and have low variability.
The primary aim of this paper is to introduce a simple and fast multivariate risk management method, \textbf{GHICA}, by implementing the IC analysis (ICA) to the high dimensional series and fitting the resulting ICs in the GH distributional framework. The GHICA 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.

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 $\text{GH}(-0.5, 1.01, 0.05, 1.11, -0.03)$ for the Allianz and $\text{GH}(-0.5, 1.21, -0.21, 1.21, 0.24)$ for the DAX portfolio. Data source: FEDC (http://sfb649.wiwi.hu-berlin.de).
The proposed GHICA method is also applicable for covariance estimation. Relative to the widely used DCC setup, the GHICA method is fast and delivers sensitive estimates. This covariance estimation comparison is illustrated based on simulated data sets. Furthermore, the 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 the risk measures calculated by alternative methods, i.e. the RiskMetrics method, the method using the exponential smoothing to estimate volatility and assuming the $t(6)$ 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 estimations using the GHICA and DCC methods. The simulated data are GH distributed with $d = 50$ components. The real data analysis in Section 4 demonstrates the quality of the GHICA method based on the 20-dimensional German DAX portfolios and a dynamic exchange rate portfolio. Several alternative methods are implemented 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. A time homogeneous model means that the covariance matrix $\Sigma_x(t)$ is a constant, i.e. $x(t) = \Sigma_x^{1/2} \varepsilon_x(t)$, where the innovations $\varepsilon_x(t) \in \mathbb{R}^d$ are assumed to be standardized with $\mathbb{E}[\varepsilon_x(t)|\mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[\varepsilon_x^2(t)|\mathcal{F}_{t-1}] = I_d$. The maximum Gaussian likelihood estimate of the time homogeneous covariance $\Sigma_x$ is the average value of the squared returns. Since the covariance is time dependent in practice, many techniques have been used to approximate the local maximum likelihood estimate (MLE) of the covariance in the conditional heteroscedastic model:

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

by specifying a “local homogeneous” interval (e.g. one year or 250 trading days). Inside the homogeneous interval, the unknown covariance is almost 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. Recall that the Gaussian assumption helps to ensure
independence of the standardized innovations, which is essential for the ML estimation and the distributional identification. As the dimension \( d \) increases, it however has 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, the distributional behaviors such as asymmetry and heavy tails are well matched, but 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 the simple linear transformation and filters out ICs: \( y(t) = Wx(t) \). The transformation matrix \( W \) is assumed to be time constant and nonsingular. The heteroscedastic model is 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 simplifies high dimensional analysis to univariate study and significantly speeds up the calculation.

In this section, the building blocks of the GHICA method are described. To be more specific, 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 calculated based on the FFT technique.

The GHICA algorithm is 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

The second usage of the GHICA method is to estimate the covariance matrix \( \Sigma_x(t) \) based on the matrix estimate \( \hat{W} \) and the variance estimates of the ICs: \( \hat{\Sigma}_x(t) = \hat{W}^{-1}\hat{D}_y(t)\hat{W}^{-1\top} \)

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 main 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 \( W = (w_1, \ldots, w_d)^	op \) is nonsingular. High order moments are essential for estimation of the ICs. In the Gaussian framework, higher 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 non-gaussian distributed. Furthermore, the above transformation holds true by simultaneously multiplying the same constants to the unknown terms \( y(t) \) and \( W: \{cy(t)\} = \{cW\}x(t) \).

To avoid the scale identification problem, it is rational to standardize the dependent series and assume that every IC has unit variance \( E(y_j) = 1 \) with \( j = 1, \cdots, d \). The Mahalanobis transformation \( \tilde{x}(t) = \tilde{\Sigma}^{-1/2}x(t) \) helps to standardize the return series and the results are analyzed using the ICA:

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

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

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

\[
I(y) = \sum_{j=1}^{d} H(y_j) - H(y) = -\sum_{j=1}^{d} \int f_{y_j}(u) \log f_{y_j}(u)du
\]

where \( H(y_j) = -\int f_{y_j}(u) \log f_{y_j}(u) du \).

