

# Forecasting Expected Shortfall: Should we use a Multivariate Model for Stock Market Factors?

Alain-Philippe Fortin, Jean-Guy Simonato and Georges Dionne\*

June 2018

## Abstract

When forecasting the market risk of stock portfolios, is a univariate or a multivariate modeling approach more effective? This question is examined in the context of forecasting the one-week-ahead Expected Shortfall for a portfolio equally invested in the Fama-French and momentum factors. Applying extensive tests and comparisons, we find that in most cases there are no statistically significant differences between the forecasting accuracy of the two approaches. This suggests that univariate models, which are more parsimonious and simpler to implement than multivariate models, can be used to forecast the downside risk of equity portfolios without losses in precision.

**Keywords:** Value-at-Risk, Expected Shortfall, Conditional Value-at-Risk, Elicibility, model comparison, backtesting, Fama-French and momentum factors

**JEL classification:** C22, C32, C52, C53, G17

---

\*Fortin, [alain-philippe.fortin@hec.ca](mailto:alain-philippe.fortin@hec.ca). Simonato, [jean-guy.simonato@hec.ca](mailto:jean-guy.simonato@hec.ca), HEC Montréal, Department of Finance. Dionne, [georges.dionne@hec.ca](mailto:georges.dionne@hec.ca), HEC Montréal, Department of Finance and Canada Research Chair in Risk Management. We thank Hugues Langlois for providing us with the original data used in Christoffersen and Langlois (2013) and for sharing some MATLAB codes.

# 1 Introduction

What is the best approach to forecast the market risk of stock portfolios? As Santos et al. (2013) emphasize, the first choice one has to make is between a univariate and a multivariate approach. Many portfolio problems, such as optimization, usually require a multivariate approach, but this is not the case for portfolio risk measurement. Indeed, risk measures are a function of the portfolio return probability distribution. This distribution can be modeled either directly via a univariate model or indirectly via a multivariate model for the vector of portfolio stock returns. Given the easier implementation and much smaller computing efforts required for univariate models, many researchers recommend this direct approach to risk measurement<sup>1</sup>. However, a case can also be made for the indirect approach, because multivariate models use more information by modeling the dependencies between portfolio stock returns. Everything else held constant, this additional information should improve risk forecasts compared with a univariate model. The main difficulties with the multivariate approach are that, as the dimension of the portfolio increases, there are more possibilities for estimation errors, and restrictions must be imposed to keep the model tractable.

The choice of which method to use has important consequences for risk managers who need not only accurate risk measures, but also flexible and convenient procedures to compute them routinely. At the same time, the question is empirical because it depends on both the data and models used. For our application, we choose to consider a stock factor portfolio. More precisely, we consider the weekly returns of an equally weighted portfolio invested in the Fama and French (1993) and Carhart (1997) momentum, or FFC, factors. We think there are three reasons that these four factors allow a rich investigation of the question at hand. First, the FFC factors are pervasive in the empirical asset pricing literature and among practitioners. Second, for multivariate portfolio analysis in high dimensions, one often has no choice but to impose a factor structure, in which case a linear model in the FFC factors is an obvious choice. Third, whereas the linear correlations between the four factors weekly returns are very small, suggesting great

---

<sup>1</sup>See for example Berkowitz and O'Brien (2002) and Christoffersen (2009).

diversification benefits and lower risk, Christoffersen and Langlois (2013) find that these measures hide important asymmetric and nonlinear dependencies that could be better captured by an adequate multivariate model than by a univariate model.

The issue has been studied many times in the context of forecasting portfolio Value-at-Risk (VaR)<sup>2</sup>. Given that VaR is typically proportional to the standard deviation of the portfolio return, the literature has focused on comparing univariate and multivariate GARCH models. Brooks and Persaud (2003) compare twelve univariate volatility models and the diagonal VEC model of Bollerslev et al. (1988) for forecasting the volatility and VaR of a portfolio comprised of UK assets at 1, 5, 10 and 20-day horizons. They find no clear improvements from using a multivariate approach compared with a univariate approach and suggest that, unless covariances are required, multivariate GARCH models are not worthwhile. Similarly, McAleer and Da Veiga (2008) compare twelve univariate and sixteen multivariate GARCH models to forecast the one-day-ahead volatility and VaR of an international equity portfolio. Although the multivariate models offer better volatility forecasts, there is no clear preference between the two approaches for VaR forecasting : eight of the univariate models and eight of the multivariate models fail the conditional coverage test of Christoffersen (1998). One aspect that makes these results hard to interpret is the fact that backtesting procedures are only meant to evaluate a model in isolation, not against another model<sup>3</sup>. To address this limitation, Santos et al. (2013) rely instead on the asymmetric tick loss function, as in Giacomini and Komunjer (2005), to compare the out-of-sample next day VaRs of different pairs of univariate vs multivariate GARCH models. Using large and diversified US stock portfolios, they find that multivariate models with dynamic conditional correlations and Student  $t$  distributed errors outperform univariate models. Further, multivariate models with constant correlations usually underperform relative to univariate models. In a recent article, Kole et al. (2017) examine the impact of different levels of temporal and portfolio aggregation on forecasting the 10-day VaR for a diversified portfolio of eight indexes, related to stocks, bonds and alternative investments. Also relying on the asymmetric tick loss function,

---

<sup>2</sup>For a recent survey on VaR forecasting and backtesting methods see Nieto and Ruiz (2016).

<sup>3</sup>See paragraph 3.1 in Acerbi and Szekely (2014).

they find that lower levels of aggregation, i.e. multivariate models for index returns or asset class returns, provide better risk forecasts relative to complete portfolio aggregation, but the differences are not large and often not significant. More specifically, multivariate models based on the three main asset classes deliver the best 1% VaR results whereas multivariate models based on the eight indexes deliver the best 5% VaR results. Further, contrary to the results of Santos et al. (2013), Kole et al.'s (2017) analysis shows that dynamic correlations do not produce improvements in risk forecasts. It is important to note that all the above-mentioned studies are based on daily returns, yet our paper considers the weekly returns of a stock only portfolio, so our aggregation level corresponds to Kole et al.'s (2017) three levels of portfolio aggregation, without the bond and alternative investment classes, and with weekly temporal aggregation. As this short survey demonstrates, the literature to date offers mixed evidence.

We distinguish ourselves from previous research by a focus on Expected Shortfall (ES), also called Conditional Value-at-Risk (CVaR). With the recent shift from VaR to ES sanctioned by the Basel Committee on Banking Supervision (BCBS)<sup>4</sup> and the growing literature on backtesting ES, this risk measure is becoming a serious alternative to VaR. The change is motivated by the fact that VaR does not capture adequately the conditional expected losses and lacks subadditivity, whereas ES avoids these limitations. Adding ES as a forecasting objective presents an interesting challenge in that, contrary to VaR, there is no loss function for which ES is the minimizer<sup>5</sup>. This property of a risk measure is known as elicibility, and gives a natural criterion for relative comparison of univariate and multivariate models. However, as Fissler et al. (2016) show, the pair VaR/ES is jointly elicitable with respect to a class of loss functions. We use this result<sup>6</sup> and the loss function proposed in Acerbi and Szekely (2014) to compare our models based on their ability to forecast VaR and ES jointly.

We apply extensive tests and comparisons between models to obtain our results. First, we perform the VaR standard backtests as well as the first two ES backtests of Acerbi

---

<sup>4</sup>See Basel Committee on Banking Supervision (2016).

<sup>5</sup>This result is proved in Gneiting (2011).

<sup>6</sup>See Dimitriadis and Bayer (2017) for an application of this result to the joint estimation of VaR and ES in a regression framework.

and Szekely (2014) for each model, at many levels. Second, we use Diebold and Mariano's (1995), or DM, test and the conditional framework of Giacomini and White (2006) to assess statistical differences between the predictive ability of each pair of models with respect to both VaR and the VaR/ES pair at the 1% and 5% level. Third, we implement the Model Confidence Set (MCS) approach of Hansen et al. (2011) to establish a set of superior models in each case.

We find that in most cases there are no significant differences between the risk forecasting accuracy of univariate and multivariate models for both VaR and the VaR/ES pair. When there is a difference, it is in favor of the univariate models. Our VaR results are in direct contrast with those of Santos et al. (2013), but partly corroborate the results of Kole et al. (2017), which are the papers most closely related to our study. Given the empirical nature of the question, this discrepancy is not incoherent. One possible explanation for the divergence with Santos et al. (2013) is the fact that they do not use univariate distributions allowing for skewness, an important feature to include, especially for measuring tail risk. Our analysis includes both univariate and multivariate skewness by considering the univariate Skewed  $t$  distribution of Hansen (1994) and the asymmetric  $t$  copula of Demarta and McNeil (2005). To our knowledge, asymmetric copula models have never been compared with asymmetric univariate models in the risk forecasting literature. Another disparity with Santos et al. (2013) is the fact that we use weekly returns whereas they use daily returns. Lower data frequency can hurt multivariate models by preventing them from adequately capturing the assets' cross-sectional and serial dependencies. The main conclusion of Kole et al. (2017) is that aggregation of daily returns into weekly or biweekly returns leads to the loss of important details in return dynamics, and affects the accuracy of VaR forecasts more than the choice between a univariate or a multivariate model. However, their results show a clearer preference for multivariate models when weekly or biweekly returns are employed, suggesting that lower levels of portfolio aggregation can compensate for fewer details in the temporal dimension. We must also emphasize that these two papers have not examined ES, so it remains unclear how data frequency impacts the relative performance of multivariate and univariate models in our

case.

The paper is organized as follows. Section 2 presents the framework and assumptions underlying the models used to forecast VaR and ES. Sections 3 and 4 are devoted to multivariate and univariate models respectively. In Section 5 we present the data and the parameter estimates of our models. Section 6 explains the procedure employed to generate out-of-sample risk forecasts. Section 7 is dedicated to VaR and ES backtests. In Section 8 we present the loss functions used to assess the relative performance of our models. Section 9 is devoted to the DM test and Section 10 is discusses the MCS approach. Section 11 concludes the paper.

## 2 Framework

Let  $\mathbf{r}_t = (r_{1t}, \dots, r_{Nt})'$  denote the random vector of the  $N$  factor returns at time  $t$  and  $r_t = \mathbf{w}'_{t-1} \mathbf{r}_t$  denote the random portfolio return at time  $t$ , where  $\mathbf{w}_{t-1}$  is the vector of predetermined equal portfolio weights at time  $t - 1$ . We assume that the time series  $\{r_t\}_{t=1}^T$  and  $\{\mathbf{r}_t\}_{t=1}^T$  are stationary. Let  $\mathbf{F}_t$  and  $F_t$  denote, respectively, the distribution function of  $\mathbf{r}_t$  and  $r_t$ , conditional on the information set available at time  $t - 1$ . Also, let  $F_{jt}$  correspond to the  $j$ th marginal distribution function of  $\mathbf{F}_t$  for  $j = 1, \dots, N$ . In this paper we assume that all distribution functions are continuous, have densities and are strictly increasing. In particular, this implies that the inverse distribution function (quantile function)  $F^{-1}(\cdot)$  is well defined.

Given a significance level  $p = 1\%$  or  $5\%$  at time  $t - 1$ , we are interested in forecasting the next period conditional VaR and ES defined by

$$VaR_t^p = F_t^{-1}(p) \tag{1}$$

and

$$ES_t^p = \mathbb{E}_{t-1}[r_t \mid r_t < VaR_t^p]. \tag{2}$$

In order to estimate these risk measures, we need a statistical model for  $F_t$ . In the uni-

variate approach, this is done by directly making assumptions on  $F_t$ . In the multivariate approach, we take an indirect path by modeling the joint conditional distribution of the vector of factor returns  $\mathbf{F}_t$ . Since  $r_t$  is a linear combination of the factor returns in  $\mathbf{r}_t$  and the weights in  $\mathbf{w}_{t-1}$  are part of the information set at time  $t-1$ ,  $F_t$  is really a function of  $\mathbf{F}_t$ . As a special case, assuming that  $\mathbf{r}_t$  follows a multivariate normal distribution implies a normal distribution for  $r_t$ . In other cases the link between the two distributions is not as clear so we rely on simulation methods to derive  $F_t$  from  $\mathbf{F}_t$  and then compute the estimated VaR and ES. Details on the procedure are given in Section 6.

