CHICAGO: A Fast and Accurate Method for Portfolio Risk Calculation

Simon A. Broda   Marc S. Paolella

First version: November 2006
Current version: February 2008

This research has been carried out within the NCCR FINRISK project on “Financial Econometrics for Risk Management”
CHICAGO: A Fast and Accurate Method for Portfolio Risk Calculation

Simon A. Broda\textsuperscript{a}\textsuperscript{*} Marc S. Paolella\textsuperscript{a}

\textsuperscript{a}Swiss Banking Institute, University of Zurich, Switzerland

First Draft: November 2006
Current Draft: February 2008

Abstract

The estimation of multivariate GARCH models remains a challenging task, even in modern computer environments. This manuscript shows how Independent Component Analysis can be used to estimate the Generalized Orthogonal GARCH model in a fraction of the time otherwise required. The proposed method is a two-step procedure, separating the estimation of the correlation structure from that of the univariate dynamics, thus facilitating the incorporation of non-Gaussian innovations distributions in a straightforward manner. The generalized hyperbolic distribution provides an excellent parametric description of financial returns data and is used for the univariate fits, but its convolutions, necessary for portfolio risk calculations, are intractable. This restriction is overcome by a saddlepoint approximation to the required distribution function, which is computationally cheap and extremely accurate — most notably in the tail, which is crucial for risk calculations. A simulation study and an application to stock returns demonstrate the validity of the procedure.

Keywords: Empirical Finance; Saddlepoint Approximation; Value at Risk

JEL Classification: C13; C16; C32; G11

\textsuperscript{*}Corresponding author. E-mail address: broda@isb.uzh.ch. Part of this research has been carried out within the National Centre of Competence in Research “Financial Valuation and Risk Management” (NCCR FINRISK), which is a research program supported by the Swiss National Science Foundation.
1 Introduction

The Value at Risk and Expected Shortfall, or VaR and ES, respectively, of a portfolio of assets have become key figures in risk management. Accurately forecasting these measures has, however, proved challenging, because returns on financial assets observed at weekly or higher frequencies exhibit excess kurtosis and mild skewness, and because the assets are not independent of each other.

The approach of fitting a univariate GARCH-type model directly to the time series of portfolio returns is effective, but bears the problem that the entire model must be re-estimated every time the composition of the portfolio changes. The conventional solution is to model the joint evolution of the assets by means of multivariate GARCH, or MGARCH models. Many different MGARCH models have been considered in the literature, with the common goal of reducing the dimensionality of the most general such model, the VEC(\(p,q\)) model of Bollerslev et al. (1988); an excellent overview is given in Bauwens et al. (2006). Among these models, the constant conditional correlation (CCC) model of Bollerslev (1990) and the dynamic conditional correlation (DCC) model of Engle (2002) bear the advantage that there exist two-step procedures for their estimation, thus effectively reducing the multivariate problem to a set of univariate ones. As such, estimation of high-dimensional models becomes feasible, albeit at the expense of flexibility. More complex models that offer a more general covariance specification typically do not allow such a simplification. Examples are the BEKK model of Engle and Kroner (1995), and the Factor-GARCH and Generalized Orthogonal (GO) GARCH models of Engle et al. (1990) and van der Weide (2002), respectively, the latter two being nested in the former. Therefore, inference in high-dimensional problems is practically infeasible, especially in a non-Gaussian setup.

This manuscript will demonstrate a two-step procedure for estimating the GO-GARCH model of van der Weide (2002) with non-Gaussian innovations. Our main tool is independent component analysis (ICA), details on which will be given below. In order to motivate the name of our method — CHICAGO, or Conditionally Heteroscedastic Independent Component Analysis of Generalized Orthogonal GARCH models, suffice it for now to say that the specific ICA algorithm employed in this paper maximizes the conditional heteroscedasticity of the estimated components. As with the CCC and DCC models, ICA allows us to decompose the problem into a set of readily estimable univariate models, while at the same time maintaining sufficient flexibility in the specification of the co-evolution of the assets.

We model the conditional distribution of the individual assets as generalized hyperbolic (GHyp); in this sense, our approach is similar to that of Chen et al. (2006), who employ the GHyp distribution and ICA in a non-parametric setting. The only missing link to an efficient calculation of portfolio Value at Risk is the evaluation of tail probabilities for a portfolio of assets. The present paper achieves this by replacing the Fast Fourier Transform used by Chen et al. (2006) to obtain the required distribution of convolutions of independent GHyp random variables.
by a saddlepoint approximation.

The remainder of this manuscript is organized as follows. Section 2 introduces the GO-GARCH model, and shows how independent component analysis can be used to estimate it in two steps. Section 3 details the univariate factor specification. Section 4 derives the saddlepoint approximation for convolutions of independent GHyp distributions. Section 5 details the results of a backtesting exercise. Section 6 concludes.

2 The GO-GARCH Model

2.1 Model Specification

Consider a set of $d$ financial assets, with associated return vector $r_t$, $t \in \{1, \ldots, T\}$, and conditional mean $\mu_t$, which is assumed to be consistently estimable by ordinary least squares. (Note that this requirement precludes GARCH-in-mean-type models.) In the GO-GARCH model of van der Weide (2002), the innovations $u_t$ are modelled as linear combinations of $d$ unobserved factors $f_t$:

$$r_t = \mu_t + u_t \quad \text{(1)}$$

$$u_t = Af_t \quad \text{(2)}$$

for some mixing matrix $A$ that is invertible and constant over time. The unobserved factors are assumed to be independent of each other, and to have unit unconditional variance. Note that the latter is an identifying restriction; any scale factors can simply been absorbed into the mixing matrix. It follows that the unconditional covariance matrix of the returns is given by

$$\Sigma = \mathbb{E}[ uu\prime] = AA\prime.$$ 

If one assumes a GARCH(1,1) process for each factor $\{f_{it}\}$, i.e., $f_t \sim (0, H_t)$, where

$$H_t = \Omega + \sum_{k=1}^{d} \alpha_k e_k e_k\prime H_{t-1} e_k e_k\prime + \sum_{i=1}^{d} \beta_k e_k e_k\prime f_{t-1} f_{t-1}\prime e_k e_k\prime,$$

$$\Omega = \sum_{k=1}^{d} (1 - \alpha_k - \beta_k) e_k e_k\prime,$$

and $e_k$ is a $d \times 1$ vector with $k$th element 1 and zeros everywhere else, then the conditional covariance of the return series becomes

$$\Sigma_t = A \Omega A\prime + \sum_{k=1}^{d} \alpha_k \lambda_k w_k \Sigma_{t-1} w_k \lambda_k + \sum_{i=1}^{d} \beta_k \lambda_k w_k \Sigma_{t-1} w_k \lambda_k,$$

