# "Construction of value-at-risk forecasts under different distributional assumptions within a BEKK framework"

Braione, Manuela ; Scholtes, Nicolas

**Abstract**

Financial asset returns are known to be conditionally heteroskedastic and generally non-normally distributed, fat-tailed and often skewed. In order to account for both the skewness and the excess kurtosis in returns, we combine the BEKK model from the multivariate GARCH literature with different multivariate densities for the returns. The set of distributions we consider comprises the normal, Student, Multivariate Exponential Power and their skewed counterparts. Applying this framework to a sample of ten assets from the Dow Jones Industrial Average Index, we compare the performance of equally- weighted portfolios derived from the symmetric and skewed distributions in forecasting out-of-sample Value-at-Risk. The accuracy of the VaR forecasts is assessed by implementing standard statistical backtesting procedures. The results unanimously show that the inclusion of fat-tailed densities into the model specification yields more accurate VaR forecasts, while the further addition of skewnes...

Document type : *Document de travail (Working Paper)*

## Référence bibliographique

Braione, Manuela ; Scholtes, Nicolas. *Construction of value-at-risk forecasts under different distributional assumptions within a BEKK framework.* CORE Discussion Papers ; 2014/59 (2014) 32 pages

2014/59

# Construction of value-at-risk forecasts under different distributional assumptions within a BEKK framework

Manuela Braione and Nicolas K. Scholtes

DISCUSSION PAPER

Center for Operations Research
and Econometrics

Voie du Roman Pays, 34
B-1348 Louvain-la-Neuve
Belgium
http://www.uclouvain.be/core

**CORE**
Voie du Roman Pays 34, L1.03.01
B-1348 Louvain-la-Neuve, Belgium.
Tel (32 10) 47 43 04
Fax (32 10) 47 43 01
E-mail: corestat-library@uclouvain.be
http://www.uclouvain.be/en-44508.html

CORE DISCUSSION PAPER
2014/59

# Construction of Value-at-Risk forecasts under different distributional assumptions within a BEKK framework

Manuela Braione[1] and Nicolas K. Scholtes[2]

November 18, 2014

**Abstract**

Financial asset returns are known to be conditionally heteroskedastic and generally non-normally distributed, fat-tailed and often skewed. In order to account for both the skewness and the excess kurtosis in returns, we combine the BEKK model from the multivariate GARCH literature with different multivariate densities for the returns. The set of distributions we consider comprises the normal, Student, Multivariate Exponential Power and their skewed counterparts. Applying this framework to a sample of ten assets from the Dow Jones Industrial Average Index, we compare the performance of equally- weighted portfolios derived from the symmetric and skewed distributions in forecasting out-of-sample Value-at-Risk. The accuracy of the VaR forecasts is assessed by implementing standard statistical backtesting procedures. The results unanimously show that the inclusion of fat-tailed densities into the model specification yields more accurate VaR forecasts, while the further addition of skewness does not lead to significant improvements.

**Keywords:** Dow Jones Industrial Average, BEKK model, Maximum likelihood, Value-at-Risk
**JEL Classification:** C01, C22, C52, C58

[1]Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium; Email: manuela.braione@uclouvain.be

[2]Université de Namur, CeReFiM, B-5000, Namur; Belgium and Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium; Email: nicolas.scholtes@unamur.be (corresponding author)

Manuela Braione and Nicolas K. Scholtes acknowledge support of the "Communauté française de Belgique" through contract "Projet d'Actions de Recherche Concertées "12/17-045" and "13/17-055", respectively granted by the Académie universitaire Louvain. We thank seminar participants at the 13[th] Journéee d'économétrie at the Université Paris Ouest Nanterre La Défense along with Luc Bauwens, Sophie Béreau, Jean-Yves Gnabo, Leonardo Iania and Florian Ielpo for useful comments. The scientific responsibility is assumed by the authors.

CORE DISCUSSION PAPER

2014/??

# Construction of Value-at-Risk forecasts under different distributional assumptions within a BEKK framework

Manuela Braione[1] and Nicolas K. Scholtes[1,2]

November 18, 2014

**Abstract**

Financial asset returns are known to be conditionally heteroskedastic and generally non-normally distributed, fat-tailed and often skewed. In order to account for both the skewness and the excess kurtosis in returns, we combine the BEKK model from the multivariate GARCH literature with different multivariate densities for the returns. The set of distributions we consider comprises the normal, Student, Multivariate Exponential Power and their skewed counterparts. Applying this framework to a sample of ten assets from the Dow Jones Industrial Average Index, we compare the performance of equally-weighted portfolios derived from the symmetric and skewed distributions in forecasting out-of-sample Value-at-Risk. The accuracy of the VaR forecasts is assessed by implementing standard statistical backtesting procedures. The results unanimously show that the inclusion of fat-tailed densities into the model specification yields more accurate VaR forecasts, while the further addition of skewness does not lead to significant improvements.

**Keywords:** Dow Jones Industrial Average, BEKK model, Maximum likelihood, Value-at-Risk

**JEL Classification:** C01, C22, C52, C58

[1]Université catholique de Louvain, CORE, B-1348 Louvain-La-Neuve, Belgium;
E-mail: manuela.braione@uclouvain.be
[2]Université de Namur, CeReFiM, B-5000, Namur, Belgium;
E-mail: nicolas.scholtes@unamur.be (corresponding author)

Manuela Braione and Nicolas K. Scholtes acknowledge support of the "Communauté francaise de Belgique" through contract "Projet d'Actions de Recherche Concertées "12/17-045" and "13/17-055", respectively granted by the Académie universitaire Louvain. We thank seminar participants at the $13^{\text{th}}$ Journée d'économétrie at the Université Paris Ouest Nanterre La Défense along with Luc Bauwens, Sophie Béreau, Jean-Yves Gnabo, Leonardo Iania and Florian Ielpo for useful comments. The scientific responsibility is assumed by the authors.

# 1 Introduction

Value-at-Risk (VaR) is a quantitative tool used to measure the maximum potential loss in value of a portfolio of assets over a defined period for a given probability. Specifically, VaR construction requires a quantile estimate of the far-left tail of the unconditional returns distribution. Though widely-used as a risk measure in the past, standard methods of VaR construction assuming iid-ness and normality have come under criticism due to their failure to incorporate three stylized facts of financial returns (*i*) the presence of volatility clustering, indicated by high autocorrelation of absolute and squared returns, (*ii*) excess kurtosis (fat tails) and (*iii*) skewness in the density of the unconditional returns distribution.

The ability to account for volatility clustering is one of the key strengths of the ARCH modelling approach developed in Engle (1982) and extended in Bollerslev (1986). Combining this approach with a non-normal conditional distribution assumption for the returns, several papers have shown that univariate GARCH models can produce reliable out-of-sample volatility forecasts. For example, Angelidis et al. (2004) combine three GARCH specifications with the univariate skew-Student and skew-GED (Generalized Error) distributions to show that these are able to produce superior VaR forecasts compared to the normal. Specifically, they apply the exponential GARCH (EGARCH) model of Nelson (1991) and the threshold ARCH (TARCH) model to five univariate returns series and find that while the choice of a skewed, heavy-tailed distribution significantly improves the forecasting performance, the choice of the volatility model appears to be irrelevant. Within the univariate distribution framework, several other papers have proposed combining VaR forecasts with non-normal distributions and GARCH-type specifications. Notably, Giot & Laurent (2003) use the skew-Student-univariate APARCH model developed in Lambert & Laurent (2001) to estimate daily VaR for stock indices, finding that it performs better than the symmetric, student APARCH.

While this literature exemplifies the need to incorporate non-normal distributions into volatility modelling, it is restricted to the univariate framework alone, thus ignoring the evidence that financial volatilities move together over time across assets and markets (Bollerslev 1990). This is a major focus of the multivariate GARCH (MGARCH) literature. Within this framework, Bauwens & Laurent (2005) develop a transformation function which allows multivariate skewed distributions to be constructed from their symmetric counterparts. By combining the Dynamic Conditional Correlation model of Engle (2002) with the Student and skew-Student distributions, they show that the skewed density outperforms the symmetric competitor in forecasting out-of-sample VaR.

Our work builds on their approach, differing along three main dimensions. First, we consider a wider set of multivariate distributional assumptions which includes both symmetric and asymmetric types of distributions. These are the normal, Student, Multivariate Exponential Power (MEP) and their skewed counterparts. This allows us to perform a direct comparison between the different candidates. Second, we estimate the multivariate BEKK model of Engle & Kroner (1995) with the aforementioned assumptions and evaluate the model from both an in- and out-of-sample perspective. Last, we construct out-of-sample portfolio VaR forecasts and assess the predictive accuracy of the models by means of statistical backtesting procedures.
The set of employed tests includes the Unconditional Coverage (UC) test, Independence (IND) test, Conditional Coverage (CC) test, Duration-Based Test of Independence (DBI), Time Until First Failure (TUFF) test and the Dynamic Quantile (DQ) test. The results of the tests are summarized using a grading scheme based on the number of acceptances of the null hypothesis which determines the distributional assumption providing the most accurate VaR forecasts. The main contribution of the paper comes from the combination of the multivariate GARCH modeling technique with alternate assumptions on the distribution of the returns in order to construct Value-at-Risk forecasts. From the literature, our paper is close in structure to Angelidis et al. (2004) and Kuester et al. (2006) who both use VaR forecast performance as a means of comparing different distributional assumptions and volatility specifications, albeit within a univariate framework and using a smaller set of distributions. Herein, we are mainly concerned with the effect of the multivariate density assumption on the model forecast accuracy, thus leaving a closer inspection of the impact of different volatility models as an open issue for further research.

The paper is organized as follows. Section 2 reviews the MGARCH modeling framework and the theoretical procedure for constructing the skewed distributions. Section 2.3 reports the Maximum Likelihood (ML) estimation procedure of the model with the multivariate distributional assumptions. Section 3 introduces the empirical methodology, comprising the portfolio construction and the VaR estimation technique while section 3.2 describes the VaR backtesting procedures. Section 4 provides estimation results and outcomes of the VaR tests and section 5 concludes with some final remarks.

# 2 Theoretical Framework

## 2.1 MGARCH Modeling

Let $y_t$ be a $N$-dimensional discrete time vector of daily returns for $t = 1, ..., T$, whose stochastic process depends on a finite dimensional parameter vector $\psi$. Conditioned on $\mathcal{F}_{t-1}$, the sigma field generated by past information until time $t-1$, $y_t$ can be rewritten as

$$y_t = \mu_t(\psi) + H_t^{1/2}(\psi) z_t, \tag{1}$$

where $\mu_t(\psi)$ is the $N \times 1$ conditional mean vector and $H_t^{1/2}(\psi)$ is a Cholesky factorization of the $N \times N$ positive definite conditional covariance matrix $H_t(\psi)$. The $N \times 1$ i.i.d. stochastic error vector $z_t$ has first and second-order moments respectively equal to $E(z_t) = 0$ and $Var(z_t) = I_N$. Since our focus is on the modeling of the covariance matrix of returns, we set $\mu_t(\psi) = 0$. We also drop $\psi$ for notational convenience.

In the multivariate GARCH (MGARCH) literature, many possible specifications for $H_t$ are available. They differ in various aspects but all have to ensure the positive definiteness of the conditional covariance matrix. In this respect, the BEKK model of Engle & Kroner (1995) guarantees the positivity of $H_t$ without imposing heavy parameter restrictions. Furthermore, the basic model structure can be easily simplified by applying its scalar parametrization, which makes the model tractable for practical applications.

**Definition 1.** *The scalar BEKK(1,1,1) model is defined as:*

$$H_t = \Omega + a y_{t-1} y_{t-1}^{'} + b H_{t-1} \tag{2}$$

*where $\Omega$ is an $N \times N$ intercept matrix and $a$ and $b$ are scalar parameters.*

The process in Eq.(2) is assured to be covariance stationary if and only if $a + b < 1$.

Following Engle & Mezrich (1996) and Francq et al. (2011), *covariance targeting* under stationarity conditions can also be applied in order to further reduce the number of parameters to be estimated. This technique consists in expressing the conditional covariance matrix as a function of the unconditional covariance and the other model parameters. A consistent estimator of the unconditional covariance matrix (to be computed before maximizing the likelihood function) is easily obtained as $\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} y_t y_t^{'}$ such that the model can be reparametrized as follows:

$$H_t = (1 - a - b)\hat{\Sigma} + a y_{t-1} y_{t-1}^{'} + b H_{t-1}. \tag{3}$$

This leaves a final number of parameters to be estimated equal to two. This specification can be applied even to large dimensional settings and, as we will see in the empirical application, significantly simplifies the computational burden during the estimation procedure.