### 3 Multivariate models

The multivariate models in this section are taken from Christoffersen and Langlois (2013), who study the joint dynamics of the FFC factors. Analysis of the data reveals that each factor marginal distribution is highly nonnormal and that the dependencies between each pair of factors are nonlinear. This suggests multivariate nonnormality for the joint distribution of the four factors. Therefore, instead of relying on the multivariate normal distribution we use copulas to fit the joint conditional distribution of the factor returns. Copulas are flexible because they enable modeling of the marginal distributions separately via Sklar's (1959) theorem. This theorem allows the decomposition of the next period joint conditional distribution of the  $N$  factor returns into their conditional marginal distributions and a conditional copula linking these marginals:<sup>7</sup>

$$\mathbf{F}_t(\mathbf{r}_t) = \mathbf{C}_t(F_{1t}(r_{1t}), \dots, F_{Nt}(r_{Nt})). \quad (3)$$

Given our assumptions in Section 2, the copula  $\mathbf{C}_t$  is uniquely determined. Each of the marginals  $F_{jt}$  contains all the univariate information on the  $j$ th factor, while the copula  $\mathbf{C}_t$  contains all the dependence information between the factors. This decomposition shows that a model for  $\mathbf{F}_t$  can be built in two steps. First model each of the marginals

---

<sup>7</sup>Patton (2006) shows that Sklar's (1959) theorem applies to conditional distributions. An important point in this decomposition is that the information set must be the same for all marginal distributions and the copula.

$F_{jt}$  for  $j = 1, \dots, N$  and then choose a copula  $\mathbf{C}_t$  to link them. We now address these two steps in turn.

Each of the four factors next period conditional marginal distribution is fitted with the following model:

$$r_{jt} = \mu_{jt} + \sigma_{jt}z_{jt} \quad \text{with} \quad z_{jt} \stackrel{i.i.d.}{\sim} F_j(0, 1) \quad (4)$$

where  $\mu_{jt}$  and  $\sigma_{jt}$  are respectively the conditional mean and standard deviation of the return on factor  $j$  and  $z_{jt}$  is an error term. The error term  $z_{jt}$  is an independent and identically distributed random variable following the unconditional standardized distribution  $F_j(0, 1)$ . This model allows for dynamic first and second moments as well as nonnormal distributions, but assumes that higher conditional moments are constant. For each factor, the conditional mean is fitted using an AR(3) model while the conditional variance is fitted using the NGARCH(1,1) model of Engle and Ng (1993):

$$\mu_{jt} = \phi_{0j} + \phi_{1j}r_{jt-1} + \phi_{2j}r_{jt-2} + \phi_{3j}r_{jt-3} \quad , \quad (5)$$

$$\sigma_{jt}^2 = \omega_j + \beta_j\sigma_{jt-1}^2 + \alpha_j(z_{jt-1} - \theta_j)^2 \quad (6)$$

where  $\omega_j > 0$ ,  $\beta_j, \alpha_j \geq 0$  and  $\alpha_j(1 + \theta_j^2) + \beta_j < 1$  for  $j = 1, \dots, N$ . A positive leverage parameter ( $\theta_j > 0$ ) implies that negative shocks ( $z_{jt-1} < 0$ ) have a larger impact on the next period variance than do positive shocks of the same magnitude. The standardized distribution of each factor  $F_j(0, 1)$  is fitted using Hansen's (1994) skewed  $t$  distribution. This standardized distribution has two parameters,  $\kappa_j$  and  $\nu_j$ , which determine the skewness and kurtosis, and its density is given in Appendix A. Marginal skewness is an important improvement from the symmetric Student distribution when measuring tail risk.

To link the marginals we consider the skewed  $t$  copula of Demarta and McNeil (2005), which is derived from the standardized multivariate skewed  $t$  distribution. This choice is motivated by the asymmetric dependence between the factors reported in Christoffersen and Langlois (2013). The skewed  $t$  copula, denoted  $\mathbf{C}_{\nu_c, \mathbf{r}_t, \boldsymbol{\lambda}}^{st}$ , is characterized by a  $N \times 1$  vector of asymmetry parameters  $\boldsymbol{\lambda}$ , a scalar degree of freedom parameter  $\nu_c$ , and a copula



correlation matrix  $\Upsilon_t$ . Its density is given in Appendix B.2.

We allow the copula correlation matrix  $\Upsilon_t$  to evolve through time. More specifically, we assume that the correlation matrix of the copula quantiles (which are defined in the appendix), denoted  $\dot{\Upsilon}_t$ , follows the cDCC model of Aielli (2013)<sup>8</sup>:

$$\mathbf{Q}_t = \mathbf{Q}(1 - \beta_c - \alpha_c) + \beta_c \mathbf{Q}_{t-1} + \alpha_c \boldsymbol{\epsilon}_{t-1}^* \boldsymbol{\epsilon}_{t-1}' \quad (7)$$

$$\dot{\Upsilon}_t = \mathbf{Q}_t^{*-\frac{1}{2}} \mathbf{Q}_t \mathbf{Q}_t^{*-\frac{1}{2}} \quad (8)$$

where  $\mathbf{Q} = E(\mathbf{Q}_t) = E(\boldsymbol{\epsilon}_t^* \boldsymbol{\epsilon}_t')$  is a positive definite correlation matrix while  $\beta_c$  and  $\alpha_c$  are non-negative scalars with  $\alpha_c + \beta_c < 1$ . Also,  $\boldsymbol{\epsilon}_t^* = \mathbf{Q}_t^{*\frac{1}{2}} \boldsymbol{\epsilon}_t$  where  $\mathbf{Q}_t^* = \text{diag}(\text{dg}(\mathbf{Q}_t))$  and  $\boldsymbol{\epsilon}_t$  is a  $N \times 1$  vector containing the standardized copula quantiles. See Appendix B.4 for details.

As benchmarks to the dynamic skewed  $t$  copula, we consider the dynamic normal copula and the dynamic student copula, i.e. both with a dynamic correlation matrix. We also implement the three copula models with a constant correlation matrix and refer to these models as static copula models. Note that in all six copula models the dynamics for the marginals stay the same. Finally, we add for reference the multivariate normal distribution with a constant and a dynamic correlation matrix, in which case each marginal follows a univariate normal distribution. That is,  $F_j(0, 1)$  corresponds to the univariate standard normal distribution for all  $j$ . This makes a total of eight multivariate models.

## 4 Univariate models

An analysis of the factor residuals  $\hat{z}_{jt}$  shows that an AR(3)-NGARCH(1,1) model with skewed  $t$  distribution approximates each factor marginal distribution  $F_{jt}$  very well. We can therefore expect this model to provide a decent fit to the equally weighted portfolio return dynamic:

---

<sup>8</sup>Aielli's (2013) cDCC model is a modification of Engle's (2002) original Dynamic Conditional Correlation (DCC) model and provides a consistent estimator for the matrix  $\mathbf{Q}$ .

$$r_t = \mu_t + \sigma_t z_t \quad \text{with} \quad z_t \stackrel{i.i.d.}{\sim} F(0, 1) \quad (9)$$

$$\mu_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \phi_3 r_{t-3} \quad (10)$$

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha (z_{t-1} - \theta)^2 \quad (11)$$

with the same parameter restrictions as before and  $F(0, 1)$  corresponding to Hansen's (1994) univariate skewed  $t$  distribution with parameters  $\kappa$  and  $\nu$ . As a reference we also consider the univariate standard normal distribution for  $F(0, 1)$ . This leads to a total of two univariate models.

## 5 Data set and parameter estimates

Similar to Christoffersen and Langlois (2013), we study the four weekly equity factors from July 5, 1963 to december 31, 2010. This corresponds to  $T = 2479$  observations. The data set comes from Kenneth French's data library<sup>9</sup>, where the details on how the FFC factors are constructed can be found. The raw weekly returns are converted into weekly log returns using the formula  $r_{ln} = \ln(1 + \frac{r_{weekly}}{100})$ . For the rest of the paper, unless otherwise indicated, the term return will mean weekly log return. Because we are working with weekly log returns, our risk forecasts are the next week log return VaR and ES.

The descriptive statistics for the factor returns are presented in Table 1. We see that the Market, Size and Momentum factors have a longer left tail, as illustrated by the negative skewness. Also, all factors display thicker tails than the normal distribution, as the kurtosis values much higher than three indicate. The nonnormality of each factor's marginal distribution is confirmed by the very large Jarque-Bera statistics in the last row, rejecting the null hypothesis of a normal distribution in all cases.

With this data set we estimate all our models by the method of Maximum Likelihood

---

<sup>9</sup>See [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). We use the original data set of Christoffersen and Langlois's (2013) study which is different from the current one in Kenneth French's data library because of retroactive adjustments.

(ML). Starting with the univariate models, let  $\boldsymbol{\theta}$  denote the vector containing the parameters of  $u_t$ ,  $\sigma_t^2$  and  $F$ . Given the sample of observations for the factor returns  $\mathbf{r}_1, \dots, \mathbf{r}_T$ , we construct a pseudo-sample of observations for the equally weighted portfolio return  $r_t = \mathbf{w}'_T \mathbf{r}_t$  for  $t = 1, \dots, T$ <sup>10</sup>. We then estimate  $\boldsymbol{\theta}$  by maximizing the conditional log-likelihood  $\ln L(\boldsymbol{\theta}) = \sum_{t=1}^T \ln f_t(x_t)$ . The ML estimates for the two univariate models are presented in Table 2, where the first column corresponds to the univariate normal distribution and the second column corresponds to the univariate skewed  $t$  distribution.

For the copula models we use a two-step estimation procedure. Differentiating both sides of equation (3) we get the conditional likelihood of  $\mathbf{r}_t$ :

$$\mathbf{f}_t(\mathbf{r}_t) = \mathbf{c}_t(F_{1t}(r_{1t}), \dots, F_{Nt}(r_{Nt})) \prod_{j=1}^N f_{jt}(r_{jt}) \quad (12)$$

where  $\mathbf{c}_t$  is the conditional copula density. Taking the log and summing over  $t$  we obtain the conditional log-likelihood function for our sample

$$\ln \mathbf{f}_t(\mathbf{r}_1, \dots, \mathbf{r}_T) = \sum_{t=1}^T \ln \mathbf{c}_t(F_{1t}(r_{1t}), \dots, F_{Nt}(r_{Nt})) + \sum_{j=1}^N \sum_{t=1}^T \ln f_{jt}(r_{jt}). \quad (13)$$

Assuming that the parameters of the marginals and the copula are all different, the last expression implies that we can maximize the log-likelihood in two steps. First, we estimate the parameters for each of the marginals  $F_{jt}$  by maximizing  $\ln L_j(\boldsymbol{\theta}_j) = \sum_{t=1}^T \ln f_{jt}(r_{jt})$  for  $j = 1, \dots, N$ . Second, using the estimated marginals  $\hat{F}_{1t}, \dots, \hat{F}_{Nt}$  we construct a pseudo-sample of observations for the copula<sup>11</sup>:

$$\hat{\mathbf{u}}_t = (\hat{F}_{1t}(r_{1t}), \dots, \hat{F}_{Nt}(r_{Nt})) \quad (14)$$

for  $t = 1, \dots, T$  and estimate its parameters by maximizing

<sup>10</sup>We are working with log returns for the factors, which we convert into arithmetic returns using the formula  $r_{weekly} = \exp(r_{ln}) - 1$  before computing their weighted sum. We then convert the arithmetic portfolio return back into log return.

