Diks, Cees; Panchenko, Valentyn; Sokolinskiy, Oleg; van Dijk, Dick

Working Paper

Comparing the Accuracy of Copula-Based Multivariate Density Forecasts in Selected Regions of Support

Tinbergen Institute Discussion Paper, No. 13-061/III

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

TI 2013-061/III
Tinbergen Institute Discussion Paper

Comparing the Accuracy of Copula-Based Multivariate Density Forecasts in Selected Regions of Support

Cees Diks[1,2]
Valentyn Panchenko[3]
Oleg Sokolinskiy[4]
Dick van Dijk[5,2]

[1] *CeNDEF, Faculty of Economics and Business, University of Amsterdam;*
[2] *Tinbergen Institute;*
[3] *University of New South Wales;*
[4] *Rutgers Business School;*
[5] *Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Comparing the Accuracy of Copula-Based Multivariate Density Forecasts in Selected Regions of Support

Cees Diks*
Tinbergen Institute and
CeNDEF, Amsterdam School of Economics
University of Amsterdam

Valentyn Panchenko†
School of Economics
University of New South Wales

Oleg Sokolinskiy‡
Rutgers Business School
Rutgers - The State University of New Jersey

Dick van Dijk§
Tinbergen Institute and
Econometric Institute
Erasmus University Rotterdam

April 18, 2013

Abstract

This paper develops a testing framework for comparing the predictive accuracy of copula-based multivariate density forecasts, focusing on a specific part of the joint distribution. The test is framed in the context of the Kullback-Leibler Information Criterion, but using (out-of-sample) conditional likelihood and censored likelihood in order to focus the evaluation on the region of interest. Monte Carlo simulations document that the resulting test statistics have satisfactory size and power properties in small samples. In an empirical application to daily exchange rate returns we find evidence that the dependence structure varies with the sign and magnitude of returns, such that different parametric copula models achieve superior forecasting performance in different regions of the support. Our analysis highlights the importance of allowing for lower and upper tail dependence for accurate forecasting of common extreme appreciation and depreciation of different currencies.

Keywords: Copula-based density forecast; Kullback-Leibler Information Criterion; out-of-sample forecast evaluation.

*Corresponding author: Center for Nonlinear Dynamics in Economics and Finance, Faculty of Economics and Business, University of Amsterdam, Roetersstraat 11, NL-1018 WB Amsterdam, The Netherlands. E-mail: c.g.h.diks@uva.nl

†School of Economics, Faculty of Business, University of New South Wales, Sydney, NSW 2052, Australia. E-mail: v.panchenko@unsw.edu.au

‡Rutgers Business School - Newark & New Brunswick, Rutgers - The State University of New Jersey, 94 Rockafeller Rd., Piscataway, New Jersey, 08901, USA. E-mail: oleg.sokolinskiy@business.rutgers.edu

§Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands. E-mail: djvandijk@ese.eur.nl

1 Introduction

The dependence between asset returns typically is nonlinear and time-varying. Traditionally, efforts to accommodate these features have focused on modeling the dynamics of conditional variances and covariances by means of multivariate GARCH and stochastic volatility (SV) models; see the surveys by Silvennoinen and Teräsvirta (2009) and Chib *et al.* (2009), respectively. Recently, copulas have become an increasingly popular tool for modeling multivariate distributions in finance (Patton, 2009; Genest *et al.*, 2009). The copula approach provides more flexibility than multivariate GARCH and SV models in terms of the type of asymmetric dependence that can be captured. In addition, an attractive property of copulas is that they allow for modeling the marginal distributions and the dependence structure of the asset returns separately.

Many parametric copula families are available, with rather different dependence properties. An important issue in empirical applications therefore is the choice of an appropriate copula specification. In practice, most often this is done by comparing alternative specifications indirectly, subjecting each of them to a battery of goodness-of-fit tests, see Berg (2009) for a detailed review. A direct comparison of alternative copulas from different parametric families has been considered by Chen and Fan (2006) and Patton (2006), adopting the approach based on pseudo likelihood ratio (PLR) tests for model selection originally developed by Vuong (1989) and Rivers and Vuong (2002). These tests compare the candidate copula specifications in terms of their Kullback-Leibler Information Criterion (KLIC), which measures the distance from the true (but unknown) copula. Similar to the goodness-of-fit tests, these PLR tests are based on the in-sample fit of the competing copulas. Diks *et al.* (2010) approach the copula selection problem from an out-of-sample forecasting perspective. Specifically, the PLR testing approach is extended to compare the predictive accuracy of alternative copula specifications, by using out-of-sample log-likelihood values corresponding with copula density forecasts. An important motivation for considering the (relative) predictive accuracy of copulas is that multivariate density forecasting is one of the main purposes in empirical applications.

Comparison of out-of-sample KLIC values for assessing relative predictive accuracy

has recently become popular for the evaluation of univariate density forecasts, see Mitchell and Hall (2005), Amisano and Giacomini (2007) and Bao *et al.* (2007). Amisano and Giacomini (2007) provide an interesting interpretation of the KLIC-based comparison in terms of scoring rules, which are loss functions depending on the density forecast and the actually observed data. In particular, the difference between the log-likelihood scoring rule for two competing density forecasts corresponds exactly to their relative KLIC values. The same interpretation continues to hold for copula-based multivariate density forecasts considered in this paper.

In many applications of density forecasts, we are mostly interested in a particular region of the density. Financial risk management is an example in case. Due to the regulations of the Basel accords, among others, the main concern for banks and other financial institutions is an accurate description of the left tail of the distribution of their portfolio's returns, in order to obtain accurate estimates of Value-at-Risk and related measures of downside risk. Correspondingly, Bao *et al.* (2004), Amisano and Giacomini (2007) and Diks *et al.* (2011) consider the problem of evaluating and comparing univariate density forecasts in a specific region of interest. Diks *et al.* (2011) demonstrate that the approach based on out-of-sample KLIC values can be adapted to this case, by replacing the full likelihood by the conditional likelihood, given that the actual observation lies in the region of interest, or by the censored likelihood, with censoring of the observations outside the region of interest.

In this paper we develop tests of equal predictive accuracy of different copula-based multivariate density forecasts in a specific region of the support. For this purpose, we combine the testing framework for comparing univariate forecasts in specific regions developed by Diks *et al.* (2011), with the logarithmic score decomposition for copula models considered in Diks *et al.* (2010). The resulting test of equal predictive accuracy can be applied to fully parametric, semi-parametric and nonparametric copula-based multivariate density models. The test is valid under general conditions on the competing copulas, which is achieved by adopting the framework of Giacomini and White (2006). This assumes that any unknown model parameters are estimated using a moving window of fixed size. The finite estimation window essentially allows us to treat competing density fore-

casts based on different copula specifications, including the time-varying estimated model parameters, as two competing *forecast methods*. Comparing scores for forecast *methods* rather than for *models* simplifies the resulting test procedures considerably, because parameter estimation uncertainty does not play a role (it simply is part of the respective competing forecast methods). In addition, the asymptotic distribution of our test statistic in this case does not depend on whether or not the competing copulas belong to nested families.

We examine the size and power properties of our copula predictive accuracy test via Monte Carlo simulations. Here we adopt fully parametric and semi-parametric copula-based multivariate dynamic models. The latter class of models (shortened as SCOMDY) developed by Chen and Fan (2005, 2006) combines parametric specifications for the conditional mean and conditional variance with a semi-parametric specification for the distribution of the (standardized) innovations, consisting of a parametric copula with non-parametric univariate marginal distributions. Our simulation results demonstrate that the predictive accuracy tests have satisfactory size and power properties in realistic sample sizes.

We consider an empirical application to daily exchange rate returns of the Canadian dollar, Swiss franc, euro, British pound, and Japanese yen against the US dollar over the period from 1992 until 2008. Based on the relative predictive accuracy of one-step-ahead density forecasts we find that different parametric copula specifications achieve superior forecasting performance in different regions of the support. Our analysis highlights the importance of accommodating positive upper (lower) tail dependence for accurate forecasting of common extreme appreciation (depreciation) of different currencies.

The paper is organised as follows. In Section 2 we briefly review copula-based multivariate density models and develop our predictive accuracy test for copulas based on out-of-sample log-likelihood scores. In Section 3 we investigate its size and power properties by means of Monte Carlo simulations. In Section 4 we illustrate our test with an application to daily exchange rate returns for several major currencies. We conclude in Section 5.

2 Methodology

The aim of this paper is to extend the copula-based density forecast evaluation tests developed by Diks *et al.* (2010), by enabling them to focus on specific regions of the copula domain. This is achieved by comparing two density forecasts which differ only in their predictive copulas, but instead of unweighted likelihood scores we take the weighted likelihood-based scores introduced in Diks *et al.* (2011) as the driving scoring rule, using weight functions defined on the copula domain.

2.1 Review of density forecast evaluation using weighted scoring rules

This subsection briefly reviews the work of Diks *et al.* (2011) on weighted likelihood-based scoring rules for density forecast evaluation. We use a more general notation, emphasising that all results extend to forecasts of a vector-valued variable $\boldsymbol{Y}_{t+1}$.

Density forecast evaluation Consider a stochastic process $\{\boldsymbol{Z}_t : \Omega \to \mathbb{R}^{k+d}\}_{t=1}^{T}$, defined on a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and identify $\boldsymbol{Z}_t$ with $(\boldsymbol{Y}_t, \boldsymbol{X}_t')'$, where $\boldsymbol{Y}_t : \Omega \to \mathbb{R}^d$ is the real-valued d-dimensional random variable of interest and $\boldsymbol{X}_t : \Omega \to \mathbb{R}^k$ is a vector of exogenous or pre-determined variables. The information set at time t is defined as $\mathcal{F}_t = \sigma(\boldsymbol{Z}_1', \ldots, \boldsymbol{Z}_t')'$. We consider the case where two competing forecast methods are available, each producing one-step ahead density forecasts, i.e. predictive densities of $\boldsymbol{Y}_{t+1}$, based on $\mathcal{F}_t$.

As in Amisano and Giacomini (2007), by 'forecast method' we mean a given density forecast in terms of past information, resulting from the choices that the forecaster makes at the time of the prediction. These include the variables $\boldsymbol{X}_t$, the econometric model (if any), and the estimation method. The only requirement that we impose on the forecast methods is that the density forecasts depend on a finite number R of most recent observations $\boldsymbol{Z}_{t-R+1}, \ldots, \boldsymbol{Z}_t$. Forecast methods of this type arise naturally, for instance, when density forecasts are obtained from time series models, for which parameters are estimated with a moving window of R observations. The advantage of comparing forecast methods rather than forecast models is that this allows for treating parameter estimation uncertainty as an integral part of the forecast methods. The use of a finite (rolling) window of R past

observations for parameter estimation considerably simplifies the asymptotic theory of tests of equal predictive accuracy, as argued by Giacomini and White (2006). It also turns out to be more convenient in that it enables comparison of density forecasts based on both nested and non-nested models, in contrast to other approaches such as that of West (1996).

Scoring rules One of the approaches that has been put forward for density forecast evaluation in general is by means of scoring rules, which are commonly used in probability forecast evaluation, see Diebold and Lopez (1996). A scoring rule is a loss function $S^*(\hat{f}_t; \boldsymbol{y}_{t+1})$ depending on the density forecast and the actually observed value $\boldsymbol{y}_{t+1}$, such that a density forecast that is 'better' receives a higher score. Note that, as argued by Diebold *et al.* (1998) and Granger and Pesaran (2000), any rational user would prefer the true conditional density p_t of $\boldsymbol{Y}_{t+1}$ over an incorrect density forecast. This suggests that it is natural to focus on scoring rules for which incorrect density forecasts $\hat{f}_t$ do not receive a higher average score than the true conditional density p_t, that is,

$$\mathsf{E}_t\left(S^*(\hat{f}_t; \boldsymbol{Y}_{t+1})\right) \leq \mathsf{E}_t\left(S^*(p_t; \boldsymbol{Y}_{t+1})\right), \qquad \text{for all } t.$$

Following Gneiting and Raftery (2007), a scoring rule satisfying this condition will be called *proper*.

