Evaluating Forecasts of Joint Extreme Events

Scoring rules for comparing copula-based multivariate density forecasts in tails

Salet, Ernst – 290131

August 10, 2010

Contents

Abstract

In this thesis we develop a testing framework for evaluating and comparing the accuracy of copula-based multivariate density models to forecast joint extreme events. The unique feature of our newly developed goodness-of-fit test is that it allows us to focus on a specific part of the copula distribution. The test has a clear intuition as it is based on likelihood based scoring rules that can be interpreted in terms of the Kullback-Leibler Information Criterion (KLIC). We show that the results based on the conventional weighted likelihood scoring rule show a clear bias towards fat-tailed models. Our proposed scoring rules based on conditional likelihood and censored likelihood do not suffer from this bias. Extensive Monte Carlo simulations and an empirical financial application based on bivariate density forecast models confirm that these two scoring rules are proper. Moreover, we show that their power increases with sample size. Among the proper scoring rules, the censored likelihood scoring rule generally shows the most powerful results in the left tail of the copula distribution. For the purpose of forecasting negative extreme events, we therefore recommend academics and risk managers to use our test based on the censored likelihood scoring rule to decide on the model that best suits the true data characteristics.

1. Introduction

Problems in quantitative finance are often multidimensional and hence require the joint modeling of random variables, including their interdependencies. Random variables may represent, for example, different risk factors such as financial returns. Their joint behavior has key impact on the riskiness of the portfolio and therefore is of paramount importance to risk managers. Empirical evidence shows that the joint behaviour of financial variables generally shows dependence, especially in the tails of the distribution (i.e. the extreme outcomes). McNeill et al (2005) mention that "dependence between extreme outcomes" and "concentration of risks" are indeed key issues in financial risk management. These phenomena are captured in three *stylized facts* of multivariate asset returns. Firstly, the tails of multivariate asset returns contain higher exceedance correlations and stronger quantile dependence as compared to the normal distribution. This phenomenon is also known as tail dependence. Moreover, the dependence between assets returns appears to be asymmetric, i.e. it is stronger in "bear" markets than in "bull" markets (Patton, 2004). Third, the dependence is time-varying; the level of (exceedance) correlation varies substantially over time. Adequately modelling these stylized facts of multivariate asset returns is of considerable importance in financial applications, including portfolio selection, option pricing, asset pricing models, Value-at-Risk (VaR) calculations and forecasting models (Chen and Fan 2006). Although we restrict this research to financial applications, it should be emphasized that the theory presented in this thesis can also be applied in other disciplines ranging from health sciences, hydrology, environmental science to macro economics.

In order to model dependence between extreme outcomes of different random variables, recent developments of *multivariate density models* are promising. Granger (2002) points out that it is most natural to study the entire conditional density, because (parts of) *predictive distributions* directly answer questions about forecasting moments.[1] In order to model the multivariate nature of the problem, copulas gained extreme popularity due to their flexibility. Once the distribution functions of the *individual* random variables are obtained (the marginals), copula models can be used to "couple" these marginals to a *joint* distribution function. The main advantage of copulas over classical families of multivariate distributions is that the selection of the copula model is independent from the choice of the marginal distributions[2] (Alexander, 2008). A wide variety of copulas exists, each with their own predescribed dependence properties. Therefore, accurate goodness-of-fit testing is of crucial importance to examine whether the selected model reflects the basic true data characteristics. In this thesis we combine the theory of existing goodness-of-fit tests on (1) univariate density forecasts in tails and (2) copula based multivariate density forecasts. This provides us with a testing framework for evaluating and comparing the accuracy of copula-based multivariate density models to forecast joint extreme events.

The literature on goodness-of-fit tests to evaluate the accuracy of *univariate* density forecasting models is expanding rapidly. An important group of these tests focuses on relative predictive accuracy and compares measures of relative distance between the competing density forecasts

---

[1] In contrast, traditional point forecasts are criticized for being rather uninformative without any indication of their uncertainty.

[2] In contrast, classical families of multivariate distributions, such as multivariate (log-)normal, or student's-t distributions require the individual variables (marginals) to be of the same parametric family of distributions. Copulas do not suffer from this limitation due to Sklar's theorem (1959), see section 2.2.

and the true (but unobserved) density. Proposed distance measures are, amongst others, the integrated squared difference, the mean squared error, or the Kullback-Leibler Information Criterion (KLIC). Most of these goodness-of-fit tests are designed for evaluating the *entire* distribution, whereas financial risk managers are mostly interested in the *left* tail of the distribution. Diks et al (2010) recently introduced two tests that compare the forecast accuracy of univariate density forecast in a specific region, such as the left tail. They adopt the existing weighted likelihood ratio (WLR) framework of Amisano and Giacomini (2007) based on KLIC-type distance measures, but replace the full likelihood by conditional and censored likelihood respectively. Diks et al (2010) emphasize that this approach does not suffer from a bias towards density forecasts with more mass in the region of interest, contrary to the tests based on WLR. In terms of Winkler and Murphy (1968), the likelihood-based tests introduced by Diks et al (2010) are analytically proven to be *proper*, which means that a correctly specified density forecast always receives a higher average score than an incorrectly specified density forecast.