<sup>11</sup>More specifically, because  $F_{jt}(r_{jt}) = F_j(z_{jt})$ , we use the residuals  $\hat{z}_{jt}$  from the first step to obtain the empirical CDF estimate  $\hat{F}_j(x) = \frac{1}{T+1} \sum_{t=1}^T \mathbf{1}_{\{\hat{z}_{jt} \leq x\}}$  and let  $\hat{u}_{jt} = \hat{F}_j(\hat{z}_{jt})$ .

$$\ln L_c(\boldsymbol{\theta}_c) = \sum_{t=1}^T \ln \mathbf{c}_t(\hat{\mathbf{u}}_t). \quad (15)$$

The parameter estimates for each of the marginals, obtained from the first step, are presented in the second part of Table 3. The first part shows the parameters estimates of the marginals under the normal distribution, obtained when estimating the static and dynamic multivariate normal distribution models <sup>12</sup>. The parameters estimates for the copulas, obtained from the second step, are given in Table 4. Our multivariate estimation results neatly replicate those of Christoffersen and Langlois (2013).

## 6 VaR and ES forecasts

The goal of the paper is to compare the out-of-sample risk forecasting accuracy of the univariate and multivariate approach. To do this we need to generate out-of-sample forecasts for the one-week-ahead VaR and ES with each of our ten models. We begin by estimating the parameters of each model using the first 20 years of weekly returns. This corresponds to the first 1,043 observations from 1963 to 1983. We re-estimate each model once a year, on July 1, using all the data available before that date, i.e. we are using an expanding estimation window. Although the parameters of the models are updated once a year, the conditional mean and variance of the equally weighted portfolio and of the FFC factors, as well as the conditional correlations between the FFC factors, are updated weekly.

In the case of multivariate models, once the one-week-ahead multivariate conditional distribution of the vector of factor returns  $\mathbf{F}_t$  is constructed at time  $t - 1$ , we simulate  $k = 100,000$  vectors of returns  $\tilde{\mathbf{r}}_t$  from this distribution. The simulated portfolio returns for the next period are then obtained by applying the current vector of equal portfolio weights  $\mathbf{w}_{t-1}$  to  $\tilde{\mathbf{r}}_t$  <sup>13</sup>:

---

<sup>12</sup>We also use a two-step estimation procedure for the static and dynamic multivariate normal distribution models. This is motivated by the likelihood decomposition given in Engle (2002).

<sup>13</sup>We do the same conversion as in footnote 10.

$$\{\tilde{r}_t^i\}_{i=1}^k = \{\mathbf{w}'_{t-1} \tilde{\mathbf{r}}_t^i\}_{i=1}^k \quad (16)$$

The  $p \cdot 100\%$  one-week-ahead estimated VaR and ES, denoted  $\widehat{VaR}_t^p$  and  $\widehat{ES}_t^p$  respectively, are computed via the following formulas :

$$\widehat{VaR}_t^p = -\text{Percentile}\left\{\{\tilde{r}_t^i\}_{i=1}^k, 100p\right\} \quad (17)$$

$$\widehat{ES}_t^p = -\frac{1}{p \cdot k} \sum_{i=1}^k \tilde{r}_t^i \cdot \mathbf{1}_{\{\tilde{r}_t^i < -\widehat{VaR}_t^p\}} \quad (18)$$

where  $\mathbf{1}_{\{\cdot\}}$  is an indicator function equal to 1 if the argument is true and 0 otherwise.

In the case of univariate models, although analytical expression for the VaR and ES are available, we still choose to rely on simulations to compute the risk measures in order to keep the estimation procedure as similar as possible between the two approaches. Thus, once the one-week-ahead univariate conditional distribution of the equally weighted portfolio return  $F_t$  is constructed at time  $t - 1$ , we simulate  $k$  portfolio returns  $\tilde{r}_t^i$  from this distribution and estimate the one-week-ahead VaR and ES directly from equations (17) and (18).

We consider two levels of significance :  $p = 1\%$  and  $p = 5\%$  for VaR and ES. Repeating the previous steps each period, we obtain, for each of the ten models, a time series of 1,436 out-of-sample forecasts for the one-week-ahead VaR and ES at both levels of significance. This out-of-sample period range from the first week of July 1983 to the end of December 2010; we lose the first 1,043 observations from the first estimation. For the rest of the paper,  $t = 1, \dots, T = 1436$  will denote the out-of-sample period.

## 7 Backtest

### 7.1 VaR backtest

We begin model comparisons by looking at the number of VaR violations at the 1% and 5% level<sup>14</sup>. A VaR violation occurs when the portfolio return drops below the estimated VaR, i.e.  $r_t < -\widehat{VaR}_t^p$ . Let  $I_t = \mathbf{1}_{\{r_t < -\widehat{VaR}_t^p\}}$  indicates whether a VaR violation occurred at time  $t$  and  $T_1 = \sum_{t=1}^T I_t$  be the total number of violations in the out-of-sample period, which is assumed positive. The number  $T_1$  is shown for each model in Table 5. Given our 1,436 observations, a correctly specified VaR model should produce around 14 and 72 violations at the 1% and 5% level respectively. The first column of Table 5 illustrates the difficulty in obtaining an adequate frequency of violations at the 1% level. Indeed, all models produce higher than 14 violations. The risk underestimation of the univariate normal distribution is particularly noticeable, producing more than 2.5 times the correct number of violations. Still, the static and dynamic skewed  $t$  copula models would fall inside the “green zone” under the Basel rules while all other models would fall inside the “yellow zone”<sup>15</sup>. The univariate skewed  $t$  distribution also performs relatively well compared with multivariate models; only the static and dynamic skewed  $t$  copula models produce a lower number of violations. At the 5% level, we see that obtaining the required frequency of violations is much more common. All the dynamic multivariate models have nearly exactly 72 violations while the static multivariate models have a slightly lower number of violations. Both univariate models have around 12 more violations than the required number.

We can test whether the observed fraction of violations is statistically different from  $p = 1\%$  or  $5\%$  using the unconditional coverage test introduced in Kupiec (1995). Details for the test are given in Appendix C.1 and the p-values appear under the UC columns of Table 5. For the 1% VaR using significance level of  $10\%$ <sup>16</sup> we reject all ten models.

---

<sup>14</sup>We also performed the VaR backtests at the 2.5% level. The results were between those at the 1% and 5% levels.

<sup>15</sup>Under the Basel rules, based on a sample of 250 observations for the 1% VaR, 4 or less violations corresponds to the green zone, between 5 and 9 violations corresponds to the yellow zone and 10 or more violations corresponds to the red zone.

<sup>16</sup>A higher significance level for the tests can be justified by the fact that failing to reject a correct

The dynamic Skewed  $t$  copula barely fails the test. Again, this highlights the challenge of adequately modeling the 1% left tail of the distribution of portfolio returns. The results are opposite when we consider the 5% VaR; all our models pass the test at the 10% significance level except for the Normal distribution and the static Skewed  $t$  copula.

Another property of a correctly specified VaR model is the independence of the violations through time. We can test for first order dependence of the violations using the independence test of Christoffersen (1998). Details on the independence test are given in Appendix C.2 and the p-values appear under the Ind columns of Table 5. For the 1% VaR, using a 10% significance level we see that only the univariate Normal distribution and the static Skewed  $t$  Copula are not rejected, with 3 and 1 consecutive violations ( $T_{11}$ ) respectively. In fact, all of our models have between 1 and 4 consecutive violations in the out-of-sample period but those with a higher frequency of violation usually display a lower discrepancy between  $\hat{\pi}_{01}$  and  $\hat{\pi}_{11}$ . For example, the dynamic Skewed  $t$  Copula has the same number of consecutive violations as the univariate normal distribution but a much lower number of violations overall, leading to a relatively high  $\hat{\pi}_{11}$  compared with  $\hat{\pi}_{01}$ . The static Skewed  $t$  Copula is an exception to the rule because it is the only model with only one pair of consecutive violations. For the 5% VaR, the results are similar; all models have 7 or 8 consecutive violations but only the univariate models pass the test.

We can jointly test for coverage and independence with the conditional coverage test of Christoffersen (1998). Details on the test can be found in Appendix C.3 and the p-values are displayed under the CC columns of Table 5. For the 1% VaR we see that the results are consistent with the conclusions of the two previous tests. Indeed, none of our models passed both of the previous tests at the 10% level and consequently none of them pass the conditional coverage test at this level either. For the 5% VaR the results are different. Although all multivariate models failed the independence test at the 10% level, most of them are not rejected in the conditional coverage test at this level. Both univariate models passed all three tests.

We can summarize this section by saying that our models fail to display both adequate

---

model (Type II error) can be very costly in risk management. See chapter 13 of Christoffersen (2012) for details.

coverage and independent violations for the 1% VaR. The Normal distribution, univariate or multivariate, clearly does not capture the 1% tail risk, whereas the univariate Skewed  $t$  distribution has a similar coverage to some multivariate models. The only model that comes close to an adequate 1% frequency of violations is the Dynamic Skewed  $t$  copula. In the case of the 5% VaR all models display adequate coverage but multivariate models fewer violations. Independence of violations is also much easier to obtain.

## 7.2 ES backtest

Contrary to VaR, there is no loss function for which ES is the unique minimizer. This important result is shown in Gneiting (2011) and often goes under the name “lack of elicibility”. This finding sparked a debate over whether it is even possible to backtest ES. Fortunately, the recent literature has clearly answered the question by proposing many ES backtests that do not rely on the elicibility property, although the procedures are not as straightforward as for VaR. Among the growing literature in this area<sup>17</sup>, we choose to implement the first two tests proposed by Acerbi and Szekely (2014). These tests do not make any assumptions about the distribution of returns (nonparametric) and are simple to execute.

For the tests we assume that, each period (week), the portfolio return  $r_t$  is distributed according to a real but unknown distribution  $F_t$  and forecast using a predictive conditional distribution  $P_t$ . Thus  $VaR_t^p$  and  $ES_t^p$  represent the true risk measures, i.e. when  $r_t \sim F_t$ , while  $\widehat{VaR}_t^p$  and  $\widehat{ES}_t^p$  represent the estimated risk measures, i.e. when  $r_t \sim P_t$ . Portfolio returns are assumed to be independent but not identically distributed. The null hypothesis for both tests is

$$H_0 : P_t^{[p]} = F_t^{[p]} \text{ for all } t \quad (19)$$

where  $P_t^{[p]}(\cdot) = \min \left\{ 1, \frac{P_t(\cdot)}{p} \right\}$  is the left tail of the distribution below the  $p$  quantile. In the first test we assume that a preliminary VaR test has been done and consider the alternative hypothesis

---

<sup>17</sup>See for example Du and Escanciano (2016) on conditional backtests for ES.



$$\begin{aligned}
H_1 : \widehat{ES}_t^p &\geq ES_t^p \text{ for all } t \text{ and } > \text{ for some } t \\
\widehat{VaR}_t^p &= VaR_t^p \text{ for all } t .
\end{aligned} \tag{20}$$

Note that the predicted VaRs are still adequate under  $H_1$ . This means that we should first accept the model for VaR before performing the test. We test  $H_0$  against  $H_1$  with the following statistic:

$$Z_1 = \frac{\sum_{t=1}^T \frac{r_t \cdot I_t}{ES_t^p}}{T_1} + 1. \tag{21}$$

As an average of VaR exceedances, the statistic is in fact completely insensitive to an excessive number of violations. It can be shown that the expected value of  $Z_1$  is zero under  $H_0$  and negative under  $H_1$ . Therefore, we expect a realized value of  $Z_1$  that is close to 0. The value signals a problem when it is negative.