It is useful to note that the correct density p_t does not depend on estimated parameters, while density forecasts typically do. This implies that even if the density forecast $\hat{f}_t$ is based on a correctly specified model, but the model includes estimated parameters, the average score $\mathsf{E}_t\left(S^*(\hat{f}_t; \boldsymbol{Y}_{t+1})\right)$ may not achieve the upper bound $\mathsf{E}_t\left(S^*(p_t; \boldsymbol{Y}_{t+1})\right)$ due to non-vanishing estimation uncertainty. As a consequence, a density forecast based on a misspecified model with limited estimation uncertainty may be preferred over a density forecast based on the correct model specification but having larger estimation uncertainty.

Null hypothesis and testing approach Given a scoring rule of one's choice, there are various ways to construct tests of equal predictive ability. Giacomini and White (2006) distinguish tests of unconditional predictive ability and conditional predictive ability. In the present paper, we focus on tests for unconditional predictive ability for clarity of ex-

position. The suggested approach can be extended to obtain tests of conditional predictive ability in a straightforward manner.

Assume that two competing density forecasts $\hat{f}_{A,t}$ and $\hat{f}_{B,t}$ and corresponding realisations of the variable $\boldsymbol{Y}_{t+1}$ are available for $t = R, R+1, \ldots, T-1$. We may then compare $\hat{f}_{A,t}$ and $\hat{f}_{B,t}$ based on their average scores, by testing formally whether their difference is statistically significantly different from zero on average. Defining the score difference

$$d^*_{t+1} = S^*(\hat{f}_{A,t}; \boldsymbol{y}_{t+1}) - S^*(\hat{f}_{B,t}; \boldsymbol{y}_{t+1}),$$

for a given scoring rule S^*, the null hypothesis of equal scores is given by

$$H_0: \quad \mathsf{E}(d^*_{t+1}) = 0, \qquad \text{for all } t = R, R+1, \ldots, T-1.$$

Let $\overline{d}^*_{R,P}$ denote the sample average of the score differences, that is, $\overline{d}^*_{R,P} = P^{-1} \sum_{t=R}^{T-1} d^*_{t+1}$ with $P = T - R$. To test the null, we may use a Diebold and Mariano (1995) type statistic

$$t_{R,P} = \frac{\overline{d}^*_{R,P}}{\sqrt{\hat{\sigma}^2_{R,P}/P}}, \tag{1}$$

where $\hat{\sigma}^2_{R,P}$ is a heteroskedasticity and autocorrelation-consistent (HAC) variance estimator of $\sigma^2_{R,P} = \mathsf{Var}\left(\sqrt{P}\,\overline{d}^*_{R,P}\right)$. The following theorem characterises the asymptotic distribution of the test statistic under the null hypothesis.

Theorem 1 *The statistic $t_{R,P}$ in* (1) *is asymptotically (as $P \to \infty$ with R fixed) standard normally distributed under the null hypothesis if: (i) $\{\boldsymbol{Z}_t\}$ is ϕ-mixing of size $-q/(2q-2)$ with $q \geq 2$, or α-mixing of size $-q/(q-2)$ with $q > 2$; (ii) $\mathsf{E}|d^*_{t+1}|^{2q} < \infty$ for all t; and (iii) $\sigma^2_{R,P} = \mathsf{Var}\left(\sqrt{P}\,\overline{d}^*_{R,P}\right) > 0$ for all P sufficiently large.*

Proof: This is the main part of Theorem 4 of Giacomini and White (2006), where a proof is provided. □

The logarithmic scoring rule Mitchell and Hall (2005), Amisano and Giacomini (2007), and Bao *et al.* (2004, 2007) focus on the logarithmic scoring rule

$$S^l(\hat{f}_t; \boldsymbol{y}_{t+1}) = \log \hat{f}_t(\boldsymbol{y}_{t+1}), \tag{2}$$

such that the score assigned to a density forecast varies positively with the value of $\hat{f}_t$ evaluated at the observation $\boldsymbol{y}_{t+1}$. Based on the P observations available for evaluation, $\boldsymbol{y}_{R+1}, \ldots, \boldsymbol{y}_T$, the density forecasts $\hat{f}_{A,t}$ and $\hat{f}_{B,t}$ can be ranked according to their average scores $P^{-1}\sum_{t=R}^{T-1}\log\hat{f}_{A,t}(\boldsymbol{y}_{t+1})$ and $P^{-1}\sum_{t=R}^{T-1}\log\hat{f}_{B,t}(\boldsymbol{y}_{t+1})$. Obviously, the density forecast yielding the highest average score would be the preferred one. The log score differences $d_{t+1}^{l} = \log\hat{f}_{A,t}(\boldsymbol{y}_{t+1}) - \log\hat{f}_{B,t}(\boldsymbol{y}_{t+1})$ may be used to test whether the predictive accuracy is significantly different, using the test statistic defined in (1). Note that this coincides with the log-likelihood ratio of the two competing density forecasts.

Weighted likelihood-based scoring rules Diks *et al.* (2011) adapt the logarithmic scoring rule for evaluating and comparing density forecasts in a specific region of interest, $M_t \subset \mathbb{R}^d$, say. As argued by Diks *et al.* (2011), this cannot be achieved by using the weighted logarithmic score $\mathrm{I}(\boldsymbol{y}_{t+1} \in M_t)\log\hat{f}_t(\boldsymbol{y}_{t+1})$, as by construction the resulting test statistic would be biased towards (possible incorrect) density forecasts with more probability mass in the region of interest. Replacing the full likelihood in (2) either by the conditional likelihood, given that the observation lies in the region of interest, or by the censored likelihood, with censoring of the observations outside M_t, does lead to scoring rules which do not suffer from this problem and remain proper. The conditional likelihood (*cl*) score function is given by

$$S^{cl}(\hat{f}_t; \boldsymbol{y}_{t+1}) = \mathrm{I}(\boldsymbol{y}_{t+1} \in M_t)\log\left(\frac{\hat{f}_t(\boldsymbol{y}_{t+1})}{\int_{M_t}\hat{f}_t(\boldsymbol{y})\mathrm{d}\boldsymbol{y}}\right), \tag{3}$$

while the censored likelihood (*csl*) score function is given by

$$S^{csl}(\hat{f}_t; \boldsymbol{y}_{t+1}) = \mathrm{I}(\boldsymbol{y}_{t+1} \in M_t)\log\hat{f}_t(\boldsymbol{y}_{t+1}) + \mathrm{I}(\boldsymbol{y}_{t+1} \in M_t^c)\log\left(\int_{M_t^c}\hat{f}_t(\boldsymbol{y})\mathrm{d}\boldsymbol{y}\right), \tag{4}$$

where M_t^c is the complement of the region of interest M_t. Note that the *cl* scoring rule does not take into account the accuracy of the density forecast for the total probability of $\boldsymbol{Y}_{t+1}$ falling into the region of interest, while the *csl* scoring rule does.

The conditional and censored likelihood scoring rules focus on a sharply defined region of interest M_t. It is possible to extend this idea by using a more general weight function $w_t(\boldsymbol{y}_{t+1})$, where the scoring rules in (3) and (4) can be recovered for the specific

choice $w_t(\boldsymbol{y}_{t+1}) = \mathrm{I}(\boldsymbol{y}_{t+1} \in M_t)$:

$$S^{cl}(\hat{f}_t; \boldsymbol{y}_{t+1}) = w_t(\boldsymbol{y}_{t+1}) \log \left(\frac{\hat{f}_t(\boldsymbol{y}_{t+1})}{\int w_t(\boldsymbol{y}) \hat{f}_t(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}} \right), \tag{5}$$

and

$$S^{csl}(\hat{f}_t; \boldsymbol{y}_{t+1}) = w_t(\boldsymbol{y}_{t+1}) \log \hat{f}_t(\boldsymbol{y}_{t+1}) + (1 - w_t(\boldsymbol{y}_{t+1})) \log \left(1 - \int w_t(\boldsymbol{y}) \hat{f}_t(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} \right). \tag{6}$$

At this point, we make the following assumptions concerning the density forecasts that are to be compared, and the weight function.

Assumption 1 *The density forecasts $\hat{f}_{A,t}$ and $\hat{f}_{B,t}$ satisfy KLIC$(\hat{f}_{A,t}) < \infty$ and KLIC$(\hat{f}_{B,t}) < \infty$, where KLIC$(h_t) = \int p_t(y) \log \left(p_t(\boldsymbol{y}) / h_t(\boldsymbol{y}) \right) \mathrm{d}\boldsymbol{y}$ is the Kullback-Leibler divergence between the density forecast h_t and the true conditional density p_t.*

Assumption 2 *The weight function $w_t(\boldsymbol{y})$ is such that (a) it is determined by the information available at time t, and hence a function of $\mathcal{F}_t$, (b) $0 \leq w_t(\boldsymbol{y}) \leq 1$, and (c) $\int w_t(\boldsymbol{y}) p_t(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} > 0$.*

Assumption 1 ensures that the expected score differences for the competing density forecasts are finite. Assumption 2 (c) is needed to avoid cases where $w_t(\boldsymbol{y})$ takes strictly positive values only outside the support of the data.

The following lemma states that the generalised *cl* and *csl* scoring rules in (5) and (6) are proper, and hence cannot lead to spurious rejections against wrong alternatives just because these have more probability mass in the region(s) of interest.

Lemma 1 *Under Assumptions 1 and 2, the generalised conditional likelihood scoring rule given in* (5) *and the generalised censored likelihood scoring rule given in* (6) *are proper.*

Proof: A proof for univariate predictive densities has been given in the appendix of Diks *et al.* (2011). All steps in that proof remain valid if the univariate density is replaced by a multivariate density, and the scalar integration variable by a vector-valued integration variable.

We may test the null hypothesis of equal performance of two density forecasts $\hat{f}_{A,t}(\boldsymbol{y}_{t+1})$ and $\hat{f}_{B,t}(\boldsymbol{y}_{t+1})$ based on the conditional likelihood score (5) or the censored likelihood score (6) in the same manner as before. That is, given a sample of density forecasts and corresponding realisations for P time periods $t = R, R+1, \ldots, T-1$, we may form the relative scores $d^{cl}_{t+1} = S^{cl}(\hat{f}_{A,t}; \boldsymbol{y}_{t+1}) - S^{cl}(\hat{f}_{B,t}; \boldsymbol{y}_{t+1})$ and $d^{csl}_{t+1} = S^{csl}(\hat{f}_{A,t}; \boldsymbol{y}_{t+1}) - S^{csl}(\hat{f}_{B,t}; \boldsymbol{y}_{t+1})$ and use these for computing the Diebold-Mariano type test statistics given in (1).

2.2 Copula comparison with weights on the copula domain

Patton's (2006) extension of Sklar's (1959) theorem to the time-series case describes how the time-dependent multivariate distribution $F_t(\boldsymbol{y}_{t+1})$ can be decomposed into conditional marginal distributions $F_{j,t}(y_j)$, $j = 1, \ldots, d$, and a conditional copula $C_t(\cdot)$, that is

$$F_t(\boldsymbol{y}) = C_t(F_{1,t}(y_1), F_{2,t}(y_2), \ldots, F_{d,t}(y_d)), \tag{7}$$

provided that the marginal conditional CDFs $F_{i,t}$ are continuous. This decomposition clearly shows the attractiveness of the copula approach for modeling multivariate distributions. Given that the marginal distributions $F_{j,t}$, $j = 1 \ldots, d$, only contain univariate information on the individual variables $Y_{j,t+1}$, their dependence is governed completely by the copula function C_t. As the choice of marginal distributions does not restrict the choice of copula, or vice versa, a wide range of joint distributions can be obtained by combining different marginals with different copulas.

