ARTICLE IN PRESS
International Journal of Forecasting xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting expected shortfall: Should we use a multivariate model for stock market factors?☆

Alain-Philippe Fortin[a], Jean-Guy Simonato[b,*], Georges Dionne[b,c]

[a] Swiss Finance Institute, Geneva, Switzerland
[b] HEC Montréal, Department of Finance, Canada
[c] Canada Research Chair in Risk Management, Canada

ARTICLE INFO

Keywords:
Fama–French and momentum factors
Value at risk
Expected shortfall
Conditional value at risk
Elicitability
Model comparison
Backtesting
Comparative predictive accuracy
Model confidence set

ABSTRACT

Is univariate or multivariate modeling more effective when forecasting the market risk of stock portfolios? We examine this question in the context of forecasting the one-week-ahead expected shortfall of a stock portfolio based on its exposure to 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 result suggests that univariate models, which are more parsimonious and simpler to implement than multivariate factor-based models, can be used to forecast the downside risk of equity portfolios without losses in precision.

1. Introduction

In finance, there is vast literature looking at the set of factors that are capable of explaining the cross-section of expected stock returns. An important result of this literature is Fama and French (1993). They find that a linear set of three factors—commonly referred to as market, size, and value—has the potential to explain most of the cross-sectional variation in stock returns. In addition to these three factors, Jegadeesh and Titman (1993) and Carhart (1997) established the relevance of a momentum factor to explain the cross-section of stock returns.

Because these factors represent the exposure of the investors to priced risks, their use by academics and practitioners is widespread, as they capture the principal risks of an investment in stocks. For example, in portfolio management, these factors can simplify an allocation process involving a large number of stocks. By estimating the stocks' sensitivity to each factor, it is possible to express the covariance matrix of the portfolio as a function of the factors and to assess its risk and expected return.

Given the importance and widespread use of these factors, a clear understanding of their properties is an important issue, especially for risk measurement and portfolio management. In Christoffersen and Langlois (2013), an empirical investigation found that the Fama and French (1993) factors and Carhart (1997) momentum factor (hereafter, FFC factors) contain key non-linear dependencies requiring multivariate models to capture their joint dynamics. In such a context, characterizing the risk of the portfolios exposed to these factors requires a multivariate approach. However, when forecasting commonly used risk quantities of portfolios such as the value at risk (VaR) or expected shortfall (ES), in many cases, univariate methods can be exploited to simplify the computations.

In this paper, as a first contribution, we examine whether a multivariate approach is better than a more straightforward univariate approach when forecasting the joint risk of a portfolio of stocks. More precisely, we consider the weekly returns of two portfolios. The first

☆ We thank Hugues Langlois for providing us with the original data used in Christoffersen and Langlois (2013) and for sharing some MATLAB codes. Dionne and Simonato acknowledge financial support from the Social Sciences and Humanities Research Council of Canada (SSHRC).
* Corresponding author.
E-mail addresses: alainphilippe.fortin@sfi.ch (A.-P. Fortin), jean-guy.simonato@hec.ca (J.-G. Simonato), georges.dionne@hec.ca (G. Dionne).

https://doi.org/10.1016/j.ijforecast.2021.11.010
0169-2070/

Please cite this article as: A.-P. Fortin, J.-G. Simonato and G. Dionne, Forecasting expected shortfall: Should we use a multivariate model for stock market factors?. International Journal of Forecasting (2022), https://doi.org/10.1016/j.ijforecast.2021.11.010.

portfolio is directly invested in the four FFC factors, as in Christoffersen and Langlois (2013). The second portfolio is invested in ten large capitalization stocks. For this stock portfolio, we project each individual stock return on the FFC factors, thereby reducing the dimension of the portfolio from ten to four. For both portfolios, we then examine whether forecasting portfolio risk using a naive univariate dynamic model provides forecasts that are equivalent to those of more sophisticated four-dimensional multivariate dynamic copula models for the FFC factors. Given the easier implementation and much smaller computing efforts required for univariate models, many researchers recommend this direct approach to risk measurement (Berkowitz & O'Brien, 2002; Christoffersen, 2009). However, a case can also be made for multivariate models, which integrate more information by characterizing the dependencies between portfolio stock returns, via their dependence on the factors . 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 that might lead to forecasting errors, and restrictions must be imposed to keep the model tractable.

As a second contribution, we examine the above issue with a focus on the ES risk measure, also called conditional value at risk (CVaR). With the recent shift from VaR to ES sanctioned by Basel Committee on Banking Supervision (BCBS) (2016, 2019) for market risk, there is growing literature on backtesting ES, as this risk measure is becoming a serious alternative to VaR; see, for example, Deng and Qiu (2021). The change is motivated by the fact that VaR does not adequately capture the conditional expected losses and lacks subadditivity, whereas ES avoids these limitations. We focus on forecasting ES at the 1% and 2.5% probability levels, the latter being the requirement under Basel III for measuring market risk. 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. This lack of elicitability for ES limits the relative comparison of univariate and multivariate models. However, as shown in Fissler and Ziegel (2016), the VaR/ES pair is jointly elicitable with respect to a class of loss functions. We use this result and the loss function proposed in Patton, Ziegel, and Chen (2019) to compare our models based on their ability to forecast VaR and ES jointly.

Our main finding is that, in most cases, there are no significant differences between the risk forecasting accuracy of univariate and multivariate factor-based models for both VaR and the VaR/ES pair.

The univariate vs. multivariate issue tackled in this paper has been studied many times in the context of forecasting portfolio VaR (Nieto & Ruiz, 2016). 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 Pers (2003) compare 12 univariate volatility models and the diagonal VEC model of Bollerslev, Engle, and Wooldridge (1988) for forecasting the VaR of a portfolio of UK assets at one-, five-, ten-, 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 12 univariate and 16 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.

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. To address this limitation, Santos, Nogales, and Ruiz (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.

Diks and Fang (2020) also compare the performance of multivariate and univariate approaches to forecast VaR. Focusing on skew-elliptical distributions with an application using daily returns, they find that better multivariate forecasts do not necessarily correspond to better aggregate portfolio return forecasts. In their study, they rely on multivariate models that are closed under affine transformations. They are thus comparing models which produce the same univariate and multivariate portfolio return distributions when comparing the performance of the forecasts. In our study, our univariate models are not necessarily available as a linear combination of the multivariate models. Although such theoretical considerations are important, we choose to favor a practitioner's point of view whose goal is to find the best risk forecasting model, and thereby not restrain ourselves to a particular class of distributions.

Finally, Kole, Markwat, Opschoor, and van Dijk (2017) examine the impact of different levels of temporal and portfolio aggregation on forecasting the ten-day VaR for a diversified portfolio of eight indexes related to stocks, bonds, and alternative investments. Also relying on the asymmetric tick loss function, 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.

As this short survey demonstrates, the literature to date offers mixed evidence, and, to the best of our knowledge, has not yet examined the ES risk measure. Our contribution to this literature, besides a close examination of a significant set of factors for financial managers, is thus to examine in more detail the multivariate vs. univariate issue in terms of the ES risk measure. Furthermore, unlike papers that have examined the issue with symmetric distributions, we rely on the asymmetric skewed t distribution in our univariate and multivariate models in order to better capture the skewness of return distributions, just like Diks and Fang (2020).

Our VaR results are in direct contrast with those of Santos et al. (2013), but partly corroborate the results of Kole et al. (2017) and Diks and Fang (2020), which are the three 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. 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. Indeed, the main conclusion of Kole et al. (2017) is that aggregation of daily returns into weekly or biweekly returns leads to the loss of details in return dynamics. Despite this potential loss, our use of weekly data (instead of daily data) is interesting in a VaR and ES computation framework, since it is close to the ten-day horizon adopted by regulators. The weekly horizon is also more relevant in our case, since we rely on the results of Christoffersen and Langlois (2013), who also use weekly data to identify the best time-series model for the FFC factors.

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. Section 8 is devoted to our tests for comparative predictive accuracy, and Section 9 discusses the model confidence set (MCS) approach. Section 10 concludes the paper.

2. Framework

We use the superscript s to denote variables related to individual stock returns, as opposed to the factors. Let $\boldsymbol{r}^s_t = \left[r^s_{1,t}, \ldots, r^s_{N^s,t}\right]'$ and $\boldsymbol{r}_t = \left[r_{1,t}, \ldots, r_{N,t}\right]'$ respectively denote the random vector of the N^s stock and N factor log returns at time t. Let $\boldsymbol{r}^s_{h,t}$ and $\boldsymbol{r}_{h,t}$ denote the vectors of holding period returns for the stocks and factors defined by $\boldsymbol{r}^s_{h,t} = \exp(\boldsymbol{r}^s_t) - 1$ and $\boldsymbol{r}_{h,t} = \exp(\boldsymbol{r}_t) - 1$, respectively. The random portfolio log return is $r_{w,t} = \ln\left(\boldsymbol{w}'_{t-1}\boldsymbol{r}^s_{h,t} + 1\right)$, where $\boldsymbol{w}_{t-1}$ is the vector of portfolio weights at time $t-1$.

We assume that all time series are stationary. Let $\boldsymbol{F}^s_t$, $\boldsymbol{F}_t$, and $F_{w,t}$ respectively denote the distribution function of $\boldsymbol{r}^s_t$, $\boldsymbol{r}_t$, and $r_{w,t}$, conditional on the information set available at time $t-1$. Also, let $F^s_{i,t}$ and $F_{j,t}$ correspond respectively to the ith and jth marginal distribution function of $\boldsymbol{F}^s_t$ and $\boldsymbol{F}_t$, for $i = 1, \ldots, N^s$ and $j = 1, \ldots, N$. In this paper, we assume that all distribution functions are continuous, have densities, and are strictly increasing. In particular, this implies that the univariate inverse distribution function (quantile function) $F^{-1}(\cdot)$ is well defined.

Given a probability level $p = 1\%$ or 2.5% at time $t-1$, we are interested in forecasting the next-period VaR and ES, defined by

$$VaR^p_t = F^{-1}_{w,t}(p)$$

and

$$ES^p_t = \mathbb{E}_{t-1}\left[r_{w,t} \mid r_{w,t} < VaR^p_t\right].$$

In order to estimate these risk measures, we need a statistical model for $F_{\boldsymbol{w},t}$.

For the univariate approach, this is done by directly making assumptions on $F_{w,t}$. Details about these assumptions and how we compute the parameter estimates, and the VaR and ES, are given in Sections 4–6.