The second test evaluates both the frequency and magnitude of VaR violations. That is, it is a joint test for VaR and ES coverage. The null hypothesis (19) is tested against the following alternative

$$\begin{aligned}
H_1 : \widehat{ES}_t^p &\geq ES_t^p \text{ for all } t \text{ and } > \text{ for some } t \\
\widehat{VaR}_t^p &\geq VaR_t^p \text{ for all } t .
\end{aligned} \tag{22}$$

The test statistic is

$$Z_2 = \sum_{t=1}^T \frac{r_t \cdot I_t}{T \cdot p \cdot ES_t^p} + 1. \tag{23}$$

Again, it can be shown that the expected value of  $Z_2$  is zero under  $H_0$  and negative under  $H_1$ .

The distributions of  $Z_1$  and  $Z_2$  under the null hypothesis are unknown but can be approximated using simulations. For each of our model, the p-values of both tests are obtained using the following steps:

1. Simulate  $M$  random portfolio returns  $\{\tilde{r}_t^i\}_{i=1}^M$  for each  $t = 1, \dots, T$ .
2. Compute  $Z_1^i$  and  $Z_2^i$  using  $\{\tilde{r}_t^i\}_{t=1}^T$  for each  $i = 1 \dots M$ .
3. Estimate  $\text{pval}_1 = \frac{\sum_{i=1}^M \mathbf{1}_{(Z_1^i < Z_1)}}{M}$  and  $\text{pval}_2 = \frac{\sum_{i=1}^M \mathbf{1}_{(Z_2^i < Z_2)}}{M}$ .

We use  $M = 5000$  and simulate the  $M$  portfolio returns each week with the predictive conditional distribution  $P_t$  used for VaR and ES forecasts.

The p-values of the two tests are shown in the columns  $Z_1$  and  $Z_2$  of Table 5. Because we rejected all models for the 1% VaR in the last section, interpreting the results of  $Z_1$  in the first part of the table seems hazardous. At least we can note that the dynamic Skewed  $t$  Copula, which is the model with the coverage closest to 1%, passes the test at the 10% significance level. The estimated  $Z_1$  p-values for the 5% ES are more revealing. Indeed, although all models have adequate 5% VaR coverage, only the univariate Skewed  $t$  distribution passes the test, and its p-value is much higher than that of all the other models, including asymmetric multivariate models. Asymmetric distributions tend to result in more accurate ES forecasts because they better capture the fat tail behavior in stock returns, but with the first test this is only apparent in the univariate case. Turning now to the p-values of  $Z_2$ , we see that we reject all models for the 1% ES at the 10% level. In the case of the 5% ES we observe that all copula models, except for the static Normal Copula, pass the test. Also, even though the univariate skewed  $t$  distribution passed the first test for the 5% ES, it is rejected in the second test. This discrepancy between the results of the two tests for the 5% ES is surprising given that all models passed the unconditional coverage test.

Overall, just like our models failed the 1% VaR backtests, they also fail the 1% ES backtests. In the case of the 5% ES, the two tests offer conflicting results.

## 8 Loss functions

As mentioned in the introduction, backtesting procedures are not the right tools to compare models. These tests provide a binary outcome, i.e. reject or do not reject a model,

whereas we would like to rank models from worst to best. In other words, backtests are useful for absolute evaluation, not relative evaluation. One could still argue that models that fail the backtest in the first stage are clearly worse than those that passed it, and should therefore be discarded before moving to the second stage. In this paper, however, we choose to keep all models irrespectively of the results of their first stage backtest. We do this because otherwise we would have no models left for the 1% VaR and ES and, as Novales Cinca and Garcia-Jorcano (2017) argue, rejected models in the first stage could still end up being top performers in the second stage when the size of VaR exceedances is taken into account.

We are interested in comparing the out-of-sample VaR and ES forecasting accuracy of our models. The standard approach to this end is to use loss functions, which compare our risk forecasts, here VaR and ES, to the realized portfolio return at each period. The idea is that a model with a given average loss will be preferred to a model with a higher average loss over the out-of-sample period. Therefore, in selecting a loss function we should make sure that our forecasting object minimizes its expected value. The existence of such a loss function for a given statistic, referred to as the elicibility property, is not automatic. Indeed, VaR is elicitable but ES is not. The  $VaR_t^p$  loss function, also known as the “check” or “tick” loss function in quantile regressions, is defined by

$$L_V(v_t, r_t) = (r_t + v_t)(p - \mathbf{1}_{\{r_t + v_t < 0\}}) \quad (24)$$

where  $v_t$  is a non-random real variable. This is an asymmetric function because the penalty given when  $\mathbf{1}_{\{r_t + v_t < 0\}} = 1$  is usually much higher than in the case where  $\mathbf{1}_{\{r_t + v_t < 0\}} = 0$ . Nonetheless, this function penalizes risk overestimation because, conditional on  $\mathbf{1}_{\{r_t + v_t < 0\}} = 1$ , a higher value for  $v_t$  leads to a higher penalty. The function  $L_V$  is the “right” loss function for VaR in the sense that

$$VaR_t^p = \arg \min_{v_t} \mathbb{E}[L_V(v_t, r_t)]. \quad (25)$$

Thus, computing  $\bar{L}_V = \frac{1}{T} \sum_{t=1}^T L_V(v_t, r_t)$  with  $v_t = \widehat{VaR}_t^p$  for each model and sorting

the numbers in increasing order allows us to rank models in descending order of VaR predictive accuracy. That is, the model with the lowest value of  $\bar{L}_V$  is deemed the most accurate VaR model, the one with the second-lowest value of  $\bar{L}_V$  is deemed the second best VaR model and so on, to the model with the highest value of  $\bar{L}_V$  which is deemed the worst VaR model.

The choice of the loss function is important because different choices can lead to very different results<sup>18</sup>. Here we take the point of view of a risk manager interested in accurately forecasting the  $p$  quantile of  $F_t$ . Other researchers have taken the point of view of the regulator and employed loss functions that penalize risk underestimation only. This view was pioneered by Lopez (1999) who proposed the following regulator loss function (RFL) for VaR:

$$RLF(v_t, r_t) = [1 + (r_t + v_t)^2] \mathbf{1}_{\{r_t + v_t < 0\}}. \quad (26)$$

However, because VaR does not minimize the expected value of  $RLF$  with respect to  $v_t$ , it is unclear why a model with a lower average loss  $\overline{RLF} = \frac{1}{T} \sum_{t=1}^T RLF(v_t, r_t)$  than another model should be deemed superior.

The VaR out-of-sample average loss  $\bar{L}_V$  for each model, in basis points, is shown in the first two columns of Table 6. We see that the univariate Skewed  $t$  distribution displays the lowest average loss for both the 1% and 5% VaR. The univariate normal distribution is also clearly not as inaccurate as the backtest might have suggested, given that its average loss is lower than all non-dynamic multivariate models for both the 1% and 5% VaR. This indicates that the frequency of VaR violations is only half of the picture in the loss function; the overestimation of risk is also important. The dynamic multivariate models, except for the multivariate normal distribution, appear more accurate than the univariate normal distribution but less accurate than the univariate Skewed  $t$  distribution. Had we used Lopez's (1999) RLF function, the results would have been completely different. The univariate Skewed  $t$  distribution would have ranked third for the 1% VaR, surpassed by the dynamic and non-dynamic Skewed  $t$  Copula models, and last for the 5% VaR.

---

<sup>18</sup>See section 5.2 in Nieto and Ruiz (2016).

Therefore, in our application, taking risk overestimation into account is a determining factor in selecting VaR models. Nonetheless, we believe that  $L_V$  is the best criterion for VaR model comparison because it is a natural loss function in the sense of equation (25), unlike  $RLF$ .

We now turn to ES model comparison. Because ES is not elicitable, there is no natural loss function available for comparing the ES forecasting accuracy of our models. However, it turns out that, because ES is a function of VaR, the pair (VaR,ES) is jointly elicitable. This fact has been noted by Acerbi and Szekely (2014) and demonstrated by Fissler et al. (2016). Thus, although we cannot compare our models based on their ability to forecast ES alone, we can compare them based on their ability to jointly forecast VaR and ES. We choose to use Acerbi and Szekely's (2014) joint loss function, which has the form

$$L_{V,E}(v_t, e_t, r_t) = \frac{p}{2}e_t^2 + \frac{\delta p}{2}v_t^2 - pe_tv_t + (e_t(v_t + r_t) + \frac{\delta}{2}(r_t^2 - v_t^2))\mathbf{1}_{\{r_t+v_t < 0\}} \quad (27)$$

where  $\delta$  is a real number. For a large class of distributions, it can be shown that

$$\{VaR_t^p, ES_t^p\} = \arg \min_{v_t, e_t} \mathbb{E}[L_{V,E}(v_t, e_t, r_t)] \quad (28)$$

under the condition that  $\delta \cdot VaR_t > ES_t$ . Therefore a natural criterion for VaR/ES model comparison is given by  $\bar{L}_{V,E} = \frac{1}{T} \sum_{t=1}^T L_{V,E}(v_t, e_t, r_t)$  with  $v_t = \widehat{VaR}_t^p$  and  $e_t = \widehat{ES}_t^p$ . We use  $\delta = 2$ , which makes the inequality  $\delta \cdot VaR_t > ES_t$  true for all  $t = 1, \dots, T$  and all models at both the 1% and 5% levels.

The out-of-sample average joint loss  $\bar{L}_{V,E}$  for each model, in basis points, is shown in the last two columns of Table 6. The results are identical to those in the first two columns. The univariate Skewed  $t$  distribution is again the most accurate model at the 1% and 5% levels. This confirmed preference for the univariate Skewed  $t$  distribution is coherent because ES is a function of VaR, which implies that more accurate VaR forecasts should also lead to more accurate ES forecasts. The univariate Normal distribution still displays a lower average joint loss than all non-dynamic multivariate models for both the

1% and 5% levels.

We can summarize this section by saying that loss functions give a very different picture of the relative performance of our models compared with the backtests. Indeed, models with the frequency of violations closest to  $p$  in the first stage do not emerge as the most accurate models in the second stage. Models that produce higher VaR and ES estimates, although closer to an adequate frequency of violation, are penalized for risk overestimation in periods where no violations occur. Given that most of the out-of-sample periods are without violations, it appears that the cost of overestimating VaR and ES can be quite high.

## 9 Testing for comparative predictive accuracy

Although there are differences between models' average losses, these differences may not be statistically significant. We can test for pairwise differences in predictive accuracy using the approach introduced by Diebold and Mariano (1995).

Let  $L$  stands for either  $L_V$  or  $L_{V,E}$  and let

$$d_{ijt} = L_{it} - L_{jt} \quad (29)$$

represent the loss differential between model  $i$  and model  $j$  at time  $t$ . The null hypothesis of equal predictive ability between two models can be formulated as

$$H_0 : \mathbb{E}(d_{ijt}) = 0 \quad . \quad (30)$$

Assuming that the loss differential  $d_{ijt}$  is stationary, the null hypothesis can be tested via a simple  $z$ -test with the test statistic

$$DM_{ij} = \frac{\bar{d}_{ij}}{\hat{\sigma}_{\bar{d}_{ij}}} \stackrel{a}{\sim} N(0, 1) \quad (31)$$

where  $\bar{d}_{ij} = \frac{1}{T} \sum_{t=1}^T d_{ijt}$  and  $\hat{\sigma}_{\bar{d}_{ij}}$  is a consistent estimate of the standard deviation of  $\bar{d}_{ij}$ . We compute  $\bar{d}_{ij}$  by regressing the loss differential on an intercept and obtain a

heteroscedasticity and autocorrelation consistent (HAC) estimate for  $\sigma_{\bar{a}_{ij}}$ .