where $\lambda_k = Ae_k$ and $w_k = (A^{-1})' e_k$. This is a special case of the factor GARCH model of Engle et al. (1990), and of the BEKK model of Engle and Kroner (1995). In contrast to the latter, however, the GO-GARCH model does not necessarily require that the underlying factors obey a GARCH structure, so that more flexible dynamic specifications such as, e.g., the Asymmetric Power ARCH, or A-PARCH, model of Ding et al. (1993), can be employed. The Full Factor Model of Vrontos et al. (2003) is based on a similar idea, but uses a triangular mixing matrix.
Using a polar decomposition, the mixing matrix $A$ can be uniquely factorized into a symmetric positive definite matrix $\Sigma^{1/2}$ and an orthogonal matrix $U$:

$$A = \Sigma^{1/2}U,$$

where $\Sigma^{1/2}$ is the symmetric positive definite square root of the unconditional covariance matrix $\Sigma$, so that

$$\text{tr}(\Sigma^{1/2}UU'^T\Sigma^{1/2}) = \Sigma,$$

because $UU' = I$. Note that, as the factors $f_t$ are unobserved, their order is not identifiable. Hence, the order of the columns of $U$ can be chosen arbitrarily to make the determinant positive, so that $U$ is a rotation matrix. As a rotation matrix, $U$ can be decomposed as the product of $\binom{d}{2}$ basic rotation matrices $R_i(\theta_i)$, where each $R_i$ is a rotation of angle $\theta_i$ in the plane spanned by one pair of axes in $\mathbb{R}^d$ (see van der Weide, 2002, Lemma 3). Thus, $U$ can be fully parameterized in terms of the Euler angles $\theta_i$. For example, in the $d = 2$ case,

$$U = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}.$$ 

The angles $\theta_i$ can be restricted to the interval $0 \leq \theta_i < \pi/2$, because the signs and order of the factors cannot be identified.

### 2.2 Estimation of the Mixing Matrix

#### 2.2.1 Estimation Strategies

If one assumes a GARCH(1,1) structure for each factor, then the GO-GARCH model contains $d^2 + 3d$ parameters. The difficulties associated with jointly estimating a parameter vector of this dimension make it desirable to have a two-step procedure available, in which the mixing matrix is estimated first, thus effectively reducing the problem to a set of $d$ univariate problems. For any two-step algorithm, it is expedient to exploit the decomposition in (3), because the $d(d+1)/2$ free parameters in $\Sigma^{1/2}$ can be consistently estimated from the (unconditional) sample covariance. As such, we will only consider the whitened and demeaned data

$$z_t = \hat{\Sigma}^{-1/2} \hat{u}_t,$$

where $\hat{u}_t = (r_t - \hat{\mu}_t)$, and $\hat{\mu}$ is a consistent estimator of $\mu$. The unconditional covariance matrix could be estimated from the sample covariance of the OLS residuals; however, though unbiased, this simple estimator may be inadequate for the high-dimensional data sets for which the method is designed. As such, we propose to use the shrinkage estimator of Ledoit and Wolf (2003), which, apart from being more efficient, guarantees positive definiteness of $\hat{\Sigma}$.

The $d(d-1)/2$ parameters in $U$, on the other hand, cannot be estimated on the basis of unconditional information alone, because from (4), any orthogonal matrix $U^*$ gives rise to the
same unconditional covariance matrix $\Sigma$. Thus, any estimation method must utilize conditional information. One such procedure is given in Boswijk and van der Weide (2006), where an estimator $\hat{U}_{BW}$ is derived as the eigenvector matrix of the symmetric matrix $\hat{B}$ solving the nonlinear least squares problem

$$
\hat{B} = \arg\min_{B: B = B'} \frac{1}{T-1} \sum_{t=2}^{T} \text{tr} \left( [z_t z'_t - I_d - B (z_{t-1} z'_{t-1} - I_d) B]^2 \right).
$$

We propose here to estimate $U$ by independent component analysis (ICA), which we briefly outline next. Details can be found in the monograph by Hyvärinen et al. (2001).

The basic ICA model assumes that a $d$-dimensional random vector $u_t \equiv [u_{1t},...,u_{dt}]'$ is observed. The $u_{it}$ are linear combinations of $d$ independent random variables $f_{it}$,

$$
u_{it} = \sum_{j=1}^{d} a_{ij} f_{jt},$$

or, in matrix form,

$$
u_t = Af_t.$$

The aim is to estimate both $f_t$ and $A$, i.e., to find a matrix $W \equiv A^{-1}$ such that $y_t \equiv Wu_t$ are independent. The naive approach of taking $W = \hat{\Sigma}^{-1/2}$ produces uncorrelated components, but yields independent components only up to an orthogonal transformation: With $U^*$ orthogonal,

$$
E[U^*y_t y'_t U^{*\prime}] = U^*U^{*\prime} = I,
$$

i.e., $U^*y_t$ is also uncorrelated, but not necessarily independent, unless the the data are i.i.d. multivariate Gaussian. In terms of the decomposition (3), the orthogonal matrix $U$ remains to be estimated, which requires information beyond that contained in the unconditional covariance matrix. Which particular additional information to use depends on the problem at hand; for example, with non-Gaussian data, the central limit theorem can be exploited: the distribution of a normalized sum of independent random variables with finite second moments converges to a Gaussian. Conversely, a linear combination $w'u_t = w'Af_t =: q'f_t$ will be “least Gaussian” if $q$ has one element equal to 1 and all others zero. The degree of non-Gaussianity can be measured by negentropy, which for a random variable $X$ with density $f_X$ is defined as

$$
J(f_X) = S(\phi_X) - S(f_X),
$$

where $S(f) = -\int f(x) \log f(x)dx$ is the differential entropy of $f$, and $\phi_X$ denotes the density of a Gaussian random variable with the same variance as $X$. The aim is then to find $w$ such that the negentropy (or an approximation thereof, as the density is typically unknown) of $w'u_t$ is maximized. The FastICA algorithm of Hyvärinen (1999) achieves this with cubic convergence.