## 2.2 Constructing skew densities

Bauwens & Laurent (2005) develop a procedure for constructing multivariate skewed densities from their symmetric counterparts. We build on their findings in order to enlarge the set of employed distributions.

The general notion of symmetry of a standardized density used herein is that of $M$-symmetry (see Definition (1) in their paper), which encompasses the class of spherically symmetric densities. These can be obtained as a special case of the general family of multivariate *elliptical* distributions, denoted as

$$g(x; \mu, \Sigma, \eta) \propto h((x-\mu)'\Sigma^{-1}(x-\mu), \eta), \tag{4}$$

where $x$ is a random vector with an integrable, positive function $h(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$, $\eta$ captures the shape parameter of the distribution. The spherically symmetric set of distributions, comprising the standard normal, Student and MEP, are obtained by setting $\mu$ and $\Sigma$ equal to zero and $I_N$, respectively.

The idea of introducing skewness into an $M$-symmetric standardized distribution revolves around scaling it differently for negative and positive values by multiplying (dividing) by a positive constant. The value of this scaling parameter (hereafter referred to as $\xi$) determines whether the resulting distribution is skewed to the left ($0 < \xi < 1$) or to the right ($\xi > 1$). As a result, the multivariate, skewed density function is obtained from:

**Definition 2.** *Given a random vector $z = (z_1, \ldots, z_N)'$ with multivariate symmetric standardized distribution, $g(z; \eta)$ following Eq. (4), the standardized skewed density $f(z|\xi; \eta)$ with vector of asymmetry parameters $\xi = (\xi_1, \ldots \xi_N)'$, can be expressed as:*

$$f(z|\xi, \eta) = 2^N \left( \prod_{i=1}^{N} \frac{\xi_i}{1+\xi_i^2} \right) g(z^\star; \eta) \tag{5}$$

*with*

$$z^\star = (z_1^\star, ..., z_N^\star)' \tag{6}$$

$$z_i^\star = z_i \xi_i^{I_i} \tag{7}$$

*and*

$$I_i = \begin{cases} -1 \ \textit{if} \ z_i \geq 0 \\ 1 \ \textit{if} \ z_i < 0 \end{cases} \tag{8}$$

The marginal $r^{th}$-order moment of the obtained skewed distribution can be computed directly from the standardized $r^{th}$ moment of the symmetric density $g(\cdot)$. This is accomplished by applying the following transformation function:

$$E(z_i^{\star r}|\xi) = M_{i,r}\frac{\xi_i^{r+1} + \frac{(-1)^r}{\xi_i^{r+1}}}{\xi_i + \frac{1}{\xi_i}} \tag{9}$$

where the $r^{th}$-order moment of the marginal $g_i(\cdot)$, truncated to the positive real values, is given by

$$M_{i,r} = \int_0^\infty 2u^r g_i(u)du. \tag{10}$$

Since only the first two moments are required in the transformation process, their analytical expression for $r = 1, 2$ in Eq.(9) is reported below:

$$m_i = \mathrm{E}(z_i^\star|\xi_i) = M_{i,1}\left(\xi_i - \frac{1}{\xi_i}\right) \tag{11}$$

$$s_i^2 = \mathrm{Var}(z_i^\star|\xi_i) = \left(M_{i,2} - M_{i,1}^2\right)\left(\xi_i^2 + \frac{1}{\xi_i^2}\right) + 2M_{i,1}^2 - M_{i,2}. \tag{12}$$

Note that the resulting skewed distribution, $f(z|\xi, \eta)$ from Definition (2) is not centered at 0 and the variance is a function of $\xi$ (and, where is the case, of the shape parameter $\eta$). Given that the elements of $z^\star$ are uncorrelated (since those of $x$ are uncorrelated by assumption), standardization of $z^\star$ is achieved by the following transformation:

$$z = (z^\star - m)./s \tag{13}$$

where $m = (m_1, ..., m_N)$ and $s = (s_1, ..., s_N)$ are the vectors of unconditional means and standard deviations of $z^\star$ computed in Equations (11) and (12) respectively and "./" denotes element-by-element division. Consequently, the standardized form of Definition (2) requires replacing Equations (7) and (8) with

$$z_i^\star = (s_i z_i + m_i)\xi_i^{I_i} \tag{14}$$

and

$$I_i = \begin{cases} -1 \text{ if } z_i \geq -\frac{m_i}{s_i} \\ 1 \text{ if } z_i < -\frac{m_i}{s_i} \end{cases} \tag{15}$$

## 2.3 Distributions

This section introduces the different distribution assumptions to be incorporated into the likelihood function. Estimation of the parameters is performed in one step by Maximum Likelihood (ML). Namely, the log-likelihood function for $T$ observations is expressed as

$$\ell_T(\psi) = \sum_{t=1}^{T} \log f(y_t|\psi, \mathcal{F}_{t-1}) \tag{16}$$

where $\psi$ is the finite-dimensional vector of model parameters and $f(y_t|\psi, \mathcal{F}_{t-1})$ denotes the assumed conditional distribution of returns. Herein, three symmetric and three asymmetric multivariate distributions will be considered. They are briefly recalled in the following. For sake of brevity, we only report the log-likelihood functions and the formulas for the moments, when needed. A detailed description of their algebraic derivations can be found in Appendix A.2.

**Multivariate normal distribution** This is the most commonly employed distribution in the literature as it is uniquely identified by its conditional first and second moments, which renders ML estimation much simpler from a computational point of view. Also, given that the score of the normal log-likelihood function has the martingale difference property when the first two conditional moments are correctly specified, the Quasi Maximum Likelihood (QML) estimates are still consistent and asymptotically normal even if the true DGP is *not* normally-distributed (Bollerslev & Wooldridge 1992). The log-likelihood function, up to a constant, is expressed as follows

$$\ell_T(\psi) = -\frac{1}{2}\sum_{t=1}^{T}\left[\log|H_t| + y_t' H_t^{-1} y_t\right]. \tag{17}$$

**Multivariate Student distribution** The Student distribution is a symmetric and bell-shaped distribution, with heavier tails than the normal. Under the multivariate Student assumption, the log-likelihood function is obtained as

$$\begin{aligned}\ell_T(\psi) &= -\frac{1}{2}\sum_{t=1}^{T}\left[\log|H_t| + (N+\nu)\log\left(1+\frac{y_t' H_t^{-1} y_t}{\nu-2}\right)\right] \\ &+ T\left[\log\Gamma\left(\frac{\nu+N}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{N}{2}\log(\nu-2)\right]\end{aligned} \tag{18}$$

where $\Gamma(\nu) = \int_0^\infty e^{-z} z^{\nu-1} dz$ denotes the Gamma function and $\nu > 2$ is the degree of freedom parameter representing the thickness of the distribution tails. As $\nu$ increases, the distribution converges to the multivariate normal.

**Multivariate Exponential Power (MEP) distribution** This distribution belongs to the Kotz family of distributions (a particular class of symmetric and elliptical distributions discussed extensively in Fang et al. (1990)) and is known to have several equivalent definitions in the literature. It can also include both the normal and the Laplace as special cases, as a function of the value of the non-normality parameter $\beta$ dictating the tail-behaviour of the distribution. Given its simple implementation, in this paper we consider the pdf given in Solaro

(2004), which gives rise to the following log-likelihood function:

$$\ell_T(\psi) = -\frac{1}{2}\sum_{t=1}^{T}\left[\log|H_t| + \left(y_t' H_t^{-1} y_t\right)^{\frac{\beta}{2}}\right] \tag{19}$$

$$- \; T\left[\log\Gamma\left(1+\frac{N}{\beta}\right) + \left(1+\frac{N}{\beta}\right)\log(2)\right] \tag{20}$$

where $\beta > 0$. When $\beta = 2$, the distribution reduces to the multivariate normal, while for $\beta = 1$ it corresponds to the multivariate Laplace. Whenever $\beta < 2$ ($> 2$), the distribution exhibits thicker (thinner) tails than the normal.

**Multivariate skew-normal distribution** Is the first non-symmetric distribution we consider herein; it accounts for the skewness of the return distribution without taking into account its kurtosis (as it does not involve a tail parameter). By means of Equations (9)–(12) and considering the univariate normal density function (i.e. assuming $N = 1$), its first and second order moments are respectively obtained as:

$$m_i = \sqrt{\frac{2}{\pi}}\left(\xi_i - \frac{1}{\xi_i}\right) \tag{21}$$

$$s_i^2 = \left(\xi_i^2 + \frac{1}{\xi_i^2} - 1\right) - m_i^2 \tag{22}$$

Applying *Definition* 2 we derive the skew-normal density function, with corresponding log-likelihood function equal to

$$\ell_T(\psi) = -\frac{1}{2}\sum_{t=1}^{T}\left[\log|H_t| + \sum_{i=1}^{N}\left(s_i\sum_{j=1}^{N} p_{ijt}y_{jt} + m_i\right)^2 \xi_i^{2I_i}\right] \tag{23}$$

$$+ \; T\left[\sum_{i=1}^{N}(\log\xi_i + \log s_i) - \log(1+\xi_i^2)\right]$$

where $p_{ijt}$ corresponds to the $j^{th}$ element of the $i^{th}$ row of $H_t^{-1/2}$ (The full derivation is provided in Appendix A.1), $\xi_i$ represents the asymmetry of each marginal and $I_i$ is defined as in Eq. (15).

**Multivariate skew-Student distribution** With the same procedure as for the skew-normal, the following equations describe the first and second order moments of the multivariate skew-Student distribution:

$$m_i = \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{\nu-1}}{\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(\xi_i - \frac{1}{\xi_i}\right) \tag{24}$$

$$s_i^2 = \left(\xi_i^2 + \frac{1}{\xi_i^2} - 1\right) - m_i^2 \tag{25}$$

The log-likelihood function for $T$ observations is given by the following expression

$$\begin{aligned}
\ell_T(\psi) &= -\sum_{t=1}^{T}\left[\frac{1}{2}\log|H_t| + \frac{\nu+N}{2}\log\left(1+\frac{\sum_{i=1}^{N}\left(s_i\sum_{j=1}^{N}p_{ijt}y_{jt}+m_i\right)^2\xi_i^{2I_i}}{\nu-2}\right)\right] \\
&+ T\left[\sum_{i=1}^{N}(\log\xi_i+\log s_i)-\log(1+\xi_i^2)\right] \\
&+ T\left[\log\Gamma\left(\frac{\nu+N}{2}\right)-\log\Gamma\left(\frac{\nu}{2}\right)-\log(\nu-2)\right]
\end{aligned} \tag{26}$$

where the parameter $\nu$ dictates the thickness of the tails and $\xi_i$ is again the asymmetry parameter of each marginal. Notice that the univariate means and standard deviations are functions of $\xi_i$ and $\nu$ and need not be estimated. Thus the skew-Student parametrization requires $N+1$ parameters to be estimated in addition to those stemming from the BEKK specification.

**Multivariate skew-MEP distribution** Finally, we consider a skew generalization of the multivariate MEP distribution which accounts for both heavy tails and skewness. Its first and second moments are obtained as

$$m_i = \frac{2^{-1+\frac{1}{\beta}}\Gamma\left(\frac{2+\beta}{\beta}\right)}{\Gamma\left(1+\frac{1}{\beta}\right)}\left(\xi_i-\frac{1}{\xi_i}\right) \tag{27}$$

$$s_i^2 = \frac{4^{\frac{1}{\beta}}\Gamma\left(\frac{3}{\beta}\right)}{\beta\Gamma\left(1+\frac{1}{\beta}\right)}\left(\xi_i^2+\frac{1}{\xi_i^2}-1\right)-m_i^2 \tag{28}$$

while the log-likelihood function to be maximized is equal to

$$\begin{aligned}
\ell_T(\psi) &= -\frac{1}{2}\sum_{t=1}^{T}\left[\log|H_t| + \left(\sum_{i=1}^{N}\left(s_i\sum_{j=1}^{N}p_{ijt}y_{jt}+m_i\right)^2\xi_i^{2I_i}\right)^{\frac{\beta}{2}}\right] \\
&+ T\left[\sum_{i=1}^{N}(\log\xi_i+\log s_i)-\log(1+\xi_i^2)\right] \\
&- T\left[\log\left(1+\frac{N}{\beta}\right)+\left(1+\frac{N}{\beta}\right)\log(2)\right].
\end{aligned} \tag{29}$$

where $\beta$ is a parameter determining tail-thickness of the density function, as in the symmetric case.