Research on the theory and applications of copulas skyrocketed in the last few decades[3]. However, goodness-of-fit tests for copulas gained considerable academic interest only recently (Embrechts, 2008). Appendix 2 shows an up to date overview of goodness-of-fit tests for copulas. It turns out that the "jury" is still not out on the best procedure. At this point, it should be emphasized that most of the available goodness-of-fit tests only allow for *indirect* comparisons of competing copula specifications and are based on *in-sample* fit. In other words, most tests evaluate the validity of a *single* copula and do not consider the *forecasting* ability of the selected model. The approaches are mostly *indirect* in the sense that comparison of alternative copula specifications is only possible by performing the same test on competing copulas, such that the model that performs best on these statistics can be selected. Goodness-of-fit tests based on a *direct* comparison of two copulas are introduced by Patton (2006) and Chen and Fan (2006). They introduce tests based on pseudo likelihood, and rely on scoring rules based on the KLIC-difference, here defined as a distance measure between the copula model under consideration and the true (but unobserved) copula. Diks et al. (2009) provide a forecasting extension of this test, and use comparable techniques based on *out-of-sample* log-likelihood scores.

In this thesis, we develop a test to *directly* evaluate the *predictive ability* of copula based multivariate dependence models focused on the *left tail* of the copula distribution.[4] As a crucial ingredient of our test, we consider three different KLIC-based scoring rules, similar to Diks et al (2009). The first scoring rule we consider, the conventional weighted likelihood (*wl*) scoring rule, is based on full likelihood and is accused of favouring models with more probability mass in the left tail. This bias towards fat-tailed models is confirmed in both our extensive Monte Carlo experiments and our practical application of the test, in which we evaluate various multivariate density models to forecast daily MSCI Total Returns of USA, Mexico, Argentina and Brazil. When using the scoring rules based on conditional likelihood (*cl*) and the censored likelihood (*csl*) however, our Monte Carlo experiments demonstrate that these spurious rejections against fat-tailed models are virtually *absent*, in particular for large sample sizes. These results indicate that, in contrary to the *wl* scoring rule, the *cl* and *csl* scoring rules are *proper* and hence are more reliable for the purpose of evaluating joint extreme events. Among the two *proper* scoring rules

---

[3] For a study of Genest et al (2009) on the explosive development of copula theory we refer to Appendix 1.
[4] This test is also generalized to other specific regions of the distribution, but since we will restrict this research to financial applications we focus on extreme negative outcomes, i.e. the left tail.

the *csl* scoring rule exhibits the most powerful results in most cases, which could be explained by the fact that it also uses information revealed by the observations outside the region of interest.

The main contribution of this thesis to existing literature is that our goodness-of-fit test focuses on a *specific region* of the copula, which can be considered as unique in the academic literature. Furthermore, it extends existing research about goodness-of-fit tests that *directly* compare the forecasting accuracy of two competing copulas, whereas most existing tests *indirectly* compare several copula specifications. At last, we extend existing literature on tests that concentrate on *out-of-sample* fit. We test the *forecasting* accuracy of two competing models, in contrast to most tests in the current literature that evaluate *in-sample* fit.

The remainder of this thesis is organized as follows. In Section 2, we will elaborate on the fundamentals of copulas, introduce several commonly used copula families, and discuss procedures to implement copula based multivariate density models. Section 3 discusses the main methods that are used in this thesis, including the scoring rules that are used as the basis for our newly developed goodness-of-fit test. In Section 4 the finite sample properties of the proposed predictive ability tests are assessed by extensive Monte Carlo simulations. The practical usefulness of the tests is illustrated with an empirical application in Section 5. We summarize and conclude in Section 6 and provide several suggestions for further research in Section 7.

2. On Copulas

2.1 Introduction

As an introduction to copulas, we will first discuss the traditional models for multivariate dependence modeling, and their major drawbacks.