For the multivariate approach, when dealing with portfolios directly invested in the FFC factors, we use multivariate copula models that capture the non-linearities and non-normalities documented in the literature for these data series. We rely on simulation methods to derive $F_{w,t}$ from $\boldsymbol{F}_t$, and then compute the estimated VaR and ES. Details about the copula models and how we compute the parameter estimates and risk measures are given in Sections 3, 5 and 6.

For the multivariate approach, when dealing with portfolios of stocks, we take a practitioner's perspective and use the factors that capture the main risks in stock returns. Specifically, we take an indirect path by first projecting each individual stock holding period return $r^s_{h,i,t}$ on the vector of holding period factor returns $\boldsymbol{r}_{h,t}$ and a constant,

$$r^s_{h,i,t} - r_{f,t} = a_i + \boldsymbol{b}'_i \boldsymbol{r}_{h,t} + \varepsilon_{i,t} \quad \text{for } i = 1, \ldots, N^s, \tag{1}$$

where $r_{f,t}$ is the weekly risk-free rate; the intercept, a_i, and the vector of loadings, $\boldsymbol{b}_i$, are assumed to be constant over time. Moreover, we assume conditional independence between the vector of error terms, $\boldsymbol{\varepsilon}_t = [\varepsilon_{1,t}, \ldots, \varepsilon_{N^s,t}]$, and $\boldsymbol{r}_{h,t}$. In the second step, we proceed by modeling the joint conditional distribution of $\boldsymbol{r}_{h,t}$ and $\boldsymbol{\varepsilon}_t$. The previous equation shows that $\boldsymbol{r}^s_{h,t}$ is a function of $\boldsymbol{r}_{h,t}$ and $\boldsymbol{\varepsilon}_t$. In turn, $r_{w,t}$ is a function of the vector of stock returns $\boldsymbol{r}^s_{h,t}$ and the weights $\boldsymbol{w}_{t-1}$, which are part of the information set at time $t-1$. In this paper, we rely on simulation methods to derive $F_{w,t}$ from the distributions of the factor returns and error terms, and then compute the estimated VaR and ES. The distribution of the factor return is obtained with similar approaches to those just described above for the case of factor portfolios. Details on the procedures are given in Sections 3, 5 and 6.

3. Multivariate models for factor returns

In this section, we provide details about the models we use to characterize the joint dynamics of the FFC factor returns. For the portfolio invested in FFC factors, these models are used directly to forecast the VaR and ES risk measures. For the portfolio invested in stocks, these models are used indirectly in the context of Eq. (1).

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's marginal distribution is highly non-normal and that the dependencies between each pair of factors are non-linear. This suggests multivariate non-normality 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 in that they can model the marginal distributions separately via Sklar's (1959) theorem. This theorem allows for 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:

$$\boldsymbol{F}_t(\boldsymbol{r}_t) = \boldsymbol{C}_t(F_{1,t}(r_{1,t}), \ldots, F_{N,t}(r_{N,t})). \tag{2}$$

Given our assumptions in Section 2, the copula $\boldsymbol{C}_t$ is uniquely determined. Each of the marginals $F_{j,t}$ contains all the univariate information on the jth factor, while the copula $\boldsymbol{C}_t$ contains all the dependence information between the factors. This decomposition shows that a model for $\boldsymbol{F}_t$ can be built in two steps. First, we choose a model for each of the marginals $F_{j,t}$ for $j = 1, \ldots, N$, and second, we choose a copula $\boldsymbol{C}_t$ to link them. We now address these two steps in turn.

As in Christoffersen and Langlois, each of the four factors of the next-period conditional marginal distribution is fitted with the following AR(3) model:

$$r_{j,t} = \mu_{j,0} + \mu_{j,1} r_{j,t-1} + \mu_{j,2} r_{j,t-2} + \mu_{j,3} r_{j,t-3} + \sigma_{j,t} z_{j,t} \quad \text{with } z_{j,t} \overset{i.i.d.}{\sim} F_j(0,1)$$

where $\mu_{j,0}$ to $\mu_{j,3}$ are the AR(3) parameters for the returns of factor j, σ_{jt} is the conditional standard deviation, and $z_{j,t}$ is an error term. The error term $z_{j,t}$ is an independent and identically distributed random variable, following the unconditional standardized distribution $F_j(0, 1)$. This model allows for constant first moments with dynamic second moments as well as non-normal distributions. For each factor, the conditional variance is fitted using the NGARCH(1,1) model of Engle and Ng (1993):

$$\sigma_{j,t}^2 = \omega_j + \beta_j \sigma_{j,t-1}^2 + \alpha_j (z_{j,t-1} - \theta_j)^2$$

where $\omega_j > 0$, $\beta_j, \alpha_j \geq 0$, and $\alpha_j(1 + \theta_j^2) + \beta_j < 1$ for $j = 1, \ldots, N$. A positive leverage parameter ($\theta_j > 0$) implies that negative shocks ($z_{j,t-1} < 0$) 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, κ_j and υ_j, which determine the skewness and kurtosis, and its density is given in Appendix A. Marginal skewness is an important improvement over 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 $\boldsymbol{C}^{st}_{\nu_c, \boldsymbol{\Upsilon}_t, \boldsymbol{\lambda}}$, is characterized by an $N \times 1$ vector of asymmetry parameters $\boldsymbol{\lambda}$, a scalar degree-of-freedom parameter ν_c, and a copula correlation matrix $\boldsymbol{\Upsilon}_t$. Its density is given in Appendix B.

We allow the copula correlation matrix $\boldsymbol{\Upsilon}_t$ to evolve through time. Specifically, we assume that the correlation matrix of the copula quantiles (which are defined in the appendix), denoted $\dot{\boldsymbol{\Upsilon}}_t$, follows the cDCC model of Aielli (2013):

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

$$\dot{\boldsymbol{\Upsilon}}_t = \boldsymbol{Q}_t^{*-\frac{1}{2}} \boldsymbol{Q}_t \boldsymbol{Q}_t^{*-\frac{1}{2}}$$

where $\boldsymbol{Q}$ (without a time index) indicates the unconditional correlation matrix of the random shocks $\boldsymbol{\epsilon}^*_t$ i.e. $\boldsymbol{Q} = E(\boldsymbol{Q}_t) = E(\boldsymbol{\epsilon}^*_t \boldsymbol{\epsilon}^{*'}_t)$, while β_c and α_c are non-negative scalars with $\alpha_c + \beta_c < 1$. Also, $\boldsymbol{\epsilon}^*_t = \boldsymbol{Q}_t^{*\frac{1}{2}} \boldsymbol{\epsilon}_t$, where $\boldsymbol{Q}^*_t = diag(dg(\boldsymbol{Q}_t))$, and $\boldsymbol{\epsilon}_t$ is an $N \times 1$ vector containing the standardized copula quantiles. Here, $dg(\cdot)$ is an operator that takes a square matrix as input and returns a vector containing the diagonal elements of the square matrix. Likewise, $diag(\cdot)$ is an operator that takes a vector as input and returns a square diagonal matrix as output.

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 $\boldsymbol{Q}$. 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 for a total of eight multivariate models.

4. Univariate models

For the univariate models, we adopt a specification similar to the one used for the marginals of the multivariate models. We assume that the portfolio return dynamics is as follows:[1]

$$r_{w,t} = \mu + \sigma_t z_t \quad \text{with} \quad z_t \overset{i.i.d.}{\sim} F(0,1)$$

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

with the same parameter restrictions as before, and $F(0, 1)$ corresponding to Hansen's (1994) univariate skewed t distribution with parameters κ and υ. As a reference, we also consider the univariate standard normal distribution for $F(0, 1)$ and its filtered historical simulation (FHS) version, the latter meaning that we simulate innovation

[1] We initially considered an AR(3) for the conditional mean, like in the multivariate models, but we found better backtesting results using a constant mean.

terms by drawing randomly into the estimated residuals $\hat{z}_t$ rather than from the standard normal distribution. This leads to a total of three univariate models.

It is important to note that our univariate models are not necessarily obtainable as a linear combination of the multivariate models in Section 3. For example, Diks and Fang (2020) consider elliptical distributions closed under affine transformations, so that their multivariate models can produce the same univariate distributions as their univariate models. Although such theoretical considerations are important, we choose to favor the point of view of a practitioner whose goal is to find the best risk forecasting model, and thereby we do not restrain ourselves to a particular class of models.

5. Data sets and parameter estimates

In this section, we first describe the data used in our study. We then explain the parameter estimation procedure for the univariate and multivariate approaches. Finally, we provide some estimation results showing the fit of these modeling approaches when applied to our full data sample.

5.1. Data sets

We study the weekly returns of the four FFC equity factors and ten large capitalization companies from July 5, 1963 to December 31, 2019. This corresponds to $T = 2,948$ observations. The factor returns are from Kenneth French's data library, where details on how the FFC factors are constructed can be found. The weekly returns are directly available in the data library for the market, size, and value factors. For the momentum factor, we aggregate daily returns into weekly log returns from July 5, 1963 to December 31, 2019. Our sample of ten large capitalization stocks consists of daily stock returns for the same period taken from CRSP. The daily returns are aggregated into weekly log returns. Table 1 lists the names of the companies entering our portfolio of stocks. In what follows, the term "return" refers to a weekly log return, unless otherwise indicated. 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 and stock log 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 indicated by kurtosis values much higher than three. The non-normality of each factor's marginal distribution is confirmed by the very large Jarque–Bera statistics in the last column, rejecting the null hypothesis of a normal distribution in all cases. Similar remarks also hold for the stock returns.

5.2. Parameter estimation for the univariate models

We estimate each univariate model by maximum likelihood (ML). Let $\boldsymbol{\theta}$ denote the vector containing the parameters. Given the sample of observations for the returns of individual elements of our portfolio (factors or stocks), we construct a pseudo-sample of observations for a given set of portfolio weights. Specifically, taking the portfolio of stocks as an example, we compute the implied pseudo-sample of portfolio returns $r_{w,t} = \ln\left(\boldsymbol{w}_{t-1}'\boldsymbol{r}_{h,t}^{s} + 1\right)$ for $t = 1, \ldots, T$. We then estimate $\boldsymbol{\theta}$ by maximizing the conditional log-likelihood $\ln L(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln f_t(r_{w,t})$, where f_t is the conditional probability density function of $r_{w,t}$. To give an idea of the fit of this approach, Table 2 presents the ML estimates for the normal and skewed t distribution for our whole sample from 1963 to 2019, for the case of an equally weighted portfolio. The left panel presents results for the portfolio of factors, while the right panel is for the portfolio of ten stocks. For both portfolios, the skewed t distribution improves the model's fit, with much higher likelihood values obtained with only two additional parameters that are both significant at the 1% level.