# 3 Empirical Application

## 3.1 Data and forecasting scheme

Our dataset (cleaned and used in the paper of Noureldin et al. (2012))[1] comprises daily open-to-close returns of 10 stocks from the Dow Jones Industrial Average: Bank of America (BAC), JP Morgan (JPM), International Business Machines (IBM), Microsoft (MSFT), Exxon Mobil (XOM), Alcoa (AA), American Express (AXP), Du Pont (DD), General Electric (GE) and Coca Cola (KO). Each univariate vector of returns is calculated as $y_t = 100 \times (\log p_t - \log p_{t-1})$ and covers a period of 2200 days, from February 2001 to November 2009. Some useful univariate descriptive statistics over the period of interest can be found in Table 1.

A preliminary inspection of the normality assumption of each series is conducted by means of two nonparametric tests, the Kolmogorov-Smirnov (KS) and the Jacques-Bera (JB) test. Their p-values are reported in the last two columns of Table 1. The KS test rejects the normality hypothesis in the vast majority of cases, with the only exceptions represented by the XOM and DD stock over the estimation sample and the KO stock during the forecasting period. The JB test builds directly on the values of skewness and kurtosis of each asset and thus rejects the normality hypothesis in all cases. Indeed, the striking feature emerging from the table is that the univariate series exhibit thick tails vis-à-vis the normal (since the kurtosis is much greater than three) and a mainly positive level of skewness over the full-sample period.

This evidence already supports the need to use distributional assumptions that are able to account for these features. More precisely, we are interested in assessing if the inclusion of more flexible distributions than the normal can lead to significant improvements in the model forecasting ability.

To this extent, one-step ahead forecasts of the conditional covariance matrix of returns need to be computed. They are recursively obtained as

$$\hat{H}_{t+1|t} = E(H_{t+1}|I_t),$$

where $I_t$ denotes the information set at time $t$ and $H_t$ is defined as in Eq. (3).

Using a rolling-fixed-window scheme, parameters are estimated over a window length of 1500 observations and used to predict the conditional covariance matrix process for the following 20 days. Each time the window is shifted forward by 20 observations and the parameters are re-estimated over the new period in order to compute the next set of forecasts. We iterate this process till the end of the dataset for a total of 35 parameter estimates and 700 one-step ahead

[1]Downloaded from `http://realized.oxford-man.ox.ac.uk/data/download`.

Table 1: Univariate descriptive statistics

| Stock | Mean | Std.dev. | Skewness | Kurtosis | KS test | JB test |
|---|---|---|---|---|---|---|
| Estimation sample: February 1, 2001 to January 23, 2007 (1500 observations) | | | | | | |
| BAC | 0.09 | 1.09 | -0.18 | 7.45 | 0.00 | 0.00 |
| JPM | 0.00 | 1.68 | 0.90 | 31.02 | 0.00 | 0.00 |
| IBM | -0.04 | 1.24 | 0.01 | 5.96 | 0.01 | 0.00 |
| MSFT | -0.01 | 1.37 | 0.37 | 6.01 | 0.00 | 0.00 |
| XOM | -0.01 | 1.13 | 0.05 | 8.27 | 0.82 | 0.00 |
| AA | 0.01 | 1.59 | 0.14 | 4.74 | 0.00 | 0.00 |
| AXP | -0.02 | 1.44 | 0.33 | 7.73 | 0.00 | 0.00 |
| DD | 0.02 | 1.21 | 0.37 | 6.76 | 0.21 | 0.00 |
| GE | -0.01 | 1.34 | 0.13 | 7.90 | 0.02 | 0.00 |
| KO | 0.01 | 0.99 | 0.16 | 5.53 | 0.00 | 0.00 |
| Forecasting sample: January 24, 2007 to October 30, 2009 (700 observations) | | | | | | |
| BAC | -0.18 | 3.95 | 0.37 | 9.36 | 0.00 | 0.00 |
| JPM | 0.01 | 3.06 | 0.36 | 8.53 | 0.00 | 0.00 |
| IBM | 0.08 | 1.45 | -0.02 | 6.31 | 0.00 | 0.00 |
| MSFT | 0.02 | 1.60 | 0.08 | 5.90 | 0.00 | 0.00 |
| XOM | 0.03 | 1.61 | -0.39 | 11.31 | 0.00 | 0.00 |
| AA | -0.04 | 2.93 | -0.83 | 7.50 | 0.00 | 0.00 |
| AXP | 0.04 | 3.06 | 0.22 | 6.96 | 0.00 | 0.00 |
| DD | -0.04 | 1.89 | -0.12 | 5.70 | 0.00 | 0.00 |
| GE | 0.02 | 2.17 | 0.21 | 8.96 | 0.00 | 0.00 |
| KO | -0.03 | 1.22 | 0.07 | 7.68 | 0.06 | 0.00 |
| Full sample: February 1, 2001 to October 30, 2009 (2200 observations) | | | | | | |
| BAC | 0.01 | 2.40 | 0.33 | 21.72 | 0.00 | 0.00 |
| JPM | 0.00 | 2.21 | 0.57 | 16.90 | 0.00 | 0.00 |
| IBM | 0.00 | 1.31 | 0.02 | 6.24 | 0.02 | 0.00 |
| MSFT | 0.00 | 1.45 | 0.25 | 6.08 | 0.00 | 0.00 |
| XOM | 0.00 | 1.30 | -0.20 | 11.56 | 0.04 | 0.00 |
| AA | 0.00 | 2.11 | -0.69 | 9.95 | 0.00 | 0.00 |
| AXP | 0.00 | 2.09 | 0.32 | 11.23 | 0.00 | 0.00 |
| DD | 0.00 | 1.46 | 0.03 | 7.25 | 0.00 | 0.00 |
| GE | 0.00 | 1.65 | 0.22 | 10.85 | 0.00 | 0.00 |
| KO | 0.00 | 1.07 | 0.11 | 6.89 | 0.00 | 0.00 |

Descriptive statistics of the stock return time series used in the empirical application. The three panels report the statistics for the in-sample period, the out-of-sample period and the full sample period, respectively. 'KS test' and 'JB test' denotes the Kolmogorov-Smirnov test and Jarque Bera test, with corresponding p-values in column.

forecasts. Table B1 in Appendix B reports the complete list of windows and forecast horizons along with their corresponding calendar dates.

The canonical approach to portfolio construction involving the minimization of the portfolio variance for a given expected return relies on the assumption of normally-distributed returns (Michaud 1989). However, as the assumption of non-normality in our paper precludes the use of the mean-variance minimization framework, we consider the equal-weighting scheme as the most appropriate choice. This has the advantage of not being affected by the specified target return as in the Markowitz framework, being only driven by the number of assets.

Thus, given the $N$-dimensional vector of weights $w = (w_1 \dots w_N)$, where $w_i = 1/N$ and $\sum_{i=1}^{N} w_i = 1$, portfolio returns and standard deviations can be respectively computed as:

$$\begin{aligned} r^p_{t+1} &= w' y_{t+1} \\ \hat{\sigma}^p_{t+1} &= \sqrt{w' \hat{H}_{t+1|t} w}, \end{aligned} \tag{30}$$

where $y_{t+1}$ denotes the $N$-dimensional vector of daily returns and $\hat{H}_{t+1|t}$ is the predicted covariance matrix of returns conditional on past information.

For each model, the portfolio VaR at confidence levels $\alpha = 5\%$ and $1\%$ is equal to

$$VaR_{t+1,\alpha} = \hat{\sigma}^p_{t+1} q_\alpha, \tag{31}$$

where $q_\alpha$ is the *left* quantile of the assumed distribution at $\alpha\%$. This implies that the predictive power of the model is linked to its ability in modeling large negative returns.

Note that the analytical formula applied for the computation of the VaR is simplified to only account for the portfolio conditional variance. Alternative approaches, as done in Bauwens et al. (2006), also assume an ARMA-type structure for the portfolio conditional mean. Ultimately we deal with *de-meaned* returns and thus specifying a more complex VaR model goes beyond the scope of this paper.

For the symmetric distributions in our analysis (normal, Student and MEP), one can easily pass from the conditional covariance matrix to the long VaR of the portfolio by applying Eq. (31) and the inverse of each CDF at $\alpha\%$.

However, for the non-symmetric distributions this is not straightforward. In order to bypass this complication, for each non-symmetric distribution we apply a simple Monte-Carlo simulation approach. Namely, we draw $j = 10.000$ random vectors from each symmetric multivariate standardized distribution $z_t$ and then we use the estimated skewness parameters to construct the corresponding skewed distribution $z_t^\star$. By assuming $r_j = \hat{H}^{1/2}_{t|t-1} z_j^\star$ as the true DGP, we obtain a set of 10,000 simulated returns over the period of interest. Finall, the simulated return distribution is used to derive the 5 and 1% quantiles for computing the VaR.

### 3.2 Testing the accuracy of VaR forecasts

The models accuracy in predicting VaR is assessed using multiple statistical backtesting methods. A common starting point for this procedure is the so-called hit function, or indicator function, which is equal to

$$I_t(\alpha) = \begin{cases} 1 \text{ if } r_t \leq VaR(\alpha) \\ 0 \text{ if } r_t > VaR(\alpha) \end{cases} \tag{32}$$

i.e. it takes the value one if the ex-post portfolio loss exceeds the VaR predicted at time $t-1$ and the value zero otherwise. According to Christoffersen (1998), in order to be accurate, the hit sequence has to satisfy two properties, namely the correct failure rate and the independence of exceptions. The former implies that the probability of realizing a VaR violation should be equal to $\alpha * 100\%$, while the latter further requires these violations to be independent of each other. These properties can be combined together into one single statement assessing that the hit function has to be an i.i.d. Bernoulli random variable with probability $p$, i.e. $I_t(p) \overset{i.i.d}{\sim} B(p)$. This represents the key foundation to many of the backtesting procedures developed in recent years and particularly to the accuracy tests being used in this paper. We focus on tests included in the following three categories:

- Evaluation of the Frequency of Violations
- Evaluation of the Independence of Violations
- Evaluation of the Duration between Violations.

Their properties are briefly described in the following paragraphs.

**Frequency of Violations** The first way of testing the VaR accuracy is to test the number or the frequency of margin exceedances. A test designed to this aim is the Kupiec test (Kupiec 1995), also known as the Unconditional Coverage (UC) test. Its null hypothesis is simply that the percentage of violated VaR forecasts or failure rate $p$ is consistent with the given confidence level $\alpha$, i.e. $H_0 : p = \alpha$.

Denoting with $F$ the length of the forecasting period and with $v$ the number of violations occurred throughout this period, the log-likelihood ratio test statistic is defined as

$$UC = -2\left(\ln\left(\frac{p^v(1-p)^{F-v}}{\hat{p}^v(1-\hat{p})^{F-v}}\right)\right), \tag{33}$$

where $\hat{p} = v/F$ is the maximum likelihood estimator under the alternative hypothesis. This ratio test statistic is asymptotically $\chi^2(1)$ distributed and the null hypothesis is rejected if the critical value at the $\alpha\%$ confidence level is exceeded.

A similar useful test is the TUFF (Time Until First Failure) test. Under the null, the probability of an exception is equal to the inverse probability of the VaR confidence level, namely $H_0 : p = \hat{p} = 1/v$. Its basic assumptions are similar to those of the Kupiec test and the t-statistic under the null is obtained as

$$TUFF = -2\left(\ln\left(\frac{p(1-p)^{v-1}}{\frac{1}{v}\left(1-\frac{1}{v}\right)^{(v-1)}}\right)\right). \tag{34}$$

The TUFF statistic is also asymptotically $\chi^2(1)$ distributed.

**Independence of Violations** A limitation of the Kupiec test is that it is only concerned with the coverage of the VaR estimates without accounting for any clustering of the violations. This aspect is crucial for VaR practitioners, as large losses occurring in rapid succession are more likely to lead to disastrous events than individual exceptions.