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 further the last term is 0 due to the orthogonality of the transformation matrix \( W \). It is clear that the mutual information \( I(\cdot) \) in (2) is equal to 0 if \( y_i \) and \( y_j \) are cross independent for \( i \neq j \) and \( i, j = 1, \ldots, d \), see Cover and Thomas (1991).
Hence for a candidate transformation \( W \), one can minimize the mutual information to achieve independence. The optimization problem is: \( \min_W \sum_{j=1}^{d} H(y_j) \) and can be further simplified to \( d \) simple optimization problems according to the inequality:

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

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. Moreover, the entropy and negentropy \( J(y_j) = H(y_0) - H(y_j) \) are in one-to-one correspondence, where \( y_0 \sim N(0, 1) \) 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). Therefore, we can also formulate the optimization problem as:

\[
\hat{w}_j = \arg\min_H H(y_j) = \arg\max J(w_j, y_j).
\]

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

\[
\hat{w}_j = \arg\min_H H(y_j) = \arg\max J(y_j)
\]

\[
J(y_j) \approx \text{const.} \{E[G(y)] - E[G(y_0)]\}^2 \\
= \text{const.} \{E[G(w_j^\top x)] - E[G(y_0)]\}^2 \\
G(y_j) = \log \cosh(y_j)
\]

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 \( W = \{w_1, \ldots, w_d\}^\top \) with \( j = 1, \ldots, d \), each has a unit norm.

2. Loop:
   - At step \( n \), Calculate \( \hat{w}_j^{(n)} = E[ x^\top(t)g \{ \hat{w}_j^{(n-1)}^\top x(t) \} ] - E[ g' \{ \hat{w}_j^{(n-1)}^\top x(t) \} ] \hat{w}_j^{(n-1)} \), where \( g \) is the first derivative of \( G(y) \) in form (3) and \( g' \) is the second derivative. The expectation \( E[\cdot] \) is approximated by the sample mean.
   - Do a symmetric orthogonalization of the estimated transformation matrix \( \hat{W}^{(n)} \):
     \[
     \tilde{W}^{(n)} = \{ \hat{W}^{(n)} \hat{W}^{(n)}^\top \}^{-1/2} \hat{W}^{(n)}
     \]
   - 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:

$$
\begin{align*}
D_y(t) &= \text{diag}\{\sigma_{y_1}^2(t), \ldots, \sigma_{y_d}^2(t)\} \\
\Sigma_x(t) &= W^{-1}D_y(t)W^{-1\top} \quad (4)
\end{align*}
$$

where $\sigma_{y_j}(t)$ is the heteroscedastic volatility of the $j$-th IC with $j = 1, \ldots, d$. 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 $\Gamma$ and the diagonal eigenvalue matrix $\Lambda$, 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 to achieve the best possible accuracy of the volatility estimation.

**Local exponential smoothing:** Given the univariate conditional heteroscedastic model:

$$
y_j(t) = \sigma_{y_j}(t)\varepsilon_{y_j}(t) \quad \text{with} \quad E[\varepsilon_{y_j}(t)|F_{t-1}] = 0 \quad \text{and} \quad E[\varepsilon_{y_j}^2(t)|F_{t-1}] = 1,
$$

we now focus on the adaptive estimation of the volatility $\sigma_{y_j}$ for $j = 1, \ldots, d$. For notational simplification, the subscripts $y_j$ in $\sigma_{y_j}$ and $j$ in $y_j$ are eliminated here. Suppose that a finite set $\{\eta_k, k = 1, \ldots, K\}$ of values of smoothing parameter is given. Every value $\eta_k$ leads to a localizing weighting scheme $\{\eta_{k-s}^m\}$ for $s \leq t$ to the local Gaussian MLE $\hat{\sigma}^{(k)}(t)$ (In practice, one truncates the smoothing window at $M_k$ such that $\eta_{M_k+1} \leq c \to 0$):