For time series data, exploiting the time structure of the data set to identify the independent components appears more natural. Financial returns data typically exhibit very strong GARCH-effects, and it suggests itself to rely on these as additional information. It is well known that the
GARCH-effects present in the sum of two (or more) series are weaker than those in the individual series themselves. It is therefore possible to separate the independent components by maximizing the autocorrelation of the squared returns. Hyvärinen et al. (2001, p. 349) devise a fixed-point algorithm based on cross cumulants that achieves this with cubic convergence. Given prewhitened data \( z_t \) as in (5), the algorithm starts with \( U_n = I \) and iterates

\[
U_{\text{temp}} = z[z'U_n \odot z'_- U_n \odot z'_- U_n]/T + z_- [z'_- U_n \odot z'_- U_n \odot z'_- U_n]/T - 2U_n - 4 \bar{C}U_n D_n
\]

\[
U_{n+1} = (U_{\text{temp}} U'_{\text{temp}})^{-1/2} U_{\text{temp}}
\]

where \( z = [z_2, \ldots, z_T], \quad z_- = [z_1, \ldots, z_{T-1}], \quad \bar{C} = (zz'_- + z_-z')/(2T), \quad D_n = \text{diag}(\text{vecd}(U_n' \bar{C} U_n)) \), vecd is the operator which forms a column vector from the diagonal elements of a matrix, and \( \odot \) denotes the Hadamard product. The iteration stops when \( 1 - c < \epsilon \), where \( c \) is the minimum over the absolute values of the diagonal elements of \( U_n' U_n \), and \( \epsilon \) is a suitable convergence threshold (we use \( 10^{-12} \)). In the rare cases that the algorithm fails to converge, one may fall back to the negentropy-based FastICA algorithm.

Figures 1 and 2 illustrate the technique. From top to bottom, the rows of each Figure plot the original components \( f_t \), the mixed components \( A f_t \), and the estimated components. It is apparent from the graphs that ICA is able to restore the original components, except for their signs and order, which are not identifiable.

### 2.2.2 Performance Comparison

Thus, there exist (at least) three different estimators of the rotation matrix \( U \): the maximum likelihood estimator (MLE), the estimator of Boswijk and van der Weide (BW), and independent component analysis (ICA). A simulation study is useful to compare their relative performance, and we detail the results of such an experiment next. We consider a bivariate model, which bears the advantage that the results can be condensed into a single statistic, namely, the estimated rotation angle \( \hat{\theta} \) of the rotation matrix \( U \). To keep matters simple, we use a GARCH(1,1) model with Gaussian innovations for each factor, with parameters \((\alpha_1, \beta_1) = (0.09, 0.9) \) and \((\alpha_2, \beta_2) = (0.04, 0.95) \). Note that for both factors, the parameters are close to the stationarity border \((\alpha_i + \beta_i = 0.99 \text{ for both factors})\), which closely mimics the characteristics typically found in actual data. Because our interest centers on the rotation angle \( \theta \), the unconditional covariance matrix was set to the identity matrix, and the length of each of the 1,000 simulated samples is \( T = 800 \). We report the root mean squared error (RMSE) and BIAS of the estimated angle \( \hat{\theta} \) for each of the three estimation methods. The MLE — like the other two estimators — was computed from demeaned and whitened data, rather than jointly estimating the unconditional covariance matrix \( \Sigma \) with the remaining parameters.

A graphical representation of the results for different values of the true rotation angle \( \theta \) is given in Figure 3. For each estimated rotation matrix \( \hat{U} = (u_{ij}) \), the rotation angle was computed
Figure 1: Example for ICA. Top row: original signals. Middle row: mixed signals. Bottom row: estimated independent components.
Figure 2: Example for ICA. Top row: original signals. Middle row: mixed signals. Bottom row: estimated independent components.
Figure 3: Performance Comparison of MLE, ICA, and Boswijk and van der Weide (BW).

as $\hat{\theta} = \cos^{-1}(u_{11})\text{sgn}(u_{12}) + k\pi/2$, where $k$ is an integer minimizing the distance between $\hat{\theta}$ and the true value, $\theta$ (note that the rotation angle is only identifiable up to multiples of $\pi/2$). For the parameter constellation under investigation, the ICA estimator displays a slight bias, whereas both the MLE and BW estimators are virtually unbiased. In terms of RMSE, the MLE is favored over the alternative estimators, and the ICA estimator has a slight yet consistent advantage over the estimator of Boswijk and van der Weide. However, it should be borne in mind that the BW and ICA estimators use only information contained in the autocorrelation of the squared returns, whereas the MLE relies on the exact specification of the factor dynamics; as such, the former two estimators are vastly more robust with respect to the assumptions on the evolution of the univariate factors.

Another important remark concerns the computation time required by each estimator: the MLE, for the sample size at hand, took on average 8.87 seconds to compute, the estimator of Boswijk and van der Weide 1.68 seconds, and the ICA estimator 0.03 seconds, rendering the latter method 297 and 56 times faster, respectively, than the former two. Lastly, the ICA method converges reliably even for high-dimensional time series, without any of the convergence problems associated with maximum likelihood estimation in such cases. Taking all aspects (accuracy, robustness, computational reliability, and speed) into account, independent component analysis is a very useful tool for estimating GO-GARCH models.

3 Univariate Factor Specification

3.1 Marginal Distribution of Assets

We assume a generalized hyperbolic law for the conditional distribution of each factor. The GHyp is an extremely flexible asymmetric and fat–tailed distribution which nests a large number
of distributions that are popular in the empirical modelling of asset returns. These include the Student’s $t$, which is ubiquitous in such applications (see, e.g., the survey article of Palm, 1996, and the references therein); the Laplace (see, e.g., Granger and Ding, 1995; Mittnik et al., 1998; Haas et al., 2006), normal inverse Gaussian (NIG), and hyperbolic (see Küchler et al., 1999; Prause, 1999). While the GHyp offers an excellent empirical fit to financial data, its use can also be motivated from a theoretical point of view. For example, Barndorff-Nielsen (1998) discusses a stochastic volatility process with normal inverse Gaussian marginals. Also, Reimann (2005) has demonstrated that a purely economic model for asset return generation results in empirical behavior strikingly similar to a hyperbolic distribution. For more information on the generalized hyperbolic, see also Paolella (2007, Ch. 9), and the references therein.

Value at Risk for portfolios of multivariate generalized hyperbolic assets has been considered in Bauer (2000) and Sadefo-Kamdem (2006), and, for the important special case of the multivariate NIG, in Aas et al. (2006). In our model, the factors conditionally obey a different, non-elliptical multivariate generalization of the GHyp, in which each component can be expressed as a linear combination of independent univariate GHyp random variables. This distribution has been analyzed in Schmidt et al. (2006), where it was found to be adequate for returns data.

