CHICAGO: A Fast and Accurate Method for Portfolio Risk Calculation

Simon Broda, Marc Paolella

University of Zürich

May 2007
**Motivation**

- **Aim:** Forecast the Value at Risk of a portfolio of $d$ assets, i.e., the quantiles of $R_t = b'r_t$, $r_t = (r_{t,1}, \ldots, r_{t,d})'$.
- **Problem:** Assets not independent, GARCH effects, heavy tails.
- **Naive approach:** Fit a univariate GARCH-type model to the time series of $R_t$.
- **Problem:** need to re-estimate the model for each $b$.
- **Solution:** Multivariate (M-) GARCH models.
- **But:** Many parameters ("curse of dimensionality"), computationally demanding.
- **We show how Conditional Heteroscedasticity-based Independent Component Analysis can be used to estimate a **GO**-GARCH model.**
- **Using Generalized Hyperbolic innovations and a saddlepoint approximation, we obtain fast and accurate portfolio VaR forecasts.**
Outline

1. The GO-GARCH Model
2. Two-Step Estimation
3. Factor Specification
4. Computing the CDF of Convolutions of Ghyp
5. Application to Stock Returns
6. Conclusions and Future Research
Outline

1. The GO-GARCH Model
2. Two-Step Estimation
3. Factor Specification
4. Computing the CDF of Convolutions of Ghyp
5. Application to Stock Returns
6. Conclusions and Future Research
MGARCH-Models

- General Setup:
  \[ r_t = \mu_t + u_t \]
  \[ u_t = \Sigma_t^{1/2} \epsilon_t \]
  \[ \epsilon_t \sim \text{i.i.d.}(0, I) \]

- Most general rendition: VEC(p,q) (Bollerslev, Engle and Wooldridge, 1988):
  \[ \text{vech} \Sigma_t = c + \sum_{i=1}^{p} B \text{vech} \Sigma_{t-p} + \sum_{i=1}^{q} A \text{vech} u_{t-q} u'_{t-q} \]
  \[ \mathcal{O}(d^4) \text{ parameters} (!) \]
The GO-Garch Model

- Need to imply restrictions to ensure $\Sigma_t > 0$, and reduce number of parameters $\Rightarrow$ BEKK (Engel and Kroner, 1995), CCC (Bollerslev, 1990), DCC (Engle, 2001), ... 
- The GO-GARCH model (van der Weide, 2002): innovations vector modelled as linear combination of $d$ unobserved factors $f_t$:

$$u_t = A f_t,$$

for some constant, invertible mixing matrix $A$.
- The unobserved factors are assumed to be independent of each other, and to have unit variance (identifying restriction).
Assuming a GARCH(1,1) process for each factor \( \{f_{it}\} \), this is similar to the Full Factor Model of Vrontos et. al. (2003), and a special case of the Factor-GARCH (Engle et. al., 1990) and BEKK (Engle and Kroner, 1995) models, with

\[
\Sigma_t = A \Omega A' + \sum_{k=1}^{d} \alpha_k \lambda_k w_k' \Sigma_{t-1} w_k \lambda'_k + \sum_{i=1}^{d} \beta_k \lambda_k w_k' u_{t-1} u_{t-1}' w_k \lambda'_k,
\]

\[
\Omega = \sum_{k=1}^{d} (1 - \alpha_k - \beta_k) e_k e_k' , \lambda_k = A e_k , w_k = A^{-T} e_k , \text{ and } e_k \text{ is a } d \times 1 \text{ vector with } k\text{th element 1 and zeros everywhere else.}
\]

- Provides a reasonable tradeoff between covariance flexibility and parsimony.
- Factors need not have a simple GARCH structure.
A polar decomposition uniquely factorizes the mixing matrix $\mathbf{A}$ into a symmetric positive definite matrix $\Sigma^{1/2}$ and a rotation matrix $\mathbf{U}$:

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

where

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

is the unconditional covariance matrix.

The $d(d + 1)/2$ free parameters in $\Sigma^{1/2}$ can be consistently estimated from the (unconditional) sample covariance.

Estimating the $d(d - 1)/2$ parameters in $\mathbf{U}$ requires conditional information, because any orthogonal matrix $\mathbf{U}^*$ gives rise to the same unconditional covariance matrix $\Sigma$ (because $\mathbf{U}^* \mathbf{U}^* = I$).
As \( \mathbf{U} \) is a rotation matrix, it can be decomposed as the product of \( \binom{d}{2} \) basic rotation matrices \( \mathbf{R}_i(\theta_i) \), where each \( \mathbf{R}_i \) is a rotation of angle \( \theta_i \) in the plane spanned by one pair of axes in \( \mathbb{R}^d \).