Traditionally, the multivariate dependence between random variables has been described using classical families of multivariate distributions, such as the multivariate (log-)normal, gamma, student-*t* or extreme value distributions. The main drawback of the use of classical families of multivariate distributions is that the individual behavior of the random variables under consideration (the marginals) must then be characterized with the *same* parametric family of distributions (Genest et al, 2007). This means that (1) the univariate distributions of the marginals and (2) the dependence between the marginals cannot be separated, although they are entirely different concepts. On the one hand, *univariate* empirical asset returns rarely show "nice" symmetric behavior, as large negative outcomes generally occur more often than large positive ones (*negative skewness*) and periods of large returns generally alternate periods with low returns (*volatility clustering*)[5]. On the other hand, the dependence between the marginals is rarely "nicely" linear, because of the stylized facts of *multivariate* asset returns as described in the introduction. One should be aware of the fact that multivariate asset returns usually show tail dependence, asymmetric dependence and time-varying dependence. Concluding, in order to model the complex multivariate dependence structure of financial assets, using classical families of multivariate distributions is often too restrictive, in particular for symmetric distributions.

As will turn out soon, *copula models* avoid this main restriction, as they isolate the dependence structure from the structure of the marginal distributions (Alexander, 2008). This allows us to model a joint distribution in two stages. The marginal distributions may be specified first, and the dependence between the marginals can be modeled hereafter by a function called a 'copula'. Due to its flexibility, copula theory and applications have gained considerable popularity in recent years and the amount of research on copulas skyrocketed starting the end of the nineties, especially in finance (see appendix 1 for details, Genest et al 2009).

As an extra motivation to use copula models, we return to the following question: if we are aware of the restrictions of using classical families of multivariate distributions, why are they (still) widely used in financial models such as Capital Asset Pricing Model (CAPM)? It turns out that if the multivariate normal distribution is used, linear correlation can be used as dependence measure[6]. Embrechts et al (2002) recall the advantages of linear correlation, as, among others, it is straightforward to work with and easy to manipulate under linear transformations. However, they also nicely point out several major pitfalls of linear correlation that can lead to misleading results. Working with linear correlation is therefore accused of being used for convenience rather than accuracy (Alexander, 2008). Appendix 3 summarizes the pitfalls of working with linear correlation and shows that copula models do not suffer from these.

---

[5] Other "stylized facts" of univariate empirical asset returns are fat tails (excess kurtosis) and no significant autocorrelations, see for example Embrechts (1997). For more details see section 5.2.1.

[6] In fact, linear correlation could be used for any other elliptical distribution, which is a generalization of the normal distribution (see section 2.3.1)

2.2 Copula theory

Throughout this section, we will show several fundamental results in copula theory using the simple case in which (X,Y) denotes a pair of random variables with joint distribution function H having margins F and G. Thus for all $x, y \in \mathbb{R}$,

$$H(x,y) = P(X \leq x, Y \leq y), \quad F(x) = P(X \leq x), \quad G(y) = P(Y \leq y)$$

If the domain is not further specified, it is meant to be simply $\mathbb{R}^2$.

The *copula function C* can be formally defined as follows:

Definition 2.1 *A (bivariate) copula is a cumulative distribution function (cdf) on* $[0,1]^2$ *whose margins are standard uniform.*

To verify whether a (bivariate) right-continuous function is a copula or not, we can make use of the requirements for a distribution function[7] and derive the following lemma:

Lemma 2.1 *A right-continuous function* $C: [0,1]^2 \rightarrow [0,1]$ *is a copula if and only if*

(i) $C(u, 0) = C(0, v) = 0$ *for all* $u, v \in [0,1]$
(ii) $C(u, 1) = u$ *and* $C(1, v) = v$ *for all* $u, v \in [0,1]$
(iii)   *C is* quasi-monotone: *i.e. for any* $0 \leq u_1 \leq u_2 \leq 1$ *and any* $0 \leq v_1 \leq v_2 \leq 1$,
$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$$

Note that a combination of requirement (i) and (ii) checks whether the marginals are standard uniform. The quasi-monotonicity condition (requirement (iii)) is also known as the rectangle property: it simply means that C assigns non-negative probability to any rectangle $[u_1, v_1] \times [u_2, v_2] \subseteq [0,0] \times [1,1]$ and ensures that the copula function is nondecreasing.