Specifically, let $\lambda \in \mathbb{R}$, $\omega > 0$, $-1 < \rho < 1$, $\mu \in \mathbb{R}$ and $\sigma > 0$. Then random variable $X$ follows a generalized hyperbolic density, written $X \sim \text{GHyp}(\lambda, \omega, \rho, \sigma, \mu)$, if its density is given by

$$ f_X(x; \lambda, \omega, \rho, \sigma, \mu) = \frac{\omega^\lambda \bar{y}^{\lambda - \frac{1}{2}}}{\sqrt{2\pi \alpha^\lambda - \frac{1}{2} \sigma K_\lambda(\omega)}} K_{\lambda - \frac{1}{2}}(\bar{\alpha} \bar{y}) e^{\rho \bar{z}}, $$

where $z \equiv \frac{x - \mu}{\sigma}$, $\bar{\alpha} \equiv \omega(1 - \rho^2)^{-1/2}$, $\bar{y} \equiv \sqrt{1 + z^2}$, and $K_\nu(x)$ is the modified Bessel function of the third kind with index $\nu$, defined as

$$ K_\nu(x) = \frac{1}{2} \int_0^\infty t^{\nu-1} e^{-\frac{1}{2}x(t+1)} \, dt. $$

Reliable numeric routines exist, for example in Matlab, for computing $K_\nu(x)$. The parameters of the generalized hyperbolic have the following interpretation: $\mu$ and $\sigma$ are genuine location and scale parameters, respectively, while $\lambda, \omega$ and $\rho$ are location–scale invariant. Parameter $\omega$ controls the tail thickness, and $\rho$ is a measure of the skewness. We will consider the special cases obtained by letting $\lambda = -\frac{1}{2}$ (normal inverse Gaussian, or NIG), and $\lambda = 1$ (hyperbolic). We found that these special cases still offer enough flexibility to capture the excess kurtosis and skewness present in financial returns data: Figures 4 and 5 show the excellent fit offered by the standardized NIG and hyperbolic distributions, respectively, for two independent components identified from the Dow Jones data set considered in Section 5.

The expected value and variance of the GHyp distribution are given by

$$ \mathbb{E}[X] = \mu + \sigma \frac{\rho}{\sqrt{1 - \rho^2}} k_1(\omega) $$

and

$$ \mathbb{V}[X] = \sigma^2 \left[ \omega^{-1} k_1(\omega) + \frac{\rho^2}{1 - \rho^2} k_2(\omega) \right], $$

9
respectively, where \( k_1(\omega) \equiv K_{\lambda+1}(\omega)/K_\lambda(\omega) \) and \( k_2(\omega) \equiv [K_\lambda(\omega)K_{\lambda+2}(\omega) - K_{\lambda+1}(\omega)^2]/K_\lambda(\omega)^2 \); see, e.g., Bibby and Sørensen (2003) and the references therein for further results and details on their derivation. For the purpose of this paper, we will standardize the generalized hyperbolic to have zero mean and unit variance and denote the standardized distribution as SGH. Its density is

\[
f_{\text{SGH}}(x; \lambda, \omega, \rho) = f_{\text{GHyp}}(x; \lambda, \omega, \rho, \hat{\delta}, \hat{\mu}),
\]

where \( \hat{\delta} = \left[ \omega^{-1}k_1(\omega) + \rho^2(1 - \rho^2)^{-1}k_2(\omega) \right]^{-1/2} \) and \( \hat{\mu} = -\rho(1 - \rho^2)^{-1/2}\hat{\delta}k_1(\omega) \).

### 3.2 Specification of Factor Volatilities

As mentioned before, the GO-GARCH model does not restrict the dynamics of the univariate factors to follow simple GARCH(1,1) processes. Rather, the model is general enough to allow for the incorporation of arbitrary univariate dynamics. As such, we assume that \( f_{it} = Z_{it}\sigma_{it} \), where the \( Z_{it} \) are independently distributed as \( \text{SGH}(\lambda_i, \omega_i, \rho_i) \), and to capture the evolution of

\[\ldots\]
the scale parameters $\sigma_{i,t}$, we use the very successful Asymmetric Power ARCH, or A-P ARCH, model proposed by Ding et al. (1993), given by

$$
\sigma_{i,t}^\delta = c_{i0} + \sum_{j=1}^{r} c_{ij} (|f_{i,t-j}| - \gamma_{ij} f_{i,t-j})^\delta_i + \sum_{j=1}^{s} d_{ij} \sigma_{i,j}^\delta_i
$$

with $c_{ij} > 0$, $d_{ij} \geq 0$, $\delta_i > 0$, and $|\gamma_{ij}| < 1$.

4 Evaluating the GHyp Distribution and its Convolutions

4.1 Tail Probabilities

Calculating tail probabilities for $R_t \equiv b' r_t = b' \mu_t + b' A f_t$, the return on a portfolio of assets with weights $b = [b_1, \ldots, b_d]'$, involves evaluating the cumulative distribution function (cdf) of a weighted sum of $d$ independent random variables, each of which in this paper is assumed to be generalized hyperbolic. That is, we have $X_i \overset{iid}{\sim} \text{GHyp}(\lambda_i, \omega_i, \rho_i, \sigma_i, \mu_i)$, $i = 1, \ldots, d$, and are interested in the cdf of $S = \sum_{i=1}^{d} a_i X_i$. For $d = 1$, numeric integration of the density (6) can be used, but for $d > 1$, no expression exists for the convolution of such random variables when the distributional parameters in (6) are allowed to differ across assets—as required in practice.

There are three approaches to resolve this. The first is inverting the characteristic function of $S$, $\varphi_S(t)$, if it is available. From Gil-Pelaez (1951), the cdf of $S$ is given by

$$
F_S(x) = \frac{1}{2} + \frac{1}{2\pi} \int_{0}^{\infty} \frac{e^{itx} \varphi_S(-t) - e^{-itx} \varphi_S(t)}{it} dt,
$$

or, equivalently and computationally more efficient,

$$
F_S(x) = \frac{1}{2} - \frac{1}{\pi} \int_{0}^{\infty} g(t) dt,
$$

where

$$
g(t) = \frac{\text{Im} z(t)}{t} \quad \text{and} \quad z(t) = e^{-itx} \varphi_S(t).
$$

However, the integrand in equation (9) tends to be oscillatory, thus rendering numerical quadrature difficult. It is therefore expedient to evaluate instead the following integral, due to Helstrom (1996):

$$
F_S(x) = H(s_0) - \frac{1}{\pi} \int_{0}^{\infty} \text{Re} \left( e^{Is} (1 - iC y) \right) dy, \quad s = s_0 + \frac{1}{2} C y^2 + i y,
$$

where $H(\cdot)$ denotes the Heaviside step function, and, with $K_S$ denoting the cumulant generating function of $S$,

$$
I(s) = K_S(s) - xs - \ln(\text{sgn}(s_0)s), \quad C = \frac{I^{(3)}(s_0)}{3I^{(2)}(s_0)},
$$

superscripts in parentheses denoting derivatives. The saddlepoint $s_0$ lies in the convergence strip of the moment generating function and satisfies

$$
\left. \frac{dI(s)}{ds} \right|_{s_0} = 0, \quad \text{sgn}(s_0) = \text{sgn}(X - \mathbb{E}[X]).
$$


The cumulant generating function, along with its first and second derivatives, is given in the Appendix. The third derivative is easily approximated numerically. The value of this method in the present context is in evaluating the accuracy of the (vastly faster) saddlepoint method described below.