The Independence test (IND) of Christoffersen (1998) uses the same likelihood ratio framework as the previous tests but is designed to explicitly detect clustering in the VaR violations. Under the null hypothesis of independence, the IND test assumes that the probability of an exeedance on a given day $t$ is not influenced by what happened the day before. Formally, $H_0 : p_{10} = p_{11}$, where $p_{ij}$ denotes the probability of an $i$ event on day $t-1$ being followed by a $j$ event on day $t$. The relevant IND test statistic can be derived as

$$IND = -2\left(\ln\left(\frac{\hat{p}^v(1-\hat{p})^{F-v}}{\hat{p}_{11}^{v_{11}}(1-\hat{p}_{11})^{v_{01}}\hat{p}_{10}^{v_{10}}(1-\hat{p}_{10})^{v_{00}}}\right)\right) \tag{35}$$

where $v_{ij}$ is the number of violations with value $i$ at time $t-1$ followed by $j$ at time $t$. Under the null, the IND statistic is also asymptotically distributed as a $\chi^2(1)$ random variable. Although the aforementioned test has received support in the literature, Christoffersen (1998) noted that it was not complete on its own. For this reason, he proposed a joint test, the Conditional Coverage (CC) test, which combines the properties of both UC and IND tests. Formally, the CC ratio statistic can be proven to be the sum of the UC and the IND statistics:

$$\begin{aligned}
CC &= -2(ln(L_0^{UC)} - ln(L_1^{IND})) \\
&= -2(ln(L_0^{UC}) - ln(L_1^{UC}) + ln(L_1^{UC}) - ln(L_1^{IND})) \\
&= -2(ln(L_0^{UC}) - ln(L_1^{UC}) + ln(L_0^{IND}) - ln(L_1^{IND})) \\
&= -2\underbrace{(ln(L_0^{UC}) - ln(L_1^{UC}))}_{UC} -2\underbrace{(ln(L_0^{IND}) - ln(L_1^{IND}))}_{IND}
\end{aligned}$$

where we added and subtracted the quantity $\ln(L_1)^{UC}$ and substituted $\ln(L_1)^{UC}$ with $\ln(L_0)^{IND}$. CC is also $\chi^2$ distributed, but with two degrees of freedom since there are two separate statistics in the test. According to Campbell (2005), in some cases it is possible that a VaR model passes the joint test while still failing either the independence test or the unconditional coverage test. Thus it is advisable to run them separately even when the joint test yields a positive result.

A second test belonging to this class is the Regression-based test of Engle & Manganelli (2004), also known as Dynamic Quantile (DQ) test. Instead of directly considering the hit sequence, the test is based on its associated quantile process $H_t(\alpha) = I_t(\alpha) - \alpha$ which assumes the following values:

$$H_t(\alpha) = \begin{cases} 1 - \alpha \text{ if } I_t = 1 \\ -\alpha \text{ if } I_t = 0 \end{cases}$$

The idea of this approach is to regress current violations on past violations in order to test for different restrictions on the parameters of the model.

Namely, we estimate the linear regression model $H_t(\alpha) = \delta + \sum_{k=1}^{K} \beta_k H_{t-k}(\alpha) + \epsilon_t$ and then we test the joint hypothesis $H_0(DQ_{cc}) : \delta = \beta_1 = ... = \beta_K = 0$. This assumption coincides with the null of Christoffersen's CC test. It is also possible to split the test and separately test the independence hypothesis and the unconditional coverage hypothesis, respectively as $H_0(DQ_{ind}) : \beta_1 = ... = \beta_K = 0$ and $H_0(DQ_{uc}) : \delta = 0$. $(DQ_{cc}), (DQ_{ind})$ and $(DQ_{uc})$ are asymptotically $\chi^2$ distributed with respectively $\{K + 1\}$, $K$ and one degrees of freedom.

**Duration between Violations** One of the drawbacks of Christoffersen's CC test is that it is not capable of capturing dependence in all forms, since it only considers the dependence of observations between two successive days. To a further extent, Christofferson & Pelletier (2004) introduced the Duration-Based test of independence (DBI), which is an improved test for both independence and coverage. Its basic intuition is that if exceptions are completely independent of each other, then the upcoming VaR violations should be independent of the time that has elapsed since the occurrence of the last exceedance (Campbell 2005). The duration (in days) between two exceptions is defined via the no-hit-duration $D_i = t_i - t_{i-1}$, where $t_i$ is the day of $i$-th violation.

A correctly specified model should have an expected conditional duration of $1/p$ days and the no-hit duration should have no memory. The authors construct the ratio statistic considering different distributions for the null and the alternative hypotheses, namely the exponential, since it is the only memory-free (continuous) random distribution, and the Weibull, which allows for duration dependence. The likelihood ratio statistic is derived as

$$DBI = -2\left(\ln\left(\frac{L_0}{L_1}\right)\right) = -2\left(\ln\left(\frac{p\exp\{-pD\}}{a^b b D^{b-1}\exp\{-(aD)^b\}}\right)\right)$$

and has a $\chi^2$ distribution with one degree of freedom.

Under the null hypothesis of independent violations, $b = 1$ and $a$ is estimated via numerical maximization of $ln(L_1)$. Whenever $b < 1$, the Weibull function has a decreasing path which corresponds to an excessive number of very long durations (very calm period) while $b > 1$ corresponds to an excessive number of very short durations, namely very volatile periods.

# 4 Results

## 4.1 Parameter estimates

The in-sample window covers the period 2001/01 – 2007/01 for a total of 1500 daily observations. Results from the in-sample estimation are reported in Table 2.

Table 2: In-sample parameters estimates

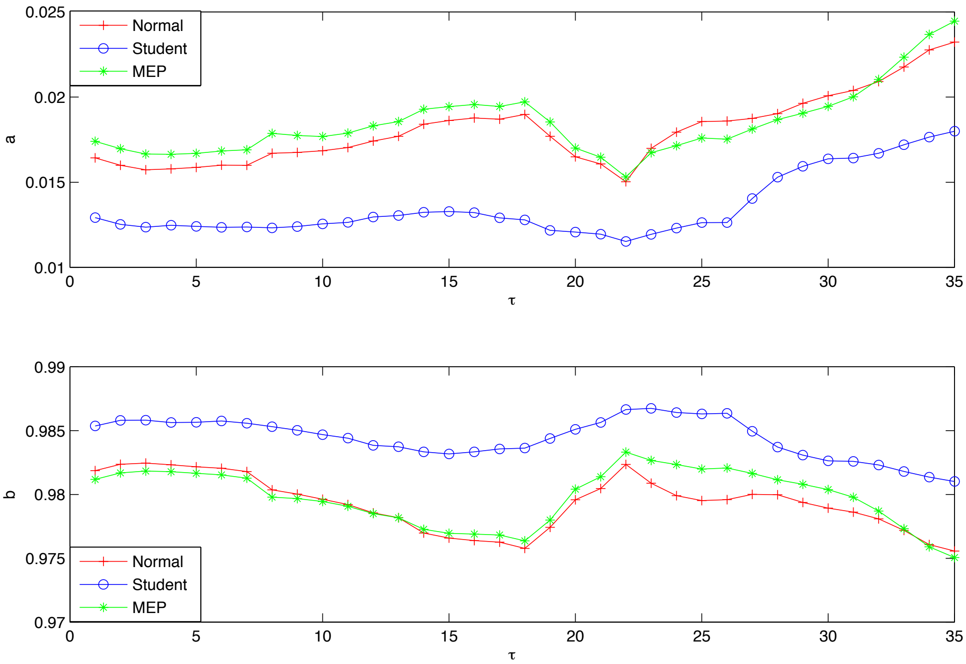
| | Normal | Student | MEP | Skew-normal | Skew-Student | Skew-MEP |
|---|---|---|---|---|---|---|
| a | 0.016 (0.00) | 0.012 (0.00) | 0.017 (0.00) | 0.016 (0.00) | 0.013 (0.00) | 0.014 (0.00) |
| b | 0.981 (0.04) | 0.985 (0.03) | 0.981 (0.04) | 0.982 (0.05) | 0.985 (0.03) | 0.985 (0.05) |
| $\nu$ | – | 9.71 (0.61) | – | – | 9.68 (0.60) | – |
| $\bar{\xi}$ | – | – | – | 1.021 (0.04) | 1.026 (0.04) | 0.998 (0.03) |
| $\beta$ | – | – | 1.96 (0.32) | – | – | 1.13 (0.25) |
| LogLik | 16423 | 19624 | 19611 | 16580 | 20511 | 20352 |
| BIC | -14.92 | -17.82 | -17.81 | -15.03 | -18.60 | -18.45 |

The table reports test statistics and robust standard errors obtained from the sBEKK model with the different distribution assumptions over the in-sample period 2001/01 – 2007/01, for T=1500. Note that the $\bar{\xi}$ parameters are averaged across univariate series and Mean Asymptotic Square Errors (MASE) are reported in brackets. The BIC is rescaled by T.

A common feature of the estimated models is that sums of $a$ and $b$ are never smaller than 0.997, thus showing a high level of persistence typical of GARCH-type models. More interestingly, the use of skewed distribution assumptions seem to be justified, as all asymmetric coefficients are significant at standard levels. Moreover, the Bayesian information criteria (BIC) and the log-likelihood values highlight the fact that the model incorporating the skew-Student and the skew-MEP distributions better fits the data than the model with the traditional normality assumption.

The estimated parameters over the out-of-sample period, $\tau = \{1, \ldots, 35\}$, are summarized by means of figures. A first interesting comparison is provided in Figure 1 between the estimated parameters of the BEKK model incorporating symmetric distribution assumptions. Parameter estimates from the normal and the MEP distributions show a similar pattern over time, suggesting that the conditional covariance matrices constructed from these models will exhibit similar temporal dynamics as well. As already mentioned, the MEP distribution collapses to a normal whenever $\beta = 2$. Figure 4 shows that this is indeed the case. By contrast, the $a$ and $b$ estimated parameters from the Student assumption have different values and a smoother temporal pattern, indicating that the use of a heavy-tailed distribution can affect the dynamics of the model.

Figure 1: BEKK parameter estimates: symmetric distributions




The introduction of skewness into the symmetric distributions significantly affects parameter estimates. As Figure 2 shows, the skew-normal and skew-MEP no longer display congruent dynamics, as the skew-MEP $a$ and $b$ estimates are now much closer in value to the skew-Student estimates. Indeed, analysis of the tail parameter in Figure 5 shows that the skew-MEP distribution is now closer to a Laplace distribution ($\beta \simeq 1$). We also report in Figure 3 the evolution of the skewness parameter $\xi$ for the three skewed distributions. The averages are computed across the 10 univariate series with corresponding ranges. Clearly, all distributions exhibit a positive level of skewness on average.

