An efficient method for market risk management under multivariate extreme value theory approach

The First International Moscow Finance Conference

Moscow - November, the 18^{th} 2011

Comments (Part 1)...

- Better description of the Filtering subsection on pages 17-20: not very clear when and where the PCA is applied to.

- Describe better the relationship with the Orthogonal-GARCH model, which it is the natural benchmark and direct competitor to your model and therefore deserves more space.

- Consider also the GO-GARCH model by Van der Weide (2002)...

- ...(if you need to save time in code programming, there is the recent nice `R` package `gogarch` by Bernhard Pfaff)

Van der Weide (2002) points out that the orthogonality condition implicitly assumed in the OGARCH model is very restrictive and raises the question "if a linkage with a set of uncorrelated economic components exists, why should the associated matrix be orthogonal ?". Therefore, if we assume that

$$Y_t|\mathcal{F}_{t-1} \sim n(\mathbf{0}, \mathbf{V}_t)$$

the observed economic process Y_t is governed by a linear combination of independent economic components $\{\mathbf{f}_t\}$

$$Y_t = \mathbf{Z}\mathbf{f}_t$$

where $\mathbf{f}_t$ are uncorrelated components, whereas $|\mathbf{Z}| \neq 0$. The unobserved components are normalized such that:

$$E[\mathbf{f}_t\mathbf{f}_t'] = \mathbf{I}_n$$

$$\mathbf{V} = E[Y_tY_t'] = \mathbf{Z}\mathbf{Z}'$$

$$\mathbf{V}_t = E_{t-1}[Y_tY_t'] = \mathbf{Z}E_{t-1}[\mathbf{f}_t\mathbf{f}_t']\mathbf{Z}' = \mathbf{Z}\mathbf{H}_t\mathbf{Z}'$$

$$\mathbf{H}_t = diag\{h_{1,t}, \ldots, h_{n,t}\}$$

$$h_{i,t} = (1 - \alpha_1 - \beta_i) + \alpha_i y_{i,t-i}^2 + \beta_i h_{i,t-1} \qquad i = 1, \ldots, n$$

$$\mathbf{V} = E[\mathbf{V}_t] = \mathbf{Z}\mathbf{H}\mathbf{Z}', \qquad \mathbf{H} \text{ diagonal}$$

The diagonal decomposition of the unconditional covariance matrix is given by:

$$\mathbf{V} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$$

where the orthogonal matrix $\mathbf{P}$, which is the O-GARCH estimator for $\mathbf{Z}$, is only guaranteed to coincide with $\mathbf{Z}$, when the diagonal elements of $\mathbf{H}$ are all distinct. Suppose that $\mathbf{H} = \mathbf{I}$, then we have that

$$\mathbf{V} = E[\mathbf{V}_t] = \mathbf{ZIZ}' = \mathbf{ZZ}'$$

The matrix $\mathbf{Z}$ is no longer identified by the eigenvector matrix of $\mathbf{V}$ as for every orthogonal matrix $\mathbf{Q}$ we have

$$(\mathbf{ZQ})(\mathbf{ZQ})' = \mathbf{I}$$

The matrix $\mathbf{Z}$ is well identified when conditional information is taken into account. Based on singular value decomposition, it follows that

$$\mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{U}_0 = \mathbf{Z}, \quad \text{where the estimator } \mathbf{U} \text{ of } \mathbf{U}_0 \text{ has } |\mathbf{U}| = 1$$

The matrices $\mathbf{P}$ and $\boldsymbol{\Lambda}$ have $\frac{n(n-1)}{2}$ and n parameters, respectively, so we have n^2 parameters for the invertible matrix $\mathbf{Z}$. The matrices $\mathbf{P}$ and $\boldsymbol{\Lambda}$ will be estimated by means of unconditional information, as they will be extracted from the sample covariance matrix $\mathbf{V}$, which has $\frac{n(n+1)}{2}$ parameters.

Instead, conditional information is required to estimate $\mathbf{U}_0$, where we have $\frac{n(n-1)}{2}$ free parameters. The O-GARCH model (when $m = n$) corresponds then to the particular choice $\mathbf{U} = I_n$.

The orthogonal matrix $\mathbf{U}_0$ is parameterized by means of *rotation matrices*. $\mathbf{G}_{ij}(\theta_{ij})$, $(n \times n)$, $i, j = 1, 2, \ldots, n$:

$$\mathbf{G}_{ij}(\theta_{ij}) = \{g_{ij}\} \qquad g_{rr} = g_{ss} = \cos(\theta_{ij})$$
$$g_{ii} = 1 \qquad i = 1, \ldots, n \qquad i \neq r, s$$
$$g_{sr} = -\sin(\theta_{ij}) \qquad g_{rs} = \sin(\theta_{ij})$$

and all other elements are zero.

