Risk Management with the Multivariate Generalized Hyperbolic Distribution

Calibrated by the Multi-Cycle EM Algorithm

by: Marcel Holtslag

The question, mentioned by Pagan (1996), is whether the more complex models to capture the nature of the conditional density of returns are better suited compared to simpler but easier to use models. Gradually, the traditional Gaussian distribution to model financial returns has been replaced by several other viable distributions suitable to capture the empirically observed heavy tail behavior, kurtosis and peakedness. This study contributes towards the further development of the effectiveness of the multivariate generalized hyperbolic distribution (MGHyp) when it is used to forecast the possible next day portfolio loss. It differentiates by introducing an asset portfolio model via DCC-MGARCH and tries to reduce the overal calibration time. By reviewing the subclasses: normal inverse gaussian (NIG), multivariate hyperbolic (Hyp) and the 10-dimensional multivariate hyperbolic distribution (KHyp) following from the MGHyp class, a recommendation is made about the overall performance.

DCC-Multivariate GARCH

The underlying portfolio model is based on the DCC(1,1)-MGARCH(1,1) of Engle (1999). It covers the time dependent correlations among the assets and as Laplante et al. (2008) mentions, statistical evidence does not support an even more complicated model if dealing with financial returns. The model assumes log return series from k assets that are conditionally multivariate normal distributed with constant mean μ and covariance matrix H_t. The information set $\mathbb{I}_{t-1}$ consists of all known information until time $t-1$.

$$r_t \mid \mathbb{I}_{t-1} \sim N(\mu, H_t) \quad \text{with} \quad H_t = D_t R_t D_t$$

where D_t is a $k \times k$ diagonal matrix of time varying standard deviations $\sqrt{h_{i,t}}$ specified by k univariate GARCH(1,1) specifications given by:

$$h_{i,t} = \omega_i + \alpha_i (r_{i,t-1} - \mu_i)^2 + \beta_i h_{i,t-1}$$

for i = 1 ... k with sufficient restrictions on parameters $\alpha + \beta < 1$ and non negative variances. The $k \times k$ dynamic correlation matrix, R_t, with ones on the i^{th} diagonal is given by:

$$R_t = diag\{Q_t\}^{-1} \; Q_t \; diag\{Q_t\}^{-1}$$

with Q_t the diagonal transformation matrix of the dynamic correlation matrix.

$$Q_t = (1-\alpha-\beta)\bar{Q} + \alpha \varepsilon_{t-1} \varepsilon'_{t-1} + \beta Q_{t-1}$$

Lastly, let the standardized residuals ε_t be denoted by

$$\varepsilon_t = D_t^{-1}(r_t - \mu)$$

The unconditional covariance matrix $\bar{Q}$ is a $k \times k$ matrix estimated from the standardized residuals ε_t. A design feature of the DCC model makes it possible to estimate it as a two step optimization problem using the (quasi) maximum log-likelihood method. As long as the first step parameter estimates are consistent, the second step parameter estimates are consistent as well assuming reasonable regularity conditions and a continuous function

Marcel Holtslag

Marcel Holtslag obtained his Master's degree in Financial Econometrics at the University of Amsterdam in May 2011 and during his study he has contributed to several VSAE committees. This article is a brief summary based on his Master's thesis written under the supervision of Dr. Simon Broda and Prof. Dr. Peter Boswijk.

in the neighborhood of the true parameter values. After the estimation process, the standardized residuals are used to calibrate the multivariate generalized hyperbolic distribution.

Multivariate generalized hyperbolic distribution

Although a vast literature has been written describing all sorts of different heavy tailed asymmetrical distributions, it is the density of Barndorff-Nielsen (1977) that is quite interesting. While the original paper concentrates its research to model mass-size distributions of aeolian sand deposits, the independent calibration of the third and fourth moments showed potential to model financial returns. Since 1992 three serious attempts were made to implement the MGHyp density for financial problems and this paper uses the latest, most general, version covered by Protassov (2004) and McNeil et al. (2005) since it handles the unwilling tractability issues.