5.3. Parameter estimation for the copula models

For the copula models of the FFC factors, we use a two-step estimation procedure. Differentiating both sides of Eq. (2), we get the conditional likelihood of $\boldsymbol{r}_t$:

$$\boldsymbol{f}_t(\boldsymbol{r}_t) = \boldsymbol{c}_t(F_{1,t}(r_{1,t}), \ldots, F_{N,t}(r_{N,t})) \prod_{j=1}^{N} f_{j,t}(r_{j,t}),$$

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

$$\ln \boldsymbol{f}_t(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T) = \sum_{t=1}^{T} \ln \boldsymbol{c}_t(F_{1,t}(r_{1,t}), \ldots, F_{N,t}(r_{N,t})) + \sum_{j=1}^{N} \sum_{t=1}^{T} \ln f_{j,t}(r_{j,t}).$$

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 $F_{j,t}$ by maximizing $\ln L_j(\boldsymbol{\theta}_j) = \sum_{t=1}^{T} \ln f_{j,t}(r_{j,t})$ for $j = 1, \ldots, N$. Second, using the estimated marginals $\hat{F}_{1,t}, \ldots, \hat{F}_{N,t}$, we construct a pseudo-sample of observations for the copula,[2]

$$\hat{\boldsymbol{u}}_t = (\hat{F}_{1,t}(r_{1,t}), \ldots, \hat{F}_{N,t}(r_{N,t}))$$

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

$$\ln L_c(\boldsymbol{\theta}_c) = \sum_{t=1}^{T} \ln \boldsymbol{c}_t(\hat{\boldsymbol{u}}_t). \tag{3}$$

To give an idea of the fit of the copula models, Table 3 presents the parameter estimates for the marginals under the normal and skewed t distribution for our whole data

[2] Specifically, because $F_{j,t}(r_{j,t}) = F_j(z_{j,t})$, we use the residuals $\hat{z}_{j,t}$ from the first step to obtain the empirical CDF estimate $\hat{F}_j(x) = \frac{1}{T+1}\sum_{t=1}^{T} \mathbf{1}_{\{\hat{z}_{j,t} \le x\}}$ and let $\hat{u}_{j,t} = \hat{F}_j(\hat{z}_{j,t})$.

Table 1
Descriptive statistics of weekly stock and factor returns (1963–2019).

	Mean	Median	Max.	Min.	SD	Skew.	Kurt.	J–B
Factors								
Market	11.82	28.00	13.46	−18.00	2.16	−0.46	8.25	3484
Size	3.32	4.50	6.99	−10.72	1.21	−0.32	8.62	3930
Value	7.08	3.00	9.94	−8.71	1.22	0.47	10.09	6280
Momentum	15.05	21.95	12.66	−16.00	1.81	−1.06	13.06	12995
Stocks								
Coca-Cola	28.53	23.14	25.95	−27.90	3.18	−0.12	8.36	3536
Exxon	23.98	20.45	17.26	−19.99	2.75	−0.13	5.16	583
GE	21.14	12.93	36.26	−20.46	3.61	0.48	10.41	6861
IBM	20.04	13.93	19.37	−17.64	3.31	0.08	6.16	1227
Chevron	26.02	20.02	19.44	−27.15	3.39	0.06	6.01	1117
UTC	30.40	30.23	19.03	−36.18	3.78	−0.24	7.64	2679
Procter	25.29	14.78	14.77	−39.36	2.90	−0.76	16.68	23284
Caterpillar	28.53	22.60	30.90	−25.48	4.20	0.22	5.86	1030
Boeing	36.57	34.81	22.50	−30.74	4.48	0.07	5.59	824
Merck	29.70	18.59	19.04	−24.47	3.53	−0.08	5.93	1057

This table presents descriptive statistics for the four factor returns using a sample from July 5, 1963 to December 27, 2019. The mean and median are in basis points, and the maximum, minimum, and standard deviation (SD) are in percentages (%). The last column presents the Jarque–Bera satistic.

Table 2
Parameter estimates for the univariate models (1963–2019).

Parameter	Port. of factors		Port. of stocks	
	Normal distribution	Skewed t distribution	Normal distribution	Skewed t distribution
μ	1.05e−03	1.01e−03	2.31e−03	2.29e−03
	(1.01e−04)	(9.96e−05)	(3.22e−04)	(3.06e−04)
β	0.709	0.733	0.771	0.779
	(0.015)	(0.025)	(0.017)	(0.021)
α	0.219	0.202	0.109	0.104
	(0.011)	(0.018)	(0.011)	(0.013)
θ	0.076	−0.038	0.763	0.775
	(0.038)	(0.057)	(0.097)	(0.117)
ν	–	6.352	–	13.355
		(0.632)		(3.015)
κ	–	−0.183	–	−0.136
		(0.025)		(0.026)
Log-likelihood	10861	10973	7273	7304

This table presents the parameter estimates of the univariate models for the case of an equally weighted portfolio. Standard errors in parentheses are computed with the outer product of gradients method. The parameter ω, the constant in the NGARCH variance dynamics, is obtained by variance targeting.

sample from 1963 to 2019. The left panel presents the parameter estimates of the marginals under the normal distribution, and the second panel shows the estimates for the skewed t. Again we see that the skewed t improves the fit of the NGARCH model with a much higher likelihood than the normal case, and with only two additional parameters.

The parameter estimates for the copulas, obtained from the second step, are given in Table 4. The left panel of the table presents the models with constant correlations, while the right panel shows the estimates for the models with dynamic correlations. We obtain parameter estimates very similar to those reported in Christoffersen and Langlois (2013).[3] Consistent with their results, the model which provides the best fit is the skewed t copula with dynamic correlations.

5.4. Parameter estimation for the stock return model

For the stock return model given by Eq. (1), we obtain the parameter estimates for the intercept $\hat{a}_i$ and factor loadings $\hat{\boldsymbol{b}}_i$ of each stock by running a standard time-series OLS regression. Table 5 gives an idea of the fit of this

[3] To verify the validity of our implementation using their original data set, we verified that our estimated parameters replicate their estimates up to the third digit.

Table 3
Parameter estimates for the marginals (1963–2019).

Parameter	Normal distribution				Skewed t distribution			
	Market	Size	Value	Mom.	Market	Size	Value	Mom.
μ_0	8.93e−04 (3.09e−04)	1.52e−04 (1.95e−04)	4.22e−04 (1.71e−04)	1.50e−03 (1.73e−04)	9.71e−04 (2.69e−04)	1.39e−04 (1.92e−04)	3.51e−04 (1.73e−04)	1.55e−03 (2.08e−04)
μ_1	0.012 (0.019)	0.076 (0.019)	0.111 (0.018)	0.091 (0.018)	−0.002 (0.019)	0.095 (0.019)	0.125 (0.019)	0.083 (0.018)
μ_2	0.053 (0.018)	0.092 (0.019)	0.029 (0.019)	0.014 (0.018)	0.043 (0.019)	0.111 (0.019)	0.039 (0.019)	0.017 (0.018)
μ_3	0.011 (0.017)	0.056 (0.018)	0.069 (0.019)	−0.020 (0.019)	0.017 (0.018)	0.060 (0.019)	0.073 (0.019)	−0.011 (0.018)
β	0.707 (0.019)	0.849 (0.015)	0.876 (0.010)	0.844 (0.010)	0.730 (0.022)	0.860 (0.020)	0.889 (0.013)	0.839 (0.015)
α	0.142 (0.012)	0.107 (0.010)	0.107 (0.008)	0.103 (0.007)	0.128 (0.015)	0.098 (0.013)	0.095 (0.011)	0.110 (0.011)
θ	0.788 (0.083)	0.160 (0.046)	0.080 (0.044)	−0.659 (0.057)	0.831 (0.115)	0.185 (0.087)	0.042 (0.071)	−0.603 (0.077)
ν	–	–	–	–	10.753 (1.684)	7.900 (0.939)	7.776 (1.017)	8.068 (1.144)
κ	–	–	–	–	−0.225 (0.025)	−0.059 (0.027)	0.036 (0.027)	−0.157 (0.027)
Log-likelihood	7512	9083	9311	8418	7587	9140	9357	8489

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

model using our whole sample from 1963 to 2019. From this table, we see that the model gives a reasonable fit. In most cases, three or four loading estimates are statistically different from zero, with an R-squared of around 30%. The model also captures well the communalities of the stock returns that we can assess with the average of the absolute value of the correlations between the stock returns, which is equal to 0.35. For the residuals filtered from the model, this average is 0.06. Finally, the Jarque–Bera statistics for the residuals in the last column indicate that the residuals are not well described by a normal distribution. As explained in the next section, these non-normalities can be handled using filtered historical simulations when computing the VaR and ES forecasts.

6. VaR and ES forecasts

The paper aims to compare the out-of-sample risk forecasting accuracy of the univariate and multivariate approaches. 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. We then re-estimate every week using a rolling estimation window of 20 years.

To avoid making too many stringent assumptions about the portfolios with which the VaR and ES will be computed, we randomly choose our sets of portfolios weights. Specifically, for each out-of-sample period in our study, we randomly generate one set of portfolio weights. For this purpose, we use the procedure described in De Giorgi, Hens, and Mayer (2007) and Rubinstein (1982). That is, we generate weights using a uniform distribution, with each weight between zero and one, and a sum of weights equal to one. The same sets of portfolio weights are applied to the univariate and multivariate models.

For the univariate models, when computing the VaR and ES, we rely on analytical formulas. The formulas for these computations are available in, for example, Christoffersen (2012).

In the case of multivariate models, for the portfolio of factors, we rely on simulations to compute the VaR and ES. Specifically, using the parameters of the one-week-ahead multivariate conditional distribution of the vector of factor returns $\boldsymbol{F}_t$ estimated at time $t-1$, we simulate a vector of factor returns $\tilde{\boldsymbol{r}}_t$ from this distribution. The vector of simulated weekly log stock returns $\tilde{\boldsymbol{r}}_t$ is converted into a vector of holding period returns $\tilde{\boldsymbol{r}}_{h,t}$. The simulated portfolio return for the next period is then obtained using $\tilde{r}_{w,t} = \ln\left(\boldsymbol{w}'_{t-1}\tilde{\boldsymbol{r}}_{h,t}+1\right)$, where $\boldsymbol{w}_{t-1}$ is the vector of portfolio weights. We repeat this $M = 200,000$ times and obtain a simulated sample $\left\{\tilde{r}^l_{w,t}\right\}_{l=1}^M$.