The second method is via the Fast Fourier transform, as used by Chen et al. (2006), a method which is time consuming if high accuracy is desired.

The third way is via the saddlepoint approximation, which is outlined in the Appendix. It can be thought of as approximate inversion of the characteristic function, but without requiring integration (thus being faster and avoiding potentially pathological integrand problems), or as an Edgeworth expansion, but vastly more accurate and without the problems associated with the latter, such as negative values of the density and poor accuracy in the tails. Its accuracy for $d > 1$ will be similar, if not higher, than for the $d = 1$ case, because, as assets are summed, a central limit effect takes place, drawing the distribution of $S$ closer to normality—for which the SPA is exact. Thus, as a worst–case scenario, we present accuracy results for $d = 1$, for which we can compute the true values via (6) for the density and numeric integration of (6) for the cdf.

Figure 6 graphically illustrates the accuracy of the saddlepoint approximation. The left panel shows the true and approximate density, while the right panel shows the relative percentage error of the cdf approximation. We see that the cdf approximation has well under one percent relative error even far into the left tail, which is where accuracy is important in Value–at–Risk calculations.

Figure 6: Left: The exact density (solid) and renormalized saddlepoint approximation (dashed) of the standardized GHyp density (8) for parameters $\lambda = 3$, $\omega = \sqrt{8}$ and $\rho = -1/3$. Right: The percentage relative error of the cdf saddlepoint approximation, defined as $100(\hat{F} - F)/\min(F, 1 - F)$, where $\hat{F}$ is given in (13) and the true cdf, $F$, was computed with numerical integration.
4.2 Quantiles and Expected Shortfall

In Value at Risk applications, the focus is on the quantiles of $S$, rather than on tail probabilities. Their computation involves numerically inverting the distribution function, i.e., solving

$$F_S(x_q) = q, \quad q \in (0, 1),$$

a computationally demanding task if exact methods are to be used for the evaluation of $F_S(\cdot)$. However, constructing a saddlepoint approximation for quantiles is no more involved than it is for tail areas, as detailed in the Appendix.

A related risk measure is the Expected Shortfall, defined, for a given VaR level $q$, as

$$\mathbb{E}\left[ S \mid S \leq x_q \right] = \frac{1}{q} \int_{-\infty}^{x_q} S f_S(s) ds,$$

which, in the case of the GHyp distribution, must be evaluated numerically, where $f_S$ can be replaced by its saddlepoint approximation. However, unless renormalized, the saddlepoint approximation to the pdf does not integrate to unity, which may jeopardize the accuracy of such a calculation. It is therefore expedient to instead evaluate

$$\mathbb{E}\left[ S \mid S \leq x_q \right] = x_q - \frac{1}{q} \int_{-\infty}^{x_q} F_S(s) ds, \quad x_q < 0,$$

which follows from (11) upon integrating by parts. In (12), $F_S(s)$ can be safely replaced by its saddlepoint approximation, which is a proper distribution function. The third option is to use the direct saddlepoint approximation given in (17) below. Figure 7 plots the Expected Shortfall computed via equations (11), (12), and (17) (dotted, dash-dot, and solid lines, respectively) against the VaR level, for a standardized GHyp distribution with $\omega = 1.5$ and $\rho = -0.1$. Approximation (12) clearly yields the highest accuracy and is therefore the recommended method.
5 Application

In order to exemplify the virtues of the CHICAGO method, we conduct a VaR backtesting exercise for a time series of Dow Jones 30 returns (3M, Alcoa, Altria, American Express, American International Group, AT&T, Boeing, Caterpillar, Citigroup, Coca-Cola, DuPont de Nemours, Exxon Mobil, General Electric, General Motors, Hewlett-Packard, Home Depot, Honeywell, Intel, IBM, Johnson & Johnson, J. P. Morgan Chase, McDonald’s, Merck, Microsoft, Pfizer, Procter & Gamble, United Technologies, Verizon, Wal-Mart, Walt Disney), which we split alphabetically into 3 \( d = 10 \)-dimensional samples. Each sample consists of daily returns for the period 9/23/92 to 3/23/07, resulting in a sample size of \( T = 3,209 \).

Univariate descriptive statistics for Samples 1, 2, and 3 are given in Tables 1, 3, and 5, respectively. Clearly, all series in the sample exhibit moderate skewness, and an amount of kurtosis incompatible with the assumption of Gaussianity. Consequently, the Jarque-Bera test rejects the null hypothesis of Gaussianity for all series; the critical value for this test at the 1% level is 9.21, and the smallest value observed in our samples is 1.060. There is also strong evidence of conditional heteroskedasticity in the series, as is indicated by Ljung-Box \( Q \) tests (with 10 lags) on the squared residuals. The critical value of this test at the 1% level is 23.21, which is exceeded by all but two series in our sample. The correlation matrices for Samples 1, 2, and 3 are given in Tables 2, 4, and 6, respectively. Within all samples, the correlation between assets is fairly high, with average correlations ranging from 0.232 to 0.316.

We apply our CHICAGO method to each of the 3 \( d = 10 \)-dimensional time series. Using a moving estimation window of 1,000 observations, we compute 1-day ahead VaR forecasts for an equally weighted portfolio \( (b_i = 1/10) \) of the stocks in each sample. The results for the NIG and hyperbolic distributions are shown in the left and right columns of Figure 8, respectively, for Samples 1, 2, and 3, from top to bottom. VaR violations at the 1% and 5% levels are indicated by marks on the bottom and top axes, respectively. We will restrict our attention to the results for Sample 1, as those for the other samples are largely similar. At the nominal 1% (5%) level, the empirical levels of the VaR forecasts obtained from the CHICAGO method are 1.13% (4.48%) for the NIG, and 1.04% (3.98%) for the hyperbolic distribution, i.e., for this (and the other two) samples, the hyperbolic distribution yields slightly more conservative VaR forecasts. The Kupiec
<table>
<thead>
<tr>
<th></th>
<th>MMM</th>
<th>AA</th>
<th>MO</th>
<th>AXP</th>
<th>AIG</th>
<th>A</th>
<th>BA</th>
<th>CAT</th>
<th>C</th>
<th>KO</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMM</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>0.339</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MO</td>
<td>0.203</td>
<td>0.145</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AXP</td>
<td>0.364</td>
<td>0.335</td>
<td>0.176</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIG</td>
<td>0.365</td>
<td>0.304</td>
<td>0.216</td>
<td>0.541</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.208</td>
<td>0.210</td>
<td>0.193</td>
<td>0.325</td>
<td>0.317</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>0.294</td>
<td>0.292</td>
<td>0.136</td>
<td>0.328</td>
<td>0.300</td>
<td>0.180</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAT</td>
<td>0.380</td>
<td>0.423</td>
<td>0.173</td>
<td>0.362</td>
<td>0.334</td>
<td>0.197</td>
<td>0.303</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.358</td>
<td>0.351</td>
<td>0.189</td>
<td>0.649</td>
<td>0.557</td>
<td>0.308</td>
<td>0.325</td>
<td>0.374</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>KO</td>
<td>0.285</td>
<td>0.162</td>
<td>0.224</td>
<td>0.317</td>
<td>0.329</td>
<td>0.255</td>
<td>0.227</td>
<td>0.218</td>
<td>0.307</td>
<td>1.000</td>
</tr>
</tbody>
</table>