The p-values for the bilateral test are displayed for each pair of models in Tables 7 to 10. Tables 7 and 8 compare each pair of models using the VaR loss function whereas Tables 9 and 10 use Acerbi and Szekely (2014) VaR/ES joint loss function. The numbers in parentheses represent the fraction of the average loss for the column model to the average loss for the row model. An up (left) arrow indicates that we reject the null hypothesis of equal predictive ability at the 5% significance level and that the column (row) model outperforms the corresponding row (column) model. Looking at the first row we see that the univariate Normal distribution model is outperformed by the univariate Skewed  $t$  model for the 1% VaR and 1% VaR/ES but there are no significant differences compared with other models. For the 5% VaR and 5% VaR/ES, the average loss differential between the two univariate models disappears and the univariate Normal distribution outperforms all static multivariate models. The second row indicates that in all cases the univariate Skewed  $t$  distribution outperforms all static multivariate models, but the average loss differential with dynamic multivariate models is not significant. We also see that dynamic multivariate models usually outperform static multivariate models.

The previous test can also be done in a conditional framework. In this case, instead of testing only which model risk forecasts are more accurate on average, we also test for predictability of the loss differential. With this perspective the null hypothesis becomes

$$H_0 : \mathbb{E}(d_{ijt} \mid \mathcal{F}_{t-1}) = 0 \ . \quad (32)$$

A statistic for testing this null hypothesis was developed by Giacomini and White (2006). Implementing their test with  $h_{t-1} = (1, d_{ijt-1})'$  as the test function, the results obtained are identical to those of the DM test and we therefore omit them.

## 10 The Model Confidence Set

Tables 7-10 contain a lot of information. We can summarize the results with the Model Confidence Set (MCS) procedure of Hansen et al. (2011). The idea is to construct a set

of models, the MCS, that will contain the best model with a given level of confidence, analogous to a confidence interval for a parameter. This set is obtained by a sequence of equal predictive ability (EPA) tests that, in case of rejection of the null hypothesis, allow us to trim the set of candidate models by eliminating the worst performing model according to an elimination rule. The steps are repeated until the EPA hypothesis fails to be rejected, in which case the set of surviving models constitute the MCS that contains the best performing model with the desired level confidence.

Let  $\mathcal{M}_0$  denote our initial set of ten models and let  $\mathcal{M} \subseteq \mathcal{M}_0$  be a nonempty subset of these models. The EPA null hypothesis is

$$H_0 : \mathbb{E}(d_{ijt}) = 0 \text{ for all } i, j \in \mathcal{M}. \quad (33)$$

Note that this null hypothesis is identical to the DM null hypothesis, except that we consider all pairs of models in  $\mathcal{M}$  instead of a single pair. As Hansen et al. (2011) discuss, a natural range statistic for testing  $H_0$  is

$$T_R = \max_{i,j \in \mathcal{M}} |DM_{ij}|. \quad (34)$$

The EPA test statistic is thus the absolute value of the DM statistic farthest away from zero among the pairs of models. Because the asymptotic distribution of  $T_R$  is unknown, Hansen et al. (2011) propose to estimate it via a circular block bootstrap scheme. This allows the computation of a bootstrap p-value for the EPA test.

The MCS procedure begins by setting  $\mathcal{M} = \mathcal{M}_0$ . We then perform the EPA test on the models in  $\mathcal{M}$ . If we reject the null hypothesis at the chosen level of confidence, we identify the worst model  $i^*$ , defined as the model with the highest loss relative to another model. In other words,

$$i^* = \arg \max_{i \in \mathcal{M}} \max_{j \in \mathcal{M}} DM_{ij}. \quad (35)$$

This model is then eliminated from  $\mathcal{M}$ . We repeat this process until we fail to reject the



EPA hypothesis, in which case we set  $MCS = \mathcal{M}$ .

We implement the MCS approach at a 95% confidence level <sup>19</sup>. We use 10 000 bootstrap re-samples with a circular bootstrap scheme and a block of one. The inputs for the procedure are the losses  $L_{it}$  of each model; the statistics  $DM_{ij}$  are computed using a bootstrap estimate of  $\sigma_{\bar{d}_{ij}}$  instead of the HAC estimate in the previous section. The results are presented in Table 11. For the 5% VaR and 5% VaR/ES, the MCS consists of all univariate and all dynamic multivariate models. That is, the procedure eliminated all static multivariate models. At the 1% level, the VaR MCS includes the univariate skewed  $t$  model as well as all dynamic multivariate models, with the exception of the dynamic multivariate normal model. Lastly, the 1% VaR/ES MCS consist of all models. This means that the EPA null hypothesis was not rejected by the first test.

These results agree with those of the last section. At the 5% level the MCS cannot identify a set of superior models among the univariate and dynamic multivariate models but eliminates all multivariate models without dynamic correlations. There is thus a significant improvement from including dynamic correlations in the multivariate models. For the 1% VaR, the MCS also eliminates the univariate normal model, but the univariate skewed  $t$  model remains in the set of superior models so there is still no clear preference between the two approaches. For the 1% VaR/ES, average loss differentials between the models are too small to distinguish superior models.

## 11 Conclusion

The goal of this paper is to compare a univariate approach to a multivariate approach for forecasting the one-week-ahead Expected Shortfall of stock portfolios. We do this comparison in the context of a portfolio equally invested in the FFC factors by considering the data and multivariate models employed by Christoffersen and Langlois (2013). In total, two univariate models are compared with eight multivariate models involving asymmetric distributions and asymmetric copulas with dynamic correlations. Using sim-

---

<sup>19</sup>We use the “mcs” MATLAB function from the MFE toolbox by Kevin Sheppard. See [https://www.kevinsheppard.com/MFE\\_Toolbox](https://www.kevinsheppard.com/MFE_Toolbox).

ulations, we generate 1,436 weekly out-of-sample VaR and ES for each model covering the period from 1983 to 2010. We analyze the relative performance of our models in two stages. In the first stage we backtest each model by comparing the ex-ante risk measures to the ex-post portfolio returns. In the second stage we rely on loss functions based on the elicibility property of VaR as well as the joint elicibility property of VaR and ES to rank models. We test for statistical differences between the average loss of models with the tests of Diebold and Mariano (1995) and Giacomini and White (2006), along with the MCS procedure of Hansen et al. (2011).

We find no significant differences between the risk forecasting accuracy of univariate models and multivariate models with dynamic correlations. However, we find significant differences in the risk forecasting accuracy of univariate models and multivariate models without dynamic correlations. These differences all support the univariate models. We also find that dynamic correlations produce a significant gain in the accuracy of multivariate models.

## References

- Acerbi, C. and B. Szekely (2014). Back-testing expected shortfall. *Risk*, 76–81.
- Aielli, G. P. (2013). Dynamic conditional correlation: On properties and estimation. *Journal of Business & Economic Statistics* 31(3), 282–299.
- Basel Committee on Banking Supervision (2016). Minimum capital requirements for market risk. Available at <https://www.bis.org/bcbs/publ/d352.pdf>.
- Berkowitz, J. and J. O’Brien (2002). How accurate are value-at-risk models at commercial banks? *The Journal of Finance* 57(3), 1093–1111.
- Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy* 96(1), 116–131.
- Brooks, C. and G. Persaud (2003). Volatility forecasting for risk management. *Journal of Forecasting* 22(1), 1–22.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance* 52(1), 57–82.
- Christoffersen, P. (2009). Value-at-risk models. *Handbook of Financial Time Series*, 753–766.
- Christoffersen, P. and H. Langlois (2013). The joint dynamics of equity market factors. *Journal of Financial and Quantitative Analysis* 48(5), 1371–1404.
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review* 39(4), 841–862.
- Christoffersen, P. F. (2012). *Elements of Financial Risk Management*. Academic Press.
- Demarta, S. and A. J. McNeil (2005). The t copula and related copulas. *International Statistical Review / Revue Internationale de Statistique* 73(1), 111–129.

- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13(3), 253–263.
- Dimitriadis, T. and S. Bayer (2017, April). A Joint Quantile and Expected Shortfall Regression Framework. *ArXiv e-prints*.
- Du, Z. and J. C. Escanciano (2016). Backtesting expected shortfall: accounting for tail risk. *Management Science* 63(4), 940–958.
- 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. F. and V. K. Ng (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance* 48(5), 1749–1778.
- Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33(1), 3–56.
- Fissler, T., J. F. Ziegel, et al. (2016). Higher order elicibility and osband’s principle. *The Annals of Statistics* 44(4), 1680–1707.
- Giacomini, R. and I. Komunjer (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economic Statistics* 23(4), 416–431.
- Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica* 74(6), 1545–1578.
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association* 106(494), 746–762.
- Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review* 35(3), 705–730.
- Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica* 79(2), 453–497.

- Kole, E., T. Markwat, A. Opschoor, and D. van Dijk (2017). Forecasting value-at-risk under temporal and portfolio aggregation\*. *Journal of Financial Econometrics* 15(4), 649–677.
- Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives* 3(2), 73–84.
- Lopez, J. A. (1999). Methods for evaluating value-at-risk estimates. *Economic Review-Federal Reserve Bank of San Francisco* (2), 3–17.
- McAleer, M. and B. Da Veiga (2008). Single-index and portfolio models for forecasting value-at-risk thresholds. *Journal of Forecasting* 27(3), 217–235.
- Nieto, M. R. and E. Ruiz (2016). Frontiers in var forecasting and backtesting. *International Journal of Forecasting* 32(2), 475 – 501.
- Novalés Cinca, A. and L. García-Jorcano (2017). Volatility specifications versus probability distributions in var forecasting. Available at [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3023885](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3023885).
- Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review* 47(2), 527–556.
- Santos, A. A. P., F. J. Nogales, and E. Ruiz (2013). Comparing univariate and multivariate models to forecast portfolio value-at-risk. *Journal of Financial Econometrics* 11(2), 400–441.
- Sklar, M. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.
- Yoshihara, T. (2018). Maximum likelihood estimation of skew-t copulas with its applications to stock returns. *Journal of Statistical Computation and Simulation*, 1–18.

## Appendix A Hansen's (1994) skewed $t$ distribution

The density of Hansen's (1994) skewed  $t$  distribution is given by

$$f(z; \kappa, \nu) = \begin{cases} bc \left( 1 + \frac{1}{\nu-2} \left( \frac{bz+a}{1-k} \right)^2 \right)^{-\frac{\nu+1}{2}} & \text{if } z < -\frac{a}{b} \\ bc \left( 1 + \frac{1}{\nu-2} \left( \frac{bz+a}{1+k} \right)^2 \right)^{-\frac{\nu+1}{2}} & \text{if } z \geq -\frac{a}{b} \end{cases} \quad (36)$$

where  $-1 < \kappa < 1$  and  $2 < \nu < \infty$ . The constants  $a, b$  and  $c$  are given by

$$a = 4\kappa c \frac{\nu-2}{\nu-1}, \quad (37)$$

$$b^2 = 1 + 3\kappa^2 - a^2, \quad (38)$$

and

$$c = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi(\nu-2)}\Gamma(\nu/2)}. \quad (39)$$

The univariate skewed  $t$  distribution has a mean of zero and a unit variance. Its skewness and kurtosis are given by

$$\mathbb{E}(z^3) = \frac{m_3 - 3am_2 + 2a^3}{b^3}, \quad (40)$$

and

$$\mathbb{E}(z^4) = \frac{m_4 - 4am_3 + 6a^2m_2 - 3a^4}{b^4} \quad (41)$$

where

$$m_2 = 1 + 3\kappa^2, \quad (42)$$

$$m_3 = 16c\kappa(1 + \kappa^2) \frac{(\nu-2)^2}{(\nu-1)(\nu-3)}, \quad \text{if } \nu > 3 \quad (43)$$

and

$$m_4 = 3 \frac{(\nu-1)}{(\nu-4)} (1 + 10\kappa^2 + 5\kappa^4), \quad \text{if } \nu > 4. \quad (44)$$