In the case of multivariate models, for portfolios of stocks, using the OLS estimates from Eq. (1) for the intercept $\hat{a}_i$ and factor loadings $\hat{\boldsymbol{b}}_i$, we extract the multivariate time series $\hat{\boldsymbol{\varepsilon}}_t = [\hat{\varepsilon}_{1,t},\ldots,\hat{\varepsilon}_{N^s,t}]$ of OLS residuals. We simulate from the distribution $\boldsymbol{F}^{\varepsilon}$ by drawing randomly into the sample of residuals $\tilde{\boldsymbol{\varepsilon}}_t, t = 1,\ldots,T$; i.e. we use a filtered historical simulation approach. Once the parameters of the one-week-ahead multivariate conditional

Table 4
Parameter estimates for the copulas (1963–2019).

Parameter	Constant correlation			Dynamic correlation		
	Normal copula	Student copula	Skewed t copula	Normal copula	Student copula	Skewed t copula
ν_c		5.192 (0.270)	5.252 (0.275)		10.387 (0.844)	10.420 (0.881)
λ_{Market}			−0.003 (0.031)			−0.077 (0.056)
λ_{Size}			−0.074 (0.032)			−0.180 (0.058)
λ_{Value}			0.065 (0.031)			0.092 (0.054)
$\lambda_{\text{Momentum}}$			−0.156 (0.038)			−0.141 (0.055)
α_c				0.083 (0.003)	0.087 (0.004)	0.086 (0.004)
β_c				0.891 (0.004)	0.888 (0.005)	0.888 (0.005)
$\rho_{\text{Market, Size}}$	0.040	0.054	0.056	0.071	0.098	0.095
$\rho_{\text{Market, Momentum}}$	0.064	0.075	0.079	0.033	0.043	0.041
$\rho_{\text{Size, Value}}$	−0.068	−0.068	−0.062	−0.104	−0.122	−0.117
$\rho_{\text{Size, Momentum}}$	0.006	0.012	−0.007	0.024	0.034	0.025
$\rho_{\text{Value, Momentum}}$	−0.139	−0.160	−0.147	−0.168	−0.197	−0.194
Log-likelihood	151.5	395.6	409.4	1107.1	1206.9	1215.6

This table presents the parameter 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).

distribution of the vector of factor returns $\boldsymbol{F}_t$ is estimated at time $t-1$, we simulate a vector of factor returns $\tilde{\boldsymbol{r}}_t$ from this distribution, and a vector of error terms $\tilde{\boldsymbol{\varepsilon}}_t = [\tilde{\varepsilon}_{1,t}, \dots, \tilde{\varepsilon}_{N^s,t}]$ from $\boldsymbol{F}^{\varepsilon}$. The vector of simulated weekly log stock returns $\tilde{\boldsymbol{r}}_t$ is converted into a vector of holding period returns $\tilde{\boldsymbol{r}}_{h,t}$. The simulated individual holding period stock returns $\tilde{r}^s_{h,i,t}$ are then computed using

$$\tilde{r}^s_{h,i,t} = r_{f,t} + \hat{a}_i + \hat{\boldsymbol{b}}_i^{'}\tilde{\boldsymbol{r}}_{h,t} + \tilde{\varepsilon}_{i,t} \quad \text{for } i = 1, \dots, N^s.$$

The simulated portfolio return for the next period is then obtained using $\tilde{r}_{w,t} = \ln\left(\boldsymbol{w}^{'}_{t-1}\tilde{\boldsymbol{r}}^s_{h,t} + 1\right)$, where $\boldsymbol{w}_{t-1}$ is the vector of portfolio weights. We repeat this $M = 200,000$ times and obtain a simulated sample $\left\{\tilde{r}^l_{w,t}\right\}^M_{l=1}$.

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

$$\widehat{VaR}^p_t = -Percentile\left\{\{\tilde{r}^l_{w,t}\}^M_{l=1}, 100p\right\} \tag{4}$$

$$\widehat{ES}^p_t = -\frac{1}{p \cdot M}\sum_{l=1}^{M} \tilde{r}^l_{w,t} \cdot \mathbf{1}_{\{\tilde{r}^l_{w,t} < -\widehat{VaR}^p_t\}} \tag{5}$$

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

We consider two probability levels for the VaR and ES: $p = 1\%$ and $p = 2.5\%$. Repeating the previous steps each period, we obtain a time series of 1905 out-of-sample forecasts for the one-week-ahead VaR and ES at both levels of significance for each of the ten models. This out-of-sample period ranges from the first week of July 1983 to the end of December 2019; we lose the first 1043 observations from the first estimation. In what follows, $t = 1, \dots, T = 1905$ denotes 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 2.5% levels. A VaR violation occurs when the portfolio return drops below the estimated VaR, i.e. $r_{w,t} < -\widehat{VaR}^p_t$. Let $I_t = \mathbf{1}_{\{r_{w,t} < -\widehat{VaR}^p_t\}}$ indicate whether a VaR violation occurred at time t, and let $T_1 = \sum_{t=1}^{T} I_t$ 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 and for the portfolios of stocks and factors in Table 6. Given our 1905 observations, a correctly specified VaR model should produce around 19 and 48 violations at the 1% and 2.5%

Table 5
Parameter estimates for the stock return model (1963–2019).

	a	b_1	b_2	b_3	b_4	R2	J–B
Coca-Cola	1.26e−03	0.770	−0.613	−0.192	0.049	0.326	1574
	(4.41e−04)	(0.041)	(0.065)	(0.065)	(0.051)		
Exxon	2.89e−04	0.779	−0.411	0.288	0.117	0.368	258
	(3.51e−04)	(0.034)	(0.045)	(0.069)	(0.042)		
GE	5.06e−07	1.162	−0.379	0.183	−0.138	0.494	10187
	(4.46e−04)	(0.054)	(0.057)	(0.087)	(0.042)		
IBM	5.73e−04	0.867	−0.186	−0.359	−0.147	0.380	4718
	(5.07e−04)	(0.029)	(0.047)	(0.053)	(0.043)		
Chevron	3.15e−04	0.930	−0.374	0.356	0.077	0.338	716
	(4.41e−04)	(0.036)	(0.051)	(0.071)	(0.054)		
UTC	7.49e−04	1.081	−0.011	0.092	0.006	0.374	1461
	(5.11e−04)	(0.042)	(0.070)	(0.074)	(0.054)		
Procter	9.89e−04	0.637	−0.543	−0.104	0.065	0.266	57687
	(4.04e−04)	(0.028)	(0.055)	(0.060)	(0.049)		
Caterpillar	4.51e−04	1.173	0.041	0.434	−0.165	0.368	2000
	(5.39e−04)	(0.043)	(0.076)	(0.101)	(0.067)		
Boeing	1.35e−03	1.116	0.051	0.154	−0.054	0.287	813
	(6.83e−04)	(0.045)	(0.090)	(0.080)	(0.069)		
Merck	1.52e−03	0.808	−0.586	−0.293	−0.031	0.300	2656
	(4.77e−04)	(0.040)	(0.053)	(0.078)	(0.056)		

This table presents the ordinary least squares (OLS) parameter estimates of the factor model on individual stock returns with Newey–West standard errors in parentheses. "R2" is the OLS R-squared, and "J–B" is the Jarque–Bera statistic for the OLS residuals.

levels, respectively. At the 1% level, the univariate normal model is particularly noticeable, producing more than 2 and 1.8 times the correct number of violations for the factor and stock portfolios, respectively. The univariate skewed t shows 21 and 22 violations, and the results for the univariate normal filtered historical simulation (FHS) are similar. In the same vein, the multivariate normal models produce too many violations at the 1% level, while the copula models produce significantly fewer violations. Our most sophisticated multivariate model, the dynamic skewed t copula, produces a number violations comparable to the univariate skewed t and univariate normal FHS. Similar remarks apply at the 2.5% level, although some models produce too few violations for the portfolio of stocks.

We can test whether the observed fraction of violations is statistically different from $p = 1\%$ or 2.5% using the unconditional coverage test introduced in Kupiec (1995). The p-values appear under the "UC" columns of Table 6. For the 1% VaR using a significance level of 5%, we reject all models based on the normal distribution, while all other models show adequate coverage. When we consider the 2.5% VaR, the results are the same, except that for the stock portfolio, the multivariate normal models now display adequate coverage.

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). The p-values appear under the "Ind". columns of Table 6. For the portfolio of factors, for the 1% VaR and 2.5% VaR, and using a 5% significance level, we see that all models pass the tests, except the normal copula. For the portfolio of stocks, all models fail the independence test. This failure can be traced back to two clusters of successive violations: a first cluster of three successive violations during the market crash of October 1987, and a second cluster of two successive violations during the financial crisis of October 2008. Fig. 1 plots the 1% VaR and realized portfolio returns from 2007 to the end of 2009, showing the second cluster of violations in October 2008. The univariate model is the skewed t distribution, while the multivariate model is the dynamic skewed t copula. We can see that both models produce similar VaR computations and produce violations at the same points in time.

We can jointly test for coverage and independence with the conditional coverage test of Christoffersen (1998). The p-values are displayed under the "CC" columns of Table 6. For both the 1% and 2.5% VaR, we see that the results are consistent with the conclusions of the two previous tests. Indeed, only the models that passed both the "UC" and "Ind". tests display adequate conditional coverage.

We can summarize this section by saying that all models pass the VaR conditional and unconditional backtests for the portfolio of factors, except those involving the normal distribution. Likewise, all models display adequate unconditional coverage for the portfolio of stocks, except those involving the normal distribution. However, they all fail the conditional tests because of the two clusters of successive violations identified above.

Table 6
VaR and ES backtests (1983–2019).