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

As discussed before, financial time series are heavy tailed distributed and can be well identified in the GH distributional framework. Under the normal inverse Gaussian (NIG) distributional assumption, (one subclass of the GH distribution, see Section 2.3 for more details), the quasi ML estimation is applicable if the exponential moment of the squared innovations $E[\exp\{\rho \varepsilon^2(t)\}]$ exists. A power transformation guarantees that:

$$
\begin{align*}
y_p(t) &= \text{sign}\{y(t)\}|y(t)|^p \\
\theta(t) &= \text{Var}\{y_p(t)|F_{t-1}\} = E\{y_p^2(t)|F_{t-1}\} = E\{|y(t)|^{2p}|F_{t-1}\} \\
&= \sigma_2^{2p}(t) E|\varepsilon(t)|^{2p} = \sigma_2^{2p}(t)C_p \quad (5)
\end{align*}
$$

where $C_p = E(|\varepsilon(t)|^{2p}|F_{t-1})$ 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 \( |y(t)|^{2p} \):

\[
\hat{\theta}^{(k)}(t) = \frac{\sum_{m=0}^{M_k} \eta_k^m |y(t-m-1)|^{2p}}{\sum_{m=0}^{M_k} \eta_k^m}
\]

with \( N_k = \sum_{m=0}^{M_k} \eta_k^m \)

(6)

The smoothing parameter \( \eta_k \) runs 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 (6) is positively related to the variation of the MLEs. This result is used to construct the sequence of the smoothing parameter \( \{\eta_k\} \):

\[
\frac{N_k}{N_k+1} \approx \frac{1 - \eta_k}{1 - \eta_{k+1}} = a > 1,
\]

(7)

where the coefficient \( a \) controls the decreasing speed of the variations.

The corresponding fitted log-likelihood ratio \( L(\eta_k, \hat{\theta}^{(k)}(t), \theta(t)) \) reads as:

\[
L(\eta_k, \hat{\theta}^{(k)}(t), \theta(t)) = L(\eta_k, \hat{\theta}^{(k)}(t)) - L(\eta_k, \theta(t)) = N_k K(\hat{\theta}^{(k)}(t), \theta(t))
\]

where the nonnegative Kullback-Leibler divergence is:

\[
K(\hat{\theta}^{(k)}(t), \theta(t)) = -\frac{1}{2} \left\{ \log \frac{\hat{\theta}^{(k)}(t)}{\theta(t)} - \frac{\hat{\theta}^{(k)}(t)}{\theta(t)} + 1 \right\}.
\]

The local MLEs \( \hat{\theta}^{(k)}(t) \) will be referred as “weak” estimates since the final estimate \( \hat{\theta}(t) \) achieves the best possible accuracy of estimation by aggregating all the weak estimates.

The procedure is sequential and starts with the estimate \( \hat{\theta}^{(1)}(t) \) that has the largest variability but small bias, i.e. we set \( \hat{\theta}^{(1)}(t) = \hat{\theta}^{(1)}(t) \). At every step \( k \geq 2 \), the new estimate \( \hat{\theta}^{(k)}(t) \) is constructed by aggregating the next “weak” estimate \( \hat{\theta}^{(k)}(t) \) 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: \( p(y,v) = p(y) \exp\{yv - d(v)\} \):

\[
\hat{v}^{(k)}(t) = \gamma_k \hat{v}^{(k)}(t) + (1 - \gamma_k) \hat{v}^{(k-1)}(t)
\]

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

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

\[
\gamma_k = K_{ag} \left\{ 2L \left( \eta_k, \bar{\theta}^{(k)}(t), \hat{\theta}^{(k-1)}(t) \right) / \lambda_k \right\}
\]

where \( K_{ag}(u) = \{1 - (u - 1/6)_+\}_+ \).

The aggregation kernel \( K_{ag} \) guarantees that the mixing coefficient \( \gamma_k \) is one if there is no essential difference between \( \bar{\theta}^{(k)}(t) \) and \( \hat{\theta}^{(k-1)}(t) \), and zero if the difference is significant. The significance level is measured by the critical value \( \zeta_k \). In Chen and Spokoiny (2006), the critical values for \( p = 0.25 \) have been calculated:

\[
\lambda_k = 0.008 + 0.005 \times (K - k).
\]

In the intermediate case, the mixing coefficient \( \gamma_k \) 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 corresponding volatility estimates \( \hat{\sigma}(t) \) are according to (5):

\[
\hat{\sigma}(t) = \{ \hat{\theta}(t)/C_p \}^{1/p},
\]

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.