**Table 2:** Correlation Matrix, Sample 1. Average: 0.297, maximum: 0.649, minimum: 0.136.

<table>
<thead>
<tr>
<th></th>
<th>DD</th>
<th>XOM</th>
<th>GE</th>
<th>GM</th>
<th>HPQ</th>
<th>HD</th>
<th>HON</th>
<th>INTC</th>
<th>IBM</th>
<th>JNJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.020</td>
<td>0.050</td>
<td>0.047</td>
<td>0.001</td>
<td>0.045</td>
<td>0.041</td>
<td>0.033</td>
<td>0.033</td>
<td>0.052</td>
<td>0.047</td>
</tr>
<tr>
<td>Std</td>
<td>1.771</td>
<td>1.473</td>
<td>1.702</td>
<td>2.108</td>
<td>2.657</td>
<td>2.181</td>
<td>2.154</td>
<td>2.154</td>
<td>1.982</td>
<td>1.473</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.001</td>
<td>0.022</td>
<td>0.046</td>
<td>0.076</td>
<td>−0.086</td>
<td>−1.208</td>
<td>−0.301</td>
<td>−0.301</td>
<td>0.011</td>
<td>−0.425</td>
</tr>
<tr>
<td>LB Stat</td>
<td>314</td>
<td>510</td>
<td>662</td>
<td>167</td>
<td>87</td>
<td>40</td>
<td>188</td>
<td>188</td>
<td>161</td>
<td>218</td>
</tr>
</tbody>
</table>

**Table 3:** Descriptive Statistics, Sample 2.

test accepts the null hypothesis of correct coverage with a \( p \)-value of 0.54 (0.26) for the NIG; for the hyperbolic model, the \( p \)-value is 0.85 (0.02), confirming that the NIG model has superior performance for this data set.
<table>
<thead>
<tr>
<th></th>
<th>DD</th>
<th>XOM</th>
<th>GE</th>
<th>GM</th>
<th>HPQ</th>
<th>HD</th>
<th>HON</th>
<th>INTC</th>
<th>IBM</th>
<th>JNJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XOM</td>
<td>0.30</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>0.41</td>
<td>0.31</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM</td>
<td>0.34</td>
<td>0.18</td>
<td>0.37</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPQ</td>
<td>0.23</td>
<td>0.16</td>
<td>0.36</td>
<td>0.26</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HD</td>
<td>0.31</td>
<td>0.22</td>
<td>0.46</td>
<td>0.34</td>
<td>0.27</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HON</td>
<td>0.40</td>
<td>0.29</td>
<td>0.45</td>
<td>0.32</td>
<td>0.29</td>
<td>0.32</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTC</td>
<td>0.40</td>
<td>0.29</td>
<td>0.45</td>
<td>0.32</td>
<td>0.29</td>
<td>0.32</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBM</td>
<td>0.24</td>
<td>0.13</td>
<td>0.42</td>
<td>0.26</td>
<td>0.44</td>
<td>0.29</td>
<td>0.29</td>
<td>0.29</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>JNJ</td>
<td>0.34</td>
<td>0.18</td>
<td>0.36</td>
<td>0.18</td>
<td>0.12</td>
<td>0.22</td>
<td>0.23</td>
<td>0.20</td>
<td>0.20</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Table 4:** Correlation Matrix, Sample 2. Average: 0.316, maximum: 0.462, minimum: 0.122.

<table>
<thead>
<tr>
<th></th>
<th>JPM</th>
<th>MCD</th>
<th>MRK</th>
<th>MSFT</th>
<th>PFE</th>
<th>PG</th>
<th>UTX</th>
<th>VZ</th>
<th>WMT</th>
<th>DIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.06</td>
<td>0.04</td>
<td>0.04</td>
<td>0.06</td>
<td>0.01</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>Std</td>
<td>1.12</td>
<td>1.72</td>
<td>1.81</td>
<td>2.16</td>
<td>1.87</td>
<td>1.66</td>
<td>1.80</td>
<td>1.76</td>
<td>1.89</td>
<td>2.02</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.42</td>
<td>-0.05</td>
<td>-1.69</td>
<td>-0.15</td>
<td>-0.22</td>
<td>-3.85</td>
<td>-1.99</td>
<td>0.08</td>
<td>0.02</td>
<td>-0.25</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>7.72</td>
<td>7.81</td>
<td>33.28</td>
<td>8.85</td>
<td>6.18</td>
<td>88.81</td>
<td>40.60</td>
<td>7.24</td>
<td>7.04</td>
<td>11.49</td>
</tr>
<tr>
<td>JB Stat</td>
<td>3.06</td>
<td>3.10</td>
<td>123.95</td>
<td>4.58</td>
<td>1.37</td>
<td>991.29</td>
<td>190.97</td>
<td>2.45</td>
<td>2.45</td>
<td>9.67</td>
</tr>
<tr>
<td>LB Q10 Stat</td>
<td>695</td>
<td>198</td>
<td>4</td>
<td>240</td>
<td>304</td>
<td>16</td>
<td>96</td>
<td>421</td>
<td>340</td>
<td>152</td>
</tr>
</tbody>
</table>