A nice general result follows from these properties of the copula: given two *arbitrary* univariate cdfs F and G, *any* copula can be used to construct a bivariate distribution $H(x,y) = C\{F(x), G(y)\}$ with margins equal to F and G[8]. Sklar (1959)'s theorem states that this general result is not restrictive:

Theorem 2.1 (Sklar (1959); following Nelsen (2006)) *Let H be a bivariate cdf with margins F and G. Then there always exists at least one copula C such that,*

$$H(x,y) = C\{F(x), G(y)\} \qquad for\ all\ \ x, y \in \mathbb{R} \qquad\qquad (2.1)$$

Furthermore, C is unique if F and G are continuous; otherwise C is uniquely determined on the $\mathrm{Ran}F \times \mathrm{Ran}G$, *where* $\mathrm{Ran}F$ *and* $\mathrm{Ran}G$ *denote the ranges of the marginal distribution functions F and G. Conversely, if C is a copula and given any distribution functions F and G, then the function H defined by (2.1) is a bivariate distribution function with margins F and G.*

Proof See Nelsen (2006)

---

[7] Nelsen (2006): A *distribution function* is a function F with domain $\mathbb{R}$ such that F is non-decreasing, $F(-\infty) = 0$ and $F(\infty) = 1$.

[8] This statement can be proved by verifying the requirements of a cumulative distribution function (cdf)

Note that C is unique *if and only if* the marginal distributions $F(x)$ and $G(y)$ are continuous, which we will assume throughout.

In addition, one can rewrite (2.1) as:

$$C(u,v) = H\{F^{-1}(u), G^{-1}(v)\} \qquad \text{for all } u \in \text{Ran}F, v \in \text{Ran}G \qquad (2.2)$$

Where $F^{-1}(u) = \min\{x: F(x) \geq u)$ and $G^{-1}(v) = \min\{y: G(y) \geq u)$ denote the generalized inverses of the marginal distribution functions F and G.

Sklar's theorem implies that the choice of the copula is not constrained by the choice of its marginals: given any set of uniform marginal distributions (F(x), G(y)) and any copula C, equation (2.1) can be used to construct a joint distribution with the given marginal distributions. This is great news for the flexibility of the copula. In stark contrast to standard multivariate distributions, every copula can couple *whatever* continuous uniform marginals together with their own predescribed interdependence. Intuitively, copulas "glue" the marginals together and "shape" them to their own taste (Embrechts, 2008).

Copulas exist in many forms, each with their own dependence structure. Tail dependence is an important distinctive characteristic of different copula specifications. To allow for asymmetric tail dependence, it is often measured by coefficients of upper and lower tail dependence. The coefficient for lower tail dependence can be defined by $\lambda_L = \lim_{q \downarrow 0} C(q,q)/q$. Note that λ_L basically corresponds to the conditional probability that the random variable X is a tail event, given that Y takes a value in its lower tail (Alexander, 2008). To measure upper tail dependence, it is convenient to work with the *survival copula* (or *rotated* copula*)* of a copula C. The survival copula can be interpreted as the mirror image of the copula C, and is defined as:

$$\bar{C}(u,v) = u + v - 1 + C(1-u, 1-v), \qquad for\ all\ u \in \text{Ran}F, v \in \text{Ran}G \qquad (2.3)$$

The coefficient for upper tail dependence can now simply be defined as $\lambda_U = \lim_{q \downarrow 0} \bar{C}(q,q)/q$. Symmetric copulas have equal tail dependence and, more generally, admit the representation $C(u_1, u_2) = \bar{C}(u_1, u_2)$. The characteristics of the popular *copula families*, such as Elliptical copulas and Archimedean copulas will be discussed in section 2.3.

The flexibility of copulas as dependence measure attracted much interest in academic literature. However, it should be noted that copulas also have their shortcomings. Mikosch (2006) for example, provides several critical notes, and reminds us of the fact that copulas are designed to describe dependence between random *variables* and not *processes*, i.e. the approach is static rather than dynamic. For the purpose of time series modelling, Patton (2006) therefore introduced a *conditional copula* specification as a multivariate distribution of uniform random variables conditional on some information set $\mathcal{F}_{t-1}$. Patton provides an extension of Sklar's theorem, allowing to decompose $H(x,y|\mathcal{F}_{t-1})$ into the two conditional marginal distributions $F(x|\mathcal{F}_{t-1})$ and $G(y|\mathcal{F}_{t-1})$ and the conditional copula $C(.|\mathcal{F}_{t-1})$:

$$H(x,y|\mathcal{F}_{t-1}) = C\{F(x|\mathcal{F}_{t-1}), G(y|\mathcal{F}_{t-1})|\mathcal{F}_{t-1}\} \qquad for\ all\ x,y \in \mathbb{R} \qquad (2.4)$$

In our implementation of multivariate density models in section 2.4 we focus on the conditional copula specification, as it is more natural to consider dynamic forecasting models in the context of time series variables.

2.3 Copula Families

In this section, we introduce several classical families of copulas and discuss their basic characteristics. For greater generality, we will consider the multivariate case, where x denotes $(x_1, ..., x_d) \in \mathbb{R}^d$ and X stands for a random vector $(X_1, ..., X_d)$ in $\mathbb{R}^d$.