The one-step-ahead predictive log-likelihood associated with $\boldsymbol{y}_{t+1}$ is seen to be given by

$$\sum_{j=1}^{d} \log f_{j,t}(y_{j,t+1}) + \log c_t(F_{1,t}(y_{1,t+1}), F_{2,t}(y_{2,t+1}), \ldots, F_{d,t}(y_{d,t+1})), \tag{8}$$

where $f_{j,t}(y_{j,t+1})$, $j = 1, \ldots, d$, are the conditional marginal densities and c_t is the conditional copula density, defined as

$$c_t(u_1, u_2, \ldots, u_d) = \frac{\partial^d}{\partial u_1 \partial u_2 \ldots \partial u_d} C_t(u_1, u_2, \ldots, u_d),$$

which we will assume to exist throughout. Using (8), the conditional likelihood and censored likelihood scoring rule of a density forecast $\hat{f}_t(\boldsymbol{y}_{t+1})$ with marginal predictive den-

sities $\hat{f}_{j,t}$, $j = 1, \ldots, d$, and copula density $\hat{c}_t$ can be decomposed as

$$S_{t+1}^{cl} = w_t(\boldsymbol{y}_{t+1}) \left(\sum_{j=1}^{d} \log \hat{f}_{j,t}(y_{j,t+1}) + \log \hat{c}_t(\hat{\boldsymbol{u}}_{t+1}) \right) - w_t(\boldsymbol{y}_{t+1}) \log \left(\int w_t(\boldsymbol{y}) \hat{f}_t(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} \right), \tag{9}$$

and

$$\begin{aligned} S_{t+1}^{csl} \quad = \quad & w_t(\boldsymbol{y}_{t+1}) \left(\textstyle\sum_{j=1}^{d} \log \hat{f}_{j,t}(y_{j,t+1}) + \log \hat{c}_t(\hat{\boldsymbol{u}}_{t+1}) \right) \\ & + (1 - w_t(\boldsymbol{y}_{t+1})) \log \left(1 - \textstyle\int w_t(\boldsymbol{y}) \hat{f}_t(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} \right), \end{aligned} \tag{10}$$

where $\hat{c}_t$ is the conditional copula density associated with the density forecast, and $\hat{\boldsymbol{u}}_{t+1} = (\hat{F}_{1,t}(y_{1,t+1}), \ldots, \hat{F}_{d,t}(y_{d,t+1}))'$ its multivariate conditional probability integral transform (PIT).

As in Diks *et al.* (2010) we assume that the two competing multivariate density forecasts differ only in their copula specifications and have identical predictive marginal densities $\hat{f}_{j,t}$, $j = 1, \ldots, d$. The two competing copula specifications are assumed to have well-defined densities $\hat{c}_{A,t}$ and $\hat{c}_{B,t}$. The null hypothesis of equal predictive ability is

$$H_0 : \quad \mathbb{E}(S_{A,t+1}^{*}) = \mathbb{E}(S_{B,t+1}^{*}),$$

where '$*$' stands for either 'cl' or 'csl'. Since the conditional marginals are identically specified under both density forecasts, the logarithms of the marginal densities in (9) and (10) cancel out, so that an equivalent formulation of the null hypothesis is

$$H_0 : \quad \mathbb{E}(\mathcal{S}_{A,t+1}^{*}) = \mathbb{E}(\mathcal{S}_{B,t+1}^{*}),$$

where, with a similar abuse of notation as above (leaving out the subscripts A and B),

$$\mathcal{S}_{t+1}^{cl}(\boldsymbol{y}_{t+1}) = w_t(\boldsymbol{y}_{t+1}) \log \left(\hat{c}(\hat{u}_{t+1}) \right) - \log \int w_t(\boldsymbol{y}) \hat{f}_t(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}$$

and

$$\mathcal{S}_{t+1}^{csl}(\boldsymbol{y}_{t+1}) = w_t(\boldsymbol{y}_{t+1}) \log \hat{c}_t(\hat{\boldsymbol{u}}_{t+1}) + (1 - w_t(\boldsymbol{y}_{t+1})) \log \left(1 - \int w_t(\boldsymbol{y}) \hat{f}_t(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} \right).$$

We use the weight function to focus on specific regions of the copula. This can be achieved by taking weight functions of the form

$$w_t(\boldsymbol{y}_{t+1}) = \tilde{w}(\hat{u}_{1,t+1}(y_{1,t+1}), \ldots, \hat{u}_{d,t+1}(y_{d,t+1})),$$

where $\tilde{w}(u_1, \ldots, u_d)$ is a weight function defined on the copula support. Note that

$$\begin{aligned} \int w_t(\boldsymbol{y}) \hat{f}_t(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} &= \int \tilde{w}_t(\hat{u}_{1,t+1}(y_1), \ldots, \hat{u}_{d,t+1}(y_d)) \hat{f}_t(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} \\ &= \int \tilde{w}_t(\hat{u}_{1,t+1}(y_1), \ldots, \hat{u}_{d,t+1}(y_d)) \, \mathrm{d}\hat{F}_t(\boldsymbol{y}) \\ &= \int \tilde{w}_t(\hat{u}_{1,t+1}, \ldots, \hat{u}_{d,t+1}) \, \mathrm{d}\hat{C}_t(\boldsymbol{u}_{t+1}) \, \mathrm{d}\boldsymbol{u}_{t+1} \\ &= \int \tilde{w}_t(\boldsymbol{u}) \hat{c}_t(\boldsymbol{u}) \, \mathrm{d}\boldsymbol{u}. \end{aligned}$$

This allows us to rewrite the scores $\mathcal{S}^*_{t+1}$ as

$$\mathcal{S}^{cl}_{t+1}(\boldsymbol{y}_{t+1}) = \tilde{w}_t(\hat{\boldsymbol{u}}_{t+1}) \left(\log \hat{c}_t(\hat{\boldsymbol{u}}_{t+1}) - \log \int \tilde{w}_t(\boldsymbol{u}) \hat{c}_t(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} \right) \tag{11}$$

and

$$\mathcal{S}^{csl}_{t+1}(\boldsymbol{y}_{t+1}) = \tilde{w}_t(\hat{\boldsymbol{u}}_{t+1}) \left(\log \hat{c}_t(\hat{\boldsymbol{u}}_{t+1}) \right) + (1 - \tilde{w}_t(\hat{\boldsymbol{u}}_{t+1})) \log \left(1 - \int \tilde{w}_t(\boldsymbol{u}) \hat{c}_t(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} \right). \tag{12}$$

Note that these 'reduced' scoring rules take the same form as the weighted likelihood-based scoring rules (5) and (6) derived before, but now involve only the density forecast copula instead of the full density forecast and the observed conditional PITs $\hat{\boldsymbol{u}}_{t+1}$ instead of the variable $\boldsymbol{y}_{t+1}$.

The weight function $\tilde{w}_t(\boldsymbol{u})$ can be chosen directly in the copula support. In the cases considered in this paper, $\tilde{w}_t$ will be time independent, and will take the form of an indicator function of a given fixed subset of the copula support. In some cases this allows for a simplification of the scoring rules. For instance, for $\tilde{w}_t(\boldsymbol{u}) = I(u_1 \leq a, \ldots u_d \leq a)$, it follows that $\int \tilde{w}_t(\boldsymbol{u}) \hat{c}_t(\boldsymbol{u}) \, \mathrm{d}\boldsymbol{u} = \hat{C}_t(a, \ldots, a)$, so that the reduced scoring rules take the form

$$\mathcal{S}^{cl}(\boldsymbol{y}_{t+1}) = I(u_{1,t+1} \leq a, \ldots u_{d,t+1} \leq a) \left(\log \hat{c}_t(\hat{\boldsymbol{u}}_{t+1}) - \log \hat{C}_t(a, \ldots, a) \right)$$

and

$$\begin{aligned} \mathcal{S}^{csl}(\boldsymbol{y}_{t+1}) &= I(u_{1,t+1} \leq a, \ldots u_{d,t+1} \leq a) \left(\log \hat{c}_t(\hat{\boldsymbol{u}}_{t+1}) \right) \\ &\quad + (1 - I(u_{1,t+1} \leq a, \ldots u_{d,t+1} \leq a)) \log \left(1 - \hat{C}_t(a, \ldots, a) \right). \end{aligned}$$

Again, the Diebold-Mariano type test statistics as given in (1) may be adopted to test the null hypothesis of equal predictive accuracy of two copula-based density forecasts $\hat{c}_{A,t}(\hat{\boldsymbol{u}}_{t+1})$ and $\hat{c}_{B,t}(\hat{\boldsymbol{u}}_{t+1})$ based on the conditional likelihood score (11) or the censored

likelihood score (12) in the same manner as before. Note that by choosing the weight function $\tilde{w}_t(\boldsymbol{u}) = 1$ we retrieve the original log scoring rule

$$\mathcal{S}^l(\boldsymbol{y}_{t+1}) = \log \hat{c}_t(\hat{\boldsymbol{u}}_{t+1}), \tag{13}$$

as considered in Diks *et al.* (2010). The Diebold-Mariano type test statistic in (1) with this log-scoring rule can be used to compare two copula-based density forecasts over the entire copula support.

Above, we have assumed that the two competing multivariate density forecasts differ only in their copula specifications and have identical predictive marginal densities $\hat{f}_{j,t}$, $j = 1, \ldots, d$. Implicitly this assumes that the parameters in the marginals and the copula can be separated from each other, so that they can be estimated in a multi-stage procedure. No other restrictions are put on the marginals. In particular, they may be specified parametrically, nonparametrically, or semi-parametrically. An important class of models that satisfies these properties is that of SCOMDY models, discussed next.

SCOMDY models The class of semi-parametric copula-based multivariate dynamic (SCOMDY) models has been introduced by Chen and Fan (2006). We discuss this class in some detail here, as we use it in the Monte Carlo simulations and the empirical application in subsequent sections. The SCOMDY models combine parametric specifications for the conditional mean and conditional variance of $\boldsymbol{Y}_t$ with a semi-parametric specification for the distribution of the standardised innovations, consisting of a parametric copula with nonparametric univariate marginal distributions. The general SCOMDY model is specified as

$$\boldsymbol{Y}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}_1) + \sqrt{H_t(\boldsymbol{\theta})}\boldsymbol{\varepsilon}_t, \tag{14}$$

where

$$\boldsymbol{\mu}_t(\boldsymbol{\theta}_1) = (\mu_{1,t}(\boldsymbol{\theta}_1), \ldots, \mu_{d,t}(\boldsymbol{\theta}_1))' = \mathbb{E}\left[\boldsymbol{Y}_t | \mathcal{F}_{t-1}\right]$$

is a specification of the conditional mean, parameterised by a finite dimensional vector of parameters $\boldsymbol{\theta}_1$, and

$$H_t(\boldsymbol{\theta}) = \text{diag}(h_{1,t}(\boldsymbol{\theta}), \ldots, h_{d,t}(\boldsymbol{\theta})),$$

where

$$h_{j,t}(\boldsymbol{\theta}) = h_{j,t}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbb{E}\left[(Y_{j,t} - \mu_{j,t}(\boldsymbol{\theta}_1))^2 | \mathcal{F}_{t-1}\right], \quad j = 1, \ldots, d,$$

is the conditional variance of $Y_{j,t}$ given $\mathcal{F}_{t-1}$, parameterised by a finite dimensional vector of parameters $\boldsymbol{\theta}_2$, where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ do not have common elements. The innovations $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{d,t})'$ are independent of $\mathcal{F}_{t-1}$ and independent and identically distributed (i.i.d.) with $\mathbb{E}(\varepsilon_{j,t}) = 0$ and $\mathbb{E}(\varepsilon_{j,t}^2) = 1$ for $j = 1, \ldots, d$. Applying Sklar's theorem, the joint distribution function $F(\boldsymbol{\varepsilon})$ of $\boldsymbol{\varepsilon}_t$ can be written as

$$F(\boldsymbol{\varepsilon}) = C(F_1(\varepsilon_1), \ldots, F_d(\varepsilon_d); \boldsymbol{\alpha}) \equiv C(u_1, \ldots, u_d; \boldsymbol{\alpha}), \tag{15}$$

where $C(u_1, \ldots, u_d; \boldsymbol{\alpha})$: $[0,1]^d \to [0,1]$ is a member of a parametric family of copula functions with finite dimensional parameter vector $\boldsymbol{\alpha}$.