**Table 5:** Descriptive Statistics, Sample 3.

<table>
<thead>
<tr>
<th></th>
<th>JPM</th>
<th>MCD</th>
<th>MRK</th>
<th>MSFT</th>
<th>PFE</th>
<th>PG</th>
<th>UTX</th>
<th>VZ</th>
<th>WMT</th>
<th>DIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPM</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCD</td>
<td>0.34</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRK</td>
<td>0.12</td>
<td>0.22</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSFT</td>
<td>0.15</td>
<td>0.17</td>
<td>0.20</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PFE</td>
<td>0.14</td>
<td>0.22</td>
<td>0.52</td>
<td>0.25</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PG</td>
<td>0.02</td>
<td>0.26</td>
<td>0.30</td>
<td>0.13</td>
<td>0.31</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTX</td>
<td>0.24</td>
<td>0.25</td>
<td>0.22</td>
<td>0.30</td>
<td>0.26</td>
<td>0.24</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VZ</td>
<td>0.12</td>
<td>0.18</td>
<td>0.21</td>
<td>0.23</td>
<td>0.29</td>
<td>0.19</td>
<td>0.23</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMT</td>
<td>0.12</td>
<td>0.27</td>
<td>0.26</td>
<td>0.29</td>
<td>0.28</td>
<td>0.26</td>
<td>0.30</td>
<td>0.26</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>DIS</td>
<td>0.23</td>
<td>0.22</td>
<td>0.20</td>
<td>0.30</td>
<td>0.23</td>
<td>0.15</td>
<td>0.34</td>
<td>0.25</td>
<td>0.28</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Table 6:** Correlation Matrix, Sample 3. Average: 0.232, maximum: 0.526, minimum: 0.029.
Figure 8: Returns, 1-day-ahead 1% and 5% VaR forecasts, and VaR violations, using NIG (left column) and hyperbolic (right column) innovations, for Samples 1, 2, and 3, from top to bottom. Empirical VaR (Kupiec p-value): Sample 1: 1.13% (0.54) and 4.48% (0.26) for NIG, 1.04% (0.85) and 3.98% (0.02) for hyperbolic. Sample 2: 0.95% (0.81) and 4.21% (0.08) for NIG, 0.81% (0.37) and 4.07% (0.04) for hyperbolic. Sample 3: 0.72% (0.17) and 4.16% (0.06) for NIG, 0.72% (0.17) and 4.21% (0.08) for hyperbolic.
6 Conclusions

The CHICAGO method developed in this paper is a fast, numerically reliable, and robust method for the estimation of Generalized Orthogonal GARCH models. The approach is completely modular: the estimation of the multivariate relationship between assets is entirely independent of the specification of the component dynamics and the distribution of innovations, thus not only simplifying model estimation, but also allowing empirical researchers to utilize whatever univariate volatility dynamics appear most adequate for the data at hand.

By virtue of the saddlepoint approximation to convolutions of independent generalized hyperbolic variates developed herein, accurate VaR forecasts for high-dimensional portfolios of assets can be obtained in a computationally efficient manner for different portfolio weights, thus allowing the procedure to be used in, e.g., real-time portfolio optimization.

A promising avenue for future research presents itself upon recognizing that the pattern of interdependence, and, hence, the mixing matrix in the GO-GARCH model, may not be stable over time. In this paper, we have addressed this possible shortcoming of the method by using a moving estimation window in our backtesting exercise. A more elaborate approach would consist in either incorporating a potential time dependence into the model itself, or in using a weighted estimator for the mixing matrix, thus giving more importance to recent observations.
A Saddlepoint Approximations

If random variable $X$ possesses a moment generating function, or, in short, mgf, given by $M_X(t) = \mathbb{E}[\exp(tX)]$, then a highly accurate approximation to the density is given by

$$
\hat{f}_X(x) = \frac{1}{\sqrt{2\pi K_X''(\hat{t})}} \exp\{K_X(\hat{t}) - x\hat{t}\}, \quad x = K_X'(\hat{t}).
$$

(13)

This is referred to as the (first order) saddlepoint density approximation to $f$, abbreviated SPA, where $\hat{t} = \hat{t}(x)$ is the solution to the saddlepoint equation and is referred to as the saddlepoint at $x$. This method of approximation is attributed to Daniels (1954), though, via its similarity to the Laplace method of approximation, there is evidence that it can be traced back to Georg Bernhard Riemann (see the discussion and references in Kass, 1988, p. 235; and also Tierney, 1988). In general, $\hat{f}_X(x)$ will not integrate to one, although it will usually be close. It is, however, easy to renormalize it, by numeric integration.

The approximate cumulative distribution function (cdf) of $X$ could be obtained by numerically integrating $\hat{f}_X$. However, in a celebrated paper, Lugannani and Rice (1980) derived a simple expression for the SPA to the cdf, given for continuous r.v.s by

$$
\hat{F}_X(x) = \Pr(X < x) = \Phi(\hat{w}) + \phi(\hat{w}) \left\{ \frac{1}{\hat{w}} - \frac{1}{\hat{u}} \right\}, \quad x \neq \mathbb{E}[X],
$$

(14)

where $\Phi$ and $\phi$ are the cdf and pdf of the standard normal distribution, respectively,

$$
\hat{w} = \text{sgn}(\hat{t}) \sqrt{2\hat{t}x - 2K(X(\hat{t}))} \quad \text{and} \quad \hat{u} = \hat{t} \sqrt{K''(\hat{t})}.
$$

In contrast to the density approximation (13), $\hat{F}_X(x)$ is such that, as $x$ extends to the left and right edges of its support, $\hat{F}_X(x)$ tends to zero and one, respectively, so that renormalization is not necessary. Using above definition of $\hat{w}$, the density approximation (13) can also be written

$$
\hat{f}_X(x) = \phi(\hat{w}) / \sqrt{K''(\hat{t})}.
$$

(15)

An approximation to the $q\%$ quantile of $X$ can be obtained by numerically solving

$$
\hat{F}_X(x_q) = q, \quad q \in (0, 1).
$$

(16)

However, a direct application of (16) would result in a nested root search, because for every evaluation of the distribution function, the saddlepoint equation has to be solved. This can be circumvented by noting that the saddlepoint equation defines a bijection between $\hat{t}$ and $x$. Hence, $\hat{F}_X(x)$ can also be viewed as a function of $\hat{t}$, and it is equivalent to solve (16) in terms of $\hat{t}$, yielding, say, $\hat{t}_{x_q}$. The quantile approximation is then given by $x_q = K_X'(\hat{t}_{x_q})$. The computational advantages stem from the facts that firstly, $x = K_X'(\hat{t})$ is in closed-form as a function of $\hat{t}$, and secondly, in many applications the convergence strip of the cgf is bounded, thus providing a bracketing interval for $\hat{t}_{x_q}$. Similarly, when integrating over the saddlepoint approximation, such
as when calculating expected shortfall from (11) or (12), one may change variables and integrate over \( \hat{t} \) instead of \( x \).

Lastly, a direct saddlepoint approximation for Expected Shortfall is also available. Martin (2006) shows that

\[
\mathbb{E}[X | X \leq c] \approx \mathbb{E}[X] - \frac{c - \mathbb{E}[X]}{\hat{t}} \frac{\hat{f}_X(c)}{\hat{F}_X(c)}
\]

(17)

Far more information on the SPA can be found in Reid (1988), Jensen (1995) and Butler (2007).