2.3.1 Elliptical Copulas

Elliptical copulas are implicit copulas, in the sense that they are built from multivariate elliptical distributions (Alexander, 2008). Since elliptical distributions generalize the multivariate normal distribution, we first introduce the normal (or Gaussian) copula.

Consider a random vector $(X_1, ..., X_d)$ from the multivariate normal distribution with mean vector μ and positive definite correlation matrix Σ. The marginal densities are standard normal $\varphi(x) = \frac{1}{2\pi} \exp{(-x^2/2)}$ while the joint density is d-variate normal:

$$h_{\Sigma}(x_1, ..., x_d) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{2}\right), \qquad x_1, ..., x_d \in \mathbb{R} \qquad (2.5)$$

Note that the distribution functions of neither $H_{\Sigma}(x_1, ..., x_d)$ nor $\Phi(x)$ are explicit, they can only be expressed as an integral. The copula of $(X_1, ..., X_d)$ is called the *Gaussian (Normal) copula* with the correlation matrix Σ for parameters and is defined as:

$$C_{\Sigma}^{Ga}(u_1, ..., u_d) = H_{\Sigma}\{\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)\}, \qquad u_1, ..., u_d \in (0,1) \qquad (2.6)$$

Where $\Phi^{-1}(u) = \min\{x: \Phi(x) \geq u)$ is the generalized inverse of the standard normal distribution function Φ. The expression of the Gaussian copula is not explicit, but can be obtained rather easily using the so called *inversion method* in (2.6).

The *Gaussian (normal) copula density* can be obtained by differentiating (2.6) and is of the form:

$$c_{\Sigma}(u_1, ..., u_d) = \frac{h_{\Sigma}\{\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)\}}{\varphi\{\Phi^{-1}(u_1)\}...\varphi\{\Phi^{-1}(u_d)\}} \qquad u_1, ..., u_d \in (0,1) \qquad (2.7)$$

Such that c_{Σ}^{Ga} boils down to:

$$c_{\Sigma}^{Ga}(u_1, ..., u_d) = \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{\xi'(\Sigma^{-1}-I)\xi}{2}\right), \qquad u_1, ..., u_d \in (0,1) \qquad (2.8)$$

Where $\xi = (\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d))$ and I the $d \times d$ identity matrix. Also for marginals that are not standard normal, the Gaussian copula density can be expressed as in (2.8). Given d continuous marginal distributions $F_i(x_i)$ and correlation matrix Σ, Alexander (2008) provides a stepwise approach to construct a Gaussian Copula density:

Algorithm to construct a Gaussian Copula density (Alexander, 2008)

Step 1: Set $u_i = F_i(x_i)$ for $i = 1 ... d$ for the (not necessarily normal) marginals

Step 2: Apply the inverse Gaussian distribution, $\xi_i = \Phi^{-1}(u_i)$ for $i = 1, ..., n$

Step 3: Use the correlation matrix Σ and the vector ξ in the copula density (2.8)

Figure 2.1: Algorithm to construct a Gaussian Copula density, given d marginal distributions $F_i(x_i)$ and correlation matrix Σ as introduced by Alexander (2008).

The derivation of a Gaussian Copula as described above can be generalized to the elliptical copula. To understand this connection, we first briefly describe the elliptical distribution.

If $\boldsymbol{X} \sim N_d(\mu, \Sigma)$ then $\boldsymbol{X} = \mu + A\boldsymbol{Z}$, where $\boldsymbol{Z} \sim N_d(\boldsymbol{0}, I_d)$ and A the Cholesky factor of Σ, i.e. $AA^T = \Sigma$. We introduce two random variables: $\boldsymbol{U} = \boldsymbol{Z}/\|\boldsymbol{Z}\|_2$ and $R = \|\boldsymbol{Z}\|_2$ with $\|z\|_2 = \sqrt{z_1^2 + \cdots + z_1^d}$, such that the elliptical distribution can now be formally defined as:

Definition 2.2 *We say that* $\boldsymbol{X}$ *follows an* elliptical distribution *with mean vector μ, positive definite dispersion matrix Σ and radial part R if and only if $\boldsymbol{X}$ admits the following representation:*

$$\boldsymbol{X} = \mu + RA\boldsymbol{U} \tag{2.9}$$

Where A is the Cholesky factor of Σ and U is a random vector independent of R and uniformly distributed on the sphere $S_2 = \{z \in \mathbb{R}^d : \|z\|_2 = 1\}$.