An important characteristic of SCOMDY models is that the univariate marginal densities $F_j(\cdot)$, $j = 1, \ldots, d$ are not specified parametrically (up to an unknown parameter vector) but are estimated nonparametrically. Specifically, Chen and Fan (2006) suggest the following three-stage procedure to estimate the SCOMDY model parameters. First, univariate quasi maximum likelihood under the assumptions of normality of the standardised innovations $\varepsilon_{j,t}$ is used to estimate the parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Second, estimates of the marginal distributions $F_j(\cdot)$ are obtained by means of the empirical CDF transformation of the residuals $\hat{\varepsilon}_{j,t} \equiv (y_{j,t} - \mu_{j,t}(\hat{\boldsymbol{\theta}}_1))/\sqrt{h_{j,t}(\hat{\boldsymbol{\theta}})}$. Finally, the parameters of a given copula specification are estimated by maximising the corresponding copula log-likelihood function.

3 Monte Carlo simulations

In this section we use Monte Carlo simulation to examine the finite-sample behaviour of our predictive accuracy test for comparing alternative copula specifications in specific regions of interest. We consider three classes of models: (1) copula model with no uncertainty in the marginal distribution (marginals are not modeled), (2) copula model with marginals specified parametrically and (3) copula model with marginals specified semi-parametrically (SCOMDY model). In all three cases copula is modeled parametrically.

Although the case without uncertainty in the marginals is of less practical relevance, it is included as a baseline.

For the model involving marginals, the true data generating process (DGP) is based on an AR(1) specification for the conditional mean and a GARCH(1,1) specification for the conditional variance, with coefficients that are typical for financial applications to exchange rates. In particular,

$$Y_{j,t} = 0.1 Y_{j,t-1} + \sqrt{h_{j,t}}\boldsymbol{\varepsilon}_{j,t} \tag{16}$$

$$h_{j,t} = 0.1 + 0.05\left(Y_{j,t-1} - 0.1\, Y_{j,t-2}\right)^2 + 0.85 h_{j,t-1}, \tag{17}$$

for $j = 1, \ldots, d$. The vector-valued innovations $\boldsymbol{\varepsilon_t}$ used in the simulations are i.i.d., but the elements $\varepsilon_{j,t}$, $j \in 1, \ldots, d$ for fixed t are not independent. For fixed t, each of the variables $\varepsilon_{j,t}$ are marginally standard normally distributed, while their dependence is either described by the Gaussian copula, the Student-t copula, the Clayton copula or the Clayton survival copula.

The Gaussian and Student-t copulas can be obtained using the so-called inversion method, that is

$$C^{\text{Ga}}(u_1, u_2, \ldots, u_d) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_d^{-1}(u_d)), \tag{18}$$

where F is the joint CDF and $F_i^{-1}(u) = \min\{x | u \leq F_i(x)\}$ is the (quasi)-inverse of the corresponding marginal CDF F_i.

The Gaussian copula is obtained from (18) by taking F to be the multivariate normal distribution with mean zero, unit variances, and correlations ρ_{ij}, $i, j = 1, \ldots, d$, and standard normal marginals F_i. The corresponding copula density is given by

$$c^{\text{Ga}}(\boldsymbol{u}; \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\Phi^{-1}(\boldsymbol{u}))'(\boldsymbol{\Sigma}^{-1} - \mathbf{I}_d)\Phi^{-1}(\boldsymbol{u})\right), \tag{19}$$

where $\mathbf{I}_d$ is the d-dimensional identity matrix, $\boldsymbol{\Sigma}$ is the correlation matrix, and $\Phi^{-1}(\boldsymbol{u}) = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d))'$, with $\Phi^{-1}(\cdot)$ denoting the inverse of the standard normal CDF. In the bivariate case $d = 2$, the correlation coefficient $\rho_{12} = \rho_{21}$ is the only parameter of the Gaussian copula.

The Student-t copula is obtained similarly, but using a multivariate Student-t distribution instead of the Gaussian. The corresponding copula density is given by

$$c^{\text{St}-t}(\boldsymbol{u}, \boldsymbol{\Sigma}, \nu) = |\boldsymbol{\Sigma}|^{-1/2} \frac{\Gamma([\nu+d]/2)\Gamma^{d-1}(\nu/2)}{\Gamma^{d}((\nu+1)/2)} \frac{\left(1 + \frac{T_\nu^{-1}(\boldsymbol{u})'\boldsymbol{\Sigma}^{-1}T_\nu^{-1}(\boldsymbol{u})}{\nu}\right)^{-(\nu+d)/2}}{\prod_{i=1}^{d}\left(1 + \frac{(T_\nu^{-1}(u_i))^2}{\nu}\right)^{-(\nu+1)/2}}, \tag{20}$$

where $T_\nu^{-1}(\boldsymbol{u}) = (T_\nu^{-1}(u_1), \ldots, T_\nu^{-1}(u_d))'$, and $T_\nu^{-1}(\cdot)$ is the inverse of the univariate Student-t CDF, $\boldsymbol{\Sigma}$ is the correlation matrix and ν is the number of degrees of freedom. In the bivariate case the Student-t copula has two parameters, the number of degrees of freedom ν and the correlation coefficient ρ_{12}. Note that the Student-t copula nests the Gaussian copula when $\nu = \infty$.

A major difference between the Gaussian copula and the Student-t copula is their ability to capture tail dependence, which may be important for financial applications. The lower tail dependence coefficient is defined as $\lambda_{\text{L}} = \lim_{q \downarrow 0} C(q, q, \ldots, q)/q$, and the upper tail dependence coefficient as $\lambda_{\text{U}} = \lim_{q \downarrow 0} C^{\text{s}}(q, q, \ldots, q)/q$, where C^{s} is the survival-copula of $\boldsymbol{\varepsilon}_t$, that is, the copula of $-\boldsymbol{\varepsilon}_t$ rather than $\boldsymbol{\varepsilon}_t$. For the Gaussian copula both tail dependence coefficients are equal to zero, while for the Student-t copula the tail dependence is symmetric and positive. Specifically, in the bivariate case $d = 2$, the tail dependence coefficients are given by

$$\lambda_{\text{L}} = \lambda_{\text{U}} = 2T_{\nu+1}\left(-\sqrt{(\nu+1)(1-\rho_{12})/(1+\rho_{12})}\right),$$

which is increasing in the correlation coefficient ρ_{12} and decreasing in the number of degrees of freedom ν.

The Clayton and Clayton survival copulas belong to the family of Archimedean copulas (see Nelsen (2006) for details). The d-dimensional Clayton copula is given by

$$C^{\text{Cl}}(\boldsymbol{u}; \alpha) = \left(\sum_{j=1}^{d} u_j^{-\alpha} - d + 1\right)^{-1/\alpha}, \qquad \text{with } \alpha > 0.$$

In contrast to the Gaussian and Student-t copulas, the Clayton copula is able to capture asymmetric tail dependence. In fact, it only exhibits lower tail dependence, while upper tail dependence is absent. In the bivariate case the lower tail dependence coefficient for

the Clayton copula is $\lambda_{\mathrm{L}} = 2^{-1/\alpha}$, which is increasing in the parameter α. The density function of the Clayton copula is

$$c^{\mathrm{Cl}}(\boldsymbol{u}, \alpha) = \left(\prod_{j=1}^{d} (1 + (j-1)\alpha) \right) \left(\prod_{j=1}^{d} u_j^{-(\alpha+1)} \right) \left(\sum_{j=1}^{d} u_j^{-\alpha} - d + 1 \right)^{-(\alpha^{-1}+d)}.$$

The Clayton survival copula is obtained as the mirror image of the Clayton copula, with density function given by

$$c^{\text{Cl-s}}(\boldsymbol{u}, \alpha) = c^{\mathrm{Cl}}(\mathbf{1} - \boldsymbol{u}, \alpha).$$

Consequently, in the bivariate case the upper tail dependence coefficient for the Clayton survival copula is $\lambda_{\mathrm{U}} = 2^{-1/\alpha}$, and is increasing in the parameter α, while the lower tail dependence coefficient is zero.

In the simulation experiments we focus on the bivariate case, i.e., $d = 2$. We set the number of observations for the moving in-sample window to $R = 1,000$ and compare the results for two different out-of-sample forecasting periods $P = 1,000$ and $P = 5,000$. Asymptotic results of the considered tests are based on the assumption that $P > R$, but in practice it is not always possible to have large P. This motivates us to consider the finite sample properties of the test for the more feasible situation $R = P = 1,000$.

All models are estimated using maximum likelihood, for models involving marginal distributions we estimate the parameters of the marginal distributions first and after obtaining the standardised innovations we transform them into PITs, either using the corresponding parametric CDF or empirical CDF (ECDF). These are then used to estimate the copula parameters.

The number of replications in each experiment is set to B =1,000.

3.1 Size

In order to assess the size properties of the test, a case is required with two competing copulas that are both 'equally (in)correct'. We achieve this with the following setup. We consider DGPs with a Student-t copula with degrees of freedom $\nu = 6$. To verify the properties for various level of dependence we vary the correlation coefficient: $\rho = 0.1, 0.5, 0.9$. We test the null hypothesis of equal predictive accuracy of Clayton and Clayton survival

copulas with their parameters being estimated using the moving in-sample window. Due to symmetry of the Student-t copula with respect to the reflections $u_i \leftrightarrow 1 - u_i$, the two competing copula specifications are equally distant from the true copula. To preserve the symmetry also when we focus on a region of interest, we focus on the central region $[0.25, 0.75]^2$, which also respects this symmetry. We report results based on two-sided tests, for censored as well as conditional scores.