The derivation is not that complicated if one starts with the assumption of the Normal-Mean-Variance-Mixture

$$\mathbf{X} \stackrel{d}{=} \mu + W\gamma + \sqrt{W}A\mathbf{Z} \quad (1)$$

with $\mathbf{Z} \sim N_k(\mathbf{0}, I_k)$, $A \in \mathbb{R}^{d\times k}$ with $AA' = \Sigma$ of dimension $k\times k$ and μ and γ are parameter vectors in $\mathbb{R}^k$. The remaining step to derive the MGHyp distribution is to assume that the mixture weight W follows a Generalized Inverse Gaussian or $N^-(\lambda,\chi,\psi)$ distribution. This results in the following MGHyp density expression:

$$f(\mathbf{x}) = \frac{\exp\{(\mathbf{x}-\mu)'\Sigma^{-1}\gamma\}\psi^{\lambda}(\sqrt{\chi\psi})^{-\lambda}}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}K_\lambda(\sqrt{\chi\psi})} \times \frac{\psi_*^{\frac{k}{2}-\lambda}}{(\sqrt{\chi_*\psi_*})^{\frac{k}{2}-\lambda}} \times K_{\lambda-\frac{k}{2}}(\sqrt{\chi_*\psi_*})$$

with χ_* and ψ_* defined by

$$\chi_* = (\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)+\chi$$
$$\psi_* = \gamma'\Sigma^{-1}\gamma+\psi$$

and $K_{\lambda-\frac{k}{2}}$ the Bessel function of the third kind. Each parameter defines one particular part or shape of the density; λ controls the shape, χ the peakedness, ψ the difference between the statistical skewness and kurtosis estimates, μ is the location vector, Σ is the dispersion matrix and γ controls the skewness vector. All six function arguments $(\lambda,\chi,\psi,\mu,\Sigma,\gamma)$ are in general unknown, hence some estimation procedure is necessary. λ is considered to be predefined since it takes quite some time to estimate its value while the difference between a predefined or estimated λ is neglectable. The other five parameters are estimated by means of the Expectation-Maximization (EM) algorithm.

Calibration by EM

The aim of the Expectation-Maximization algorithm is to maximize the conditional expectation of the full model log-likelihood function such that if the dataset is incomplete, consistent parameters could still be estimated. Each iteration of the EM algorithm consists of two steps, called the *Expectation step* that deals with the missing values and the *Maximization step* to estimate the parameters.

Defining the E-step

Let $(p) \in \mathbb{R}$, strictly non-negative, denote the current EM cycle and let $\Theta^{(p)}$ denote the collection of parameters λ, μ, Σ, χ, ψ and γ at cycle *(p)* such that the E-step is defined as

$$Q(\Theta\,|\,\Theta^{(p)}) = \mathbb{E}\,[\,\log f(\mathbf{x}_{complete}\,|\,\Theta)\,|\,\mathbf{x}_{observed},\Theta^{(p)}\,]$$

Unfortunately, the complete data specification depends not only on the observations $\mathbf{x}$, but also on the missing variables $\mathbf{w}$ since the observational data for the GIG distribution is unavailable. Estimating the joint density $f(\mathbf{x},\mathbf{w})$ is therefore quite difficult in its present form but if somehow it is known that $f(\mathbf{w}|\Theta)$ has been realized, this knowledge can provide information whether $f(\mathbf{x}|\mathbf{w};\Theta)$ has also been realized. Assume $f(\mathbf{w}|\Theta) > 0$ such that

$$f(\mathbf{x}_{complete}|\Theta) = \prod_{i=1}^{T} f(\mathbf{x}_i|w_i;\Theta)f(w_i|\Theta)$$