Similar to the multivariate normal distribution, cdfs of elliptical distribution are (rarely) explicit. The density however often is. If $\boldsymbol{X}$ has density h then there exists a non-negative function g, the so-called *density generator*, such that:

$$h(x) = \frac{1}{|\Sigma|^{1/2}} g\{(\boldsymbol{x} - \mu)'\Sigma^{-1}(\boldsymbol{x} - \mu)\}, \qquad \boldsymbol{x} \in \mathbb{R}^d \tag{2.10}$$

For a proof and further details, see Fang et al (1990). For example, the *density generator* to obtain the multivariate normal distribution is $g = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{x}{2}\right)$. Indeed, the d-variate normal distribution in (2.5) is obtained by substituting this specific density generator in function $h(x)$ of (2.10). In table 1 we list three common elliptical distributions that we use in this thesis, in terms of their density generator g. The Student-t Copula $C_{\Sigma,v}^T$ and the Cauchy Copula C_{Σ}^{Ca} can also be obtained using *inversion method* in (2.5). One should however substitute the inverse Gaussian distribution ($\Phi^{-1}(u)$) for $T_v^{-1}(u)$ and $T_1^{-1}(u)$ respectively, where $T_v^{-1}(.)$ represents the inverse Student-t distribution. The densities of the Student-t and Cauchy Copula can also be obtained similarly to the Gaussian copula, by adopting the algorithm in figure 1 with two modifications. First, one should again use $T_v^{-1}(u)$ and $T_1^{-1}(u)$ instead of $\Phi^{-1}(u)$. Second, in Step 3 the more general representation in (2.7) should be used instead of the Gaussian copula density in (2.8).

Family	Elliptical density generator $g(x)$	Tail dependence coefficients
Gaussian	$\dfrac{1}{(2\pi)^{d/2}} \exp\left(-\dfrac{x}{2}\right)$	$\lambda_L = \lambda_U = 0$
Student-t	$\dfrac{\Gamma\{(d+v)/2\}}{(\pi v)^{d/2}\Gamma(v/2)}\left(1 + \dfrac{x}{v}\right)^{-(d+v)/2}$	$\lambda_L = \lambda_U = 2T_{v+1}\left(-\sqrt{(v+1)\dfrac{(1-\rho_{12})}{(1+\rho_{12})}}\right)$
Cauchy	$\dfrac{\Gamma\left\{\dfrac{d+1}{2}\right\}}{(\pi)^{\frac{d+1}{2}}}(1+x)^{-\frac{d+1}{2}}$	$\lambda_L = \lambda_U = 2T_2\left(-\sqrt{\dfrac{2(1-\rho_{12})}{(1+\rho_{12})}}\right)$

Table 2.1: Three common families of Archimedean copulas with their Archimedean generator and tail dependence coefficients. NB: the density generator is given in a general *d*-dimensional setting, whereas the tail dependence coefficients are given for the bivariate copula, that is fixing d=2.

The Gaussian and Cauchy Copula only have the correlation matrix Σ as parameters, while the Student-t copula has an additional degrees of freedom parameter v to control the degree of tail dependence. Note that as a general result, the Student-t copula nests the Gaussian Copula when $v = \infty$ and the Student-t Copula is equal to the Cauchy Copula when $v = 1$. The relation between the copulas can be nicely interpreted in terms of tail dependence. On one side of the spectrum, there is the Gaussian Copula with zero ("minimum") tail dependence, that is $\lambda_L = \lambda_U = 0$, as described in section 2.1. At the other side of the spectrum we have the Cauchy copula with ("maximum") positive tail dependence, see table 1. The tail dependence of the Student-t Copula can attain all values within this spectrum, and therefore is always positive and decreasing in the degrees of freedom parameter v. The Student-t Copula is thus more flexible than both the Cauchy copula and the Gaussian Copula, at the cost of estimating an extra parameter v.

Figure 2.2 visualizes the differences between the three copulas, showing the densities of the Gaussian, Student-t and Cauchy copula and a sample of size $n = 10,000$ from the bivariate copulas C_{Σ}^{Ga}, $C_{\Sigma,v}^{T}$ and C_{Σ}^{Ca}. From left to right we observe an increasing probability mass in the corners near (0,0) and (1,1) (which in the remainder we will call the *left* and the *right* tail respectively), implying increasing tail dependence.