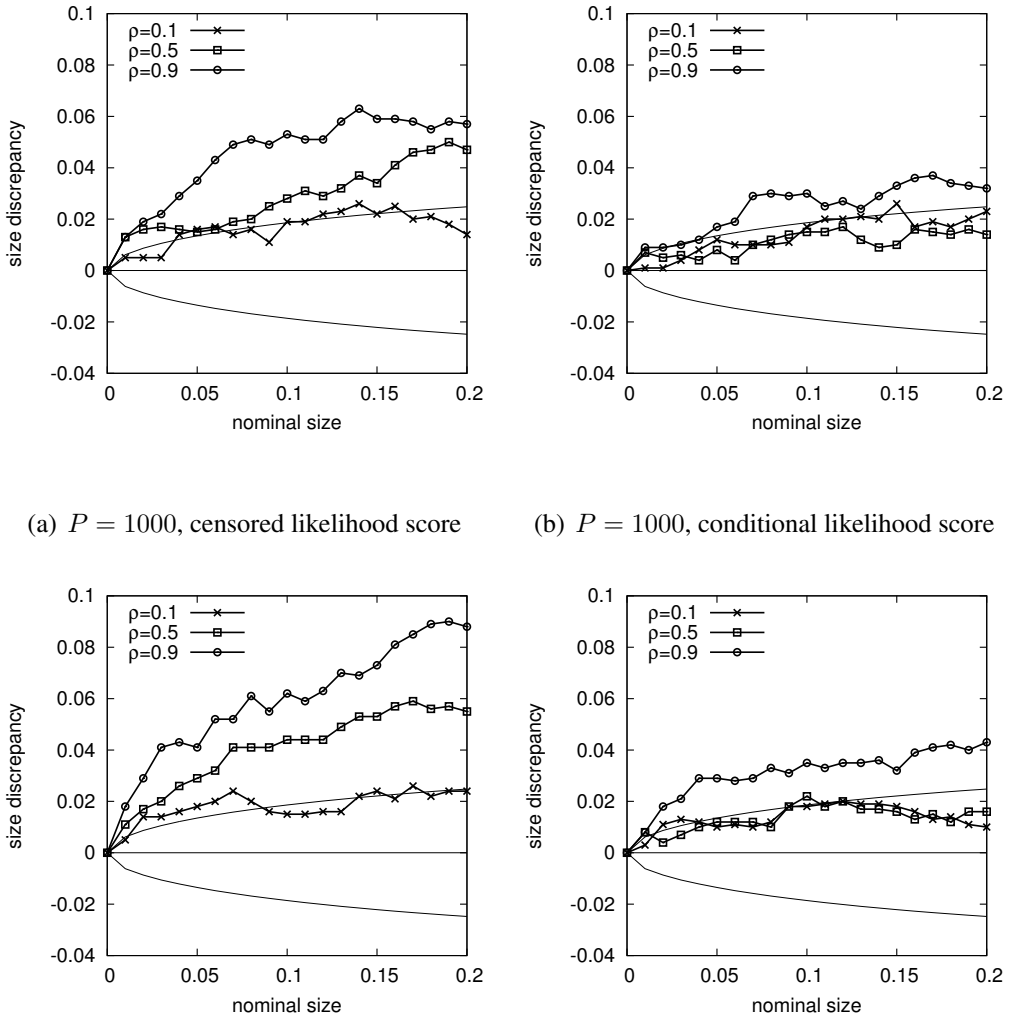
The discrepancy between the actual size (or observed rejection rate) and the nominal size of the test are shown in Figs 1 and 2. Fig. 1 compares the results for the DGPs with varying level of dependence. The higher the dependence the more deviation is observed from the nominal size. The deviations are mainly caused by poor finite sample performance of the HAC estimator which attempts to capture dependence between scores.[1] The tests based on censored likelihood scores tend to over-reject somewhat more often than the tests based on conditional likelihood scores, but overall they exhibit similar properties. As expected, the size distortion can be seen to become smaller for a larger number of out-of-sample evaluations P.

Fig. 2 compares the results for three DGPs with varying models for the marginal distributions: no uncertainty about the marginals, parametric marginals and semi-parametric marginals. The three different estimators exhibit similar size properties.

We also considered other DGPs and higher dimensions to verify the size properties of the test. We do not report detailed results here for the sake of brevity; they can be briefly summarised as follows. The unifying theme was that the size distortion crucially depends on the strength of dependence in the considered DGP. In some cases the semi-parametric models showed somewhat larger size distortions than the parametric models and models with known marginals. The size properties of both tests are comparable to those found by Diks *et al.* (2011) for the tests of copula predictive accuracy for the full support.

[1] We have tried various implementations of the HAC estimator available in the literature (see den Haan and Levin (1996) for review), but no implementation was fully satisfactory for considered DGPs and sample sizes.

Figure 1: *Size discrepancy plots for the test of equal predictive accuracy for various levels of dependence in the DPG*



(a) $P = 1000$, censored likelihood score

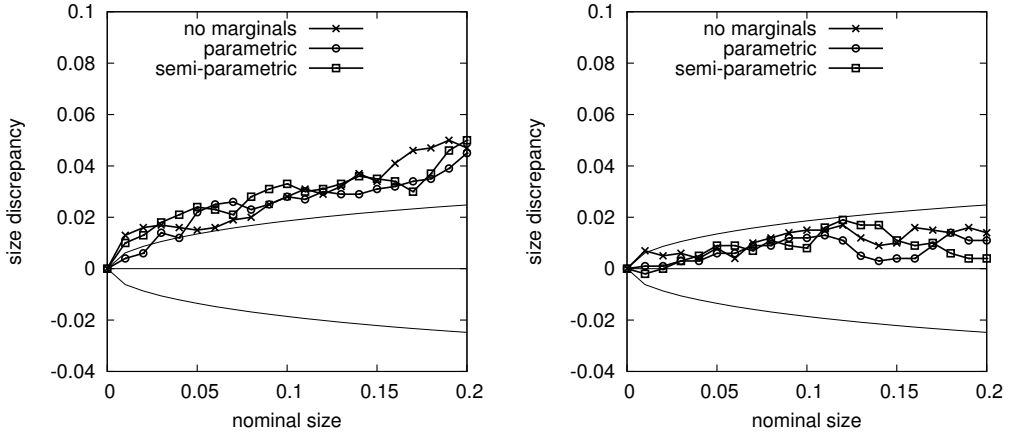
(b) $P = 1000$, conditional likelihood score

(c) $P = 5000$, censored likelihood score

(d) $P = 5000$, conditional likelihood score

The panels display the actual size-nominal size discrepancy of a two-sided test of equal performance of the Clayton and Clayton survival copulas. The DGPs are based on a Student-t copula with $\nu = 6$ degrees of freedom and varying levels of correlation ρ. There is no uncertainty about the marginal distributions. The tests are based on the central copula region $[0.25, 0.75]^d$ ($d = 2$) and use either censored (left) or conditional scores (right). The number of observations in the moving in-sample estimation window is $R = 1,000$ and the number of out-of-sample evaluations is $P = 1,000$ and $P = 5,000$. The reported results are based on $1,000$ replications. The thin lines indicate the 95% point-wise confidence bounds.

Figure 2: *Size discrepancy plots for the test of equal predictive accuracy for various DPGs*



(a) $P = 1000$, censored likelihood score (b) $P = 1000$, conditional likelihood score

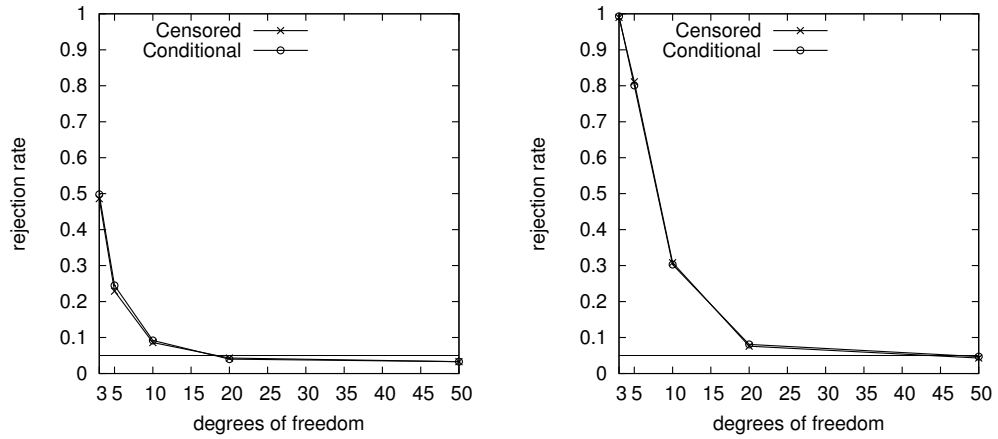
The panels display the actual size - nominal size discrepancy of a two-sided test of equal performance of the Clayton and Clayton survival copulas. The DGPs use a Student-t copula with $\nu = 6$ degrees of freedom and correlation $\rho = 0.5$. We consider three processes for the marginals: no uncertainty about the marginals (labeled as 'no marginals'), parametric marginals with Normally distributed innovations and semi-parametric marginals with ECDF-based standardised innovation. In the latter two cases the marginals follow an AR(1)-GARCH(1,1) process. The tests are based on the central copula region $[0.25, 0.75]^d$ ($d = 2$) and use either censored (left) or conditional scores (right). The number of observations in the moving in-sample estimation window is $R = 1,000$ and the number of out-of-sample evaluations is $P = 1,000$. The reported results are based on $1,000$ replications. The thin lines indicate the 95% point-wise confidence bounds.

3.2 Power

We evaluate the power of the test of equal predictive accuracy by performing a simulation experiment where one of the competing copula specifications corresponds with that of the DGP, while the distance of the alternative, incorrect copula specification to the DGP varies depending on a certain parameter of the DGP. Specifically, the DGP is a Student-t copula of dimension $d = 2$ with correlation coefficient $\rho = 0.3$. The number of degrees of freedom ν is varied over the interval $[3, 50]$. We compare the predictive accuracy of the correct Student-t copula specification (with both parameters ρ and ν being estimated) against an incorrect Gaussian copula specification (with only the correlation coefficient ρ being estimated) in the left tail region $[0, 0.25]^d$. Hence, we focus on the question whether the proposed tests can distinguish between copulas with and without tail dependence. Note, however, that the Student-t copula approaches the Gaussian copula as ν increases, and the tail dependence disappears with the coefficients λ_{L} and λ_{U} converging to zero. Intuitively, the higher the value of ν in the DGP, the more difficult it becomes to distinguish between these two copula specifications.

The results are shown in Fig. 3 in the form of power plots, showing the observed rejection rates (for a nominal size of 0.05) as a function of the degrees of freedom parameter, ν, in the DGP, for tests based on the censored and conditional scoring rule. The results displayed are for the null hypothesis that the Gaussian and Student-t copulas perform equally well, against the one-sided alternative hypothesis that the correctly specified Student-t copula has a higher average score. Intuitively, since the true DGP uses the Student-t copula, we might expect the Student-t copula to perform better. Note, however, that as the number of degrees of freedom ν in the copula describing the DGP becomes large, the Gaussian copula might outperform the Student-t copula. This is a consequence of the fact that the Gaussian copula is very close to to the Student-t copula for large values of ν, but requires one parameter less to be estimated. Indeed, the rejection rates become smaller than the nominal size for very large values of ν. Consequently, Fig. 3 shows that the test has higher power for smaller values of ν. Naturally, the test based on the larger number of out-of-sample evaluations, P, shows higher rejection rates.