# Appendix B Demarta and McNeil's (2005) skewed $t$ copula

## B.1 Multivariate skewed $t$ distribution density

Demarta and McNeil's (2005) skewed  $t$  copula is based on a version of the multivariate skewed  $t$  distribution. We say that the  $N \times 1$  random vector  $\mathbf{x}$  follows a multivariate skewed  $t$  distribution, denoted  $\mathbf{x} \sim \mathbf{F}_{\nu_c, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}}^{st}$ , if it has the following density function:

$$\mathbf{f}^{st}(\mathbf{x}; \nu_c, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = c \frac{K_{\frac{\nu_c+N}{2}} \left( \sqrt{(v_c + (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})) \boldsymbol{\lambda}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\lambda}} \right) \exp(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} \boldsymbol{\lambda}}{\left( \sqrt{(v_c + (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})) \boldsymbol{\lambda}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\lambda}} \right)^{-\frac{\nu_c+N}{2}} \left( 1 + \frac{(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}{v_c} \right)^{\frac{\nu_c+N}{2}}} \quad (45)$$

where  $K_\gamma(\cdot)$  is the modified Bessel function of the third kind and  $c$  is a constant given by

$$c = \frac{2^{\frac{2-(\nu_c+N)}{2}}}{\Gamma\left(\frac{\nu_c}{2}\right) (\pi v_c)^{\frac{N}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}}. \quad (46)$$

Also,  $\nu_c$  is scalar degree of freedom parameter,  $\boldsymbol{\mu}$  is a  $N \times 1$  vector of location parameters,  $\boldsymbol{\Sigma}$  is a  $N \times N$  symmetric positive definite dispersion matrix and  $\boldsymbol{\lambda}$  is a  $N \times 1$  vector of asymmetry parameters. The first two moments of  $\mathbf{x}$  are given by

$$\mathbb{E}(\mathbf{x}) = \boldsymbol{\mu} + \frac{v_c}{v_c - 2} \boldsymbol{\lambda} \quad (47)$$

and

$$\text{Cov}(\mathbf{x}) = \frac{v_c}{v_c - 2} \boldsymbol{\Sigma} + \frac{2v_c^2}{(v_c - 2)^2(v_c - 4)} \boldsymbol{\lambda} \boldsymbol{\lambda}'. \quad (48)$$

The multivariate skewed  $t$  distribution has the following stochastic representation

$$\mathbf{x} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{w} \mathbf{y} + \boldsymbol{\lambda} w \quad (49)$$

where  $w$  is an inverse gamma random variable,  $w \sim IG(\nu_c/2, \nu_c/2)$ ,  $\mathbf{y}$  a  $N \times 1$  vector of

normal variables,  $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ , and  $\mathbf{y}$  and  $w$  are independent.

## B.2 Skewed $t$ copula density

The skewed  $t$  copula is derived using the standardized multivariate skewed  $t$  distribution

$\mathbf{F} := \mathbf{F}_{\nu_c, \mathbf{0}, \Upsilon, \boldsymbol{\lambda}}^{st}$  with  $j$ th marginal  $F_j$  and correlation matrix  $\Upsilon$  :

$$\mathbf{C}_{\nu_c, \Upsilon, \boldsymbol{\lambda}}^{st}(\mathbf{u}) = \mathbf{F}(F_1^{-1}(u_1), \dots, F_N^{-1}(u_N)). \quad (50)$$

Let  $\mathbf{f} := \mathbf{f}_{\nu_c, \mathbf{0}, \Upsilon, \boldsymbol{\lambda}}^{st}$  be the density of  $\mathbf{F}$  with  $j$ th marginal  $f_j$ . The copula density is obtained by differentiating both sides of equation (50):

$$\begin{aligned} \mathbf{c}^{st}(\mathbf{u}; \nu_c, \Upsilon, \boldsymbol{\lambda}) &= \frac{\mathbf{f}(F_1^{-1}(u_1), \dots, F_N^{-1}(u_N))}{\prod_{j=1}^N f_j(F_j^{-1}(u_j))} \\ &= \frac{2^{\frac{(v_c-2)(N-1)}{2}} K_{\frac{v_c+N}{2}} \left( \sqrt{(v_c + \boldsymbol{\eta}' \Upsilon^{-1} \boldsymbol{\eta}) \boldsymbol{\lambda}' \Upsilon^{-1} \boldsymbol{\lambda}} \right) \exp(\boldsymbol{\eta}' \Upsilon^{-1} \boldsymbol{\lambda})}{\Gamma\left(\frac{v_c}{2}\right)^{1-N} |\Upsilon|^{\frac{1}{2}} \left( \sqrt{(v_c + \boldsymbol{\eta}' \Upsilon^{-1} \boldsymbol{\eta}) \boldsymbol{\lambda}' \Upsilon^{-1} \boldsymbol{\lambda}} \right)^{-\frac{v_c+N}{2}} \left( 1 + \frac{\boldsymbol{\eta}' \Upsilon^{-1} \boldsymbol{\eta}}{v_c} \right)^{\frac{v_c+N}{2}}} \\ &\quad \times \prod_{j=1}^N \frac{\left( \sqrt{(v_c + \eta_j^2) \lambda_j^2} \right)^{-\frac{v_c+1}{2}} \left( 1 + \frac{\eta_j^2}{v_c} \right)^{\frac{v_c+1}{2}}}{K_{\frac{v_c+1}{2}} \left( \sqrt{(v_c + \eta_j^2) \lambda_j^2} \right) \exp(\eta_j \lambda_j)} \end{aligned} \quad (51)$$

where  $\eta_j := F_j^{-1}(u_j)$  is defined as the  $j$ th copula quantile for  $j = 1, \dots, N$ . The density in equation (51) is used to obtain the skewed  $t$  copula parameter estimates in the second step of the estimation, i.e. when maximizing the log-likelihood in equation (15).

## B.3 Copula quantiles

One difficulty in the the second step of the skewed  $t$  copula ML estimation is obtaining the copula quantiles  $\eta_{jt} = F_j^{-1}(u_{jt})$  because the inverse marginals  $F_j^{-1}$  are not known in closed form. Christoffersen and Langlois (2013) address this problem by using empirical



quantiles from a large number simulations with representation (49). Yoshihara (2018) simulation results suggest that using a monotone interpolator is faster and more accurate than using empirical quantiles so we instead choose this approach to compute the copula quantiles. We apply following procedure with  $m = 150$  interpolating points:

1. Let  $u_{min} = \min_{j=1, \dots, T} u_{jt}$  and  $u_{max} = \max_{j=1, \dots, T} u_{jt}$ .
2. Compute  $\eta_{min} = F_j^{-1}(u_{min})$  and  $\eta_{max} = F_j^{-1}(u_{max})$  using an accurate quantile function. This is done by finding the quantile such that numerical integration of the univariate density  $f_j$  until that point equals  $u_{min}$  or  $u_{max}$  respectively. We use the bisection method for this.
3. Calculate  $\eta_k = \eta_{min} + \frac{(\eta_{max} - \eta_{min})}{m-1}$  and compute  $u_k = F_j(\eta_k)$  by numerical integration of  $f_j$  for  $k = 2, \dots, m-1$ .
4. Use a monotone interpolator with the data points  $\{(\eta_{min}, u_{min}), \dots, (\eta_{max}, u_{max})\}$  to obtain  $\eta_{jt} = F_j^{-1}(u_{jt})$  for all other values of  $u_{jt} \in [u_{min}, u_{max}]$ . We apply MATLAB piecewise cubic Hermite interpolating polynomial (“pchip”) to this end.
5. Repeat for  $j = 1, \dots, N$ .

## B.4 Aielli’s (2013) cDCC

In the dynamic skewed  $t$  copula model we allow the copula correlation matrix  $\mathbf{\Upsilon}$  to evolve through time. Let  $\mathbf{\Upsilon}_t$  be the copula correlation matrix at time  $t$  and let  $\dot{\mathbf{\Upsilon}}_t$  be the copula quantiles correlation matrix at time  $t$ . Also, let

$$\bar{\eta}_j := \mathbb{E}(\eta_j) = \frac{v_c}{v_c - 2} \lambda_j \quad (52)$$

and

$$s_j := \text{var}(\eta_j) = \frac{v_c}{v_c - 2} + \frac{2v_c^2}{(v_c - 2)^2(v_c - 4)} \lambda_j^2 \quad (53)$$

be respectively the expectation and the standard deviation of the  $j$ th copula quantile. This allows us to define the standardized copula quantiles  $\epsilon_{jt} = \frac{\eta_{jt} - \bar{\eta}_j}{s_j}$  for  $j = 1, \dots, N$

and the vector  $\boldsymbol{\epsilon}_t = (\epsilon_{jt}, \dots, \epsilon_{Nt})'$  containing them. The link between the two correlation matrices  $\boldsymbol{\Upsilon}_t$  and  $\dot{\boldsymbol{\Upsilon}}_t$  is given by equation (48):

$$\boldsymbol{\Upsilon}_t = \frac{\nu_c - 2}{\nu_c} \left( \boldsymbol{D} \dot{\boldsymbol{\Upsilon}}_t \boldsymbol{D} - \frac{2\nu_c^2}{(\nu_c - 2)(\nu_c - 4)} \boldsymbol{\lambda} \boldsymbol{\lambda}' \right) \quad (54)$$

where  $\boldsymbol{D} := \text{diag}(s_1, \dots, s_N)$  is a  $N \times N$  diagonal matrix containing the standard deviations of the copula quantiles. Thus the dynamic of  $\boldsymbol{\Upsilon}_t$  comes from the dynamic of  $\dot{\boldsymbol{\Upsilon}}_t$ , which we assume is the cDCC model of Aielli (2013) :

$$\boldsymbol{Q}_t = \boldsymbol{Q}(1 - \beta_c - \alpha_c) + \beta_c \boldsymbol{Q}_{t-1} + \alpha_c \boldsymbol{\epsilon}_{t-1}^* \boldsymbol{\epsilon}_{t-1}^{*'} \quad (55)$$

$$\dot{\boldsymbol{\Upsilon}}_t = \boldsymbol{Q}_t^{*-1/2} \boldsymbol{Q}_t \boldsymbol{Q}_t^{*-1/2} \quad (56)$$

where  $\boldsymbol{Q} = \mathbb{E}(\boldsymbol{Q}_t) = \mathbb{E}(\boldsymbol{\epsilon}_t^* \boldsymbol{\epsilon}_t^{*'})$  is a positive definite correlation matrix while  $\beta_c$  and  $\alpha_c$  are non-negative scalars with  $\alpha_c + \beta_c < 1$ . Also,  $\boldsymbol{\epsilon}_t^* = \boldsymbol{Q}_t^{*1/2} \boldsymbol{\epsilon}_t$  where  $\boldsymbol{Q}_t^* = \text{diag}(dg(\boldsymbol{Q}_t))$  and  $dg(\cdot)$  is an operator that returns a vector containing the diagonal elements of a square matrix argument. The matrix  $\boldsymbol{Q}$  is obtained by targeting using the cDCC estimator proposed in Aielli (2013) (See definition 3.3).

## B.5 Simulation

The skewed  $t$  copula can be simulated with the following steps, where  $k$  represents the chosen number of simulations:

1. Simulate  $k$  vectors  $\{\tilde{\boldsymbol{x}}^i\}_{i=1}^k$  using the stochastic representation (49) and the forecast correlation matrix for the next period.
2. Compute  $\{\tilde{\boldsymbol{u}}^i\}_{i=1}^k = \left\{ \left( F_1(\tilde{x}_1^i), \dots, F_N(\tilde{x}_N^i) \right)' \right\}_{i=1}^k$ . Given a high number of simulations, instead of using numerical integration we can use the empirical CDF estimate of  $F_j$  with the simulated serie  $\{\tilde{x}_j^i\}_{i=1}^k$  to calculate  $\left\{ F_j(\tilde{x}_j^i) \right\}_{i=1}^k$  for  $j = 1, \dots, N$ .

Once the copula uniform variables  $\{\tilde{\boldsymbol{u}}^i\}_{i=1}^k$  have been simulated, we can apply Hansen's (1994) inverse distribution function to obtain the next period simulated standardized

returns  $\{z_j^i\}_{i=1}^k$  for each factor  $j$ . We then compute the simulated  $j$ th factor returns by multiplying each element of  $\{z_j^i\}_{i=1}^k$  by the  $j$ th factor forecast conditional volatility and adding to it the  $j$ th factor forecast conditional mean.