Let $\mathbf{x}_i$ be a vector of dimension k containing standardized residuals of the DCC(1,1)-MGARCH(1,1) model of k assets at some time i, where $i \in [1\ ...\ T]$, assume that all observation vectors $\mathbf{x}_i$ are captured in a $1\times kT$ vector $(\mathbf{x}_1, ..., \mathbf{x}_i, ..., \mathbf{x}_T)$ and let the latent variables $\mathbf{w} = (\mathbf{w}_1, ..., \mathbf{w}_i, ..., \mathbf{w}_T)$ be drawn by $N^-(\lambda,\chi,\psi)$ given as

$$f(w_i;\lambda,\chi,\psi) = \frac{\chi^{-\lambda}(\sqrt{\chi\psi})^{\lambda}}{2K_\lambda(\sqrt{\chi\psi})} w_i^{\lambda-1} e^{-\frac{1}{2}(\frac{\chi}{w_i}+\psi w_i)}$$

with $K_\lambda(\sqrt{\chi\psi})$ as the modified Bessel function of the third kind with index λ. The expression for the conditional density $f(\mathbf{x}_i|w_i,\Theta)$ is derived using the Normal-Mean-Variance-Mixture (1).

$$f(\mathbf{x}_i|w_i,\mu,\Sigma,\gamma) = \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}} w_i^{\frac{k}{2}}} e^{(\mathbf{x}_i-\mu)'\Sigma^{-1}\gamma} e^{-\frac{(\mathbf{x}_i-\mu)'\Sigma^{-1}(\mathbf{x}_i-\mu)}{2w_i}} e^{-\frac{w_i}{2}\gamma'\Sigma^{-1}\gamma}$$

By simply substituting the found density expressions, one eventually finds the completely defined E-step as

$$Q(\Theta|\Theta^{(p)}) = Q_1(\mathbf{x}_i,\mu,\Sigma,\gamma)+Q_2(\lambda,\chi,\psi)$$

$$
\begin{aligned}
Q_1(\cdot) = & -\frac{T}{2}\log|\Sigma| - \frac{k}{2}\sum_{i=1}^{T}\mathbb{E}\left[\log w_i \mid \mathbf{x}_i, \Theta^{(p)}\right] + \sum_{i=1}^{T}(\mathbf{x}_i - \mu)'\Sigma^{-1} \\
& -\frac{1}{2}\sum_{i=1}^{T}\mathbb{E}\left[w_i^{-1} \mid \mathbf{x}_i, \Theta^{(p)}\right](\mathbf{x}_i - \mu)'\Sigma^{-1}(\mathbf{x}_i - \mu) \\
& -\frac{1}{2}\gamma'\Sigma^{-1}\gamma\sum_{i=1}^{T}\mathbb{E}\left[w_i \mid \mathbf{x}_i, \Theta^{(p)}\right] - \frac{Tk}{2}\log(2\pi)
\end{aligned}
$$

$$
\begin{aligned}
Q_2(\cdot) = & (\lambda - 1)\sum_{i=1}^{T}\mathbb{E}\left[\log w_i \mid \mathbf{x}_i, \Theta^{(p)}\right] - \frac{\chi}{2}\sum_{i=1}^{T}\mathbb{E}\left[w_i^{-1} \mid \mathbf{x}_i, \Theta^{(p)}\right] \\
& -\frac{\psi}{2}\sum_{i=1}^{T}\mathbb{E}\left[w_i \mid \mathbf{x}_i, \Theta^{(p)}\right] - \frac{\lambda T}{2}\log\chi \\
& +\frac{\lambda T}{2}\log\psi - T\log\left[2K_\lambda(\sqrt{\chi\psi})\right]
\end{aligned}
$$

It can be shown that all three conditional expectations following from $Q_1(\cdot)$ and $Q_2(\cdot)$ are actually defined by the first moment of the Generalized Inverse Gaussian distribution if one utilizes Bayes rule.