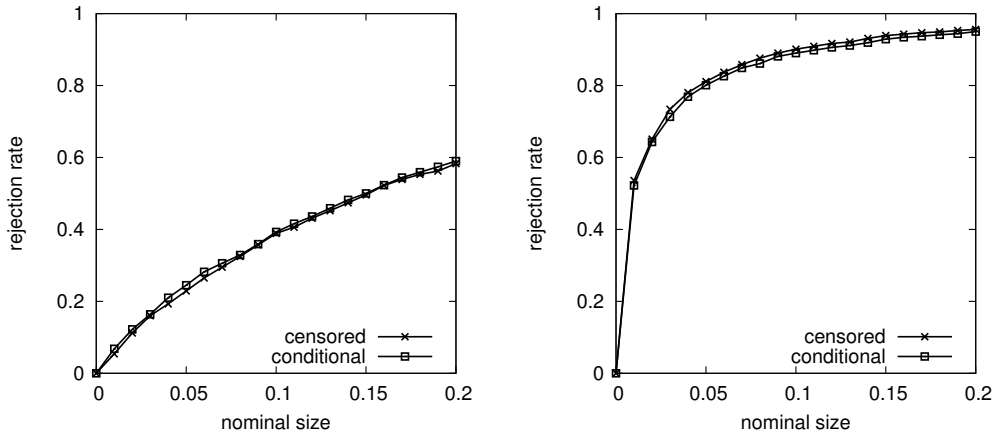
Figure 3: *Power of the test of equal predictive accuracy for various levels of tail dependence*



(a) P=1000 (b) P=5000

The panels show the observed rejection rates (on the vertical axis) of a one-sided test of equal performance of the Gaussian and Student-t copulas, against the alternative hypothesis that the correctly specified Student-t copula has a higher average score. The tests are based on the left tail copula region $[0, 0.25]^d$ ($d = 2$) and use either censored or conditional scores. The horizontal axis displays the degrees of freedom parameter of the Student-t copula characterising the DGP. The off-diagonal correlation coefficients ρ_{ij}, $i \neq j$, are all set to 0.3. There is no uncertainty about the marginals. The test of equal predictive accuracy compares a Student-t copula (with both parameters ρ and ν estimated, rather than known) against a Gaussian copula with the parameter ρ also being estimated. The nominal size is 0.05, indicated by the thin horizontal lines. The number of observations in the moving in-sample estimation window is $R = 1,000$ and the number of out-of-sample evaluations is $P = 1,000$ and $P = 5,000$, repsectively. The reported results are based on $1,000$ replications.

Figure 4: *Size-power plots for the test of equal predictive accuracy*



(a) $P = 1,000$ (b) $P = 5,000$

The panels display the observed rejection rates (on the vertical axis) as a function of nominal size (on the horizontal axis) for a one-sided version of our test of equal performance of the Gaussian and Student-t copulas, against the alternative hypothesis that the correctly specified Student-t copula has a higher average score. The tests are based on the lower tail copula region $[0, 0.25]^d$ *($d = 2$) and based on either censored or conditional scores. The DGP is characterised by the Student-t copula with correlation coefficient* $\rho = 0.3$ *and degrees of freedom* $\nu = 5$*. There is no uncertainty about the marginals. The test of equal predictive accuracy compares a Student-t copula (with both parameters* ρ *and* ν *estimated, rather than known) against a Gaussian copula with the parameter* ρ *also being estimated. The nominal sizes is* 0.05*, indicated by horizontal lines. The number of observations in the moving in-sample estimation window is* $R = 1,000$ *and the number of out-of-sample evaluations is* $P = 1,000$ *and* $P = 5,000$*. Reported results are based on* $1,000$ *replications.*

Fig. 4 shows the observed rejection rates of the test as a function of the nominal size for a fixed number of degrees of freedom, $\nu = 5$, which may be observed in financial applications. The test exhibits nontrivial power for all sizes considered.

We also considered power properties for models with parametric and semi-parametric marginal specifications, which we do not report here due to space considerations. Additional estimation uncertainty due to the marginals slightly decreased the power, but the overall pattern was very similar to the case with no marginal uncertainty.

Finally, comparison with the results of Diks *et al.* (2010) shows that the tests based on the full copula support have higher power than the tests focusing on the left tail only. This is to be expected, as the tests for predictive accuracy in a given region of support attempt to solve a much more difficult statistical problem (the observations outside of the targeted region are of limited value to the testing of the hypothesis).

In summary, although the suggested tests of predictive accuracy in the selected region of support exhibit moderate discrepancy from the nominal size, they have satisfactory statistical power.

4 Empirical application

We examine the empirical usefulness of our predictive accuracy test with an application to exchange rate returns for several major currencies. Specifically, we consider daily returns on the US dollar exchange rates of the Canadian dollar (CAD), euro (EUR), and Japanese yen (JPY). The data are noon buying rates in New York and are obtained from the Federal Reserve Bank of New York. We base our analysis on the daily FX returns over the period from January 2, 1980 until July 21, 2008. Up to December 31, 1998, the euro series actually concerns the exchange rate of the German Deutschmark (DM), while the euro is used as of January 4, 1999.

We employ a GARCH framework with Student-t innovations to model the marginal characteristics of the daily exchange rate returns. For the conditional mean and the conditional variance of the return on currency j we use an AR(5)-GARCH(1,1) specification,

given by

$$Y_{j,t} = c_j + \sum_{l=1}^{5} \phi_{j,l} Y_{j,t-l} + \sqrt{h_{j,t}}\varepsilon_{j,t} \tag{21}$$

$$h_{j,t} = \omega_j + \alpha_j \left(Y_{j,t-1} - c_j - \sum_{l=1}^{5} \phi_{j,l} Y_{j,t-1-l} \right)^2 + \beta_j h_{j,t-1}, \tag{22}$$

where $\omega_j > 0$, $\beta_j \geq 0$, $\alpha_j > 0$ and $\alpha_j + \beta_j < 1$.

The joint distribution of the standardised innovations $\varepsilon_{j,t}$ combines Student-t univariate marginal distributions F_j with a parametric copula C. We consider a substantial number of different copula specifications. In particular, we consider the Gaussian (Ga) and Student-t (St-t) elliptic copulas and the classic Archimedean copulas and their mixtures, that is, the Clayton (Cl), Clayton survival (Cl-s), mixture of Clayton and Clayton survival (Cl/Cl-s), Gumbel (Gu), given by

$$C^{\mathrm{Gu}}(\boldsymbol{u}; \alpha) = \exp\left(\left(-\sum_{i=1}^{d} \left(-\ln u_i \right)^{\alpha} \right)^{1/\alpha} \right)$$

Gumbel survival (Gu-s), and mixture of Gumbel and Gumbel survival (Gu/Gu-s). The Gumbel copula for $\alpha > 1$ has upper tail dependence with upper tail dependence index $\lambda_{\mathrm{u}} = 2 - 2^{1/\alpha}$. Likewise, the Gumbel survival copula has lower tail dependence $\lambda_{\mathrm{L}} = 2 - 2^{1/\alpha}$.

We compare the one-step ahead density forecasting performance of the different copula specifications using a rolling window scheme. The length of the rolling estimation window is set to $R = 2,000$ observations, such that $P = 2,772$ observations (from December 23, 1987 until December 31, 1998) during the pre-euro sub-period and $P = 2,406$ observations (from January 4, 1999 until July 21, 2008) during the post-euro sub-period are left for out-of-sample forecast evaluation. For comparing the accuracy of the resulting copula-based density forecasts we use the Diebold-Mariano type test based on the conditional likelihood in (11) and the censored likelihood in (12). As both scoring rules give qualitatively similar results, to save space we only report results of the tests based on the censored likelihood.[2]

[2]Detailed results based on the conditional likelihood are available upon request.

We focus on three specific regions of the copula support. The first region, labeled *D*, corresponds to all currencies suffering a simultaneous depreciation against the USD, and is defined as

$$D = \{(u_1, \ldots, u_d) | u_j < r \text{ for all } j = 1, \ldots, d\},$$

where the threshold $r \in \{0.20, 0.25, 0.30\}$. Note that we only consider regions with identical thresholds for all exchange rates. This obviously is an arbitrary choice, but it is made in order to limit the number of regions under consideration. Below, we present detailed results for $r = 0.25$ only, but include the two alternative values in the discussion to address the sensitivity of the results to the specific choice of the threshold value. The second region, denoted *U*, is the mirror image of *D* in the sense that it represents a simultaneous appreciation against the USD and is defined as

$$U = \{(u_1, \ldots, u_d) | u_j > 1 - r \text{ for all } j = 1, \ldots, d\},$$

where again the threshold $r \in \{0.20, 0.25, 0.30\}$. The third region concerns the central part of the copula support. This region *M* is defined as

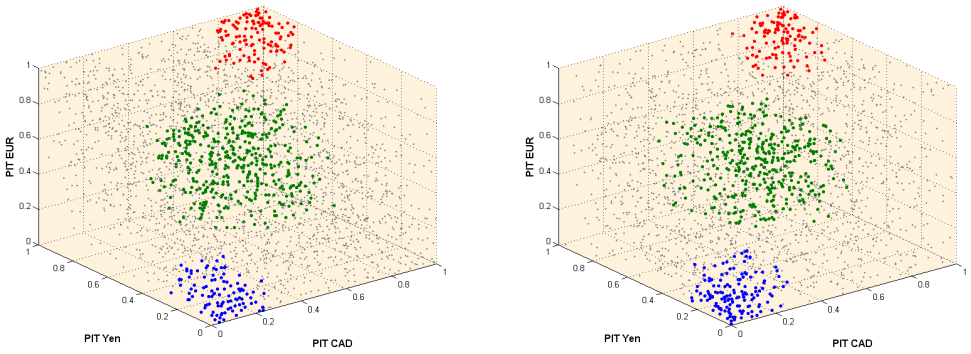
$$M = \{(u_1, \ldots, u_d) | r < u_j < 1 - r \text{ for all } j = 1, \ldots, d\},$$

where we use the same values of r as for the regions *D* and *U*. Region *M* corresponds to 'regular' trading conditions.

Fig. 5 illustrates the three regions for the conditional PITs from one-step ahead density forecasts for the daily CAD/USD-JPY/USD-EUR/USD exchange rates return innovations. The PITs of observations are color-coded in accordance with their attribution to a particular region of copula support: observations in region *D* are in blue, observations in region *U* are in red, the ones in region *M* are in green, and, finally, the observations in the compliment to the union of the above regions are in grey.

As an additional diagnostic tool, we use the model confidence set (MCS) concept of Hansen *et al.* (2011) to identify the collection of models which includes the best copula specification with a certain level of confidence. Starting with the full set of models, at each iteration we test the null hypothesis that all the considered models have equal predictive ability according to the selected scoring rule. If the null hypothesis is rejected, the worst

Figure 5: *Conditional PITs from one-step ahead density forecasts for daily CAD/USD-JPY/USD-EUR/USD returns*



(a) Pre-euro sub-period: December 23, 1987 - December 31, 1998

(b) Post-euro sub-period: January 4, 1999 - July 21, 2008

performing model is omitted, and equal predictive ability is tested again for the remaining models. This procedure is repeated until the null hypothesis cannot be rejected and the collection of models that remains at this point is defined to be the MCS. In our application of the MCS procedure, we always exclude the worst performing model and repeat the algorithm until only one model remains in the MCS. This allows us to obtain a complete ranking of the competing models. We report the MCS p-values at every iteration. The implementation of the MCS is based on bootstrapping. To accommodate the possibility of autocorrelation in the scoring rules, we use the stationary bootstrap methodology of Politis and Romano (1994). We report the results for the probability of sampling the consecutive observation equal to 0.9. The ranking of the different copula specifications and the MCS p-values are robust to the choice of this probability.

Due to space considerations we focus the density forecast comparison on the post-euro sub-period in the discussion below. That is not to say that the pre-euro sub-period does not render interesting results. In fact, in line with Patton (2006), we find striking differences in the dependence characteristics of the three exchange rates between the pre- and post-euro sub-periods. These differences do not become apparent when the predictive accuracy of the density forecasts arising from different copula specifications is compared over the whole copula support, using the test statistic of Diks *et al.* (2010). This follows from Table 1, which reports the values of the pairwise $Q_{R,P}$ test statistic based on the log-

scoring rule $\mathcal{S}^l(\boldsymbol{y}_{t+1})$ as given in (13) for the pre- and post-euro sub-periods. Obviously, the matrices in the two panels are antisymmetric, that is $Q_{R,P}(i,j) = -Q_{R,P}(j,i)$ for copula specifications i and j. We nevertheless report the full matrices as this allows for an easy assessment of the relative performance of the various copulas. Given that in each panel the (i,j)th entry is based on the score difference $d^l_{t+1} = \mathcal{S}^l_{j,t+1}(\boldsymbol{y}_{t+1}) - \mathcal{S}^l_{i,t+1}(\boldsymbol{y}_{t+1})$, positive values of the test statistic indicate that the copula in column j achieves a higher average score than the one in row i. Hence, the more positive values in a given column, the higher the ranking of the corresponding copula specification.