For example, in the \( d = 2 \) case,

\[
\mathbf{U} = \begin{bmatrix}
\cos \theta & \sin \theta \\
-\sin \theta & \cos \theta
\end{bmatrix}.
\]

Thus, \( \mathbf{U} \) can be fully parameterized in terms of the Euler angles \( \theta_i \).
Related Literature

- Hyvärinen, Karhunen and Oja (2001): monograph on ICA
- Boswijk and van der Weide (2006): 2-Step estimation of GO-GARCH model
Outline

1 The GO-GARCH Model
2 Two-Step Estimation
3 Factor Specification
4 Computing the CDF of Convolutions of Ghyp
5 Application to Stock Returns
6 Conclusions and Future Research
Estimation

- Estimation by ML is difficult, especially under non-normality.
- We would like to have a two-step approach (as in, e.g., the CCC and DCC models): Estimate $A = \Sigma^{1/2}U$ first, then fit $d$ univariate GARCH models.
- $\Sigma^{1/2}$ can be estimated from unconditional information. Thus, consider the demeaned and whitened data $z$:

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

where $\hat{u}_t \equiv (r_t - \hat{\mu}_t)$ and $\hat{\Sigma} \equiv (\hat{u}_t\hat{u}_t')/T$.

- The goal is then to estimate $U$. We propose to use Independent Component Analysis.
ICA is a relatively young technique used in signal processing.

Assume we observe a $d$ dimensional random vector $\mathbf{u}_t$.

The $u_{it}$ are linear combinations of $d$ independent random variables $f_{it}$:

$$\mathbf{u}_t = \mathbf{A}f_t.$$

Only $\mathbf{u}_t$ is observed. Aim: estimate $\mathbf{A}$, i.e., find a matrix $\mathbf{A}$ such that the linear combinations $\mathbf{y}_t \equiv \mathbf{A}^{-1}\mathbf{u}_t$ are independent.

Naive approach of taking $\mathbf{A} = \mathbb{E}[(\mathbf{u}_t - \mathbb{E}[\mathbf{u}_t])(\mathbf{u}_t - \mathbb{E}[\mathbf{u}_t])']^{1/2}$ yields uncorrelated components, but yields independent components only up to an orthogonal transformation: With $\mathbf{U}^*$ orthogonal,

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

i.e., $\mathbf{U}^{ast}\mathbf{y}$ is also uncorrelated $\Rightarrow$ have to use more information

Whitening is still useful, need to consider only orthogonal matrices.
Using Additional Information

- Which additional information to use depends on the problem at hand.
  - Example: If the data are non-gaussian, that information can be exploited.
  - Central Limit Theorem: Sums of independent r.v.s are "more gaussian".
  - The ICs are those linear combinations of the data that are "least gaussian", as measured by, e.g., excess kurtosis.

- For financial data, it suggests itself to use the GARCH-effects as extra information: the sum of two (or more) "garchy" series will typically be less "garchy". Thus, the independent components are those linear combinations of the data that have the most pronounced GARCH-effects, as measured by, e.g., autocorrelation of the squared series.
ICA using Variance Nonstationarity

- The algorithm of Hyvärinen, Karhunen and Oja (2001, p.349) achieves this with cubic convergence. Given prewhitened data $z_t$, the algorithm starts with $U_n = I$ and iterates

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

$$U_{n+1} = (U_{\text{temp}} U_{\text{temp}}')^{-1/2} U_{\text{temp}}$$

where $z = [z_2, \ldots, z_T]$, $z_- = [z_1, \ldots, z_{T-1}]$, $\bar{C} = (zz'_- + z_- z')/(2T)$, $D_n = \text{diag} \left( \text{vecd} \left( U'_n \bar{C} U_n \right) \right)$, 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 $U'_{n+1} U_n$, 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.
ICA Example 1
ICA Example 2
Boswijk and van der Weide (2006) derive an estimator $\hat{U}_{BW}$ as the eigenvector matrix of $\hat{B}$, where

$$\hat{B} = \arg\min_B B_0 \sum_{t=1}^T \frac{1}{T} \text{tr} \left( [z_t z_t' - I_d - B (z_{t-1} z_{t-1}' - I_d) B]^2 \right),$$

which has to be solved numerically.

- We compare the performance of the two methods with a small simulation study.
- We use $d = 2$, so only one parameter to estimate (rotation angle $\theta$).
- $T = 800$, data are Normal-GARCH(1,1), with $a_1 = 0.09$, $b_1 = 0.9$, $a_2 = 0.04$, $b_2 = 0.95$, $\Sigma = I$. 