Defining the M-step

The updated parameters $\Theta^{(p+1)}$ are found via the second step by separately maximizing $Q_1(\mathbf{x}_i, \mu^{(p)}, \Sigma^{(p)}, \gamma^{(p)})$ and $Q_2(\chi^{(p)}, \psi^{(p)}, \lambda)$. Set the derivative of $Q_1(\cdot)$ with respect to $\mu^{(p)}$, $\gamma^{(p)}$ and $\Sigma^{(p)}$ equal to zero and simply solve the system of unknowns.

$$
\begin{aligned}
\gamma^{(p+1)} &= \frac{\frac{1}{T}\sum_{i=1}^{T}\delta_i^{(p)}(\overline{\mathbf{x}} - \mathbf{x}_i)}{\overline{\delta}^{(p)}\overline{\eta}^{(p)}} \\
\mu^{(p+1)} &= \frac{\frac{1}{T}\sum_{i=1}^{T}\mathbf{x}_i\delta_i^{(p)} - \gamma^{(p+1)}}{\overline{\delta}^{(p)}} \\
\Sigma^{(p+1)} &= \frac{1}{T}\sum_{i=1}^{T}\delta_i^{(p)}(\mathbf{x}_i - \mu^{(p+1)})(\mathbf{x}_i - \mu^{(p+1)})' \\
& \quad + \overline{\eta}^{(p)}\gamma^{(p+1)}(\gamma')^{(p+1)}
\end{aligned}
$$

Maximizing the quasi log-likelihood function $Q_2(\cdot)$ with respect to χ and ψ is performed by a numerical maximization method.

$$
\begin{aligned}
\max_{\chi,\psi} & (\lambda - 1)\sum_{i=1}^{T}\xi_i - \frac{\chi}{2}\sum_{i=1}^{T}\delta_i - \frac{\psi}{2}\sum_{i=1}^{T}\eta_i \\
& -\frac{\lambda T}{2}\log(\chi) + \frac{\lambda T}{2}\log(\psi) - T\log[2K_\lambda(\sqrt{\chi\psi})]
\end{aligned}
$$

Since only $Q_2(\cdot)$ is optimized, instead of the complete log-likelihood, it is required to recalculate the weights $\delta_i^{(p)}$, $\eta_i^{(p)}$ and $\xi_i^{(p)}$ using the updated parameters μ, Σ and γ of the current cycle *(p)*. Furthermore, to address the near singularity problem for Σ, the dispersion matrix is scaled using the determinant of the sample covariance matrix.

$$
\Sigma^{(p+1)} = \frac{|\text{cov}(\mathbf{x})|^{\frac{1}{k}}\Sigma^{(p+1)}}{|\Sigma^{(p+1)}|^{\frac{1}{k}}}
$$

The E- and M-step keeps updating the parameters until the difference between the cycles *(p)* and *(p-1)* is neglectable.

Risk assessment

The one day ahead forecast portfolio return is estimated by the underlying DCC-MGARCH portfolio model with the residuals following the MGHyp distribution, calibrated by the previous 500 observations using the conditional VaR approach with the nominal coverage levels 95% and 99%. A rolling window using the previous 1000 observations has been tested, but no significant differences compared with the rolling window of 500 were noticeable. Let x be the asset weights, sum up to one and let H_{t+1} be the forecast DCC-MGARCH covariance matrix such that the forecast portfolio return is denoted by

$$
x'r_{t+1} = x'\mu + x'H_{t+1}^{\frac{1}{2}}\varepsilon
$$

with ε as the MGHyp distribution. To estimate the conditional VaR using a multivariate density model is rather complicated due to the integrand, but by introducing weights the problem translates to an ordinary univariate case by which we all know how to solve. The proof is easy. Assume a multivariate linear function $\mathbf{B}X + \mathbf{b}$, assume that the intercept vector $\mathbf{b} = \mathbf{0}$ and that matrix $\mathbf{B}$ is actually a weighting vector $x \in \mathbb{R}^k$ such that the sum of the weights equals one. Finally, let X denote the Normal-Mean-Variance-Mixture and W distributed by the GIG distribution such that the multivariate case translates to the univariate one.