The algorithm is described as follows:

1. Initialization: \( \hat{\theta}^{(1)}(t) = \bar{\theta}^{(1)}(t) \).
2. Loop: for \( k \geq 2 \)

\[
\hat{\theta}^{(k)}(t) = \left( \frac{\gamma_k}{\theta^{(k)}(t)} + \frac{1 - \gamma_k}{\theta^{(k-1)}(t)} \right)^{-1}
\]

where the mixing coefficient \( \gamma_k \) is computed as:

\[
\gamma_k = K_{ag} \left\{ 2L \left( \eta_k, \bar{\theta}^{(k)}(t), \hat{\theta}^{(k-1)}(t) \right) / \lambda_k \right\}
\]

3. Final estimate: if \( k = K \), \( \hat{\theta}(t) = \hat{\theta}^{(K)}(t) \) or \( \gamma_k = 0 \), \( \hat{\theta}(t) = \hat{\theta}^{(k-1)}(t) \).
4. Compute \( C_p \) such that \( \text{Var}\{\hat{\varepsilon}(t)\} = \text{Var}\left[y(t)\{C_p/\hat{\theta}(t)\}^{1/p}\right] = 1 \).
5. Compute volatility estimate \( \hat{\sigma}(t) = \{\hat{\theta}(t)/C_p\}^{1/p} \).

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:

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

where $\hat{R}_x$ 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 $\theta_1$ and $\theta_2$ are identified by the ML estimation.

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

After the ICA, one obtains ICs that 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:

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

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:

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

The characteristic function of the NIG variable is:

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

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

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

\[
f_{NIG}(y'; \alpha', \beta', \delta', \mu') = f_{NIG}(cy; \alpha'|c|, \beta'/c, |c|\delta, c\mu).
\]  

**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:

\[
f(y') = \frac{1}{|c|} f_{y}(\frac{y}{c}) = \frac{\alpha' \delta'}{\pi} \frac{K_1 \left\{ \alpha' \sqrt{\delta'^2 + (y' - \mu')^2} \right\}}{\sqrt{\delta'^2 + (y' - \mu')^2}} \exp\{\delta' \sqrt{\alpha'^2 - \beta'^2} + \beta'(y' - \mu')\}
\]

\[
= f_{NIG}(y'; \alpha', \beta', \delta', \mu').
\]

To calculate risk measures, it requires the identification of the portfolio returns’ density. Based on the GHICA model, the portfolio returns are formulated 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 \( \alpha \)-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 (8):

\[
\zeta_j \sim \text{NIG}(\zeta_j, \alpha_j, \beta_j, \delta_j, \mu_j) = \text{NIG}(\zeta_j, |a_j|, \beta_j/a_j, |a_j|\delta_j, a_j\mu_j).
\]
The characteristic function of the return \( r = \sum_{j=1}^{d} \zeta_j \) at time \( t \) is:

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

The density function is approximated by the Fourier transformation:

\[
f(r) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itr)\psi(z)dt \approx \frac{1}{2\pi} \int_{-s}^{s} \exp(-itr)\psi(z)dt
\]