As the general description of elliptical distributions suggests, there exist many other elliptical distributions such as the Hyperbolic or the Pearson Type II distribution. Furthermore, one could consider mixtures of elliptical distributions, or introduce *skewed* elliptical distributions. For the purpose of this thesis however, we restrict ourselves to the symmetric elliptical distributions as given in table 2.1. We consider these three specific distributions as they are often used in practice (in particular the Gaussian and Student-t copula), and they can be easily compared.
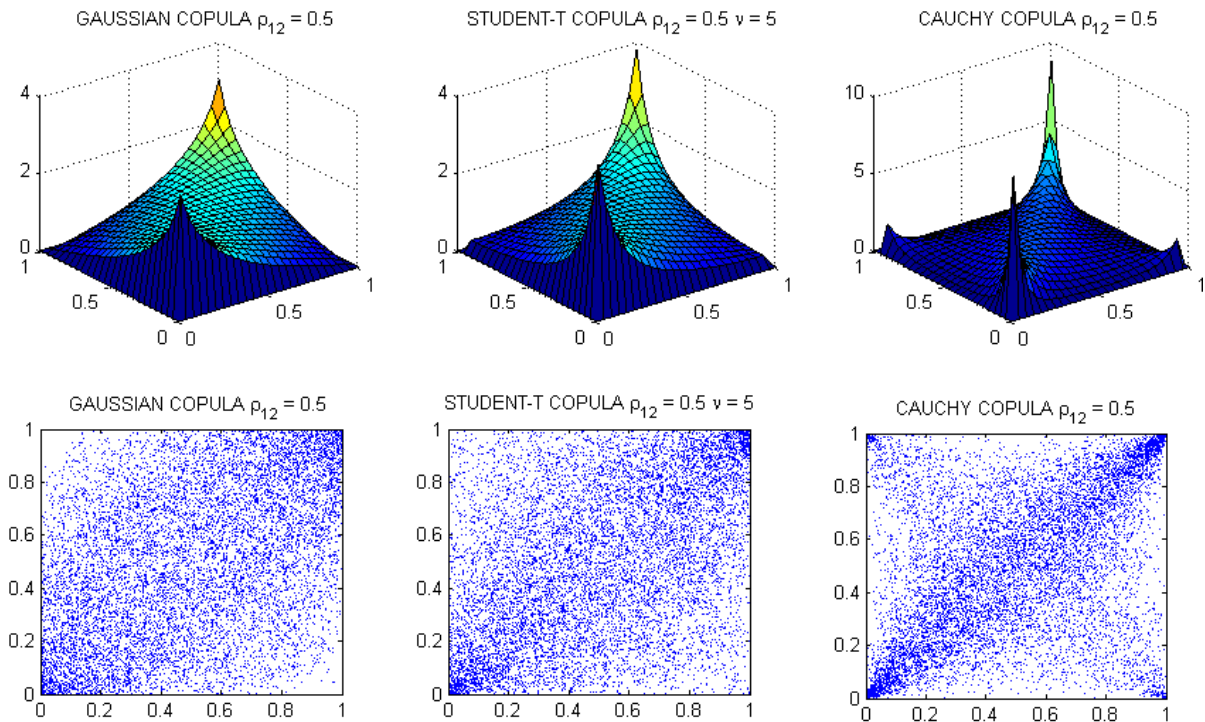

Figure 2.2: The three upper graphs show the density of the (bivariate) Gaussian Copula, the Student-t Copula and the Cauchy copula respectively. The lower scatter plots show a simulation sample of *n*=10,000 of the three corresponding copula specifications. In all graphs, the parameters are fixed to $\rho_{12} = 0.5$ and $v = 5$ (for the student-t Copula).

2.3.2 Archimedean Copulas

Where elliptical copulas are derived from multivariate distributions, Archimedean copulas are based on an *Archimedean generator* function. If this generator function, which we denote by $\psi(u)$ throughout, satisfies certain conditions, Archimedean copulas can be built. This concept is formalized in the following definition.

Definition 2.3 *A non-decreasing and continuous function* $\psi: [0,\infty) \to [0,1]$ *which satisfies the conditions* $\psi(0) = 1$ *and* $\lim_{x\to\infty}\psi(x) = 0$ *and is strictly decreasing on* $[0, \inf\{x: \psi(x) = 0\}]$ *is called an* Archimedean generator. *A d-dimensional copula C is called* Archimedean *if it permits the representation*

$$C(u_1, \dots, u_d) = \psi^{-1}\{\psi(u_1) + \cdots + \psi(u_d)\}, \qquad u_1, \dots, u_d \in 0,1) \qquad (2.11)$$

For some Archimedean generator ψ *and its (pseudo) inverse* $\psi^{-1}: (0,1] \to [0,\infty)$ *where by convention* $\psi(\infty) = 0$ *and* $\psi^{-1}(0) = \inf\{u: \psi(u) = 0\}$.

The density function of the Archimedean copula as given in (2.11) is

$$c(u_1, \dots, u_d) = \psi^{-1}_{(n)}\{\psi(u_1) + \cdots + \psi(u_d)\}\prod_{i=1}^{d}\psi'(u_i) \qquad u_1, \dots, u_d \in 0,1) \qquad (2.12)$$

Since any convex, monotonic decreasing function can be used as Archimedean generator, a large variety of Archimedean copulas exists. In this thesis we consider three common families of Archimedean copulas, the Clayton, Gumbel and Frank copula. Their corresponding Archimedean generator functions and tail dependence coefficients are summarized in table 2.2.