Application and remarks

The equally weighted portfolio is constructed by the S&P 500 top ten constituents by market cap; Apple Inc (AAPL), Chevron Corp (CVX), General Electric (GE), Intl Business Machines Corp (IBM), JP Morgan Chase & Co (JPM), Microsoft Corp (MSFT), Procter and Gamble (PG), AT&T (T), Wells Fargo & Co (WFC) and Exxon Mobile Corp (XOM). The finite sample $T = 2{,}766$, for the period 01/01/2000 to 01/01/2011, is formed by taking daily negative log returns of the adjusted daily close price. It is apparent from descriptive statistics that eight out of the ten assets endured a loss over the covered period due to the liquidity crisis. Furthermore, all ten asset returns exhibit heavy tail and asymmetrical properties. Finally, Mardia's test to test whether the portfolio is gaussian distributed is rejected at a 1% significance level with p-value 0.0000.

Results

By reviewing the subclasses: normal inverse gaussian (NIG), multivariate hyperbolic (Hyp) and the 10-dimensional multivariate hyperbolic (KHyp) a comparison is made which of the hyperbolic subclasses performs the best to handle the observed heavy tail and asymmetry. Figure 1 shows two of the twelve outcomes of the time series analysis illustrating the risk violations by + markings while table 1 and 2 presents the statistical data.

It is apparent from table 1 and 2 that Christoffersen unconditional coverage test indicates strong evidence to reject the null of correct coverage for all three symmetric distributions using the nominal 95% coverage level. Of these three distributions, only the NIG is found to be statistically significant for the Christoffersen conditional test. While the latter distributions all underestimate risk, no significant problems are found when estimating the risk using the asymmetrical distributions. For the nominal 99% coverage level no strong statistical significance is found of under or overestimated risk for the symmetrical as well as the asymmetrical distributions.

It seems suspicious that all three symmetrical distributions outperform the asymmetrical distribution based on the MSE value. Firstly, the portfolio is heavily skewed according to Mardia's test. Secondly, the asymmetrical distribution nests the symmetrical distribution such that one should expect that at least one asymmetrical subclass could outperform a symmetrical

Figure 1. The one day ahead conditional Value at Risk using the Hyperbolic distribution for the nominal 95% (2nd line from top) and 99% (3rd line from top) coverage level. Conditional VaR violations are indicated by + markings with 95% below the 99% coverage level.

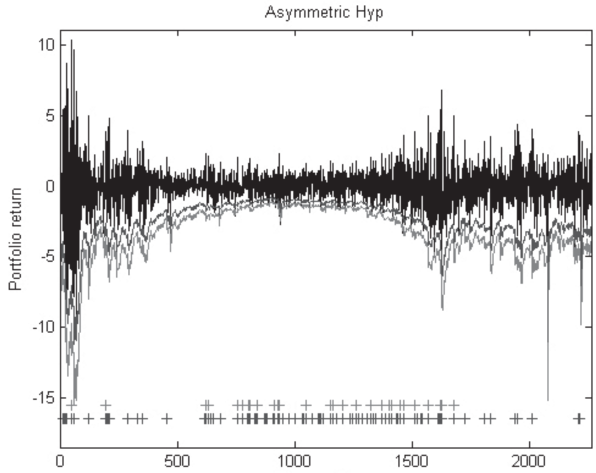

Table 1. Backtesting test statistics based on the 95% coverage level. The p-values are denoted between the parentheses and * (**) indicates a rejection at 1% (5%) significance level. All three symmetric subclasses are rejected by the Kupiec test at 1% significance level.