Simon Broda, Marc Paolella
CHICAGO: A Fast and Accurate Method for Portfolio Risk Calculation
Comparison of Estimators

Figure: Performance Comparison of MLE, ICA, and Boswijde and van der Weide (BW). Average computation time for each of the 1000 replications: 8.87 sec for MLE, 1.68 for BW, and 0.03 sec for ICA, 297 and 56 times faster, respectively.
Outline

1. The GO-GARCH Model
2. Two-Step Estimation
3. Factor Specification
4. Computing the CDF of Convolutions of Ghyp
5. Application to Stock Returns
6. Conclusions and Future Research
*The Model*

- Our aim is to obtain VaR forecasts for $R_{t+1} = b' Af_{t+1}$.
- Full specification of the model requires us to specify:
  1. the univariate factor dynamics and
  2. the innovations distribution.
- Our two-step procedure allows us to keep these separate from the covariance specification.
**Specification of IC Dynamics**

- We model the $i$th IC as $f_{i,t} = \sigma_{i,t} Z_{i,t}$.
- To capture the evolution of the scale parameters $\sigma_{i,t}$, we use the A-PARCH model proposed by Ding, Granger and Engle (1993), given by

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

where $z_{i,t} = \epsilon_{i,t} \sigma_{i,t}$, $c_{i,j} > 0$, $d_{i,j} \geq 0$, $\delta_i > 0$, and $|\gamma_{i,j}| < 1$.

- We set $r_i = s_i = 1$ and $\delta_i = 1$, which was shown in Mittnik and Paolella (2000) to produce very accurate VaR forecasts for fx data.
We model the innovations as Generalized Hyperbolic: 
\[ Z_{i,t} \sim \text{GHyp}(\lambda_i, \omega_i, \rho_i, \sigma_i, \mu_i), \] i.e., with \( \lambda \in \mathbb{R}, \omega > 0, |\rho| < 1, \mu \in \mathbb{R} \) and \( \delta > 0 \), their density is 

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

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

\[
K_{\nu}(x) = \frac{1}{2} \int_0^\infty t^{\nu-1} e^{-\frac{1}{2}x(t+t^{-1})} \, dt.
\]
Moments of GHyp

- The expected value and variance of the GH distribution are given by

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

and

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

where \( k_1(\omega) \equiv K_{\lambda+1}(\omega)/K_\lambda(\omega) \) and \( k_2(\omega) \equiv [K_\lambda(\omega)K_{\lambda+2}(\omega) - K_{\lambda+1}(\omega)^2]/K_\lambda(\omega)^2 \).

- We standardize the generalized hyperbolic to have zero mean and unit variance and denote the standardized distribution as SGH:

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

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

- GHyp is an extremely flexible distribution: It nests, among others, the Laplace, Student’s $t$, Normal, …
- It possesses semi-heavy tails, i.e., its log-density decays roughly linearly, a common feature in financial data.
- $\delta$ and $\mu$ are location and scale parameters, respectively.
- $\rho$ is a skewness parameter.
- $\omega$ dictates the tail-heaviness.
- The shape parameter $\lambda$ is notoriously difficult to estimate.
- Solution: consider special cases
  1. $\lambda = -\frac{1}{2}$ (Normal Inverse Gaussian (NIG))
  2. $\lambda = 1$ (Hyperbolic)
The cumulant generating function of $X$ is

$$K_X (t) = \mu t + \ln K_\lambda (\omega Q) - \ln K_\lambda (\omega) - \lambda \ln (Q),$$

where $Q = Q (t) \equiv \sqrt{1 - (2\beta t + t^2)/\psi}$, $\beta \equiv \omega \delta^{-1} \rho (1 - \rho^2)^{-1/2}$, and $\psi \equiv \omega^2 \delta^{-2}$.

The convergence strip $\mathcal{K}$ of $K_X$ is

$$-\beta - \sqrt{\beta^2 + \psi} < t < -\beta + \sqrt{\beta^2 + \psi}.$$
Outline

1. The GO-GARCH Model
2. Two-Step Estimation
3. Factor Specification
4. Computing the CDF of Convolutions of Ghyp
5. Application to Stock Returns
6. Conclusions and Future Research
In order to compute VaR, we require the (conditional) inverse cdf of $R_{t+1} = \mathbf{b}' \mathbf{Af}_{t+1}$, i.e., a weighted sum of independent GHyp random variables.

Simulation is one straightforward possibility, but because tail values are required, an extremely large number of replications will be required to obtain reasonable and reliable accuracy.

Numerical inversion of the cf is another option, but the requisite integrand is oscillatory, rendering numerical quadrature difficult.

We therefore propose use of a Saddlepoint Approximation.
Saddlepoint Approximations: Introduction

- The SPA can be thought of as
  1. approximate inversion of the characteristic function, but without requiring integration
  2. 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.
The saddlepoint approximation to the density is given by

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

where $\hat{t} = \hat{t}(x)$ is the solution to the saddlepoint equation and is referred to as the saddlepoint at $x$.