	$p = 1\%$									$p = 2.5\%$								
	T_1	UC	Ind.	CC	Z_1	Z_2	Z_{es}	DE_u	DE_c	T_1	UC	Ind.	CC	Z_1	Z_2	Z_{es}	DE_u	DE_c
	Portfolios of factors																	
Univariate models																		
Normal dist.	41	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.09	73	0.00	0.90	0.00	0.00	0.00	0.00	0.00	0.08
Skew. t dist.	21	0.66	0.49	0.72	0.33	0.29	0.26	0.52	0.23	51	0.62	0.21	0.41	0.31	0.27	0.26	0.55	0.12
Filt. hist. simu.	22	0.51	0.47	0.62	0.67	0.36	0.57	0.73	0.53	47	0.93	0.14	0.34	0.36	0.55	0.55	0.79	0.17
Multivariate models																		
Normal dist.	42	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.07	62	0.04	0.20	0.06	0.00	0.00	0.00	0.00	0.05
Normal cop.	32	0.01	0.02	0.00	0.09	0.00	0.00	0.00	0.94	53	0.44	0.07	0.15	0.00	0.05	0.00	0.02	0.06
Sym. t cop.	25	0.19	0.41	0.31	0.21	0.07	0.06	0.19	0.30	52	0.53	0.07	0.15	0.05	0.14	0.04	0.15	0.06
Skew. t cop.	21	0.66	0.49	0.72	0.18	0.24	0.14	0.44	0.40	50	0.73	0.05	0.14	0.13	0.26	0.13	0.39	0.12
Dyn. normal dist.	35	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.64	63	0.03	0.95	0.10	0.00	0.00	0.00	0.00	0.50
Dyn. normal cop.	27	0.08	0.40	0.16	0.04	0.01	0.01	0.01	0.63	56	0.23	0.78	0.47	0.03	0.04	0.01	0.12	0.57
Dyn. sym. t cop.	23	0.38	0.45	0.51	0.06	0.10	0.03	0.07	0.24	53	0.44	0.67	0.68	0.07	0.12	0.05	0.28	0.59
Dyn. skew. t cop.	22	0.51	0.47	0.62	0.10	0.16	0.07	0.21	0.27	50	0.73	0.57	0.80	0.10	0.24	0.12	0.53	0.83
	Portfolios of stocks																	
Univariate models																		
Normal dist.	36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	68	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Skew. t dist.	22	0.51	0.00	0.01	0.12	0.18	0.10	0.12	0.00	55	0.29	0.00	0.00	0.23	0.11	0.10	0.23	0.00
Filt. hist. simu.	23	0.38	0.00	0.01	0.41	0.26	0.30	0.42	0.00	59	0.11	0.00	0.00	0.71	0.09	0.27	0.29	0.00
Multivariate models																		
Normal dist.	31	0.01	0.01	0.00	0.25	0.01	0.03	0.01	0.00	52	0.53	0.00	0.01	0.02	0.22	0.04	0.12	0.00
Normal cop.	22	0.51	0.00	0.01	0.77	0.37	0.63	0.89	0.00	44	0.59	0.00	0.00	0.43	0.78	0.68	0.61	0.00
Sym. t cop.	22	0.51	0.00	0.01	0.80	0.38	0.66	1.00	0.00	44	0.59	0.00	0.00	0.45	0.78	0.69	0.58	0.00
Skew. t cop.	22	0.51	0.00	0.01	0.69	0.34	0.56	0.85	0.00	45	0.70	0.00	0.00	0.36	0.71	0.60	0.81	0.00
Dyn. normal dist.	29	0.03	0.01	0.00	0.25	0.02	0.05	0.03	0.00	53	0.44	0.00	0.01	0.05	0.20	0.07	0.19	0.00
Dyn. normal cop.	21	0.66	0.00	0.00	0.32	0.36	0.31	0.43	0.00	49	0.84	0.00	0.01	0.34	0.50	0.42	0.96	0.00
Dyn. sym. t cop.	21	0.66	0.00	0.00	0.32	0.36	0.31	0.52	0.00	49	0.84	0.00	0.01	0.34	0.50	0.43	0.97	0.00
Dyn. skew. t cop.	21	0.66	0.00	0.00	0.27	0.34	0.27	0.44	0.00	49	0.84	0.00	0.01	0.27	0.48	0.37	0.93	0.00

This table presents the VaR and ES backtest results. The first panel is for the portfolio of factors, and the second panel is for the portfolio of stocks. Each panel presents results for the 1% level and the 2.5% level. Column T_1 indicates the total number of violations in the out-of-sample period. Columns UC, Ind., and CC contain the p-values for the unconditional coverage, the independence, and the conditional coverage VaR tests, respectively. Columns Z_1 and Z_2 contain the p-values for the first two ES tests of Acerbi and Szekely (2014), and column Z_{es} is for the ES test proposed in Acerbi and Szekely (2017). DE_u and DE_c are the unconditional and conditional ES tests proposed in Du and Escanciano (2017).

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 elicitability". This finding sparked a debate over whether it is even possible to backtest ES. Fortunately, recent literature has answered the question by proposing many ES backtests that do not rely on the elicitability property, although the procedures are not as straightforward as for VaR. From the available research in this area, we choose to implement the first two tests proposed by Acerbi and Szekely (2014), a third test proposed in Acerbi and Szekely (2017), and the unconditional and conditional ES tests of Du and Escanciano (2017). These tests do not make any assumptions about the distribution of returns (non-parametric), and are computed by simulation. The details of all of these backtests (assumptions and computations) are given in Appendix C. When computing these tests, we use $M = 50{,}000$ paths and simulate the M portfolio returns each week with the predictive conditional distribution used for the VaR and ES forecasts.

The p-values of all five tests are shown in Table 6, in the columns Z_1, Z_2, and Z_{es} (the three Acerbi and Sczekely tests), and DE_u and DE_c (the two Du and Escanciano tests), for both the factor and stock portfolios. We notice that the univariate models pass all backtests for the portfolio of stocks, except for the normal case. The multivariate models, with the exception of those involving the normal distribution, also pass the conditional and unconditional tests. We also note that the conclusions for the Acerbi and Szekely tests are not always consistent with those of the Du and Escanciano tests. For example, for 2.5% case, the dynamic normal copula model fails the Acerbi and Szekely tests, but passes the conditional and unconditional Du and Escanciano tests. For the stock portfolio, a picture similar to the one obtained for the VaR backtest emerges. Most models pass the unconditional backtests and fail the conditional test.

8. Testing for comparative predictive accuracy

8.1. Loss functions

As mentioned in the introduction, backtesting procedures are not necessarily the right tools to compare models. These tests provide a binary outcome, i.e. they 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 evaluations, but not for relative evaluations.

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 the risk forecasts, here VaR and ES, to the realized

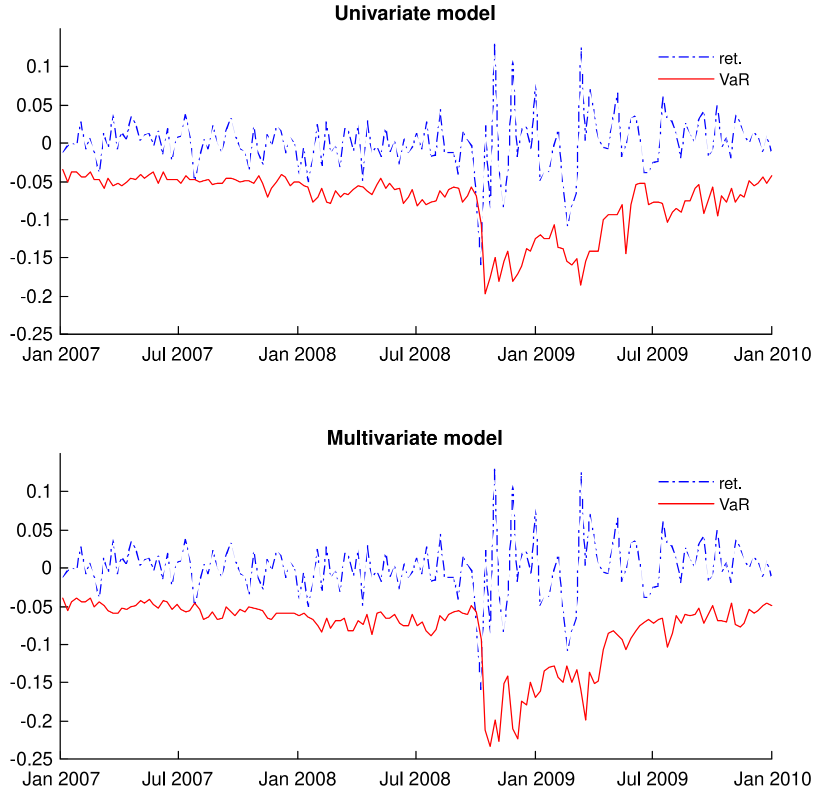

Fig. 1. 1% VaR and realized portfolio returns for univariate and multivariate models (2007–2009).

portfolio return at each period. The idea is that a model with a given average loss will be preferred over 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 elicitability property, is not automatic, since 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

$$L_V(v_t, r_{w,t}) = (r_{w,t} + v_t)(p - \mathbf{1}_{\{r_{w,t}+v_t<0\}})$$

where v_t is a non-random real variable. This is an asymmetric function because the penalty given when $\mathbf{1}_{\{r_{w,t}+v_t<0\}} = 1$ is usually much higher than in the case where $\mathbf{1}_{\{r_{w,t}+v_t<0\}} = 0$. Nonetheless, this function penalizes risk overestimation because, conditional on $\mathbf{1}_{\{r_{w,t}+v_t<0\}} = 1$, 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

$$VaR_t^p = \underset{v_t}{\operatorname{argmin}}\, \mathbb{E}[L_V(v_t, r_{w,t})]. \tag{6}$$

Thus, by computing $\bar{L}_V = \frac{1}{T}\sum_{t=1}^{T} L_V(v_t, r_{w,t})$ with $v_t = \widehat{VaR}_t^p$ for each model and sorting the numbers in increasing order, we can rank the 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.

For both factor and stock portfolios, the VaR out-of-sample average loss $\bar{L}_V$ of each model, in basis points, is shown in Table 7. For the stock portfolio, we see that the univariate skewed t distribution displays the lowest average loss for the 1% VaR, while the dynamic skewed t copula displays the lowest average loss for the 2.5% VaR. For the factor portfolio, the univariate normal FHS model displays the lowest average loss for both VaR levels.

We now turn to the ES model comparison. As discussed above, 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 Patton et al. (2019) joint loss function, given by

$$L_{V,E}(v_t, e_t, r_{w,t}) = -\frac{1}{pe_t}\mathbf{1}_{\{r_{w,t}+v_t<0\}}(r_{w,t} + v_t) + \frac{v_t}{e_t} + \ln(e_t) - 1.$$

Under the assumption of positive VaR and ES, Patton et al. (2019) show that

$$\{VaR_t^p, ES_t^p\} = \underset{v_t, e_t}{\operatorname{argmin}}\, \mathbb{E}[L_{V,E}(v_t, e_t, r_{w,t})], \tag{7}$$

and that $L_{V,E}$ generates loss differences that are homogeneous of degree zero, a property that has been associated with higher power for the Diebold and Mariano (1995) tests; see also Nolde and Ziegel (2017). Therefore,

Table 7
VaR and VaR/ES average loss (1983–2019).