Figure 2: BEKK parameters: skewed distributions

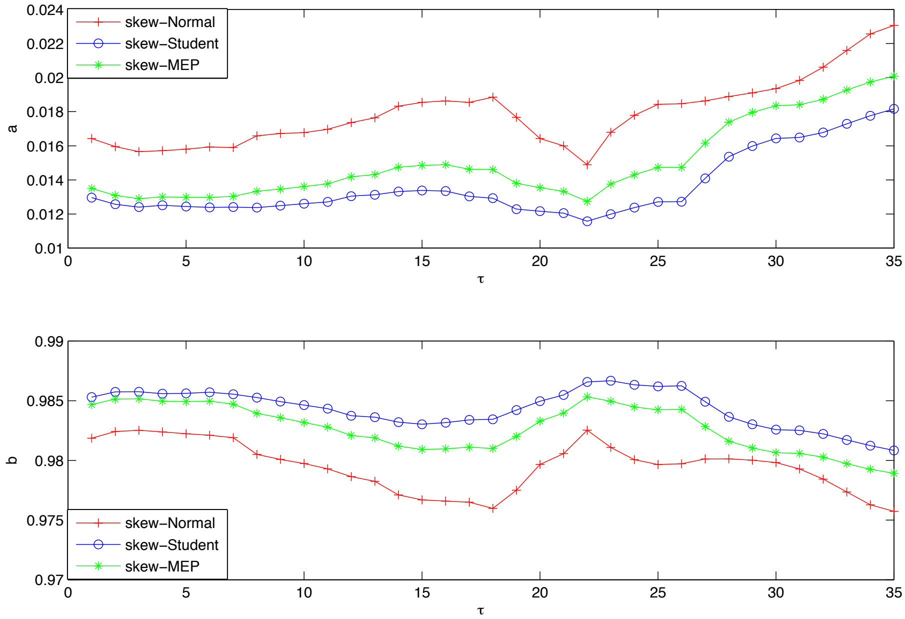



Figure 3: Skew parameter averages

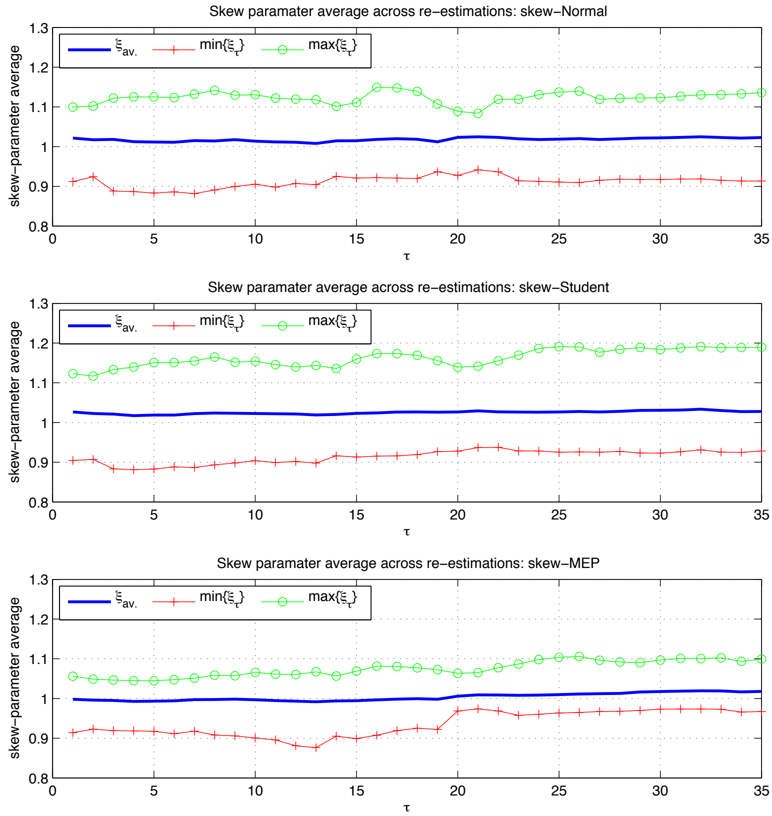


Figure 4: Tail parameter $\beta$: MEP distribution

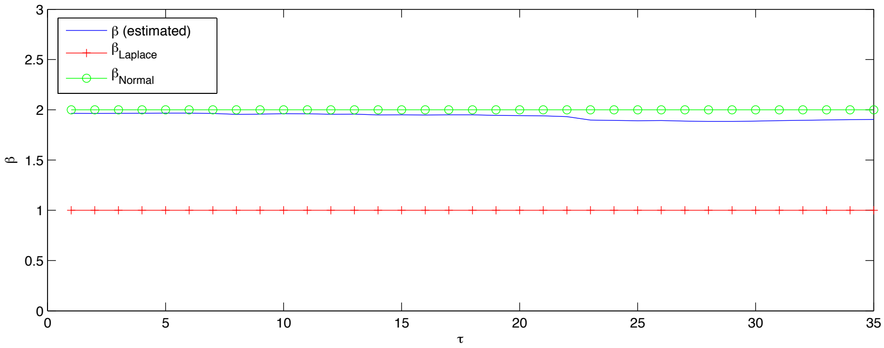
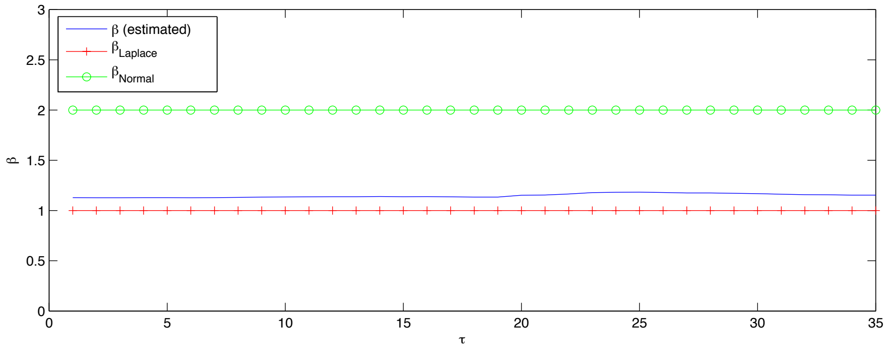
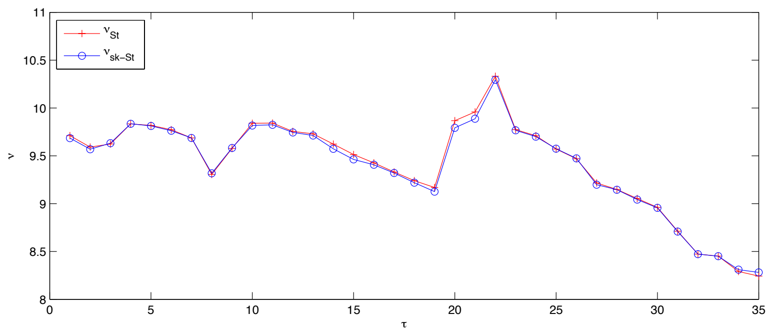



Figure 5: Tail parameter $\beta$: skew-MEP distribution




As a general finding, BEKK parameter estimates exhibit similar movements across time. Specifically, $a$ increases until $\tau = 18$, followed by a drop in value that occurs over re-estimations 18-22 after which it increases at a faster rate than before. Obviously the opposite effect is incurred for $b$ under all distribution assumptions. Consulting Table B1 in the appendix, we see that those windows include the periosd corresponding to the onset of the US subprime mortgage crisis. A similar effect is observed for the tail parameter of Student and skew-Student distributions (Figure 6); prior to the crisis, there was a gradual reduction in the tail-thickness of the returns distribution, followed by a sharp spike in $\nu_{St}$ and $\nu_{sk-St}$ as the rolling window begins to include the crisis period (during which there was a marked increase in the downside risk of assets, as shown in our results).

Figure 6: Tail parameter: Student and skew-Student distribution



## 4.2 Out-of-sample evidence

Given the set of estimated model parameters, a series of 700 conditional covariance forecasts are obtained. Each model one-step ahead covariance prediction, denoted as $\hat{H}_{t+1} = E(H_{t+1}|I_t)$, can be compared with the ex-post realization of the true conditional covariance matrix, denoted as $\Sigma_t$. Given that the latter is a latent object, we use an unbiased proxy represented by the 5-minutes realized covariance estimator, $\hat{\Sigma}_t$, which is proven to be a more efficient estimator than the one based on the outer product of returns under the assumption of absence of microstructure noise and other biases; see Barndorff-Nielsen & Shephard (2002) and Aït-Sahalia et al. (2005) among others.

We follow Ledoit et al. (2003) and assess the predictive accuracy of the models using the root-mean-square error (RMSE) based on the Frobenius norm of the forecast error.

This is computed by

$$F_{T_h} = \frac{1}{T_h} \sum_{t}^{T_h} ||\hat{\Sigma}_t - \hat{H}_t|| \tag{36}$$

where $T_h$ denotes the out-of-sample length.

Table 3 contains the results on the forecasting accuracy of the model incorporating the different distributions measured by the Frobenius norm. It appears that the sBEKK model with the Student distribution outperforms all the others, even if the improvement over the skew-Student is rather negligible. However, symmetric heavy-tailed distributions achieve smaller values of the average Frobenius norm than the normal and the inclusion of skewness leads to further improvements, as the skew-normal and the skew-MEP unequivocally outperform their symmetric counterparts.

Table 3: Evaluation of Forecasting Accuracy in Terms of RMSE

| Frobenius norm of forecast error | | | | | |
|---|---|---|---|---|---|
| Normal | Student | MEP | Skew-normal | Skew-Student | Skew-MEP |
| 44.58 | 44.09 | 44.56 | 44.57 | 44.1 | 44.47 |

Table reports the average Frobenius norm of the forecast error as given by Eq. (36).

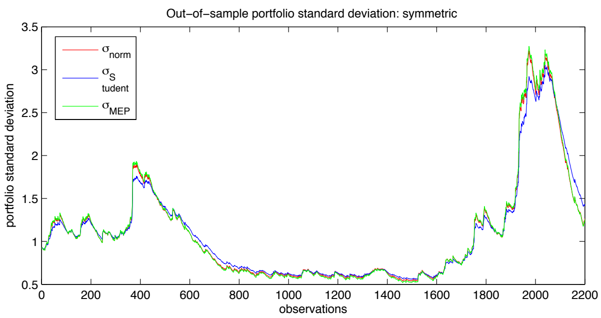
Finally, the out-of-sample covariance matrix predictions are used to construct equally-weighted portfolios for the computation of the daily VaR. Table 4 compares portfolios standard deviation for both the in- and out-of-sample periods.

Table 4: Portfolios descriptive statistics

| | Normal | Student | MEP | Skew-normal | Skew-Student | Skew-MEP |
|---|---|---|---|---|---|---|
| Estimation sample: February 1, 2001 to January 23, 2007 (1500 observations) | | | | | | |
| $\bar{\sigma}^p$ | 0.900 | 0.909 | 0.897 | 0.900 | 0.909 | 0.907 |
| $\min\{\sigma^p\}$ | 0.537 | 0.558 | 0.519 | 0.537 | 0.558 | 0.558 |
| $\max\{\sigma^p\}$ | 1.897 | 1.759 | 1.932 | 1.897 | 1.760 | 1.792 |
| Forecasting sample: January 24, 2007 to October 30, 2009 (700 observations) | | | | | | |
| $\bar{\sigma}^p$ | 1.473 | 1.454 | 1.487 | 1.472 | 1.454 | 1.463 |
| $\min\{\sigma^p\}$ | 0.537 | 0.558 | 0.519 | 0.537 | 0.558 | 0.558 |
| $\max\{\sigma^p\}$ | 3.221 | 3.052 | 3.272 | 3.219 | 3.053 | 3.120 |

Table reports average, minimum and maximum value of portfolio standard deviation over the in- and the out-of-sample period.

As already noted, the financial crisis features heavily in the summary statistics. Since this period is included in the forecasting sample (corresponding to observations 1921-1940 according to Table B1), we notice a sharp increase in the portfolio standard deviation of all the models (see Figure 7). Apparently, the heavy-tailed *and* skewed distributions (skew-Student, skew-MEP) have a slightly higher average portfolio variance than the thin-tailed distributions in the in-sample period. This pattern is reversed in the forecasting period, as the skew-Student and skew-MEP exhibit a lower portfolio standard deviation than their symmetric counterparts.



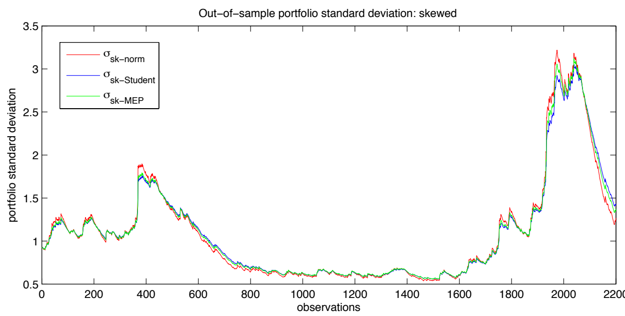




Figure 7: Portfolio standard deviation for the sBEKK model with symmetric distributions (left figure) and skewed distributions (right figure).

## 4.3 VaR backtesting results

Table 5 reports the results from the UC, TUFF, IND, CC and DBI tests while Table 6 contains results from the DQ test. All statistical tests are computed for the 5 and 1% VaR confidence level. We report test statistics along with their corresponding p-values in brackets. Since the applied tests measure the models accuracy in forecasting VaR along several dimensions (as detailed in Section 3.2), the overall results are summarized using a performance measure which considers the percentage of acceptances of the null hypothesis across the different tests.

Table 5: VaR backtesting results

| | Normal | Student | MEP | Skew-normal | Skew-Student | Skew-MEP |
|---|---|---|---|---|---|---|
| | | | 5% VaR | | | |
| # violation/frequency | 53<br>0.075 | 40<br>0.057 | 52<br>0.074 | 50<br>0.071 | 38<br>0.054 | 35<br>0.050 |
| UC | 8.475<br>(**0.003**) | 0.720<br>(0.396) | 7.611<br>(**0.005**) | 6.008<br>(**0.014**) | 0.263<br>(0.607) | 1.000<br>(0.000) |
| TUFF | 5.991<br>(**0.014**) | 0.021<br>(0.883) | 5.991<br>(**0.014**) | 5.991<br>(**0.014**) | 0.021<br>(0.883) | 0.021<br>(0.883) |
| IND | 9.861<br>(**0.001**) | 0.763<br>(0.380) | 8.828<br>(**0.000**) | 6.921<br>(**0.008**) | 0.273<br>(0.601) | 0.054<br>(0.814) |
| CC | 18.336<br>(**0.000**) | 1,483<br>(0.476) | 16.440<br>(**0.000**) | 12.929<br>(**0.0016**) | 0.536<br>(.764) | 0.054<br>(0.972) |
| DBI | 0.857<br>(0.354) | 1.499<br>(0.221) | 0.796<br>(0.372) | 1.776<br>(0.182) | 2.624<br>(0.105) | 2.535<br>(0.111) |
| Grade | 20% | 100% | 20% | 20% | 100% | 100% |
| | | | 1% VaR | | | |
| # violation/frequency | 19<br>0.027 | 9<br>0.012 | 17<br>0.024 | 19<br>0.027 | 7<br>0.010 | 8<br>0.011 |
| UC | 14.15<br>(**0.002**) | 0.529<br>(0.466) | 10.31<br>(**0.001**) | 14.15<br>(**0.000**) | 0.000<br>(1.000) | 0.137<br>(0.710) |
| TUFF | 1.425<br>(0.232) | 1.425<br>(0.232) | 1.425<br>(0.232) | 1.425<br>(0.232) | 1.425<br>(0.232) | 1.425<br>(0.232) |
| IND | 0.054<br>(**0.000**) | 0.770<br>(0.382) | 10.978<br>(**0.003**) | 14.626<br>(**0.000**) | 0.141<br>(0.706) | 0.326<br>(0.568) |
| CC | 28.72<br>(**0.000**) | 1.299<br>(0.522) | 21.291<br>(**0.000**) | 28.779<br>(**0.000**) | 0.141<br>(0.931) | 0.464<br>(0.793) |
| DBI | 3.108<br>(0.077) | 0.636<br>(0.424) | 1.573<br>(0.209) | 2.164<br>(0.141) | 0.331<br>(0.564) | 0.282<br>(0.595) |
| Grade | 40% | 100% | 40% | 40% | 100% | 100% |

The table reports statistics and corresponding p-values obtained from the statistical backtesting tests described in Section 4.3. VaR computed at 5% and 1% confidence levels. Rejections of the null highlighted in bold.