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 \( z_j = -s + j \cdot h \) with \( j = 0, \ldots, N \).
  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 \( f(r) = \frac{1}{N} C_k \text{DFT}(y)_k \) with \( C_k = \frac{2s}{N} (-1)^k \exp(-ik\pi) \) with \( k = 0, \ldots, N-1 \). 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 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 \( \Sigma_1 \), meaning uncorrelatedness, and two symmetric and semi-positive defined matrices \( \Sigma_2 \) and \( \Sigma_3 \). (Here we first generate \( d \times d \) matrix \( U_1 \) whose elements are uniform random variables for \( \Sigma_2 \) and standard Gaussian variables for \( \Sigma_3 \), then calculate a new matrix \( U_2 = U_1 \ast U_1^t \) 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 ordered eigenvalues of these two matrices are displayed in Figure 2, by which the eigenvalues are distributed in \([5.92e-004, 3.779]\) (\( \Sigma_2 \)) and \([0.002, 3.573]\) (\( \Sigma_3 \)) respectively. The off-diagonal values span over \([-0.433, 0.468]\) in the first self-correlated matrix (\( \Sigma_2 \)) and \([-0.447, 0.464]\) in the second one (\( \Sigma_3 \)). Temporal stationarity is assumed to have a maximal length of 400 and a minimal length of 100. The
Fig. 2: Ordered eigenvalues of the generated covariance matrices.
structure shifts of the generated covariance are illustrated in Figure 3. The level of the shifts is either small with a shift from one self-correlated matrix ($\Sigma_2$ or $\Sigma_3$) to the identity matrix or contrariwise, e.g. at the point 700, or large with a change between the two self-correlated matrices, e.g. at the point 1800.

Furthermore, two distributional parameters $\mu$ and $\beta$ 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 $\alpha$ and $\delta$:

\[
E(\varepsilon_x) = \mu + \frac{\beta \delta}{\sqrt{\alpha^2 - \beta^2}} = 0
\]

\[
Var(\varepsilon_x) = \frac{\delta}{\sqrt{\alpha^2 - \beta^2}} + \frac{\beta^2}{\delta^3 \sqrt{\alpha^2 - \beta^2}} = \frac{\delta}{\alpha} = 1
\]

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:

\[ x(t) = W^{-1}y(t) = W^{-1}D_{y/2}(t) \varepsilon_y(t), \]

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.

\[ \Sigma_x(t) = D_x(t)R_x(t)D_x(t)^\top. \]

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

Figure 4 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 \( \Sigma_2 \) and \(-0.124 \) in \( \Sigma_3 \). 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
Fig. 4: Realized estimates of $\Sigma(2, 5)$ based on the GHICA and DCC methods. The generated data consists of 50 NIG distributed components.
estimates w.r.t. the true covariance are calculated pointwise.

\[ RAE(i,j) = \frac{\sum_{t=301}^{T} |\hat{\Sigma}_{GHICA}(i,j)(t) - \Sigma(i,j)(t)|}{\sum_{t=301}^{T} |\hat{\Sigma}_{DCC}(i,j)(t) - \Sigma(i,j)(t)|} \]

If \( RAE(i,j) \leq 1 \), it means that the GHICA method reaches higher accuracy in the estimation 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. \( \sum_{d=1}^{d} \sum_{j=1}^{d} \frac{1}{d \times d} 1(RAE(i,j) \leq 1) \) 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 5 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 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-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:

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

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

\[
\text{Risk charge (RC)} = \text{mean}(\text{VaR}_{t,pr})
\]

**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, expected shortfall (ES) has been considered together with the VaR to measure and control the market risks.

\[
\text{ES} = \mathbb{E}\{-r(t)| - r(t) > \text{VaR}_{t,pr}\}
\]

Investors suffer loss once bankruptcy happens. Even in the best case, 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 profits and hence the small value of risk charge favors their requirements.

**Internal supervisory:** It is important for internal supervisory to exactly measure the market risk exposures before risk controlling. For this reason, internal supervisory prefers

<table>
<thead>
<tr>
<th>No. exceptions</th>
<th>Increase of $M_f$</th>
<th>Zone</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 bis 4</td>
<td>0</td>
<td>green</td>
</tr>
<tr>
<td>5</td>
<td>0.4</td>
<td>yellow</td>
</tr>
<tr>
<td>6</td>
<td>0.5</td>
<td>yellow</td>
</tr>
<tr>
<td>7</td>
<td>0.65</td>
<td>yellow</td>
</tr>
<tr>
<td>8</td>
<td>0.75</td>
<td>yellow</td>
</tr>
<tr>
<td>9</td>
<td>0.85</td>
<td>yellow</td>
</tr>
<tr>
<td>More than 9</td>
<td>1</td>
<td>red</td>
</tr>
</tbody>
</table>