	Portfolios of factors				Portfolios of stocks			
	Mean loss VaR		Mean loss VaR/ES		Mean loss VaR		Mean loss VaR/ES	
	$p = 1\%$	$p = 2.5\%$	$p = 1\%$	$p = 2.5\%$	$p = 1\%$	$p = 2.5\%$	$p = 1\%$	$p = 2.5\%$
Univariate models								
Normal dist.	3.517	6.660	−3.443	−3.752	8.192	16.319	−2.576	−2.810
Skew. t dist.	3.341	6.594	−3.561	−3.786	7.907	16.127	−2.636	−2.834
Filt. hist. simu.	3.240	6.493	−3.583	−3.807	7.989	16.189	−2.621	−2.831
Multivariate models								
Normal dist.	3.772	6.938	−3.299	−3.669	7.905	16.046	−2.626	−2.836
Normal cop.	3.564	6.853	−3.458	−3.718	7.699	15.950	−2.658	−2.844
Sym. t cop.	3.524	6.849	−3.492	−3.727	7.701	15.949	−2.658	−2.844
Skew. t cop.	3.500	6.826	−3.504	−3.735	7.715	15.957	−2.656	−2.843
Dyn. normal dist.	3.562	6.605	−3.428	−3.763	7.795	15.918	−2.644	−2.847
Dyn. normal cop.	3.437	6.536	−3.522	−3.790	7.942	16.082	−2.625	−2.836
Dyn. sym. t cop.	3.408	6.526	−3.546	−3.798	7.948	16.065	−2.624	−2.837
Dyn. skew. t cop.	3.397	6.518	−3.554	−3.802	7.941	16.063	−2.625	−2.837

This table presents each model's average loss for VaR ($\bar{L}_V$) and the for the VaR/ES pair ($\bar{L}_{V,E}$) in the out-of-sample period.

our criterion for VaR/ES model comparison is $\bar{L}_{V,E} = \frac{1}{T}\sum_{t=1}^{T} L_{V,E}(v_t, e_t, r_{w,t})$ with $v_t = \widehat{VaR}_t^p$ and $e_t = \widehat{ES}_t^p$.

The out-of-sample average joint loss $\bar{L}_{V,E}$ for each model, in basis points, is shown in Table 7. Note that the losses are negative. For the stock portfolio, the univariate FHS is the most accurate model at the 1% and 2.5% levels. For the portfolio of stocks, the function values are, in general, very close to each other. The lowest value is achieved by the static normal and symmetric t copula models at the 1% level, while the dynamic normal distribution reaches the lowest value at the 2.5% level.

In the next subsection, we examine whether these differences in values are statistically significant, with tests of comparative predictive accuracy.

8.2. Comparative predictive accuracy

Although there are differences between the models in terms of average losses, these differences may not be statistically significant. We can test for pairwise differences in predictive accuracy using the conditional predictive ability (CPA) test of Giacomini and White (2006), which is a conditional version of the Diebold and Mariano (1995) test.

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

$$d_{i,j,t} = L_{i,t} - L_{j,t}$$

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

$$H_0 : \mathbb{E}(d_{i,j,t} \mid \mathcal{F}_{t-1}) = 0 \qquad \text{for all } t.$$

which is equivalent to saying that $d_{i,j,t}$ is a martingale difference sequence. Giacomini and White (2006) propose the following Wald statistic:

$$GW_{i,j} = T\bar{Z}'\hat{\Omega}^{-1}\bar{Z},$$

where $\bar{Z} = \frac{1}{T}\sum_{t=2}^{T} Z_t$, $\hat{\Omega}^{-1} = \frac{1}{T}\sum_{t=2}^{T} Z_t Z_t'$, $Z_t = h_{t-1}d_{i,j,t}$, and h_{t-1} is a vector with variables measurable at time $t-1$. We use $h_{t-1} = (1, d_{i,j,t-1})'$. Under some regularity conditions, Giacomini and White (2006) show that the asymptotic distribution of the statistic $GW_{i,j}$ is χ_q^2, where q is the dimension of h_{t-1}, here two. We compute p-values in the right tail of this distribution and reject when $GW_{i,j}$ is sufficiently large.

The p-values for the unilateral tests are displayed, for each pair of models,[4] in Tables 8 and 9. Table 8 corresponds to 1% VaR comparisons using the "tick" loss function, while Table 9 corresponds to 1% VaR/ES comparisons using Patton et al.'s (2019) joint loss function. 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 VaR loss function tests for the portfolio of factors (the first panel of Table 8), we see that the two univariate models show significantly lower function values when compared to the static multivariate copula models (but not the dynamic ones). We also observe that the static multivariate normal and static normal copula models tend to be dominated by the multivariate static and dynamic models. However, other than these, the loss function values are not statistically different from one another. Hence, there are no significant differences between the univariate approaches and the dynamic multivariate approaches for portfolios of factors in terms of loss function values.

Looking at the VaR loss function tests for the portfolio of stocks (the second panel of Table 8), we do not see any arrows indicating significant differences. Based on these results, we also conclude that, in the context of multivariate VaR calculations with FFC factors, there are no significant differences between the univariate approaches and multivariate approaches for portfolios of stocks in terms of loss function values.

The 1% VaR/ES comparisons in Table 9 are similar to those obtained from the VaR loss function. For the

[4] We exclude the univariate normal model from the comparison, since it failed most backtests for both portfolios.

Table 8
GW test for 1% VaR loss function (1983–2019).

	Uni. filt. hist. sim.	Multi. normal	Normal cop.	Sym. *t* cop.	Skew. *t* cop.	Dyn. multi. normal	Dyn. normal cop.	Dyn. sym. *t* cop.	Dyn. skew. *t* cop.
Portfolio of factors									
Uni. skewed *t*	0.10	0.09	0.01←	0.02←	0.02←	0.14	0.34	0.26	0.30
Uni. filt. hist. sim.		0.06	0.01←	0.01←	0.01←	0.10	0.19	0.13	0.15
Multi. normal			0.04↑	0.04↑	0.05↑	0.15	0.02↑	0.02↑	0.02↑
Normal cop.				0.21	0.27	0.59	0.04↑	0.03↑	0.03↑
Sym. *t* cop.					0.06	0.89	0.38	0.07	0.05
Skew. *t* cop.						0.89	0.75	0.10	0.08
Dyn. multi. normal							0.16	0.13	0.21
Dyn. normal cop.								0.15	0.32
Dyn. sym. *t* cop.									0.06
Portfolios of stocks									
Uni. skew. *t*	0.10	0.48	0.61	0.61	0.62	0.36	0.27	0.28	0.29
Univ. filt. hist. sim.		0.43	0.50	0.51	0.52	0.29	0.27	0.27	0.27
Multi. normal			0.32	0.31	0.33	0.23	0.93	0.87	0.83
Normal cop.				0.57	0.49	0.47	0.26	0.26	0.30
Sym. *t* cop.					0.47	0.47	0.27	0.26	0.30
Skew. *t* cop.						0.48	0.28	0.27	0.32
Dyn. multi. normal							0.34	0.30	0.32
Dyn. normal cop.								0.55	0.17
Dyn. sym. *t* cop.									0.60

This table presents the *p*-values of the GW test comparing each row model to a column model for the 1% VaR for portfolios of factors. The loss function used for pairwise comparisons is L_V, the VaR tick loss function defined in Section 8.1. 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
GW test for 1% VaR and ES joint loss function (1983–2019).

	Uni. filt. hist. sim.	Multi. normal	Normal cop.	Sym. *t* cop.	Skew. *t* cop.	Dyn. multi. normal	Dyn. normal cop.	Dyn. sym. *t* cop.	Dyn. skew. *t* cop.
Portfolio of factors									
Uni. skewed *t*	0.39	0.06	0.01←	0.05←	0.06	0.12	0.21	0.10	0.15
Uni. filt. hist. sim.		0.05←	0.01←	0.02←	0.03←	0.11	0.17	0.07	0.14
Multi. normal			0.01↑	0.01↑	0.02↑	0.13	0.02↑	0.01↑	0.01↑
Normal cop.				0.17	0.24	0.70	0.02↑	0.02↑	0.02↑
Sym. *t* cop.					0.03↑	0.62	0.45	0.05	0.04↑
Skew. *t* cop.						0.56	0.77	0.05↑	0.03↑
Dyn. multi. normal							0.05↑	0.05↑	0.08
Dyn. normal cop.								0.14	0.27
Dyn. sym. *t* cop.									0.06
Portfolios of stocks									
Uni. skew. *t*	0.10	0.54	0.60	0.59	0.61	0.36	0.22	0.24	0.24
Univ. filt. hist. sim.		0.50	0.51	0.51	0.53	0.31	0.25	0.25	0.26
Multi. normal			0.30	0.29	0.31	0.24	0.50	0.56	0.55
Normal cop.				0.53	0.45	0.52	0.28	0.27	0.32
Sym. *t* cop.					0.39	0.52	0.29	0.28	0.32
Skew. *t* cop.						0.54	0.32	0.31	0.35
Dyn. multi. normal							0.52	0.51	0.52
Dyn. normal cop.								0.29	0.17
Dyn. sym. *t* cop.									0.94

This table presents the *p*-values of the GW test comparing each row model to a column model for the 1% VaR/ES pair for portfolios of factors. The loss function used for pairwise comparisons is $L_{V,E}$, Patton et al.'s (2019) VaR/ES joint loss function defined in Section 8.1. 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.

portfolio of factors (the first panel), we again notice some significant differences between the univariate and static multivariate models. However, the differences between the univariate approaches and the dynamic multivariate approaches are not significant. For the portfolio of stocks, we again do not find any significant differences between

Table 10
MCS results (1983–2019) for portfolios of factors.

	MCS	
	$p = 1\%$	$p = 2.5\%$
VaR	All models	All models
VaR/ES	All models except multivariate normal	All models except multivariate normal, normal copula, and sym. t copula

This table presents the model confidence set (MCS) obtained using the procedure of Hansen, Lunde, and Nason (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 2.5% level. All models refer to the initial set of all models examined in the MCS testing procedure.

the models. Finally, although we do not report them here, similar calculations were made for the 2.5% case. The results from these calculations were qualitatively similar to those obtained for the 1% case.

We can summarize this section by saying that, for both portfolios, there are, in general, no significant differences in the loss functions between the best multivariate models and the two univariate models. This result is more striking for the ten stock portfolios. We do not find any significant differences between any multivariate factor-based models and our two univariate models. For the portfolios of factors, we find that static models tend to be dominated by univariate and multivariate dynamic models.

9. Model confidence set

We summarize the results in Tables 8 and 9 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 the 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 of confidence.

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

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

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

$$T_R = \max_{i,j \in \mathcal{M}} |DM_{i,j}|,$$

where $DM_{i,j}$ denotes the Diebold and Mariano (1995) test statistic. The EPA test statistic is thus the absolute value of the DM statistic furthest away from zero among the pairs of models. Because the asymptotic distribution of T_R is unknown, Hansen et al. (2011) propose estimating it via a circular block bootstrap scheme. This allows for 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,

$$i^* = \operatorname*{argmax}_{i \in \mathcal{M}} \max_{j \in \mathcal{M}} DM_{i,j}.$$

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.[5] We use 10,000 bootstrap resamples with a circular bootstrap scheme and a block of one.