## Appendix C VaR backtests

### C.1 Kupiec's (1995) unconditional coverage test

Let  $\pi$  denote the true probability of breaching VaR for a particular model and let  $p$  denote the chosen significance level for the VaR. The null hypothesis adequate coverage  $H_0 : \pi = p$  is tested using a likelihood ratio test:

$$LR_{uc} = -2 \ln \left[ \frac{L(p)}{L(\hat{\pi})} \right] = -2 \ln \left[ \frac{(1-p)^{T_0} p^{T_1}}{(1-\hat{\pi})^{T_0} \hat{\pi}^{T_1}} \right] \quad (57)$$

where  $T$  is the number of out-of-sample observations,  $T_1$  is defined as in Section 7.1,  $T_0 = T - T_1$  and  $\hat{\pi} = \frac{T_1}{T}$ . This test statistic asymptotically follows a  $\chi^2$  distribution with one degree of freedom, corresponding to one restriction ( $\pi = p$ ).

### C.2 Christoffersen's (1998) independence test

Let  $T_{00}$  represent the number of periods with no violation followed by a period with no violation,  $T_{10}$  the number of periods with violation followed by a period with no violation,  $T_{01}$  the number of periods with no violation followed by a period with violation and  $T_{11}$  the number of periods with violation followed by a period with violation. Define  $\pi_{11}$  the probability of a violation next period conditional on a violation in the previous period and  $\pi_{01}$  the probability of a violation next period conditional on no violation in the previous period. The null hypothesis of first order independence  $H_0 : \pi_{11} = \pi_{01}$  can be tested with a likelihood ratio test :

$$LR_{ind} = -2 \ln \left[ \frac{L(\hat{\Pi})}{L(\hat{\Pi}_1)} \right] = -2 \ln \left[ \frac{(1-\hat{\pi})^{T_0} \hat{\pi}^{T_1}}{(1-\hat{\pi}_{01})^{T_{00}} \hat{\pi}_{01}^{T_{01}} (1-\hat{\pi}_{11})^{T_{10}} \hat{\pi}_{11}^{T_{11}}} \right] \quad (58)$$

where  $\hat{\pi}_{01} = \frac{T_{01}}{T_{00}+T_{01}}$  and  $\hat{\pi}_{11} = \frac{T_{11}}{T_{10}+T_{11}}$ . This test statistic asymptotically follows a  $\chi^2$  distribution with one degree of freedom, corresponding to one restriction ( $\pi_{11} = \pi_{01}$ ).

### C.3 Christoffersen's (1998) conditional coverage test

The conditional coverage null hypothesis  $H_0 : \pi_{11} = p$  and  $\pi_{01} = p$  is tested by combining the likelihood ratios of the two previous test:

$$LR_{cc} = LR_{uc} + LR_{ind} \tag{59}$$

and asymptotically follows a  $\chi^2$  distribution with two degrees of freedom, corresponding to the two restrictions  $\pi_{11} = p$  and  $\pi_{01} = p$ .

Table 1: Descriptive Statistics of Weekly Factor Returns (1963–2010)

	Market	Size	Value	Momentum
Mean (BPS)	6.98	3.48	8.28	14.52
Median (BPS)	25.97	6.00	6.00	22.96
Max (%)	12.52	6.24	9.37	11.89
Min (%)	-20.35	-9.90	-7.13	-17.17
SD (%)	2.21	1.18	1.21	1.88
Skewness	-0.75	-0.44	0.18	-1.44
Kurtosis	10.01	8.04	8.16	15.38
J-B	5311	2701	2765	16681

This table presents the descriptive statistics for the four factor returns using the sample from July 5, 1963 to december 31, 2010. The mean and median are in basis points (BPS) while the maximum, the minimum and the standard deviation (SD) are in percentages (%). The last line presents the Jarque-Bera statistic.

Table 2: Parameter Estimates for the Univariate Models (1963–2010)

Parameter Estimates	Normal Distribution	Skewed $t$ Distribution
$\phi_0$	8.06e-04 (1.11e-04)	8.16e-04 (1.09e-04)
$\phi_1$	0.132 (0.021)	0.116 (0.021)
$\phi_2$	0.067 (0.021)	0.061 (0.021)
$\phi_3$	0.026 (0.020)	0.045 (0.020)
$\beta$	0.709 (0.016)	0.703 (0.028)
$\alpha$	0.230 (0.012)	0.231 (0.021)
$\theta$	0.062 (0.041)	0.045 (0.065)
$\nu$	-	5.725 (0.570)
$\kappa$	-	-0.160 (0.028)
Log-likelihood	9226	9333

This table presents the parameters estimates of the univariate models. Standard errors in parentheses are computed with the outer product of gradients method. The parameter  $\omega$  is obtained by variance targeting.

Table 3: Parameter Estimates for the Marginals (1963–2010)

Parameter Estimates	Normal Distribution				Skewed $t$ Distribution			
	Market	Size	Value	Momentum	Market	Size	Value	Momentum
$\phi_0$	6.34e-04 (3.34e-04)	1.54e-04 (2.05e-04)	5.42e-04 (1.83e-04)	1.55e-03 (1.86e-04)	7.30e-04 (2.91e-04)	1.45e-04 (2.03e-04)	4.79e-04 (1.83e-04)	1.58e-03 (2.22e-04)
$\phi_1$	0.030 (0.020)	0.086 (0.021)	0.137 (0.019)	0.107 (0.019)	0.011 (0.021)	0.101 (0.021)	0.147 (0.020)	0.101 (0.019)
$\phi_2$	0.053 (0.021)	0.108 (0.021)	0.037 (0.022)	0.007 (0.020)	0.046 (0.021)	0.124 (0.021)	0.042 (0.021)	0.013 (0.019)
$\phi_3$	0.022 (0.019)	0.073 (0.019)	0.069 (0.021)	-0.003 (0.020)	0.024 (0.020)	0.076 (0.020)	0.076 (0.021)	0.006 (0.020)
$\beta$	0.752 (0.018)	0.849 (0.017)	0.862 (0.011)	0.832 (0.011)	0.793 (0.020)	0.849 (0.022)	0.877 (0.015)	0.831 (0.016)
$\alpha$	0.127 (0.012)	0.112 (0.012)	0.118 (0.009)	0.112 (0.008)	0.105 (0.014)	0.109 (0.015)	0.105 (0.013)	0.117 (0.012)
$\theta$	0.758 (0.091)	0.124 (0.048)	-0.057 (0.051)	-0.646 (0.056)	0.797 (0.131)	0.152 (0.086)	-0.073 (0.077)	-0.596 (0.080)
$\nu$	-	-	-	-	10.000 (1.583)	8.776 (1.307)	8.392 (1.302)	7.455 (1.078)
$\kappa$	-	-	-	-	-0.221 (0.027)	-0.060 (0.029)	0.021 (0.029)	-0.161 (0.029)
Log-likelihood	6290	7714	7860	7098	6357	7752	7891	7164

This table presents the parameter estimates for the AR(3)-NGARCH(1,1) model with Skewed  $t$  distribution fitted on each factor. Standard errors in parentheses are computed with the outer product of gradients method. The parameter  $\omega$  is fixed by variance targeting.

Table 4: Parameter Estimates for the Copulas (1963–2010)

Parameter Estimates	Constant Correlation			Dynamic Correlation		
	Normal Copula	Student Copula	Skewed $t$ Copula	Normal Copula	Student Copula	Skewed $t$ Copula
$\nu_c$		4.508 (0.228)	4.603 (0.216)		9.020 (0.735)	8.939 (0.756)
$\lambda_{\text{Market}}$			-0.023 (0.023)			-0.077 (0.052)
$\lambda_{\text{Size}}$			-0.067 (0.027)			-0.145 (0.053)
$\lambda_{\text{Value}}$			0.037 (0.023)			0.070 (0.047)
$\lambda_{\text{Momentum}}$			-0.140 (0.035)			-0.169 (0.051)
$\beta_c$				0.886 (0.004)	0.884 (0.006)	0.885 (0.006)
$\alpha_c$				0.089 (0.003)	0.094 (0.005)	0.092 (0.005)
$\rho_{\text{Market,Size}}$	-0.014	0.001	-0.003	0.013	0.039	0.035
$\rho_{\text{Market,Value}}$	-0.358	-0.352	-0.348	-0.378	-0.378	-0.371
$\rho_{\text{Market,Momentum}}$	0.107	0.115	0.108	0.082	0.089	0.087
$\rho_{\text{Size,Value}}$	-0.047	-0.047	-0.036	-0.086	-0.109	-0.103
$\rho_{\text{Size,Momentum}}$	0.013	0.022	-0.021	0.032	0.040	0.028
$\rho_{\text{Value,Momentum}}$	-0.077	-0.094	-0.077	-0.110	-0.139	-0.136
Log-likelihood	188.7	435.1	450.2	1051.0	1151.9	1161.4

This table presents the parameters estimates of each copula obtained in the second step of the estimation. Standard errors in parentheses are computed with the outer product of gradients method. The copula unconditional correlation matrix is fixed by correlation targeting (except for the normal copula).



Table 5: VaR and ES backtests (1983-2010)

	$p = 1\%$						$p = 5\%$					
	$T_1$	UC	Ind	CC	$Z_1$	$Z_2$	$T_1$	UC	Ind	CC	$Z_1$	$Z_2$
<i>Univariate Models</i>												
Normal Distribution	39	0.00	0.11	0.00	0.00	0.00	84	0.15	0.17	0.14	0.00	0.00
Skewed $t$ Distribution	26	0.01	0.09	0.00	0.94	0.03	85	0.12	0.38	0.20	0.36	0.06
<i>Multivariate Models</i>												
Normal Distribution	33	0.00	0.04	0.00	0.00	0.00	69	0.73	0.02	0.06	0.00	0.02
Normal Copula	30	0.00	0.02	0.00	0.01	0.00	69	0.73	0.06	0.16	0.00	0.06
Symmetric $t$ Copula	26	0.01	0.09	0.00	0.19	0.00	68	0.64	0.05	0.14	0.00	0.21
Skewed $t$ Copula	23	0.04	0.38	0.07	0.17	0.01	65	0.40	0.03	0.07	0.00	0.33
Dynamic Normal Distribution	34	0.00	0.01	0.00	0.00	0.00	73	0.88	0.04	0.12	0.00	0.03
Dynamic Normal Copula	27	0.00	0.01	0.00	0.10	0.00	73	0.88	0.04	0.12	0.00	0.11
Dynamic Symmetric $t$ Copula	26	0.01	0.01	0.00	0.34	0.00	74	0.79	0.05	0.13	0.02	0.16
Dynamic Skewed $t$ Copula	21	0.10	0.00	0.00	0.21	0.03	73	0.88	0.04	0.12	0.05	0.23

This table presents the VaR and ES backtest results. The first four columns are at the 1% level and the last four columns are at the 5% level. Columns  $T_1$  indicates the total number of violations in out-of-sample period. Columns UC, Ind and CC contains the p-values for the unconditional coverage, the independence and the conditional coverage VaR tests respectively. Columns  $Z_1$  and  $Z_2$  contains the p-values for the first two ES tests of Acerbi and Szekely.

Table 6: VaR and VaR/ES Average Loss (1983-2010)

	Average Loss VaR		Average Loss VaR/ES	
	$p = 1\%$	$p = 5\%$	$p = 1\%$	$p = 5\%$
<i>Univariate Models</i>				
Normal Distribution	3.001	8.467	0.058	0.111
Skewed $t$ Distribution	2.561	8.452	0.044	0.107
<i>Multivariate Models</i>				
Normal Distribution	3.422	9.205	0.072	0.125
Normal Copula	3.321	9.236	0.068	0.125
Symmetric $t$ Copula	3.196	9.249	0.063	0.124
Skewed $t$ Copula	3.150	9.253	0.061	0.124
Dynamic Normal Distribution	3.012	8.459	0.063	0.114
Dynamic Normal Copula	2.812	8.468	0.055	0.112
Dynamic Symmetric $t$ Copula	2.763	8.461	0.054	0.111
Dynamic Skewed $t$ Copula	2.748	8.474	0.053	0.111

This table presents each model's average loss for VaR ( $\bar{L}_V$ ) and the for pair VaR/ES ( $\bar{L}_{V,E}$ ) in the out-of-sample period. Columns one and three are at the 1% level while columns two and four are at the 5% level.