 Remarks

1. The saddlepoint equation must be solved (numerically) anew for each value of $x$.
2. The approximate pdf will not, in general, integrate to one, but can be renormalized.
3. SPA requires derivatives of the cgf $\Rightarrow$ see above.
CDF Saddlepoint Approximation

- The approximate cdf of $X$ could be obtained by numerically integrating $\hat{f}$.

- 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

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

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

$$\hat{w} = \text{sgn}(\hat{t}) \sqrt{2\hat{t}x - 2K_X(\hat{t})} \quad \text{and} \quad \hat{u} = \hat{t} \sqrt{K''_X(\hat{t})}.$$
Application to the GHyp

- Application to Ghyp is straightforward; requires only derivatives of the cgf (see above).
- In the special case of the NIG distribution, the SPA is in closed form!
- Specifically,

\[
\begin{align*}
K_X(t) &= \mu t - \omega Q + \omega, \\
K'_X(t) &= \mu + \omega \frac{\beta + t}{Q\psi}, \\
K''_X &= \frac{\omega}{Q\psi} + \frac{\omega(\beta + t)^2}{Q^3\psi^2},
\end{align*}
\]

and the saddlepoint is given by

\[
\hat{t} = (x - \mu) \frac{\bar{\alpha}}{\bar{y}\bar{\delta}^2} - \beta.
\]

- The \(d > 1\) case is straightforward: From independence, the cgf of \(S = \sum_{i=1}^{n} a_i X_i, \ a_i \neq 0\), is

\[
K_S(t) = \sum_{i=1}^{n} K_{X_i}(a_i t).
\]
Illustration of Accuracy

Figure: Left: The exact pdf (solid) and renormalized spa (dashed) for a GHyp with $\lambda = 3$, $\omega = \sqrt{8}$, $\rho = -1/3$, $\sigma = 1$ and $\mu = 0$. Right: The relative percentage error (RPE) of the cdf saddlepoint approximation.
Illustration of Accuracy (2)

Figure: Quantiles of $X_1 + X_2$, $X_i \sim \text{GHyp}(\lambda, \omega_i, \rho_i, \delta_i, \mu_i)$, $\lambda = -0.5$, $\omega_1 = 1.93$, $\omega_2 = 0.9$, $\rho_1 = 0.02$, $\rho_2 = 0.06$, $\delta_1 = 1.22$, $\delta_1 = 0.81$, $\mu_1 = -0.06$, $\mu_2 = -0.03$. Exact numbers (solid) took 80s to compute, SPA (dotted) 1s.
Outline

1. The GO-GARCH Model
2. Two-Step Estimation
3. Factor Specification
4. Computing the CDF of Convolutions of Ghyp
5. Application to Stock Returns
6. Conclusions and Future Research
Application to VaR Forecasting (1)

- $d = 10$ dimensional time series of Dow Jones stock returns (3M, Alcoa, Altria, American Express, American International Group, AT&T, Boeing, Caterpillar, Citigroup, Coca-Cola)
- Equally weighted portfolio: $b_i = 1/10$.
- Using a moving estimation window of 1,000 observations, we compute 1-day ahead VaR forecasts.
**Figure:** Returns, 1-day-ahead 1% and 5% VaR forecasts, and VaR violations, using NIG innovations. Empirical VaR: 0.95% and 3.85%
Application to VaR Forecasting (3)

Figure: Kernel density (solid) of filtered residuals and fitted NIG densities (dashes) for two ICs.
Application to VaR Forecasting (4)

Figure: Returns, 1-day-ahead 1% and 5% VaR forecasts, and VaR violations, using hyperbolic innovations. Empirical VaR: 0.91% and 3.40%
Application to VaR Forecasting (5)

Figure: Kernel density (solid) of filtered residuals and fitted HYP densities (dashes) for two ICs.
Outline

1 The GO-GARCH Model
2 Two-Step Estimation
3 Factor Specification
4 Computing the CDF of Convolutions of Ghyp
5 Application to Stock Returns
6 Conclusions and Future Research
Conclusions

- CHICAGO method is capable of producing fast and accurate portfolio VaR forecasts.
- Modular approach: Estimation of factor loadings independent of specification of IC dynamics and distribution of innovations.
- After parameter estimates have been computed, SPA allows VaR forecasts to be evaluated extremely quickly for different portfolio weights, while maintaining outstanding accuracy, and thus allows the procedure to be used in, e.g., real-time portfolio optimization.
Future Research

- Use of Shrinkage / Bayesian estimators for $\mu_t$ and $\Sigma$
- Use of weighted ICA to account for parameter changes / misspecification
- Use of marginal models with better predictive power (mixture models with pseudo long memory, better leverage effect modelling, and time-varying skewness and kurtosis)
- Use of several variations of CHICAGO and other easily estimated models for use with optimal combinations of density forecasts