The results for the portfolios of factors are presented in Table 10. For the 1% and 2.5% VaR, the MCS consists of all models. This means that the first test did not reject the EPA null hypothesis. It is thus not possible to rule out any of the models for the VaR. The MCS also consists of all models for the 1% joint VaR/ES function, but without the static multivariate normal. Finally, for the 2.5% joint VaR/ES loss function, the test rules out three models: the static multivariate normal, the static normal copula, and the static symmetric t copula.

For the case of the portfolio of stocks, perhaps unsurprisingly given the results from the previous section, the MCS consists of all models in all cases, i.e. for both VaR and VaR/ES loss functions at both levels of significance.

We can summarize this section by saying that the MCS set includes both univariate and multivariate models. The test thus indicates that, in the context of multivariate VaR and ES computations with the FFC models, risk forecasts from the univariate and multivariate approaches are not different from a statistical point of view.

10. Conclusion

Using the Fama and French (1993) and Carhart (1997) stock market factors, we examined whether a multivariate modeling stock factor-based approach is more effective than a simple univariate approach to forecast the

[5] We use the “mcs” MATLAB function from the MFE toolbox by Kevin Sheppard; see https://www.kevinsheppard.com/MFE_Toolbox.

market risk of a random weighted portfolio of stocks and a portfolio of FFC factors. Our contribution to literature comparing univariate and multivariate approaches, besides a close examination of a significant data set for financial managers, is that we examined in more detail the issue in terms of the expected shortfall risk measure. In total, three univariate models were compared with eight multivariate models involving asymmetric distributions and asymmetric copulas with dynamic correlations.

Using simulations, we generated 1905 weekly out-of-sample VaR and ES for each model, covering the period from 1983 to 2019. We analyzed the relative performance of our models in two stages. In the first stage, we backtested each model by comparing the ex ante risk measures to the ex post portfolio returns. In the second stage, we relied on loss functions based on the elicitability property of VaR as well as the joint elicitability property of VaR and ES to rank models. We tested for statistical differences between the average loss of models with the CPA test of Giacomini and White (2006), and with the MCS procedure of Hansen et al. (2011).

We found no significant differences between the risk forecasting accuracy of univariate models and multivariate factor-based models with dynamic correlations. Moreover, we found that the risk forecasting accuracy of multivariate models without dynamic correlations is significantly lower than that of their dynamic counterparts, and also lower than that of non-normal univariate models.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Hansen's (1994) skewed *t* distribution

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

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

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

$$a=4\kappa c\frac{\nu-2}{\nu-1},$$

$$b^2=1+3\kappa^2-a^2,$$

and

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

The univariate skewed t distribution has a mean of zero and a unit variance.

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 $\boldsymbol{x}$ follows a multivariate skewed t distribution, denoted $\boldsymbol{x}\sim\boldsymbol{F}^{st}_{\nu_c,\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda}}$, if it has the following density function (see the unnumbered equation given in Box I). where $K_{(\cdot)}(\cdot)$ is the modified Bessel function of the third kind, and c is a constant given by

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

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

$$\mathbb{E}(\boldsymbol{x})=\boldsymbol{\mu}+\frac{\nu_c}{\nu_c-2}\boldsymbol{\lambda}$$

and

$$\mathrm{Cov}(\boldsymbol{x})=\frac{\nu_c}{\nu_c-2}\boldsymbol{\Sigma}+\frac{2\nu_c^2}{(\nu_c-2)^2(\nu_c-4)}\boldsymbol{\lambda}\boldsymbol{\lambda}'. \tag{8}$$

The multivariate skewed t distribution has the following stochastic representation:

$$\boldsymbol{x}\overset{d}{=}\boldsymbol{\mu}+\sqrt{w}\boldsymbol{y}+\boldsymbol{\lambda}w \tag{9}$$

where w is an inverse gamma random variable, $w\sim IG(\nu_c/2,\nu_c/2)$, $\boldsymbol{y}$ is an $N\times 1$ vector of normal variables, $\boldsymbol{y}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{\Sigma})$, and $\boldsymbol{y}$ and w are independent.

B.2. Skewed *t* copula density

The skewed t copula is derived using the standardized multivariate skewed t distribution $\boldsymbol{F}:=\boldsymbol{F}^{st}_{\nu_c,\boldsymbol{0},\boldsymbol{\Upsilon},\boldsymbol{\lambda}}$ with jth marginal F_j and correlation matrix $\boldsymbol{\Upsilon}$:

$$\boldsymbol{C}^{st}_{\nu_c,\boldsymbol{\Upsilon},\boldsymbol{\lambda}}(\boldsymbol{u})=\boldsymbol{F}(F_1^{-1}(u_1),\ldots,F_N^{-1}(u_N)). \tag{10}$$

Let $\boldsymbol{f}:=\boldsymbol{f}^{st}_{\nu_c,\boldsymbol{0},\boldsymbol{\Upsilon},\boldsymbol{\lambda}}$ be the density of $\boldsymbol{F}$ with jth marginal f_j. The copula density (see Eq. (11) given in Box II), is obtained by differentiating both sides of Eq. (10), where $\eta_j:=F_j^{-1}(u_j)$ is defined as the jth copula quantile for $j=1,\ldots,N$. The density in Eq. (11) in Box II 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 Eq. (3).

B.3. Copula quantiles

One difficulty in the second step of the skewed t copula ML estimation is obtaining the copula quantiles $\eta_{j,t}=F_j^{-1}(u_{j,t})$ 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 (9). Yoshiba's (2018) simulation results suggest that using a monotone interpolator is faster and more accurate than using empirical quantiles, so we instead choose this approach

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

Box I.

$$\begin{aligned}\boldsymbol{c}^{st}(\boldsymbol{u};\nu_c,\boldsymbol{\Upsilon},\boldsymbol{\lambda}) &= \frac{\boldsymbol{f}(F_1^{-1}(u_1),\ldots,F_N^{-1}(u_N))}{\prod_{j=1}^N f_j(F_j^{-1}(u_j))}\\ &= \frac{2^{\frac{(\nu_c-2)(N-1)}{2}}K_{\frac{\nu_c+N}{2}}\left(\sqrt{\left(\nu_c+\boldsymbol{\eta}'\boldsymbol{\Upsilon}^{-1}\boldsymbol{\eta}\right)\boldsymbol{\lambda}'\boldsymbol{\Upsilon}^{-1}\boldsymbol{\lambda}}\right)\exp\left(\boldsymbol{\eta}'\boldsymbol{\Upsilon}^{-1}\boldsymbol{\lambda}\right)}{\Gamma\left(\frac{\nu_c}{2}\right)^{1-N}|\boldsymbol{\Upsilon}|^{\frac{1}{2}}\left(\sqrt{\left(\nu_c+\boldsymbol{\eta}'\boldsymbol{\Upsilon}^{-1}\boldsymbol{\eta}\right)\boldsymbol{\lambda}'\boldsymbol{\Upsilon}^{-1}\boldsymbol{\lambda}}\right)^{-\frac{\nu_c+N}{2}}\left(1+\frac{\boldsymbol{\eta}'\boldsymbol{\Upsilon}^{-1}\boldsymbol{\eta}}{\nu_c}\right)^{\frac{\nu_c+N}{2}}}\\ &\quad\times\prod_{j=1}^N\frac{\left(\sqrt{\left(\nu_c+\eta_j^2\right)\lambda_j^2}\right)^{-\frac{\nu_c+1}{2}}\left(1+\frac{\eta_j^2}{\nu_c}\right)^{\frac{\nu_c+1}{2}}}{K_{\frac{\nu_c+1}{2}}\left(\sqrt{\left(\nu_c+\eta_j^2\right)\lambda_j^2}\right)\exp\left(\eta_j\lambda_j\right)}\end{aligned} \tag{11}$$

Box II.

to compute the copula quantiles. We apply following procedure with $m = 150$ interpolating points:

(1) Let $u_{min} = \min_{j=1,\ldots,T} u_{j,t}$ and $u_{max} = \max_{j=1,\ldots,T} u_{j,t}$.
(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 the 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 the numerical integration of f_j for $k = 2,\ldots,m-1$.
(4) Use a monotone interpolator with the data points $\{(\eta_{min}, u_{min}),\ldots,(\eta_{max}, u_{max})\}$ to obtain $\eta_{jt} = F_j^{-1}(u_{j,t})$ for all other values of $u_{j,t} \in [u_{min}, u_{max}]$. We apply MATLAB's piecewise cubic Hermite interpolating polynomial ("pchip") to this end.
(5) Repeat for $j = 1,\ldots,N$.

B.4. Aielli's (2013) cDCC

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

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

and

$$s_j := \mathrm{var}(\eta_j) = \frac{\nu_c}{\nu_c-2} + \frac{2\nu_c^2}{(\nu_c-2)^2(\nu_c-4)}\lambda_j^2$$

respectively be the expectation and the standard deviation of the jth copula quantile. This allows us to define the standardized copula quantiles $\epsilon_{j,t} = \frac{\eta_{j,t}-\bar{\eta}_j}{s_j}$ for $j = 1,\ldots,N$ and the vector $\boldsymbol{\epsilon}_t = (\epsilon_{j,t},\ldots,\epsilon_{Nt})'$ containing them. The link between the two correlation matrices $\boldsymbol{\Upsilon}_t$ and $\dot{\boldsymbol{\Upsilon}}_t$ is given by Eq. (8):

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

where $\boldsymbol{D} := diag(s_1,\ldots,s_N)$ is an $N\times N$ 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):

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

$$\dot{\boldsymbol{\Upsilon}}_t = \boldsymbol{Q}_t^{*-\frac{1}{2}}\boldsymbol{Q}_t\boldsymbol{Q}_t^{*-\frac{1}{2}}$$

where $\boldsymbol{Q} = \mathbb{E}(\boldsymbol{Q}_t) = \mathbb{E}(\boldsymbol{\epsilon}_t^*\boldsymbol{\epsilon}_t^{*'})$ is a positive definite correlation matrix, while β_c and α_c are non-negative scalars with $\alpha_c + \beta_c < 1$. Also, $\boldsymbol{\epsilon}_t^* = \boldsymbol{Q}_t^{*\frac{1}{2}}\boldsymbol{\epsilon}_t$, where $\boldsymbol{Q}_t^* = diag(dg(\boldsymbol{Q}_t))$, 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 M vectors $\{\tilde{\boldsymbol{x}}^l\}_{l=1}^M$ using the stochastic representation (9) and the forecast correlation matrix for the next period.
(2) Compute $\left\{\tilde{\boldsymbol{u}}^l\right\}_{l=1}^M = \left\{\left(F_1(\tilde{x}_1^l), \ldots, F_N(\tilde{x}_N^l)\right)'\right\}_{l=1}^M$. Given a high number of simulations, instead of using numerical integration we can use the empirical CDF estimate of F_j with the simulated series $\{\tilde{x}_j^l\}_{l=1}^M$ to calculate $\left\{F_j(\tilde{x}_j^l)\right\}_{l=1}^M$ for $j = 1, \ldots, N$.