According to Table 5, at both confidence levels the BEKK model with the Student assumption outperforms the other symmetric distributions which appear to be rejected in a vast majority of cases. Even if we turn to the skewed distributions, the heavy-tailed skew-Student and skew-MEP (recall that the skew-MEP approximates the Laplace, which is a heavy-tailed distribution) perform better than the model under the skew-normal assumption.

This suggests that the inclusion of heavy-tails in the distribution specification already allows for a significant improvement in the VaR forecasts accuracy.

By contrast, moving from symmetric to skewed distributions yields ambiguous results. Clearly, a more pronounced effect is observed in the MEP case, while the transition from normal to skew-normal does not result in an increase of the grade. This might suggest that incorporating skewness alone *without* allowing for heavy-tails is not sufficient for increasing the model accuracy. However, though moving from the Student to the skew-Student distribution does not increase the overall grade, closer inspection of the p-values shows that, in 3/5 cases, the results for the Student distribution are closer to the critical value at the 5% level (this increases to 4/5 cases at the 1% level). This suggests that when computing VaR for extreme events, i.e. much further in the tail than 5 and 1%, including skewness would improve the accuracy of VaR forecasts.

These findings are further confirmed by looking at the results of the DQ test for 1 and 2 VaR lagged values reported in Table 6.

Table 6: Dynamic Quantile test results

| | Normal | Student | MEP | Skew-normal | Skew-Student | Skew-MEP |
|---|---|---|---|---|---|---|
| | | | K=1 | | | |
| | | | 5% VaR | | | |
| $DQ_{UC}$ | 10.48<br>(**0.001**) | 0.775<br>(0.378) | 9.319<br>(0.002) | 7.204<br>(**0.007**) | 0.280<br>(0.596) | 0.001<br>(0.991) |
| $DQ_{IND}$ | 1.65<br>(0.198) | 0.030<br>(0.861) | 1.433<br>(0.231) | 0.306<br>(1.046) | 0.001<br>(0.992) | 0.057<br>(0.810) |
| $DQ_{CC}$ | 11.463<br>(**0.003**) | 0.798<br>(0.670) | 10.188<br>(**0.006**) | 7.868<br>(**0.019**) | 0.280<br>(0.869) | 0.057<br>(0.971) |
| | | | 1% VaR | | | |
| $DQ_{UC}$ | 13.358<br>(**0.000**) | 0.598<br>(0.439) | 13.358<br>(**0.000**) | 19.552<br>(**0.000**) | 0.001<br>(0.996) | 0.150<br>(0.697) |
| $DQ_{IND}$ | 1.278<br>(0.258) | 0.152<br>(0.696) | 2.095<br>(0.147) | 1.502<br>(0.220) | 0.071<br>(0.789) | 0.107<br>(0.743) |
| $DQ_{CC}$ | 22.121<br>(**0.000**) | 0.736<br>(0.691) | 16.57<br>(**0.003**) | 22.34<br>(**0.000**) | 0.071<br>(0.964) | 0.254<br>(0.880) |
| | | | K=2 | | | |
| | | | 5% VaR | | | |
| $DQ_{UC}$ | 8.568<br>(**0.003**) | 0.731<br>(0.392) | 7.508<br>(**0.006**) | 5.623<br>(**0.017**) | 0.268<br>(0.604) | 0.001<br>(0.976) |
| $DQ_{IND}$ | 0.014<br>(0.904) | 0.672<br>(0.412) | 0.069<br>(0.792) | 0.299<br>(0.584) | 1.166<br>(0.280) | 2.160<br>(0.141) |
| $DQ_{CC}$ | 12.202<br>(**0.006**) | 2.595<br>(0.458) | 11.324<br>(**0.010**) | 9.879<br>(**0.019**) | 2.659<br>(0.447) | 3.476<br>(0.323) |
| | | | 1% VaR | | | |
| $DQ_{UC}$ | 16.440<br>(**0.000**) | 0.490<br>(0.483) | 12.406<br>(**0.000**) | 16.524<br>(**0.000**) | 0.000<br>(0.995) | 0.121<br>(0.728) |
| $DQ_{IND}$ | 10.288<br>(**0.001**) | 3.396<br>(0.065) | 4.039<br>(**0.044**) | 11.402<br>(**0.000**) | 5.440<br>(**0.021**) | 4.308<br>(**0.0378**) |
| $DQ_{CC}$ | 34.002<br>(**0.000**) | 9.598<br>(**0.022**) | 18.568<br>(**0.000**) | 35.283<br>(**0.000**) | 12.655<br>(**0.005**) | 10.772<br>(**0.013**) |

The table reports statistics and corresponding p-values obtained from the Dinamic Quantile (DQ) tests with number of lags $K = 1, 2$ as described in Section 4.3. VaR computed at 5% and 1% confidence levels. Rejections of the null highlighted in bold.

As opposed to other backtesting methods, the DQ test takes into account a more general temporal dependence between the series of violations and is considered the most reliable in assessing the VaR accuracy. For both regression specifications, the normal, skew-normal and MEP distributions again underperform compared to the other distributions, mostly due to failures of the Unconditional (UC) and Conditional Coverage (CC) hypothesis. Despite the fact that the $DQ_{CC}$ nests the $DQ_{UC}$ and $DQ_{IND}$ tests, the latter is passed in all cases for $K = 1$, indicating that VaR violations are not dependent over time. By augmenting the number of lags to $K = 2$ and moving to the most extreme quantile, the Student distribution is the only one to pass the test at the 5% level (skew-Student and the skew-MEP not rejected at the 1% level). However, in this setting the overall performance of the models is found to be considerably inferior as they all fail the $DQ_{CC}$ test at the 5% level.

As already outlined by the previous tests, transforming from a normal to a skew-normal distribution does not affect the grade. By contrast, moving from a normal-approximating MEP ($B \simeq 2$) to a Laplace-approximating skew-MEP ($B \simeq 1$) results in a remarkably better performance of the model. This may lend further support to the notion that inclusion of a heavy-tailed distribution assumption is crucial in constructing accurate VaR forecasts.
To conclude, while the empirical application provide a clear evidence that the thin-tailed distributions deliver poor VaR forecasts compared to the corresponding heavy-tailed and skewed counterparts, it is not possible to fully assess weather the inclusion of skewness on top of heavy-tails is strictly necessary to improve the models forecasting accuracy.

# 5 Concluding remarks

As empirical evidence suggests, financial asset returns are conditionally heteroskedastic and generally non-normally distributed, fat-tailed and often skewed. It is also widely known that financial volatility tends to move together across assets and markets, exhibiting strong comovements over time. This requires an accurate modeling of the time-varying covariances of asset returns, which is at least as challenging as modeling univariate volatility alone. On the contrary, usual practice is to relying on multivariate GARCH specifications coupled with the normality assumption of the return distribution, which does not accommodate the stylized facts listed above and can have serious implications for portfolio diversification and risk management.
In this article we examined the economic and statistical impact of using a more flexible distribu-

tion model for asset allocation decisions in an out-of-sample setting. Specifically, we estimated a multivariate BEKK model coupled with three symmetric and three skewed distributional assumptions (i.e. normal, Student, MEP and their skewed counterparts) and evaluated the models accuracy in predicting equally-weighted portfolio Value-at-Risk (VaR).

We employed a series of standard backtesting methods to compare the distribution-based model performance and they unanimously showed that the inclusion of a *heavy-tailed* distribution is crucial for constructing accurate VaR forecasts, while the further addition of skewness fails to make a significant difference. This is shown in the large improvement in all test results when moving from a MEP to a skew-MEP distributional assumption compared to the marginal difference when moving from the Student to skew-Student distribution. However, we also found evidence that introducing skewness could lead to improvements in VaR forecast accuracy for extreme events located further than standard 5 and 1% confidence level in the left-tail of the returns distribution. This may warrant further investigations.

There are several possible avenues of research extending from this work. First, we only dealt with the BEKK parametrization. In spite of the multiple advantages of this model, an extension to multivariate GARCH specifications that also consider asymmetric past return-to-volatility feedbacks could lead to interesting results. Another possibility would be to consider higher forecast horizons for the VaR in order to check if the inclusion of skewness and asymmetric forms of dependence can lead to significant improvements in the long run. Finally, in a VaR perspective, despite the fact that the quantile regression method represents a marked improvement over the existing backtesting alternatives, other methods could also be investigated. For example, extreme value theory-based approaches which focus only on the tails of the returns distribution, represent already a valid starting point in this direction.

# References

Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2005), 'How often to sample a continuous-time process in the presence of market microstructure noise', *Review of Financial studies* **18**(2), 351–416.

Angelidis, T., A., B. & Degiannakis, S. (2004), 'The use of garch models in var estimation', *Statistical Methodology* **1**, 105–128.

Barndorff-Nielsen, O. E. & Shephard, N. (2002), 'Estimating quadratic variation using realized variance', *Journal of Applied Econometrics* **17**(5), 457–477.

Bauwens, L. & Laurent, S. (2005), 'A new class of multivariate skew densities, with application to garch models', *Journal of Business and Economic Statistics* **23**, 346–354.

Bauwens, L., Omrane, W. & Rengifo, E. (2006), 'Intra-daily fx optimal portfolio allocation', *CORE Discussion Paper 2006/10* .

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**, 307–327.

Bollerslev, T. (1990), 'Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model', *The Review of Economics and Statistics* pp. 498–505.

Bollerslev, T. & Wooldridge, J. (1992), 'Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances', *Econometric Reviews* **1**, 143–172.

Campbell, S. (2005), 'A review of backtesting and backtesting procedures', *Finance and Economics Discussion Series, Division of Research and & Statistics and Monetary Affairs, Federal Reserve Board* .

Christoffersen, P. (1998), 'Evaluating interval forecasts', *International Economic Review* **39**, 841–862.

Christofferson, P. & Pelletier, D. (2004), 'Backtesting value-at-risk: A duration-based approach', *Journal of Financial Econometrics* **2**, 84–108.

Engle, R. (1982), 'Autoregressive conditional heteroskedasticity with estimates of united kindgom heteroskedasticity', *Econometrica* **50**, 987–1007.

Engle, R. (2002), 'Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models', *Journal of Business & Economic Statistics* **20**(3), 339–350.

Engle, R. & Kroner, K. (1995), 'Multivariate simultaneous generalized arch', *Econometric Theory* **11**, 122–150.

Engle, R. & Manganelli, S. (2004), 'Caviar: Conditional autoregressive value at risk by regression quantiles', *Journal of Business and Economic Statistics* **22**, 367–381.

Engle, R. & Mezrich, J. (1996), 'Garch for groups', *Risk* pp. 36–40.

Fang, K., Kotz, S. & Ng, K. (1990), *Symmetric Multivariate and Related Distributions*, Chapman and Hall.

Francq, C., Horváth, L. & Zakoïan, J.-M. (2011), 'Merits and drawbacks of variance targeting in garch models', *Journal of Financial Econometrics* **9**, 305–327.

Giot, P. & Laurent, S. (2003), 'Market risk in commodity markets: A var approach', *Energy Economics* **25**, 435–457.

Kuester, K., Mittnik, S. & Paolella, M. (2006), 'Value-at-risk prediction: A comparison of alternative strategies', *Journal of Financial Econometrics* **4**(1), 53–89.