**B Application to the GHyp Distribution**

We will require derivatives of the Bessel function. It is straightforward to verify that

\[-2K_\nu'(x) = K_{\nu-1}(x) + K_{\nu+1}(x), \quad \nu \in \mathbb{R}, \quad x \in \mathbb{R}_{>0}.\]

(18)

First consider the \( d = 1 \) case. Let \( X \sim \text{GHyp}(\lambda, \omega, \rho, \sigma, \mu) \) with density (6). With \( \beta \equiv \omega \delta^{-1} \rho(1 - \rho^2)^{-1/2} \) and \( \psi \equiv \omega^2 \delta^{-2} \), the moment generating function of \( X \) is given by

\[
\mathbb{M}_X(t) = e^{\mu t} \frac{K_\lambda \left( \omega Q \right)}{K_\lambda (\omega)} \frac{\left( 1 - \frac{2\beta t + t^2}{\psi} \right)^{-\lambda/2}}{\left( 1 - \frac{2\beta t + t^2}{\psi} \right)^{\lambda/2}},
\]

(19)

with convergence strip given by those values of \( t \) such that

\[
1 - \frac{2\beta t + t^2}{\psi} > 0 \quad \Leftrightarrow \quad t^2 + 2\beta t - \psi < 0.
\]

The solutions of \( t^2 + 2\beta t - \psi = 0 \) are \( t = -\beta \pm \sqrt{\beta^2 + \psi} \), so that the convergence strip is

\[-\beta - \sqrt{\beta^2 + \psi} < t < -\beta + \sqrt{\beta^2 + \psi}.\]

(20)

The cumulant generating function, or cgf, is defined as \( \mathbb{K}_X(t) = \ln \mathbb{M}_X(t) \). The cgf corresponding to (19) is

\[
\mathbb{K}_X(t) = \mu t + \ln K_\lambda (\omega Q) - \ln K_\lambda (\omega) - \lambda \ln (Q),
\]

(21)

where \( Q = Q(t) := \sqrt{1 - (2\beta t + t^2)/\psi} \). It is easy to see that

\[
\frac{dQ(t)}{dt} = \frac{1}{2} \left( 1 - \frac{2\beta t + t^2}{\psi} \right)^{-1/2} \left( -\frac{2\beta t + 2t}{\psi} \right) = \frac{-\beta + t}{Q \psi},
\]

so, via (18) and some simplification,

\[
\mathbb{K}'_X(t) = \mu + \frac{\beta + t}{Q \psi} \left( \frac{\omega}{2} \frac{K_{\lambda-1} (\omega Q)}{K_\lambda (\omega Q)} + \frac{\lambda}{Q} \right).
\]

Numerically solving \( \mathbb{K}'_X(t) = x \) in the range (20) then gives the saddlepoint \( \hat{t} \). It can be proven that there exists one, and only one, solution to the saddlepoint equation for \( t \) restricted to
the convergence strip of the mgf, so that this root search is well–defined and computationally straightforward.

Tedious but straightforward algebra then shows that

$$
\mathbb{K}'(t) = \frac{\beta + t}{Q^2\psi} \mathbb{P}_1 + \left[ \frac{\omega}{2} \frac{K_{\lambda-1}(\omega Q) + K_{\lambda+1}(\omega Q)}{K_{\lambda}(\omega Q)} + \frac{\lambda}{Q} \right] \times \left[ \frac{1}{Q^2\psi} \left( 1 + \frac{(\beta + t)^2}{Q^2\psi} \right) \right],
$$

where

$$
P_1 = \frac{\omega}{2} P_2 + \frac{\lambda(\beta + t)}{Q^4\psi}
$$

and

$$
K_{\lambda}^2(\omega Q) P_2 = K_{\lambda}(\omega Q) \times \left[ \frac{\omega}{2} \left( \frac{\beta + t}{Q^2}\psi \right) (K_{\lambda-2}(\omega Q) + 2K_{\lambda}(\omega Q) + K_{\lambda+2}(\omega Q)) \right]
$$

$$
- [K_{\lambda-1}(\omega Q) + K_{\lambda+1}(\omega Q)] \times \left[ \frac{\omega}{2} (K_{\lambda-1}(\omega Q) + K_{\lambda+1}(\omega Q)) \left( \frac{\beta + t}{Q^2}\psi \right) \right].
$$

With these expressions, the first–order SPA to the pdf and cdf can be computed from (13) and (14), respectively.

The application to the \(d > 1\) case is now straightforward: Because of independence, the cgf of

$$
S = \sum_{i=1}^{d} a_i X_i, \quad a_i \neq 0,
$$

is just \(\mathbb{K}_S(t) = \sum_{i=1}^{d} \mathbb{K}_{X_i}(a_i t)\), where the cgf of each \(X_i\) is given in (21). The singularities of the cgf lie on both sides of the origin at the points \((-\beta_i \pm \sqrt{\beta_i^2 + \psi_i})/a_i\). In the special case of the NIG distribution, matters simplify considerably. Specifically,

$$
\mathbb{K}_X(t) = \mu t - \omega Q + \omega, \quad \mathbb{K}'_X(t) = \mu + \omega(\beta + t)/Q^2\psi, \quad \mathbb{K}''_X = \frac{\omega}{Q^2\psi} + \frac{\omega(\beta + t)^2}{Q^4\psi^2},
$$

and the saddlepoint is now given in closed form as

$$
\hat{t} = z \frac{\bar{\alpha}}{\bar{y}^2} - \beta,
$$

where \(z, \bar{y} \) and \(\bar{\alpha} \) are as in (6). The other quantities entering and (14) and (15) simplify to

$$
\mathbb{K}'_X(\hat{t}) = \bar{y}^2 \delta^2 \bar{\alpha} \quad \text{and} \quad \hat{w} = \text{sgn}(\hat{t}) \sqrt{2(\bar{y} \bar{\alpha} - z \rho \bar{\alpha} - \omega)}.
$$

For completeness, we also give the simplified expressions for the quantities \(\hat{\kappa}_i \equiv \mathbb{K}_X^{(i)}(\hat{t})\mathbb{K}''_X(\hat{t})^{-i/2}, \quad i \in \{3, 4\}\), required for evaluating the second-order equivalents of (13) and (14):

$$
\hat{\kappa}_3 = 3 - \frac{z}{\sqrt{\bar{y} \bar{\alpha}}} \quad \text{and} \quad \hat{\kappa}_4 = 3 \frac{1 + 5z^2}{\bar{y} \bar{\alpha}}.
$$
References


Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10:626–634. 4


23