From Table 1, we observe that for both the pre- and post-euro sub-periods, the Student-t copula performs best when the full copula support is taken into account in the comparison. Based on the pairwise $Q_{R,P}$ test statistic, the null of equal predictive accuracy is rejected at better than the 1% significance level for all competing copula specifications, except the Gumbel-Gumbel survival and the Clayton-Clayton survival mixtures during the post-euro period. This is confirmed by the MCS results. During the pre-euro sub-period, the MCS at conventional significance levels only consists of the Student-t copula, while for the post-euro period the two mixture copulas may be included as well. Results are very different, however, when we focus on sub-regions. Specifically, unreported results show that during the pre-euro sub-period the Gaussian copula is the superior choice for capturing the dependence structure in each of the three considered regions of support *D*, *U* and *M* with the threshold $r = 0.25$. The corresponding p-values for the pairwise tests of equal accuracy of a Student-t copula and a Gaussian copula are 0.001 for region *D*, 0.113 for region *M*, and 0.056 for region *U*. This suggests the lack of positive upper and lower tail dependence in the case of joint depreciation or appreciation of the three exchange rates against the US dollar. It follows that the strength of the Student-t copula specification when the full support is considered stems from its performance in the complement to the union of the three selected regions of support. The latter corresponds to strong movements in opposite directions of the three currencies values against the US dollar. For the post-euro sub-period we again find rather different results for the selected regions of support.

Table 2 shows results of the pairwise $Q_{R,P}$ test statistic based on the censored likeli-

Table 1: *Daily CAD/USD-JPY/USD-EUR/USD returns: Pair-wise tests of equal predictive accuracy of copulas for full copula support, based on the test suggested by Diks et al. (2010).*

	Ga	St-t	Cl	Cl-s	Cl/Cl-s	Gu	Gu-s	Gu/Gu-s
Panel A: Pre-euro sub-period (December 23, 1987 - December 31, 1998)								
Ga		2.97	−7.00	−6.74	−6.24	−6.47	−7.04	−6.34
St-t	−2.97		−7.99	−7.61	−7.37	−7.40	−7.96	−7.45
Cl	7.00	7.99		0.21	3.30	0.79	−0.50	2.14
Cl-s	6.74	7.61	−0.21		3.02	0.50	−0.75	1.85
Cl/Cl-s	6.24	7.37	−3.30	−3.02	0.00	−3.09	−4.29	−2.45
Gu	6.47	7.40	−0.79	−0.50	3.09		−1.11	1.82
Gu-s	7.04	7.96	0.50	0.75	4.29	1.11		3.35
Gu/Gu-s	6.34	7.45	−2.14	−1.85	2.45	−1.82	−3.35	
MCS order	7		4	3	6	2	1	5
MCS p-val	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: Post-euro sub-period (January 4, 1999 - July 21, 2008)								
Ga		5.27	−2.38	−1.50	1.05	−0.78	−1.55	1.15
St-t	−5.27		−4.60	−3.74	−1.47	−2.99	−3.69	−1.24
Cl	2.38	4.60		1.05	5.51	2.00	1.68	5.29
Cl-s	1.50	3.74	−1.05		4.76	1.63	−0.11	4.46
Cl/Cl-s	−1.05	1.47	−5.51	−4.76		−4.15	−4.45	0.59
Gu	0.78	2.99	−2.00	−1.63	4.15		−1.04	4.37
Gu-s	1.55	3.69	−1.68	0.11	4.45	1.04		5.03
Gu/Gu-s	−1.15	1.24	−5.29	−4.46	−0.59	−4.37	−5.03	
MCS order	5		1	2	6	4	3	7
MCS p-val	0.14	1.00	0.00	0.00	0.245	0.07	0.01	0.27

Note: Values of the Diks *et al.* (2010) test statistic. The test statistic is based on one-step ahead density forecasts for daily CAD/USD, JPY/USD and EUR/USD returns during the corresponding sub-period, with the length of the rolling estimation window set equal to $R = 2,000$ observations in both sub-periods. Consequently, the number of forecasts is $P = 2,772$ for *Panel A: Pre-euro sub-period (December 23, 1987 - December 31, 1998)*, and $P = 2,406$ for *Panel B: Post-euro sub-period (January 4, 1999 - July 21, 2008)*. In each panel the (i, j)th entry is based on the score differences such that positive values of the test statistic indicate that the model in column j achieves a higher average score than the model in row i. Acronyms used for referring to copula specifications: Ga - Gaussian; St-t - Student-t; Cl - Clayton; Cl-s - Clayton survival; Cl/Cl-s - Clayton-Clayton survival mixture; Gu - Gumbel; Gu-s - Gumbel survival; Gu/Gu-s - Gumbel-Gumbel survival mixture. MCS order is the iteration, at which the model is omitted from the MCS, while the MCS p-val is the corresponding p-value.

hood scoring rule for regions D, U, and M with $r = 0.25$. These demonstrate that after 1999, the Gumbel survival, the Gumbel, and the Student-t specifications perform best for these three individual regions of copula support. Next, we describe and analyze these results in more detail.

The dependence structure in region *D*, corresponding to common significant depreciation of the currencies against the US dollar, strongly favors the Gumbel survival copula specification (see Panel A in Table 2). The Gumbel-Gumbel survival mixture copula is a close second-best choice. Notably, the Gumbel survival copula outperforms the Gaussian copula with an approximate p-value of 0.04. This signifies a major departure from the lack of positive lower tail dependence observed in the pre-euro sub-period. Other copula specifications with positive lower tail dependence, including the Gumbel-Gumbel survival mixture, Student-t, and Clayton copulas also demonstrate good performance in region *D*. Copula specifications with positive upper tail dependence but no lower tail dependence, like the Clayton Survival and Gumbel copulas, yield the worst results and are the first specifications to be dropped from the MCS. This strongly suggests that the ability to reflect positive lower tail dependence is a highly desirable feature for modeling extreme joint depreciation of the three currencies against the US dollar. Accommodating only positive upper tail dependence is of no use in region *D*.

Let us now investigate the details of the comparison between the Gumbel Survival and Student-t copulas. The choice of the Student-t copula as the main competitor originates from the superiority of the Student-t copula when the full copula support is considered. Define an *unbalanced* observation from region *D* as the one which to corresponds to a PIT for which one of the components is substantially smaller or larger than the other two. More precisely, we define an observed PIT as unbalanced when its largest component is more than 25 times larger than its smallest component. All other observations are labeled as *balanced*. This type of classification is particularly useful in explaining the relative performance of different copula specifications in any given region of support. We visualise the balanced and unbalanced PIT observations and their effect on the test statistic by means of specially-designed scatter plots, with the following features. All observations favouring the winning copula (in the case of region *D*, this is the Gumbel Survival cop-

Table 2: *Daily CAD/USD-JPY/USD-EUR/USD returns: Pair-wise tests of equal predictive accuracy of copulas for selected regions of support, based on the censored likelihood scoring rule*

	Ga	St-t	Cl	Cl-s	Cl/Cl-s	Gu	Gu-s	Gu/Gu-s
Panel A: region *D*								
Ga		2.27	1.88	−4.54	2.03	−4.36	1.77	1.98
St-t	−2.27		0.60	−4.26	0.02	−4.02	1.29	1.08
Cl	−1.88	−0.60		−4.17	−0.84	−3.96	1.15	0.73
Cl-s	4.54	4.26	4.17		4.57	4.48	3.56	4.08
Cl/Cl-s	−2.03	−0.02	0.84	−4.57		−4.48	1.27	1.17
Gu	4.36	4.02	3.96	−4.48	4.48		3.21	3.81
Gu-s	−1.77	−1.29	−1.15	−3.56	−1.27	−3.21		−1.00
Gu/Gu-s	−1.98	−1.08	−0.73	−4.08	−1.17	−3.81	1.00	
MCS order	3	5	6	1	4	2		7
MCS p-val	0.24	0.385	0.45	0.00	0.35	0.00	1.00	0.45
Panel B: region *U*								
Ga		2.86	−3.84	1.87	2.19	2.09	−3.72	2.25
St-t	−2.86		−3.79	0.36	1.31	1.61	−3.69	1.55
Cl	3.84	3.79		3.43	3.57	3.14	3.63	3.44
Cl-s	−1.87	−0.36	−3.43		1.41	1.82	−3.23	1.56
Cl/Cl-s	−2.19	−1.31	−3.57	−1.41		1.25	−3.39	0.64
Gu	−2.09	−1.61	−3.14	−1.82	−1.25		−2.93	−1.22
Gu-s	3.72	3.69	−3.63	3.23	3.39	2.93		3.24
Gu/Gu-s	−2.25	−1.55	−3.44	−1.56	−0.64	1.22	−3.24	
MCS order	3	5	1	4	6		2	7
MCS p-val	0.08	0.21	0.00	0.13	0.355	1	0.00	0.355
Panel C: region *M*								
Ga		3.26	−0.44	−2.43	2.07	0.69	0.62	2.51
St-t	−3.26		−2.51	−3.30	−2.03	−2.49	−2.34	−1.90
Cl	0.44	2.51		−2.16	2.89	1.36	1.50	2.60
Cl-s	2.43	3.30	2.16		4.47	4.12	4.21	4.24
Cl/Cl-s	−2.07	2.03	−2.89	−4.47		−3.61	−3.18	1.64
Gu	−0.69	2.49	−1.36	−4.12	3.61		−0.06	3.47
Gu-s	−0.62	2.34	−1.50	−4.21	3.18	0.06		2.84
Gu/Gu-s	−2.51	1.90	−2.60	−4.24	−1.64	−3.47	−2.84	
MCS order	2		4	1	6	3	5	7
MCS p-val	0.01	1.00	0.015	0.00	0.06	0.01	0.04	0.08