Kupiec, P. (1995), 'Techniques for verifying the accuracy of risk management models', *Journal of Derivatives* **23**, 73–84.

Lambert, P. & Laurent, S. (2001), Modelling financial time series using garch-type models with a skewed student distribution for the innovations, Technical report, Discussion paper 0125, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Ledoit, O., Santa-Clara, P. & Wolf, M. (2003), 'Flexible multivariate garch modeling with an application to international stock markets', *Review of Economics and Statistics* **85**(3), 735–747.

Michaud, R. O. (1989), 'The markowitz optimization enigma: is' optimized'optimal?', *Financial Analysts Journal* pp. 31–42.

Nelson, D. B. (1991), 'Conditional heteroskedasticity in asset returns: A new approach', *Econometrica* pp. 347–370.

Noureldin, D., Shephard, N. & Sheppard, K. (2012), 'Multivariate high-frequency-based volatility (heavy) models', *Journal of Applied Econometrics* **27**, 907–933.

Solaro, N. (2004), 'Random variate generation from multivariate exponential power distribution', *Statistica Applicazioni II* **2**, 25–44.

# Appendix A Derivations

## Appendix A.1 Transformation

The transformation $z_t = H_t^{-1/2} y_t$ is incorporated into the symmetric, standardised pdfs as follows:

$$
\begin{aligned}
\kappa^{\star\prime}\kappa^{\star} &= (\kappa_1^{\star}, ..., \kappa_N^{\star})'(\kappa_1^{\star}, ..., \kappa_N^{\star}) \\
&= \left(\ldots (s_i z_i + m_i)\xi_i^{I_i} \ldots\right)' \left(\ldots (s_i z_i + m_i)\xi_i^{I_i} \ldots\right) \\
&= \left(\ldots (s_i \sum_{j=1}^{N} p_{ij} y_j + m_i)\xi_i^{I_i} \ldots\right)' \left(\ldots (s_i \sum_{j=1}^{N} p_{ij} y_j + m_i)\xi_i^{I_i} \ldots\right) \\
&= \sum_{i=1}^{N} \left( s_i \sum_{j=1}^{N} p_{ij} y_j + m_i \right) \xi_i^{2I_i}
\end{aligned}
$$

where $p_{ij}$ corresponds to the $j^{th}$ element of the $i^{th}$ row of $H_t^{-1/2}$. Note that the $t$ subscript is dropped for simplicity. The matrix square root operation is carried out by applying the Cholesky decomposition of $H_t$ such that $BB' = H_t$. As a result, each $z_i$ is obtained by multiplying the row vector of $H_t^{-1/2}$ corresponding to asset $i$ with the demeaned return vector (giving us the inner summation above) which is then multiplied by the univariate standard deviation and added to the univariate mean. The presence of skewness is factored in by the term $\xi_i^{I_i}$, where the factor $I_i$ is defined as in Eq. (15).

## Appendix A.2 Distributions moments

We report the first two moments of the univariate symmetric normal, Student and MEP distributions along with the formulas for the derivation of the univariate moments of their skewed counterparts. These are used to compute the log-likelihood function as given in Section 2.3.

**Skew-Normal** Symmetric normal first and second moments:

$$
\begin{aligned}
M_{i,1} &= \int_0^\infty \frac{2}{\sqrt{2\pi}} u \exp\left\{-\frac{1}{2}u^2\right\} du \\
&= \sqrt{\frac{2}{\pi}} \int_0^\infty u \exp\left\{-\frac{1}{2}u^2\right\} du \\
&= \sqrt{\frac{2}{\pi}} \\
M_{i,2} &= \int_0^\infty \frac{2}{\sqrt{2\pi}} u^2 \exp\left\{-\frac{1}{2}u^2\right\} du \\
&= \sqrt{\frac{2}{\pi}} \int_0^\infty u^2 \exp\left\{-\frac{1}{2}u^2\right\} du \\
&= 1
\end{aligned}
$$

The skewed moments are computed using Equations (11) and (12) as follows:

$$
\begin{aligned}
m_i &= M_{i,1}\left(\xi_i - \frac{1}{\xi_1}\right) \\
&= \sqrt{\frac{2}{\pi}}\left(\xi_i - \frac{1}{\xi_1}\right) \\
s_i^2 &= \left(M_{i,2} - M_{i,1}^2\right)\left(\xi_i^2 + \frac{1}{\xi_i^2}\right) + 2M_{i,1}^2 - M_{i,2} \\
&= \left(1 - \frac{2}{\pi}\right)\left(\xi_i^2 + \frac{1}{\xi_i^2}\right) + \frac{4}{\pi} - 1 \\
&= \frac{\pi - 2}{\pi}\left(\xi_i^2 + \frac{1}{\xi_i^2}\right) + \frac{4 - \pi}{\pi}
\end{aligned}
$$

**Skew-Student** Symmetric Student distribution first and second moments:

$$\begin{aligned}
M_{i,1} &= \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi(\nu-2)}}\int_0^\infty u\left(1+\frac{u^2}{\nu-2}\right)^{-\frac{1+\nu}{2}} du \\
&= \frac{2\sqrt{\nu-2}\left(\frac{\nu-1}{2}\right)\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{\pi}(\nu-1)\Gamma\left(\frac{\nu}{2}\right)} \\
&= \frac{2\sqrt{\nu-2}\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}(\nu-1)\Gamma\left(\frac{\nu}{2}\right)} \\
&= \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{\nu-1}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} \\
M_{i,2} &= \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi(\nu-2)}}\int_0^\infty u^2\left(1+\frac{u^2}{\nu-2}\right)^{-\frac{1+\nu}{2}} du \\
&= \frac{(\nu-2)\Gamma\left(\frac{\nu}{2}-1\right)}{2\Gamma\left(\frac{\nu}{2}\right)} \\
&= \frac{(\nu-2)\Gamma\left(\frac{\nu}{2}-1\right)}{2\left(\frac{\nu-2}{2}\right)\Gamma\left(\frac{\nu}{2}-1\right)} \\
&= 1
\end{aligned}$$

First and second order moments of the skewd distribution are expressed as follows; specifically the second skewed moment is obtained as a function of the first:

$$m_i = M_{i,1}\left(\xi_i - \frac{1}{\xi_i}\right) \Rightarrow M_{i,1}^2 = m_i^2\left(\frac{\xi_i^2}{(\xi_i^2-1)^2}\right)$$

Substituting into Eq.(12) gives:

$$\begin{aligned}
s_i^2 &= M_{i,1}^2\left(-\xi_i^2 - \frac{1}{\xi_i^2} + 2\right) + M_{i,2}\left(\xi_i^2 + \frac{1}{\xi_i^2} - 1\right) \\
&= \frac{\xi_i^2}{(\xi_i^2-1)^2}\left(\frac{-\xi_i^4 + 2\xi_i^2 + 2}{\xi_i^2}\right)m_i^2 + M_{i,2}\left(\frac{\xi_i^4 - \xi_i^2 + 1}{\xi_i^2}\right) \\
&= \frac{\xi_i^2}{(\xi_i^2-1)^2}\frac{-(\xi_i^2-1)^2}{\xi_i^2}m_i^2 + M_{i,2}\frac{\xi_i^2\left(\xi_i^2 - 1 + \frac{1}{\xi_i^2}\right)}{\xi_i^2} \\
&= M_{i,2}\left(\xi_i^2 + \frac{1}{\xi_i^2} - 1\right) - m_i^2
\end{aligned}$$

Eq.(25) is obtained by substituting $M_{i,2} = 1$ into the above result.

**Skew-MEP** Symmetric MEP first and second moments:

$$
\begin{aligned}
M_{i,1} &= \frac{2}{\Gamma\left(1+\frac{1}{\beta}\right)} 2^{1+\frac{1}{\beta}} \int_0^\infty u \exp\left\{-\frac{1}{2}u^\beta\right\} du \\
&= \frac{2^{-1+\frac{1}{\beta}}\Gamma\left(\frac{2+\beta}{\beta}\right)}{\Gamma\left(1+\frac{1}{\beta}\right)} \\
M_{i,2} &= \frac{2}{\Gamma\left(1+\frac{1}{\beta}\right)} 2^{1+\frac{1}{\beta}} \int_0^\infty u^2 \exp\left\{-\frac{1}{2}u^\beta\right\} du \\
&= \frac{4^{\frac{1}{\beta}}\Gamma\left(\frac{3}{\beta}\right)}{\beta\Gamma\left(1+\frac{1}{\beta}\right)}
\end{aligned}
$$

Skewed moments obtained as:

$$
\begin{aligned}
m_i &= \frac{2^{-1+\frac{1}{\beta}}\Gamma\left(\frac{2+\beta}{\beta}\right)}{\beta\Gamma\left(1+\frac{1}{\beta}\right)} \left(\xi_i - \frac{1}{\xi_1}\right) \\
s_i^2 &= \frac{4^{\frac{1}{\beta}}\Gamma\left(\frac{3}{\beta}\right)}{\beta\Gamma\left(1+\frac{1}{\beta}\right)} \left(\xi_i^2 + \frac{1}{\xi_i^2} - 1\right) - m_i^2.
\end{aligned}
$$

# Appendix B Tables

Table B1: Windows length and corresponding calendar time

| | Rolling fixed-window | | Forecast horizon | |
|---|---|---|---|---|
| It. | Observations | Days | Observations | Days |
| 1 | 1-1500 | 2/1/01 - 1/23/07 | 1501-1520 | 1/24/07 - 2/21/07 |
| 2 | 21-1520 | 3/2/01 - 2/21/07 | 1521-1540 | 2/22/07 - 3/21/07 |
| 3 | 41-1540 | 3/30/01 - 3/21/07 | 1541-1560 | 3/22/07 - 4/19/07 |
| 4 | 61-1560 | 4/30/01 - 4/19/07 | 1561-1580 | 4/20/07 - 5/17/07 |
| 5 | 81-1580 | 5/29/01 - 5/17/07 | 1581-1600 | 5/18/07 - 6/15/07 |
| 6 | 101-1600 | 6/26/01 - 6/15/07 | 1601-1620 | 6/18/07 - 7/16/07 |
| 7 | 121-1620 | 7/25/01 - 7/16/07 | 1621-1640 | 7/17/07 - 8/13/07 |
| 8 | 141-1640 | 8/22/01 - 8/13/07 | 1641-1660 | 8/14/07 - 9/11/07 |
| 9 | 161-1660 | 9/26/01 - 9/11/07 | 1661-1680 | 9/12/07 - 10/9/07 |
| 10 | 181-1680 | 10/24/01 - 10/9/07 | 1681-1700 | 10/10/07 - 11/6/07 |
| 11 | 201-1700 | 11/21/01 - 11/6/07 | 1701-1720 | 11/7/07 - 12/5/07 |
| 12 | 221-1720 | 12/20/01 - 12/5/07 | 1721-1740 | 12/6/07 - 1/4/08 |
| 13 | 241-1740 | 1/22/02 - 1/4/08 | 1741-1760 | 1/7/08 - 2/4/08 |
| 14 | 261-1760 | 2/20/02 - 2/4/08 | 1761-1780 | 2/5/08 - 3/4/08 |
| 15 | 281-1780 | 3/20/02 - 3/4/08 | 1781-1800 | 3/5/08 - 4/2/08 |
| 16 | 301-1800 | 4/18/02 - 4/2/08 | 1801-1820 | 4/3/08 - 4/30/08 |
| 17 | 321-1820 | 5/16/02 - 4/30/08 | 1821-1840 | 5/1/08 - 5/29/08 |
| 18 | 341-1840 | 6/14/02 - 5/29/08 | 1841-1860 | 5/30/08 - 6/26/08 |
| 19 | 361-1860 | 7/15/02 - 6/26/08 | 1861-1880 | 6/27/08 - 4/2/08 |
| 20 | 381-1880 | 8/12/02 - 7/25/08 | 1881-1900 | 7/28/08 - 8/22/08 |
| 21 | 401-1900 | 9/10/02 - 8/22/08 | 1901-1920 | 8/25/08 - 9/22/08 |
| 22 | 421-1920 | 10/8/02 - 9/22/08 | 1921-1940 | 9/23/08 - 10/20/08 |
| 23 | 441-1940 | 11/5/02 - 10/20/08 | 1941-1960 | 10/21/08 - 11/17/08 |
| 24 | 461-1960 | 12/4/02 - 11/17/08 | 1961-1980 | 11/18/08 - 12/16/08 |
| 25 | 481-1980 | 1/3/03 - 12/16/08 | 1981-2000 | 12/17/08 - 1/15/09 |
| 26 | 501-2000 | 2/3/03 - 1/15/09 | 2001-2020 | 1/16/09 - 2/13/09 |
| 27 | 521-2020 | 3/4/03 - 2/13/09 | 2021-2040 | 2/17/09 - 3/16/09 |
| 28 | 541-2040 | 4/1/03 - 3/16/09 | 2041-2060 | 3/17/09 - 4/22/05 |
| 29 | 561-2060 | 4/30/03 - 4/14/09 | 2061-2080 | 4/15/09 - 5/12/09 |
| 30 | 581-2080 | 5/29/03 - 5/12/09 | 2081-2100 | 5/13/09 - 6/10/09 |
| 31 | 601-2100 | 6/26/03 - 6/10/09 | 2101-2120 | 6/11/09 - 7/9/09 |
| 32 | 621-2120 | 7/25/03 - 7/9/09 | 2121-2140 | 7/10/09 - 8/6/09 |
| 33 | 641-2140 | 8/22/03 - 8/6/09 | 2141-2160 | 8/7/09 - 9/3/09 |
| 34 | 661-2160 | 9/22/03 - 9/3/09 | 2161-2180 | 9/4/09 - 10/2/09 |
| 35 | 681-2180 | 10/20/03 - 10/2/09 | 2181-2200 | 10/5/09 - 10/30/09 |