Once the copula uniform variables $\left\{\tilde{\boldsymbol{u}}^l\right\}_{l=1}^M$ have been simulated, we can apply Hansen's (1994) inverse distribution function to obtain the next-period simulated standardized returns $\{\tilde{z}_j^l\}_{l=1}^M$ for each factor j. We then compute the simulated jth factor returns by multiplying each element of $\{\tilde{z}_j^l\}_{l=1}^M$ by the jth factor forecast conditional volatility and adding to it the jth factor forecast conditional mean.

Appendix C. Expected shortfall backtests

For the Acerbi and Szekely tests, we assume that each period (week), the portfolio return $r_{w,t}$ has an unknown distribution F_t and is forecasted using a predictive conditional distribution P_t. Here, VaR_t^p and ES_t^p represent the true risk measures, i.e. when $r_{w,t} \sim F_t$, while $\widehat{VaR}_t^p$ and $\widehat{ES}_t^p$ represent the estimated risk measures, i.e. when $r_{w,t} \sim P_t$. Portfolio returns are assumed to be independent but not identically distributed. The null hypothesis for the three tests is

$$H_0 : P_t^{[p]} = F_t^{[p]} \quad \text{for all} \quad t \tag{12}$$

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:

$$H_1 : \begin{array}{l} \widehat{ES}_t^p \geq ES_t^p \quad \text{for all} \quad t \quad \text{and} \quad > \quad \text{for some} \quad t \\ \widehat{VaR}_t^p = VaR_t^p \quad \text{for all} \quad t \quad . \end{array}$$

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 again H_1 with the following statistic:

$$Z_1 = \frac{\sum_{t=1}^T \frac{r_{w,t} \cdot I_t}{ES_t^p}}{T_1} + 1.$$

As an average of VaR exceedances, the statistic is 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 (12) is tested against the following alternative:

$$H_1 : \begin{array}{l} \widehat{ES}_t^p \geq ES_t^p \quad \text{for all} \quad t \quad \text{and} \quad > \quad \text{for some} \quad t \\ \widehat{VaR}_t^p \geq VaR_t^p \quad \text{for all} \quad t \quad . \end{array}$$

The test statistic is

$$Z_2 = \frac{1}{T} \sum_{t=1}^T \frac{r_{w,t} \cdot I_t}{p \cdot ES_t^p} + 1.$$

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

The third test evaluates both the frequency and magnitude of VaR violations. The null hypothesis (12) is tested against the same alternative as the second test. The test statistic is

$$Z_{es} = \frac{1}{T} \sum_{t=1}^T \frac{p \cdot (ES_t^p - VaR_t^p) + (r_{w,t} + VaR_t^p) \cdot I_t}{p \cdot ES_t^p}$$

and the expected value of Z_{es} is zero under H_0 and negative under H_1.

The distributions of Z_1, Z_2, and Z_{es} under the null hypothesis are unknown but can be approximated using simulations. For each of our models, the p-values of the three tests are obtained using the following steps:

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

We use $M = 50{,}000$ and simulate the M portfolio returns each week with the predictive conditional distribution P_t used for VaR and ES forecasts.

The last two tests are from Du and Escanciano (2017). The idea is to consider the cumulative violation process,

$$H_t(p) = \frac{1}{p}(p - u_t)\mathbf{1}_{\{u_t \leq p\}} \quad t = 1, \ldots, T,$$

with $u_t = F_t(r_{w,t})$. The name "cumulative violation process" comes from the fact that, as shown in Du and Escanciano (2017), $H_t(p) = \frac{1}{p}\int_0^p I_t(p)dp$, where $I_t(p)$ is the violation process at level p. Using the fact that $\left\{I_t(p) - p\right\}_{t=1}^T$ is a martingale difference sequence under a correct specification of VaR at level p, Du and Escanciano (2017) show that the process $\left\{H_t(p) - p/2\right\}_{t=1}^T$, is a martingale difference sequence under a correct specification of the left p tail. This leads to the following unconditional and conditional null hypotheses:

$$H_{0u} : \mathbb{E}[H_t(p)] = p/2 \quad \text{for all } t,$$

$$H_{0c} : \mathbb{E}[H_t(p) \mid \mathcal{F}_{t-1}] = p/2 \quad \text{for all } t.$$

For the unconditional test of ES at level p, Du and Escanciano (2017) propose the following t-test statistic:

$$DE_u = \frac{\sqrt{T}(\bar{H}(p) - p/2)}{\sqrt{p(1/3 - p/4)}},$$

where $\bar{H} = \frac{1}{T}\sum_{t=1}^T \hat{H}_t(p)$, $\hat{H}_t(p) = \frac{1}{p}(p - \hat{u}_t)\mathbf{1}_{\{\hat{u}_t \leq p\}}$, and $\hat{u}_t = \hat{F}_t(r_{w,t})$. We estimate $\hat{F}_t(r_{w,t})$ at each date t by applying the empirical CDF computed from $M = 50{,}000$ simulations of $r_{w,t}$, as in Step 1, above. Under some regularity conditions, Du and Escanciano (2017) show that the asymptotic distribution of DE_u is $N(0, \sigma^2_{DE_u})$, with the asymptotic variance $\sigma^2_{DE_u}$ depending on the estimation error in $\hat{F}_t$. We estimate $\sigma^2_{DE_u}$ by simulation methods. Specifically, from the simulations of the M portfolio returns at each date t, we are able to compute M statistics DE_u, as in Step 2, above. The estimated variance $\hat{\sigma}^2_{DE_u}$ is then set equal to the sample variance of the M simulated statistics. Once the standardized test statistic $DE_u/\hat{\sigma}_{DE_u}$ is obtained, we perform a bilateral test based on the standard normal distribution.

For the conditional test, Du and Escanciano (2017) propose the following the Box–Pierce test statistic:

$$DE_c = T \sum_{r=1}^{R} \hat{\rho}_r^2,$$

where $\hat{\rho}_r = \frac{\hat{\gamma}_r}{\hat{\gamma}_0}$ and $\hat{\gamma}_r = \frac{1}{T-r} \sum_{t=1+r}^{T} \big(\hat{H}_t(p) - p/2\big)\big(\hat{H}_{t-r}(p) - p/2\big)$. We reject H_{0c} when DE_c is sufficiently large. Du and Escanciano (2017) show that DE_c has a weighted chi-squared asymptotic distribution with weights depending on $\hat{F}_t$. We approximate the distribution of DE_c and compute the p-value of the test using $M = 50{,}000$, as in Steps 1, 2, and 3, above.

References

Acerbi, C., & Szekely, B. (2014). Back-testing expected shortfall. In *Risk, Vol. 7* (pp. 6–81).

Acerbi, C., & Szekely, B. (2017). General properties of backtestable statistics. In *Working paper*, MSCI.

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 (BCBS) (2016). *Minimum capital requirements for market risk* (pp. 1–92). Bank for International Settlements (BIS), publication no. 352, (Jan-2016).

Basel Committee on Banking Supervision (BCBS) (2019). *Minimum capital requirements for market risk* (pp. 1–136). Bank for International Settlements (BIS), publication no. 457.(Jan-2019).

Berkowitz, J., & O'Brien, J. (2002). How accurate are value-at-risk models at commercial banks? *The Journal of Finance*, *57*(3), 1093–1111.

Bollerslev, T., Engle, R. F., & Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, *96*(1), 116–131.

Brooks, C., & Pers, G. (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*, 57–82.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, *39*(4), 841–862.

Christoffersen, P. (2009). Value–at–risk models. In *Handbook of financial time series, Vol. 75* (pp. 3–766).

Christoffersen, P. F. (2012). *Elements of financial risk management*. Academic Press..

Christoffersen, P., & Langlois, H. (2013). The joint dynamics of equity market factors. *Journal of Financial and Quantitative Analysis*, *48*(5), 1371–1404.

De Giorgi, E., Hens, T., & Mayer, J. (2007). Computational aspects of prospect theory with asset pricing applications. *Computational Economics*, *29*, 267–281.

Demarta, S., & McNeil, A. J. (2005). The t copula and related copulas. *International Statistical Review / Revue Internationale De Statistique*, *73*(1), 111–129.

Deng, K., & Qiu, J. (2021). Backtesting expected shortfall and beyond. *Quantitative Finance*, *21*(7), 1109–1125.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

Diks, C., & Fang, H. (2020). Comparing density forecasts in a risk management context. *International Journal of Forecasting*, *36*(2), 531–551.

Du, Z., & Escanciano, J. C. (2017). 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., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, *48*(5), 1749–1778.

Fama, E. F., & French, K. R. (1993). Common risk factors in there turns on stocks and bonds. *Journal of Financial Economics*, *33*, 3–56.

Fissler, T., & Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics*, *44*(4), 1680–1707.

Giacomini, R., & Komunjer, I. (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economic Statistics*, *23*(4), 416–431.

Giacomini, R., & White, H. (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., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, *48*, 65–91.

Kole, E., Markwat, A., Opschoor, A., & van Dijk, D. (2017). Forecasting value-at-risk under temporal and portfolio aggregation. *Journal of Financial Econometrics*, *15*, 649–677.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, *3*(2), 73–84.

McAleer, M., & Da Veiga, B. (2008). Single-index and portfolio models for forecasting value-at-risk thresholds. *Journal of Forecasting*, *27*(3), 217–235.

Nieto, M. R., & Ruiz, E. (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting*, *32*(2), 475–501.

Nolde, N., & Ziegel, J. F. (2017). Elicitability and backtesting: perspectives for banking regulation. *The Annals of Applied Statistics*, *11*(4), 1833–1874.

Patton, A. J., Ziegel, J. F., & Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, *211*(2), 388–413.

Rubinstein, R. Y. (1982). Generating random vectors uniformly distributed inside and on the surface of different regions. *European Journal of Operational Research*, *10*(2), 205–209.

Santos, A. A. P., Nogales, F. J., & Ruiz, E. (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.

Yoshiba, T. (2018). Maximum likelihood estimation of skew-t copulas with its applications to stock returns. *Journal of Statistical Computation and Simulation*, *26*, 1–18.