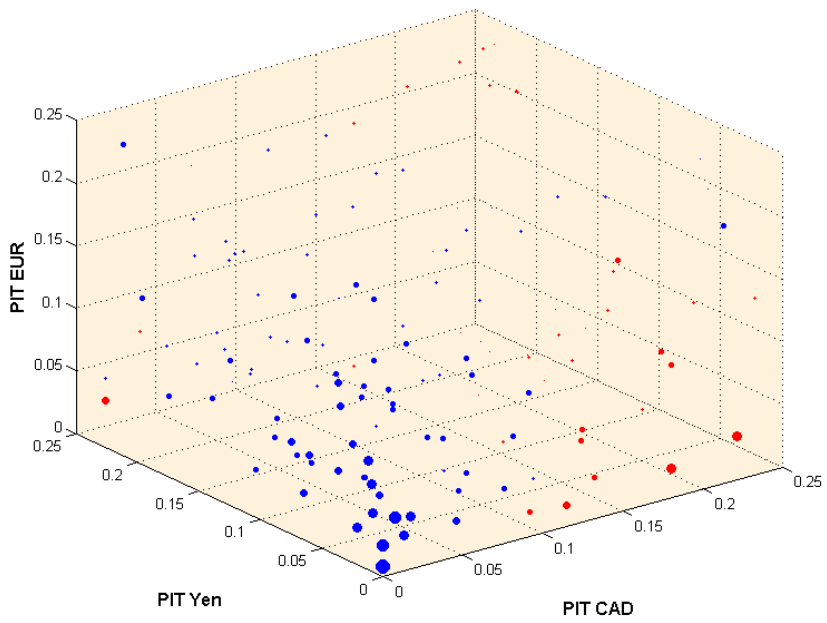
Note: Values of the Diebold-Mariano type test statistic $t_{R,P}$ defined in (1) based on the censored likelihood score (12) for the regions *D*, *U* and *M* with the threshold $r = 0.25$. The test statistic is based on one-step ahead density forecasts for daily CAD/USD, JPY/USD and EUR/USD returns during the period January 4, 1999 - June 21, 2008 ($P = 2,406$), with the length of the rolling estimation window set equal to $R = 2,000$ observations. In each panel the (i,j)th entry is based on the score difference $d_{t+1}^{csl} = \mathcal{S}_{j,t+1}^{csl}(\boldsymbol{y}_{t+1}) - \mathcal{S}_{i,t+1}^{csl}(\boldsymbol{y}_{t+1})$ such that positive values of the test statistic indicate that the model in column j achieves a higher average score than the model in row i. Acronyms used for referring to copula specifications: Ga - Gaussian; St-t - Student-t; Cl - Clayton; Cl-s - Clayton survival; Cl/Cl-s - Clayton-Clayton survival mixture; Gu - Gumbel; Gu-s - Gumbel survival; Gu/Gu-s - Gumbel-Gumbel survival mixture. MCS order is the iteration, at which the model is omitted from the MCS, while the 'MCS p-val' is the corresponding p-value.

ula) are colored blue, while all observations favouring the competing specification (the Student-t copula in region *D*) are colored red. The size of the marker corresponds to the absolute value of the weighted score difference the larger the marker, the greater the absolute value of the weighted score difference. Therefore, the most influential observations are coded by large markers. The upper panel of Fig. 6 contains such a scatter-plot of PITs falling in the *D* region of copula support where the corresponding weighted score differences determine the colour and shape of the markers. It is evident from the top panel of Fig. 6 that balanced observations strongly support the Gumbel survival copula, while the Student-t copula benefits almost exclusively from unbalanced observations. Balanced observations are also much more numerous with a substantial number of these corresponding to large values of the weighted score differences.

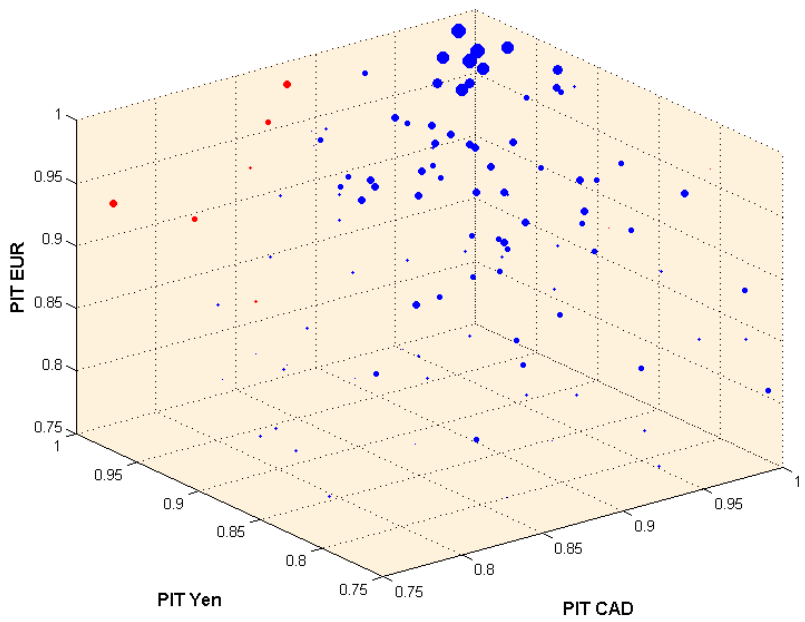
For region *U*, associated with a common significant appreciation of the currencies against the US dollar, we essentially find a mirror image of the above results. Now the Gumbel copula emerges as the best option for capturing the dependence structure of strongly appreciating currencies, see Panel B in Table 2. It is followed by the Gumbel-Gumbel survival and Clayton-Clayton survival mixture copulas. We should remark that the Gumbel copula also outperforms the Student t-copula specification with an approximate p-value of 0.05. The Gumbel copula, dominant in the *U* region, is the mirror image of the Gumbel survival copula, being the best choice for the *D* region. Thus, also the same type of the dependence structure (Gumbel copula and its mirror image) emerges as the best choice for the two regions of support. The regions themselves are naturally "mirror reflections" of one another. This confers a reassuring symmetry to the results. The MCS results indicate that the top three copulas that excel in capturing the dependence structure in region *U* are the Gumbel, Gumbel-Gumbel survival and Clayton-Clayton survival mixture copulas. All three copula specifications allow for positive upper tail dependence. At the same time, the Gumbel survival and Clayton copula are performing worst and are the first to be excluded from the MCS. This again is the mirror image of the MCS results for region *D*; accommodating positive upper tail dependence is crucial in the *U* region of copula support, but allowing for lower tail dependence is of no use.

Looking deeper at the causes of the superiority of the Gumbel copula over the Student-

Figure 6: *Conditional PITs from one-step ahead density forecasts for daily CAD/USD-JPY/USD-EUR/USD returns for the period January 4, 1999 - July 21, 2008*



(a) Region *D*



(b) Region *U*

t copula, we come to conclusions analogous to the ones reached for region *D*. The bottom panel of Fig. 6 depicts PITs falling in region *U*, where the corresponding weighted scores differences between the Gumbel and Student-t copulas determine the color and shape of the markers: blue markers correspond to positive differences (favouring the Gumbel copula), while red markers correspond to negative differences (favouring the Student-t copula). Just as in the case of region *D*, balanced observations favour almost exclusively the Gumbel copula. The only observations that sufficiently reinforce the Student-t copula specification are away from the main diagonal (from $\{0.75, 0.75, 0.75\}$ to $\{1, 1, 1\}$).

The final studied *M* region of copula support corresponds to modest movements of the exchange rates. The Student-t copula decisively outperforms all competing copula specifications in this region of support. The significantly better performance of the Student-t copula than the Gumbel-Gumbel survival mixture copula (with a p-value of 0.03) is of particular importance because it yet again illustrates the practical importance of separately analysing the dependence structure in different regions of the copula support. At the 0.10 significance level, only the Student-t copula is in the MCS model set.

Finally, we examine the robustness of our results. First, our results are robust to variations in the volume of *D*, *M*, and *U* regions as parameterised by the threshold value r. The ranking of relative performance of top copula specifications in the selected regions of support is unchanged for $r = 0.20$ and 0.30. Second, estimation of the PITs using the SCOMDY model with non-parametric marginal probability distributions (rather than the Student-t marginal distributions) yields copula rankings which are, in general, consistent with the results obtained with the fully parametric model. SCOMDY-generated PITs favour the Gumbel-Gumbel survival mixture and pure Gumbel survival copulas in region D, the Student-t and Gumbel copulas in region *U*, and, finally, the Gumbel-Gumbel survival mixture and Student-t copulas in region *M*. While the conclusions about the importance of the positive lower and upper tail dependence features continue to hold, the differences between the top performing copula specifications are less pronounced when using the SCOMDY model.

5 Conclusions

Many practical applications involving joint density forecasts of multivariate asset returns focus on a particular part of the domain of support. Given that the dependence structure may vary, for example, with the sign and magnitude of returns, it becomes imperative to identify the best forecast method for the targeted part of the distribution. Copula modeling allows for a straightforward construction of a flexible multivariate distribution via their decomposition into the dependence structure, represented by a copula function, and marginal distributions. In this paper, we develop Kullback-Leibler Information Criterion (KLIC) based test of equal (out-of-sample) forecasting accuracy of different copula specifications in a selected region of the support. The test combines the approaches suggested by Diks *et al.* (2010) and Diks *et al.* (2011), making use of censored and conditional logarithmic scoring rules.

Monte Carlo simulation shows that the tests possess satisfactory power properties and moderate size distortions. The application of the tests to daily exchange rate returns clearly demonstrates that the best copula specification varies with the targeted region of the support. This finding highlights the practical usefulness of the suggested test.

References

Amisano, G. and Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics*, **25,** 177–190.

Bao, Y., Lee, T.-H. and Saltoğlu, B. (2004). A test for density forecast comparison with applications to risk management. Working paper 04-08, UC Riverside.

Bao, Y., Lee, T.-H. and Saltoğlu, B. (2007). Comparing density forecast models. *Journal of Forecasting*, **26,** 203–225.

Berg, D. (2009). Copula goodness-of-fit testing: An overview and power comparison. *European Journal of Finance*, **15,** 675–701.

Chen, X. and Fan, Y. (2005). Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. *Canadian Journal of Statistics*, **33,** 389–414.

Chen, X. and Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, **135,** 125–154.

Chib, S., Omori, Y. and Asai, M. (2009). Multivariate stochastic volatility. In *Handbook of Financial Time Series* (eds T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch), pp. 365–400. Springer-Verlag, Berlin.

den Haan, Wouter J. and Levin, Andrew T. (1996). A practitioner's guide to robust covariance matrix estimation. NBER Technical Working Papers 0197. National Bureau of Economic Research, Inc.

Diebold, F.X., Gunther, T.A. and Tay, A.S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, **39,** 863–883.

Diebold, F.X. and Lopez, J.A. (1996). Forecast evaluation and combination. In *Handbook of Statistics, Vol. 14* (eds G.S. Maddala and C.R. Rao), pp. 241–268. Amsterdam: North-Holland.

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

Diks, C., Panchenko, V. and van Dijk, D. (2010). Out-of-sample comparison of copula specifications in multivariate density forecasts. *Journal of Economic Dynamics and Control*, **34,** 1596–1609.

Diks, C., Panchenko, V. and van Dijk, D. (2011). Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics*, **163,** 215–230.

Genest, C., Gendron, M. and Bourdeau-Brien, M. (2009). The advent of copulas in finance. *European Journal of Finance*, **15,** 609–618.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, **74,** 1545–1578.

Gneiting, Tilmann and Raftery, Adrian E (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102,** number 477, 359–378.

Granger, C.W.J. and Pesaran, M.H. (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, **19,** 537–560.

Hansen, P. R., Lunde, A. and Nason, J. M. (2011). The model confidence set. *Econometrica*, **79,** 453–497.

Mitchell, J. and Hall, S.G. (2005). Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR 'fan' charts of inflation. *Oxford Bulletin of Economics and Statistics*, **67,** 995–1033.

Nelsen, R.B. (2006). *An Introduction to Copulas*, Second edn. Springer Verlag, New York.

Patton, A.J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, **47,** 527–556.

Patton, A.J. (2009). Copula-based models for financial time series. In *Handbook of Financial Time Series* (eds T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch), pp. 767–786. Springer-Verlag, Berlin.

Politis, D.N. and Romano, J.P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, **89,** 1303–1313.

Rivers, Douglas and Vuong, Quang (2002). Model selection tests for nonlinear dynamic models. *The Econometrics Journal*, **5,** number 1, 1–39.

Silvennoinen, A. and Teräsvirta, T. (2009). Multivariate GARCH models. In *Handbook of Financial Time Series* (eds T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch), pp. 201–229. Springer-Verlag, Berlin.

Sklar, A. (1959). Fonction de répartitions à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, **8,** 229–231.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, **57,** 307–333.

West, Kenneth D (1996). Asymptotic inference about predictive ability. *Econometrica*, **64,** 1067–1084.