Example: $n = 3$,

$$\mathbf{G}_{12} = \begin{bmatrix} \cos(\theta_{12}) & \sin(\theta_{12}) & 0 \\ -\sin(\theta_{12}) & \cos(\theta_{12}) & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{G}_{13} = \begin{bmatrix} \cos(\theta_{13}) & 0 & \sin(\theta_{13}) \\ 0 & 1 & 0 \\ -\sin(\theta_{13}) & 0 & \cos(\theta_{13}) \end{bmatrix}$$

and $\mathbf{G}_{23}$ has the block with $\cos(\theta_{23})$ and sin $\sin(\theta_{23})$ functions in the right low corner. The $n(n-1)/2$ rotation angles are parameters to be estimated.

Every n-dimensional orthogonal matrix $\mathbf{U}$ with $\det(\mathbf{U}) = 1$ can be represented as a product of $\frac{n(n-1)}{2}$ rotation matrices:

$$\mathbf{U} = \prod_{i<j} \mathbf{G}_{ij}(\theta_{ij}) \qquad -\pi \leq \theta_{ij} \leq \pi$$

$\mathbf{G}_{ij}(\theta_{ij})$ performs a rotation in the plane spanned by the i-th and the j-th vectors of the canonical basis of $\mathbb{R}^n$ over an angle δ_{ij}. In compact notation, for $n = 3$, we have $\mathbf{U} = \mathbf{G}_{12}\mathbf{G}_{13}\mathbf{G}_{23}$.

The implied conditional correlation matrix can be expressed as follows:

$$\mathbf{R}_t = \mathbf{D}_t^{-1}\mathbf{V}_t\mathbf{D}_t^{-1}$$

$$\mathbf{D}_t = (\mathbf{V}_t \odot \mathbf{I}_m)^{1/2}$$

For example, for $n = 2$, we would have

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 \\ \cos\theta & \sin\theta \end{bmatrix}$$

so that the conditional correlations matrix is:

$$\begin{aligned}
\mathbf{R}_t &= \mathbf{D}_t^{-1}\mathbf{V}_t\mathbf{D}_t^{-1} \\
&= \mathbf{D}_t^{-1}\mathbf{Z}\mathbf{H}_t\mathbf{Z}'\mathbf{D}_t^{-1} \\
&= \begin{bmatrix} h_{1t}^{-1/2} & 0 \\ 0 & h_{2t}^{-1/2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \cos\theta & \sin\theta \end{bmatrix} \begin{bmatrix} h_{1t} & h_{12t} \\ h_{12t} & h_{2t} \end{bmatrix} \times \\
&\quad \begin{bmatrix} 1 & \cos\theta \\ 0 & \sin\theta \end{bmatrix} \begin{bmatrix} h_{1t}^{-1/2} & 0 \\ 0 & h_{2t}^{-1/2} \end{bmatrix}
\end{aligned}$$

and

$$\begin{aligned}
\rho_t &= \frac{Cov_{t-1}(y_{1t}, y_{2t})}{\sqrt{Var_{t-1}(y_{1t})}\sqrt{Var_{t-1}(y_{2t})}} = \frac{h_{1t}\cos(\theta)}{\sqrt{h_{1t}}\sqrt{h_{1t}\cos^2(\theta) + h_{2t}\sin^2(\theta)}} \\
&= \frac{\cos(\theta)}{\sqrt{\cos^2(\theta) + \frac{h_{1t}}{h_{2t}}\sin^2(\theta)}} = \frac{1}{\sqrt{1 + z_t\tan^2(\theta)}}
\end{aligned}$$

where $z_t = \frac{h_{1t}}{h_{2t}}$.

ESTIMATION: Let define $(\alpha', \beta', \theta')'$ the parameters to be estimated by means of the conditional information. Then

$$\alpha = (\alpha_1, \ldots, \alpha_n) \quad \beta = (\beta_1, \ldots, \beta_n) \quad \theta = (\theta_1, \ldots, \theta_m), \qquad m = \frac{n(n-1)}{2}$$

The log-likelihood per each observation t can be expressed:

$$\begin{aligned} l_t &= -\frac{1}{2}\left(n\log(2\pi) + \log|\mathbf{V}_t| + Y_t'\mathbf{V}_t^{-1}Y_t\right) \\ &= -\frac{1}{2}\left(n\log(2\pi) + \log|\mathbf{Z}_\theta \mathbf{H}_t \mathbf{Z}_\theta| + Y_t'(\mathbf{Z}_\theta \mathbf{H}_t \mathbf{Z}_\theta')^{-1}Y_t\right) \\ &= -\frac{1}{2}\left(n\log(2\pi) + \log|\mathbf{Z}_\theta \mathbf{Z}_\theta'| + \log|\mathbf{H}_t| + Y_t'(\mathbf{Z}_\theta \mathbf{H}_t \mathbf{Z}_\theta')^{-1}Y_t\right) \end{aligned}$$

while the global log-likelihood is given by $\log L_T = \sum_t l_t$. The estimation of this model can be performed in two steps:

1. Estimate $\mathbf{P}$ and $\mathbf{\Lambda}$ from the sample covariance matrix, so to obtain:

$$\widehat{\mathbf{V}} = \widehat{\mathbf{P}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{P}}'$$

$$\mathbf{Z} = \widehat{\mathbf{P}}\widehat{\mathbf{\Lambda}}^{1/2}\mathbf{U}, \quad \text{where} \quad \mathbf{U} = \prod_{i<j} \mathbf{G}_{ij}(\theta_{ij}) \qquad -\pi \le \theta_{ij} \le \pi$$

2. Estimate $(\alpha', \beta', \theta')'$ by maximizing the $\log L_T$.

- *Backtesting section need to improved*: consider also the ***Christoffersen's (1998) conditional coverage test***, which simultaneously examines if the total number of failures is equal to the expected one and the VaR failure process is independently distributed. The statistic is computed as follows:

$$LR_{CC} = -2\ln[(1-p)^{T-N}p^{N}] + 2\ln[(1-\pi_{01})^{n00}\pi_{01}^{n01}(1-\pi_{11})^{n10}\pi_{11}^{n11}] \quad (1)$$

  where n_{ij} is the number of observations with value i followed by j for $i, j = 0, 1$ and

$$\pi_{ij} = \frac{n_{ij}}{\sum_j n_{ij}} \quad (2)$$

  are the corresponding probabilities, while N are the losses in excess of VaR out of T observations. Under the H_0, this test is distributed as a $\chi^2(2)$.

- Compare the different multivariate models by looking at their VaR forecasts by using the ***Hansen's Superior Predictive Ability (SPA) test*** (2005) together with Giacomini and Komunjer (2005) asymmetric loss function....

The Hansen's (2005) Superior Predictive Ability (SPA) test compares the performances of two or more forecasting models, by evaluating the forecasts with a pre-specified loss function. The 'best' forecast model is the model that produces the smallest expected loss.

The SPA tests for the best standardized forecasting performance with respect to a benchmark model, and the null hypothesis is that none of the competing models is better than the benchmark one.

Since the object of interest is the conditional p-quantile of the portfolio loss distribution, we use the asymmetric linear loss function proposed in Gonzalez-Rivera et al. (2006) and Giacomini and Komunjer (2005), and defined as

$$\mathcal{T}_p(e_{t+1}) \equiv (p - \mathbf{1}(e_{t+1} < 0)e_{t+1} \tag{3}$$

where $e_{t+1} = L_{t+1} - \widehat{VaR_{t+1|t}}$, $\mathbf{1}(\cdot)$ is the indicator function, L_{t+1} is the realized loss, while $\widehat{VaR_{t+1|t}}$ is the VaR forecast at time $t+1$ on information available at time t. See also Fantazzini (2009) for a recent application.

Comments (Part 3)

- Attention in your final comments in the empirical section. You say on page 47: "*The results of the Pearson's test for the currency portfolio...show that models based on conditionally normal or t-distributed residuals, as well as the HS model, can be rejected in favor of the proposed multivariate EVT alternative.*"

  I am sorry, but for what I read on table 10 this is not true

Method	Lower tail	Upper tail
EVT	0.4170	1.3142
	(0.0189)	(0.1410)
Normal	38.5252	7.5773
	($\sim$ 1.0)	(0.8917)
t	2.2298	2.0934
	(0.3064)	(0.2814)
HS	123.1067	146.7974
	($\sim$ 1.0)	($\sim$ 1.0)

  Moreover, for extreme quantiles (99% and 99.9%), t and EVT have almost identical out-of-sample performances.

  $\Rightarrow$ *However this is case for the four FX rates up to 2008. In the updated version of the paper, considering Dow Jones stocks up to the end of 2010, EVT is much better and seems to statistically outperform the competitors.*

References

Christoffersen, P. (1998). Evaluating Interval Forecats, International Economic Review, 39, 841-862.

Fantazzini, D. (2009). The Effects of Misspecified Marginals and Copulas on Computing the Value at Risk: A Monte Carlo Study, Computational Statistics and Data Analysis, 53(6), 2168-2188.

Giacomini, R., Komunjer, I. (2005). Evaluation and Combination of Conditional Quantile Forecasts, Journal of Business and Economic Statistics, 23, 416-431.

Gonzalez-Rivera, G., Lee, T., Santosh, M. (2006). Forecasting volatility: A reality check based on option pricing, utility function, Value at Risk, and predictive likelihood, International Journal of Forecasting, 20, 629-645.

Hansen, P. (2005). A Test for Superior Predictive Ability, Journal of Business and Economic Statistics, 23(4), 365-380.

Van derWeide, Roy (2002), GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model, Journal of Applied Econometrics, 17(5), 549 - 564.