Table 7: DM Test for 1% VaR (1983-2010)

	Univariate Skewed $t$	Multivariate Normal	Normal Copula	Symmetric $t$ Copula	Skewed $t$ Copula	Dynamic Multivariate Normal	Dynamic Normal Copula	Dynamic Symmetric $t$ copula	Dynamic Skewed $t$ copula
Univariate Normal	0.01 $\uparrow$ ( 0.85)	0.08 ( 1.14)	0.14 ( 1.11)	0.35 ( 1.07)	0.46 ( 1.05)	0.95 ( 1.00)	0.26 ( 0.94)	0.18 ( 0.92)	0.15 ( 0.92)
Univariate Skewed $t$		0.01 $\leftarrow$ ( 1.34)	0.01 $\leftarrow$ ( 1.30)	0.01 $\leftarrow$ ( 1.25)	0.01 $\leftarrow$ ( 1.23)	0.10 ( 1.18)	0.30 ( 1.10)	0.39 ( 1.08)	0.41 ( 1.07)
Multivariate Normal			0.03 $\uparrow$ ( 0.97)	0.02 $\uparrow$ ( 0.93)	0.02 $\uparrow$ ( 0.92)	0.03 $\uparrow$ ( 0.88)	0.00 $\uparrow$ ( 0.82)	0.00 $\uparrow$ ( 0.81)	0.00 $\uparrow$ ( 0.80)
Normal Copula				0.03 $\uparrow$ ( 0.96)	0.03 $\uparrow$ ( 0.95)	0.08 ( 0.91)	0.00 $\uparrow$ ( 0.85)	0.00 $\uparrow$ ( 0.83)	0.00 $\uparrow$ ( 0.83)
Symmetric $t$ Copula					0.04 $\uparrow$ ( 0.99)	0.29 ( 0.94)	0.02 $\uparrow$ ( 0.88)	0.01 $\uparrow$ ( 0.86)	0.00 $\uparrow$ ( 0.86)
Skewed $t$ Copula						0.43 ( 0.96)	0.03 $\uparrow$ ( 0.89)	0.01 $\uparrow$ ( 0.88)	0.01 $\uparrow$ ( 0.87)
Dynamic Multivariate Normal							0.01 $\uparrow$ ( 0.93)	0.01 $\uparrow$ ( 0.92)	0.02 $\uparrow$ ( 0.91)
Dynamic Normal Copula								0.05 ( 0.98)	0.14 ( 0.98)
Dynamic Symmetric $t$ copula									0.49 ( 0.99)

This table presents the p-values of the DM test comparing each row model to a column model for the 1% VaR. The loss function used for pairwise comparisons is  $L_V$ . Numbers in parentheses indicate the ratio of the column model average loss to the row model average loss in the out-of-sample period. An up (left) arrow indicates that we reject the null hypothesis of equal predictive ability at the 5% significance level and that the column (row) model outperforms the corresponding row (column) model.

Table 8: DM Test for 5% VaR (1983-2010)

	Univariate Skewed $t$	Multivariate Normal	Normal Copula	Symmetric $t$ Copula	Skewed $t$ Copula	Dynamic Multivariate Normal	Dynamic Normal Copula	Dynamic Symmetric $t$ copula	Dynamic Skewed $t$ copula
Univariate Normal	0.67 ( 1.00)	0.01← ( 1.09)	0.00← ( 1.09)	0.00← ( 1.09)	0.00← ( 1.09)	0.96 ( 1.00)	1.00 ( 1.00)	0.97 ( 1.00)	0.97 ( 1.00)
Univariate Skewed $t$		0.01← ( 1.09)	0.00← ( 1.09)	0.00← ( 1.09)	0.00← ( 1.09)	0.97 ( 1.00)	0.93 ( 1.00)	0.96 ( 1.00)	0.90 ( 1.00)
Multivariate Normal			0.01← ( 1.00)	0.00← ( 1.00)	0.00← ( 1.01)	0.00↑ ( 0.92)	0.00↑ ( 0.92)	0.00↑ ( 0.92)	0.00↑ ( 0.92)
Normal Copula				0.10 ( 1.00)	0.19 ( 1.00)	0.00↑ ( 0.92)	0.00↑ ( 0.92)	0.00↑ ( 0.92)	0.00↑ ( 0.92)
Symmetric $t$ Copula					0.73 ( 1.00)	0.00↑ ( 0.91)	0.00↑ ( 0.92)	0.00↑ ( 0.91)	0.00↑ ( 0.92)
Skewed $t$ Copula						0.00↑ ( 0.91)	0.00↑ ( 0.92)	0.00↑ ( 0.91)	0.00↑ ( 0.92)
Dynamic Multivariate Normal							0.69 ( 1.00)	0.95 ( 1.00)	0.50 ( 1.00)
Dynamic Normal Copula								0.48 ( 1.00)	0.70 ( 1.00)
Dynamic Symmetric $t$ copula									0.30 ( 1.00)

This table presents the p-values of the DM test comparing each row model to a column model for the 5% VaR. The loss function used for pairwise comparisons is  $L_V$ . Numbers in parentheses indicate the ratio of the column model average loss to the row model average loss in the out-of-sample period. An up (left) arrow indicates that we reject the null hypothesis of equal predictive ability at the 5% significance level and that the column (row) model outperforms the corresponding row (column) model.

Table 9: DM Test for 1% VaR and ES (1983-2010)

	Univariate Skewed $t$	Multivariate Normal	Normal Copula	Symmetric $t$ Copula	Skewed $t$ Copula	Dynamic Multivariate Normal	Dynamic Normal Copula	Dynamic Symmetric $t$ copula	Dynamic Skewed $t$ copula
Univariate Normal	0.04 $\uparrow$ ( 0.76)	0.06 ( 1.23)	0.13 ( 1.16)	0.43 ( 1.08)	0.59 ( 1.05)	0.40 ( 1.09)	0.51 ( 0.94)	0.37 ( 0.92)	0.30 ( 0.91)
Univariate Skewed $t$		0.02 $\leftarrow$ ( 1.62)	0.03 $\leftarrow$ ( 1.53)	0.03 $\leftarrow$ ( 1.42)	0.03 $\leftarrow$ ( 1.39)	0.11 ( 1.43)	0.24 ( 1.25)	0.30 ( 1.21)	0.30 ( 1.20)
Multivariate Normal			0.05 $\uparrow$ ( 0.94)	0.05 ( 0.87)	0.06 ( 0.85)	0.07 ( 0.88)	0.00 $\uparrow$ ( 0.77)	0.01 $\uparrow$ ( 0.75)	0.01 $\uparrow$ ( 0.74)
Normal Copula				0.06 ( 0.93)	0.07 ( 0.91)	0.40 ( 0.94)	0.01 $\uparrow$ ( 0.82)	0.01 $\uparrow$ ( 0.79)	0.01 $\uparrow$ ( 0.78)
Symmetric $t$ Copula					0.12 ( 0.98)	0.93 ( 1.01)	0.06 ( 0.88)	0.03 $\uparrow$ ( 0.85)	0.01 $\uparrow$ ( 0.84)
Skewed $t$ Copula						0.76 ( 1.03)	0.15 ( 0.90)	0.06 ( 0.87)	0.03 $\uparrow$ ( 0.86)
Dynamic Multivariate Normal							0.02 $\uparrow$ ( 0.87)	0.03 $\uparrow$ ( 0.85)	0.04 $\uparrow$ ( 0.83)
Dynamic Normal Copula								0.12 ( 0.97)	0.13 ( 0.96)
Dynamic Symmetric $t$ copula									0.34 ( 0.99)

This table presents the p-values of the DM test comparing each row model to a column model for the 1% pair VaR/ES. The loss function used for pairwise comparisons is  $L_{V,E}$ . Numbers in parentheses indicate the ratio of the column model average loss to the row model average loss in the out-of-sample period. An up (left) arrow indicates that we reject the null hypothesis of equal predictive ability at the 5% significance level and that the column (row) model outperforms the corresponding row (column) model.

Table 10: DM Test for 5% VaR and ES (1983-2010)

	Univariate Skewed $t$	Multivariate Normal	Normal Copula	Symmetric $t$ Copula	Skewed $t$ Copula	Dynamic Multivariate Normal	Dynamic Normal Copula	Dynamic Symmetric $t$ copula	Dynamic Skewed $t$ copula
Univariate Normal	0.08 ( 0.96)	0.03← ( 1.13)	0.02← ( 1.13)	0.03← ( 1.12)	0.03← ( 1.11)	0.56 ( 1.03)	0.87 ( 1.01)	0.95 ( 1.00)	0.98 ( 1.00)
Univariate Skewed $t$		0.02← ( 1.17)	0.01← ( 1.17)	0.01← ( 1.16)	0.01← ( 1.16)	0.29 ( 1.07)	0.42 ( 1.05)	0.46 ( 1.04)	0.47 ( 1.04)
Multivariate Normal			0.69 ( 1.00)	0.39 ( 0.99)	0.39 ( 0.99)	0.02↑ ( 0.91)	0.01↑ ( 0.89)	0.01↑ ( 0.89)	0.00↑ ( 0.89)
Normal Copula				0.25 ( 0.99)	0.29 ( 0.99)	0.02↑ ( 0.91)	0.01↑ ( 0.89)	0.01↑ ( 0.89)	0.00↑ ( 0.89)
Symmetric $t$ Copula					0.41 ( 1.00)	0.04↑ ( 0.92)	0.01↑ ( 0.90)	0.01↑ ( 0.90)	0.01↑ ( 0.90)
Skewed $t$ Copula						0.05 ( 0.92)	0.02↑ ( 0.90)	0.01↑ ( 0.90)	0.01↑ ( 0.90)
Dynamic Multivariate Normal							0.10 ( 0.98)	0.11 ( 0.98)	0.15 ( 0.98)
Dynamic Normal Copula								0.20 ( 1.00)	0.34 ( 0.99)
Dynamic Symmetric $t$ copula									0.64 ( 1.00)

This table presents the p-values of the DM test comparing each row model to a column model for the 5% pair VaR/ES. The loss function used for pairwise comparisons is  $L_{V,E}$ . Numbers in parentheses indicate the ratio of the column model average loss to the row model average loss in the out-of-sample period. An up (left) arrow indicates that we reject the null hypothesis of equal predictive ability at the 5% significance level and that the column (row) model outperforms the corresponding row (column) model.

Table 11: MCS Results (1983-2010)

		MCS	
		$p = 1\%$	$p = 5\%$
VaR		Univariate Skewed $t$ Dynamic Normal Copula Dynamic Student Copula Dynamic Skewed $t$ Copula Dynamic Skewed $t$ Copula Dynamic Skewed $t$ Copula	Univariate Normal Univariate Skewed $t$ Dynamic Multivariate Normal Dynamic Normal Copula Dynamic Student Copula Dynamic Skewed $t$ Copula
VaR/ES	All models		Univariate Normal Univariate Skewed $t$ Dynamic Multivariate Normal Dynamic Normal Copula Dynamic Student Copula Dynamic Skewed $t$ Copula

This table presents the Model Confidence Set obtained using the procedure of Hansen et al. (2011) at a 95% level of confidence. The first row corresponds to VaR model comparisons with the loss function  $L_V$  and the second corresponds to VaR/ES model comparisons with the loss function  $L_{V,E}$ . The first column is at the 1% level while the second column is at the 5% level.