# Appendix C Figures

Figure C1: VaR: normal and skew-normal

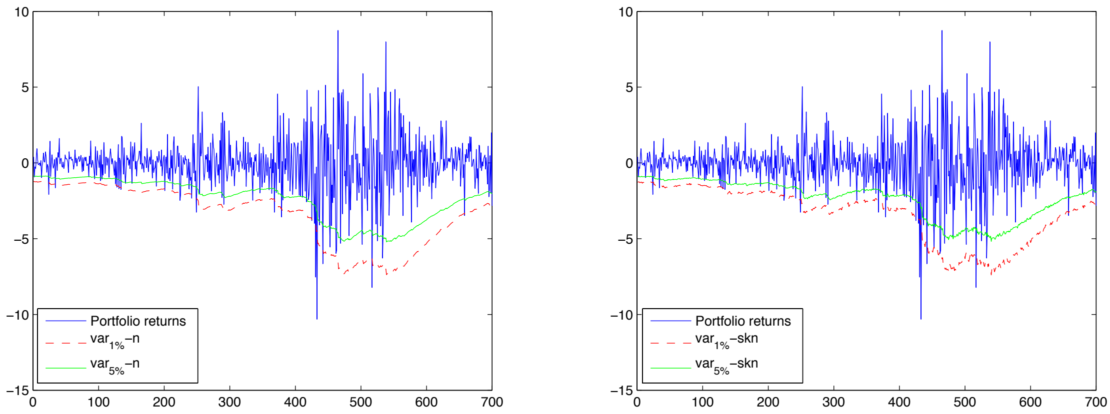



Figure C2: VaR: Student and skew-Student

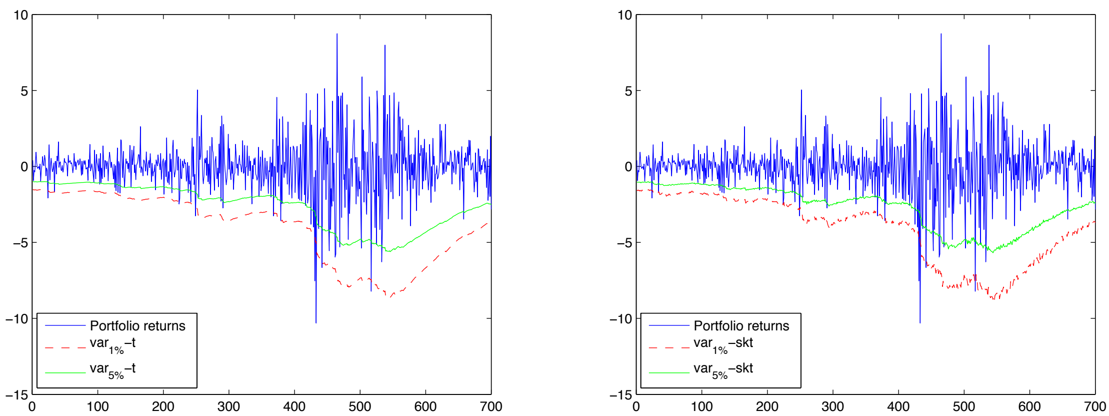
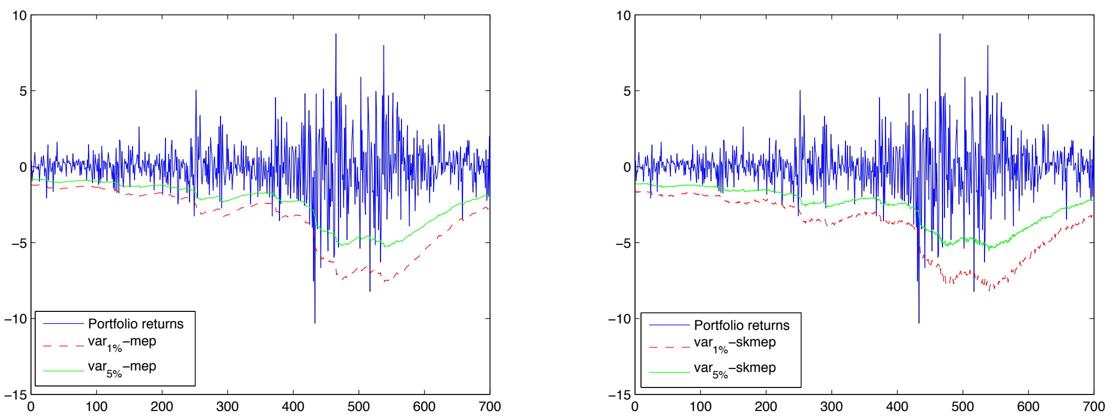



Figure C3: VaR: MEP and skew-MEP



# Recent titles

## CORE Discussion Papers

2014/18 Koen DECANCQ, Marc FLEURBAEY and Erik SCHOKKAERT. Inequality, income, and well-being.
2014/19 Paul BELLEFLAMME and Martin PEITZ. Digital piracy: an update.
2014/20 Eva-Maria SCHOLZ. Licensing to vertically related markets.
2014/21 N. Baris VARDAR. Optimal energy transition and taxation of non-renewable resources.
2014/22 Benoît DECERF. Income poverty measures with relative poverty lines.
2014/23 Antoine DEDRY, Harun ONDER and Pierre PESTIEAU. Aging, social security design and capital accumulation.
2014/24 Biung-Ghi JU and Juan D. MORENO-TERNERO. Fair allocation of disputed properties.
2014/25 Nguyen Thang DAO. From agriculture to manufacture: How does geography matter ?
2014/26 Xavier Y. WAUTHY. From Bertrand to Cournot via Kreps and Scheinkman: a hazardous journey.
2014/27 Gustavo BERGANTIÑOS and Juan MORENO-TERNERO. The axiomatic approach to the problem of sharing the revenue from bundled pricing.
2014/28 Jean HINDRIKS and Yukihiro NISHIMURA. International tax leadership among asymmetric countries.
2014/29 Jean HINDRIKS and Yukihiro NISHIMURA. A note on equilibrium leadership in tax competition models.
2014/30 Olivier BOS and Tom TRUYTS. Auctions with prestige motives.
2014/31 Juan D. MORENO-TERNERO and Lars P. ØSTERDAL . Normative foundations for equity-sensitive population health evaluation functions.
2014/32 P. Jean-Jacques HERINGS, Ana MAULEON and Vincent VANNETELBOSCH. Stability of networks under Level-*K* farsightedness.
2014/33 Lionel ARTIGE, Laurent CAVENAILE and Pierre PESTIEAU. The macroeconomics of PAYG pension schemes in an aging society.
2014/34 Tanguy KEGELART and Mathieu VAN VYVE. A conic optimization approach for SKU rationalization.
2014/35 Ulrike KORNEK, Kei LESSMANN and Henry TULKENS. Transferable and non transferable utility implementations of coalitional stability in integrated assessment models.
2014/36 Ibrahim ABADA, Andreas EHRENMANN and Yves SMEERS. Endogenizing long-term contracts in gas market models.
2014/37 Julio DAVILA. Output externalities on total factor productivity.
2014/38 Diane PIERRET. Systemic risk and the solvency-liquidity nexus of banks.
2014/39 Paul BELLEFLAMME and Julien JACQMIN. An economic appraisal of MOOC platforms: business models and impacts on higher education.
2014/40 Marie-Louise LEROUX, Pierre PESTIEAU and Grégory PONTHIERE. Longévité différentielle et redistribution: enjeux théoriques et empiriques.
2014/41 Chiara CANTA, Pierre PESTIEAU and Emmanuel THIBAULT. Long term care and capital accumulation: the impact of the State, the market and the family.
2014/42 Gilles GRANDJEAN, Marco MANTOVANI, Ana MAULEON and Vincent VANNETELBOSCH. Whom are you talking with ? An experiment on credibility and communication structure.
2014/43 Julio DAVILA. The rationality of expectations formation.
2014/44 Florian MAYNERIS, Sandra PONCET and Tao ZHANG. The cleaning effect of minimum wages. Minimum wages, firm dynamics and aggregate productivity in China.
2014/45 Thierry BRECHET, Natali HRITONENKOVA and Yuri YATSENKO. Domestic environmental policy and international cooperation for global commons.
2014/46 Mathieu PARENTI, Philip USHCHEV and Jacques-François THISSE. Toward a theory of monopolistic competition.

# Recent titles

## CORE Discussion Papers - continued

2014/47 Takatoshi TABUCHI, Jacques-François THISSE and Xiwei ZHU. Does technological progress affect the location of economic activity?
2014/48 Paul CASTANEDA DOWER, Victor GINSBURGH and Shlomo WEBER. Colonial legacy, linguistic disenfranchisement and the civil conflict in Sri Lanka.
2014/49 Victor GINSBURGH, Jacques MELITZ and Farid TOUBAL. Foreign language learnings: An econometric analysis.
2014/50 Koen DECANCQ and Dirk NEUMANN. Does the choice of well-being measure matter empirically? An illustration with German data.
2014/51 François MANIQUET. Social ordering functions.
2014/52 Ivar EKELAND and Maurice QUEYRANNE. Optimal pits and optimal transportation.
2014/53 Luc BAUWENS, Manuela BRAIONE and Giuseppe STORTI. Forecasting comparison of long term component dynamic models for realized covariance matrices.
2014/54 François MANIQUET and Philippe MONGIN. Judgment aggregation theory can entail new social choice results.
2014/55 Pasquale AVELLA, Maurizio BOCCIA and Laurence A. WOLSEY. Single-period cutting planes for inventory routing problems.
2014/56 Jean-Pierre FLORENS and Sébastien VAN BELLEGEM. Instrumental variable estimation in functional linear models.
2014/57 Abdelrahaman ALY and Mathieu VAN VYVE. Securely solving classical networks flow problems.
2014/58 Henry TULKENS. Internal vs. core coalitional stability in the environmental externality game: A reconciliation.
2014/59 Manuela BRAIONE and Nicolas K. SCHOLTES. Construction of Value-at-Risk forecasts under different distributional assumptions within a BEKK framework.

## Books

V. GINSBURGH and S. WEBER (2011), *How many languages make sense? The economics of linguistic diversity*. Princeton University Press.
I. THOMAS, D. VANNESTE and X. QUERRIAU (2011), *Atlas de Belgique – Tome 4 Habitat*. Academia Press.
W. GAERTNER and E. SCHOKKAERT (2012), *Empirical social choice*. Cambridge University Press.
L. BAUWENS, Ch. HAFNER and S. LAURENT (2012), *Handbook of volatility models and their applications*. Wiley.
J-C. PRAGER and J. THISSE (2012), *Economic geography and the unequal development of regions*. Routledge.
M. FLEURBAEY and F. MANIQUET (2012), *Equality of opportunity: the economics of responsibility*. World Scientific.
J. HINDRIKS (2012), *Gestion publique*. De Boeck.
M. FUJITA and J.F. THISSE (2013), *Economics of agglomeration: cities, industrial location, and globalization.* (2nd edition). Cambridge University Press.
J. HINDRIKS and G.D. MYLES (2013). *Intermediate public economics.* (2nd edition). MIT Press.
J. HINDRIKS, G.D. MYLES and N. HASHIMZADE (2013). *Solutions manual to accompany intermediate public economics.* (2nd edition). MIT Press.

## CORE Lecture Series

R. AMIR (2002), Supermodularity and complementarity in economics.
R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.
A. SHAPIRO (2010), Stochastic programming: modeling and theory.