	$\hat{\beta}$	Coverage	Independence	Conditional	MSE
asym NIG	0.0419	3.2674 (0.0707)	0.08577 (0.7696)	3.3532 (0.1870)	0.00167
asym HYP	0.0424	2.9100 (0.0880)	0.0880 (0.7668)	2.9980 (0.2234)	0.00172
asym 10dim Hyp	0.0442	1.6959 (0.1928)	0.1347 (0.7136)	1.8306 (0.4004)	0.00261
sym NIG	0.0358	10.6869 (0.0011)*	0.4789 (0.4890)	11.1658 (0.0038)*	0.00064
sym HYP	0.0371	8.7006 (0.0032)*	0.3249 (0.5685)	9.0256 (0.0110)**	0.00071
sym 10dim Hyp	0.0384	6.9369 (0.0084)*	0.2102 (0.6466)	7.1472 (0.0281)**	0.00177

Table 2. Backtesting test statistics based on the 99% coverage level. The p-values are denoted between the parentheses and * (**) indicates a rejection at 1% (5%) significance level.

	$\hat{\beta}$	Coverage	Independence	Conditional	MSE
asym NIG	0.0079	1.0373 (0.3085)	0.3045 (0.5812)	1.3417 (0.5113)	0.00442
asym HYP	0.0102	0.0054 (0.9412)	0.4925 (0.4828)	0.4980 (0.7796)	0.00507
asym 10dim Hyp	0.0126	1.1873 (0.2759)	0.7262 (0.3941)	1.9134 (0.3842)	0.01255
sym NIG	0.0128	1.6519 (0.1987)	0.7784 (0.3776)	2.4303 (0.2967)	0.01422
sym HYP	0.0143	0.2383 (0.6255)	0.5805 (0.4461)	0.8188 (0.6641)	0.01434
sym 10dim Hyp	0.0137	2.7884 (0.0949)	0.8883 (0.3459)	3.6768 (0.1591)	0.02164

subclass. Due to the parsimonious behavior of the MSE, more observations are explained by the simpler and less complex symmetrical distribution. This results in lower MSE values and falsely indicates the better suited model. It probably could explain the odd ranking, and therefore its adviced not to rank the models based on the MSE value. If one only compares the setups if the same assumptions are used, e.g. symmetry and nominal coverage level, its follows from tables 1 and 2 that in these four cases the NIG distribution outperforms. This result is not surprising and has been documented in other papers, for instance by Protassov (2004) and McNeil et al. (2005).

Calibration time improvement

The calibration of the MGHyp density by EM is considered to be in general a slow optimization process. Let the time to optimize one cycle of the backtesting analysis be given as five minutes, such that estimating the full backtesting sample (2,266 cycles)[1] takes a shocking eight days to complete. It would simply take too much time for an empirical study with twelve different distributions, namely 96 days.This study proposes the use of parallel processing using multi core desktops. The effectiveness of the parallel processing unit is noticeable since it reduces the running time by 2600%. Using one of the symmetric distributions results a running time of merely six hours while the asymmetric distributions takes seven hours to complete. This is perfectly explainable because the symmetric case assumes $\gamma = 0$ such that it isn't estimated by the EM algorithm.

References

Barndorff-Nielsen, O. (1977). "Normal inverse gaussian distributions and stochastic volatility modelling." *Scandinavian Journal of Statistics*, 24:01-13.

Engle, R. (1999). "Dynamic conditional correlation - a simple class of multivariate garch models." *Econometrica*, 50:987-1007.

Laplante, J., Desrochers, J., and J.Prefontaine (2008). "The garch(1,1) model as a risk predictor for international portfolios." *International Business and Economic Research Journal*, 7:23-34.

McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton University Press, Princeton, New Jersey.

Pagan, A. (1996). "The econometrics of financial markets." *Journal of Empirical Finance*, 3:15-102.

Protassov, R. (2004). "Em-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ." *Statistics and Computing*, 14:67-77.

[1] Full sample reduced by the first 500 calibration observations.