Tab. 1: Traffic light as a factor of the exceeding amount, cited from Frank, Härdle and Hafner (2004).
the model with the empirical probability \( \hat{p}_r \) as close to the expected values as possible:

\[
\hat{p}_r = \frac{\text{No. exceptions}}{\text{No. total observations}}
\]

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. \( p_r = 1\% \) for regulatory reason and \( p_r = 0.5\% \) 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 Gaussian distributional assumption and a modification with \( t(6) \) 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 \( \eta = 0.94 \):

\[
\begin{align*}
  r(t) &= b^\top x(t) = \sigma_r(t)\varepsilon_r(t) \\
  \sigma_r^2(t) &= \frac{\left\{ \sum_{m=0}^{M} \eta^m r^2(t - m - 1) \right\}}{\left( \sum_{m=0}^{M} \eta^m \right)}
\end{align*}
\]

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 further it often suffers from a low accuracy of estimation.

Figure 6 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 \( p_r = 0.5\% \) 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
Exponential smoothing $t(6)$ 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 RiskMetrics 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 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:

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

<table>
<thead>
<tr>
<th>$h$</th>
<th>$b(t)$</th>
<th>pr</th>
<th>GHICA</th>
<th>RiskMetrics N$(\mu, \sigma^2)$</th>
<th>Exponential smoothing $t(6)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>pr</td>
<td>RC</td>
<td>ES</td>
<td>pr</td>
</tr>
<tr>
<td>1</td>
<td>$b^{(1)}$</td>
<td>1%</td>
<td>0.55%</td>
<td>0.0264</td>
<td>0.0456</td>
</tr>
<tr>
<td></td>
<td>$b^{(1)}$</td>
<td>0.5%</td>
<td>0.44%*</td>
<td>0.0297</td>
<td>0.0472*</td>
</tr>
<tr>
<td></td>
<td>$b^{(2)}$</td>
<td>1%</td>
<td>0.59%</td>
<td>0.0265</td>
<td>0.0448</td>
</tr>
<tr>
<td></td>
<td>$b^{(2)}$</td>
<td>0.5%</td>
<td>0.42%*</td>
<td>0.0298</td>
<td>0.0476*</td>
</tr>
<tr>
<td>5</td>
<td>$b^{(1)}$</td>
<td>1%</td>
<td>0.83%</td>
<td>0.0550</td>
<td>0.0841</td>
</tr>
<tr>
<td></td>
<td>$b^{(1)}$</td>
<td>0.5%</td>
<td>0.51%*</td>
<td>0.0612</td>
<td>0.0939*</td>
</tr>
<tr>
<td></td>
<td>$b^{(2)}$</td>
<td>1%</td>
<td>0.83%*</td>
<td>0.0554</td>
<td>0.0828*</td>
</tr>
<tr>
<td></td>
<td>$b^{(2)}$</td>
<td>0.5%</td>
<td>0.50%*</td>
<td>0.0617</td>
<td>0.0943*</td>
</tr>
</tbody>
</table>

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 *. The method preferred by investor is marked by †. For the internal supervisory, the method marked by ‡ is recommended.

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:
Fig. 6: 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.
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 \( ^s \).

where \( x(t) = \{ x_1(t), \ldots, x_d(t) \}^\top \). 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.

\[
r(t) = b(t)^\top x(t) = b(t)^\top \Sigma^{(1/2)}(t) \varepsilon_x(t)
\]

where \( \varepsilon_x \sim N(\mu, \Sigma_x) \) with the diagonal covariance matrix \( \Sigma_x \). Notice that the quantile vector with \( \text{pr} \)-quantiles of individual innovation does not necessarily correspond to the \( \text{pr} \)-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:

\[
\begin{align*}
\mathbb{E}\{r(t)\} &= b(t)^\top \Sigma_x^{(1/2)}(t) \mathbb{E}\{\varepsilon_x(t)\} \\
\text{Var}\{r(t)\} &= b(t)^\top \Sigma_x^{(1/2)}(t) \text{Var}\{\varepsilon_x(t)\} \Sigma_x^{(1/2)}(t) b(t)
\end{align*}
\]

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 fast procedure given either static or dynamic portfolios but also produces better results than several alternative risk management methods.
References