The Clayton copula is given as follows:

$$C_\theta^{Cl}(u_1, \dots, u_d) = (u_1^{-\theta} + \cdots + u_d^{-\theta} - d + 1)^{-1/\theta} \qquad (2.13)$$

The Clayton copula contains asymmetric dependence, as it does not contain upper tail dependence, i.e. $\lambda_U = 0$, but positive lower tail dependence of $\lambda_L = 2^{-1/\theta}$, which is increasing in its parameter θ. This implies that the Clayton copula is best suited for applications in which negative outcomes are likely to occur together.

Family	Archimedean generator $\psi(x)$	Parameter Range	Tail dependence coefficients
Clayton	$\dfrac{1}{\theta}(x^{-\theta} - 1)$	$\theta \geq -1$	$\begin{cases}\lambda_L = 2^{-1/\theta}, \lambda_U = 0 & \text{if } \theta > 0 \\ \lambda_L = \lambda_U = 0 & \text{if } \theta \leq 0\end{cases}$
Gumbel	$(-\log x)^\theta$	$\theta \geq 1$	$\lambda_L = 0,\ \lambda_U = 2 - 2^{1/\theta}$
Frank	$-\log\left(\dfrac{e^{-\theta x} - 1}{e^{-\theta} - 1}\right)$	$\theta \in \mathbb{R}$	$\lambda_L = \lambda_U = 0$

Table 2.2: Three common families of Archimedean copulas with their Archimedean generator and tail dependence coefficients.

The expression of the Gumbel copula is given as:

$$C_\theta^{Gu}(u_1, \ldots, u_d) = \exp\left(-\left[(-\log u_1)^\theta + \cdots + (-\log u_d)^\theta\right]^{1/\theta}\right) \tag{2.14}$$

The Gumbel copula also contains asymmetric tail dependence. It has lower tail independence, i.e. $\lambda_L = 0$, but upper tail dependence of $\lambda_U = 2 - 2^{1/\theta}$. Consequently, standard Gumbel copulas are mostly used for applications in which extreme positive outcomes are likely to occur together. One can also use the survival Gumbel copula (or *rotated* Gumbel copula), which is a mirror image of the standard Gumbel copula. Consequently, tail dependence is reversed, such that $\lambda_L = 2 - 2^{1/\theta}$ and $\lambda_U = 0$.

The Frank copula is given as:

$$C_\theta^{Fr}(u_1, \ldots, u_d) = -\frac{1}{\theta}\log\left(1 + \frac{(\exp(-\theta u_1)-1)\ldots(\exp(-\theta u_d)-1)}{\exp(-\theta)-1}\right) \tag{2.15}$$

The Frank copula is symmetric and shows tail independence, similar to the elliptical Gaussian copula. As compared to the Gaussian copula however, it turns out that for increasing parameters the dependence in the center of the distribution is getting relatively stronger. The Frank copula is therefore best suited for applications in which tail dependence is weak.

The differences between the different Archimedean copulas become apparent in Figure 2.3. We display both the Copula densities and a sample of $n = 10,000$ from C_θ^{Cl}, C_θ^{Gu} and C_θ^{Fr}. In order to make the parameter values of the different copulas comparable, we express the parameter values in terms of the Kendall's tau τ, a commonly used measure of concordance for copulas (see Appendix 3 for details). The relation between Kendall's tau τ and the Archimedean generator function ψ can be given as:

$$\tau(C) = 1 + 4\int_0^1 \frac{\psi_\theta(t)}{\psi_\theta'(t)}\,dt \tag{2.16}$$

Using this result we may derive the following expressions:

Clayton copula: $\theta = \frac{2\tau}{1-\tau}$ (2.17)

Gumbel copula: $\theta = \frac{1}{1-\tau}$ (2.18)

Frank copula: $\frac{[D(\theta)-1]}{\theta} = \frac{1-\tau}{4}$ with $D(\theta) = \frac{1}{\tau}\int_0^\theta \frac{t}{e^t-1}\,dt$ (2.19)

Figure 2.3 shows the results for $\tau = 0.5$. Note that $\tau = 0.5$ corresponds to the linear correlation parameter of $\rho_{12} = 0.71$, as the relation between Kendall's tau and ρ is given by $\rho = \sin\left(\frac{\pi}{2}\tau\right)$. Appendix 4 provides figures of the six copulas considered in this thesis for different values of τ, such that the dependence structures of the copula specification can be directly compared.
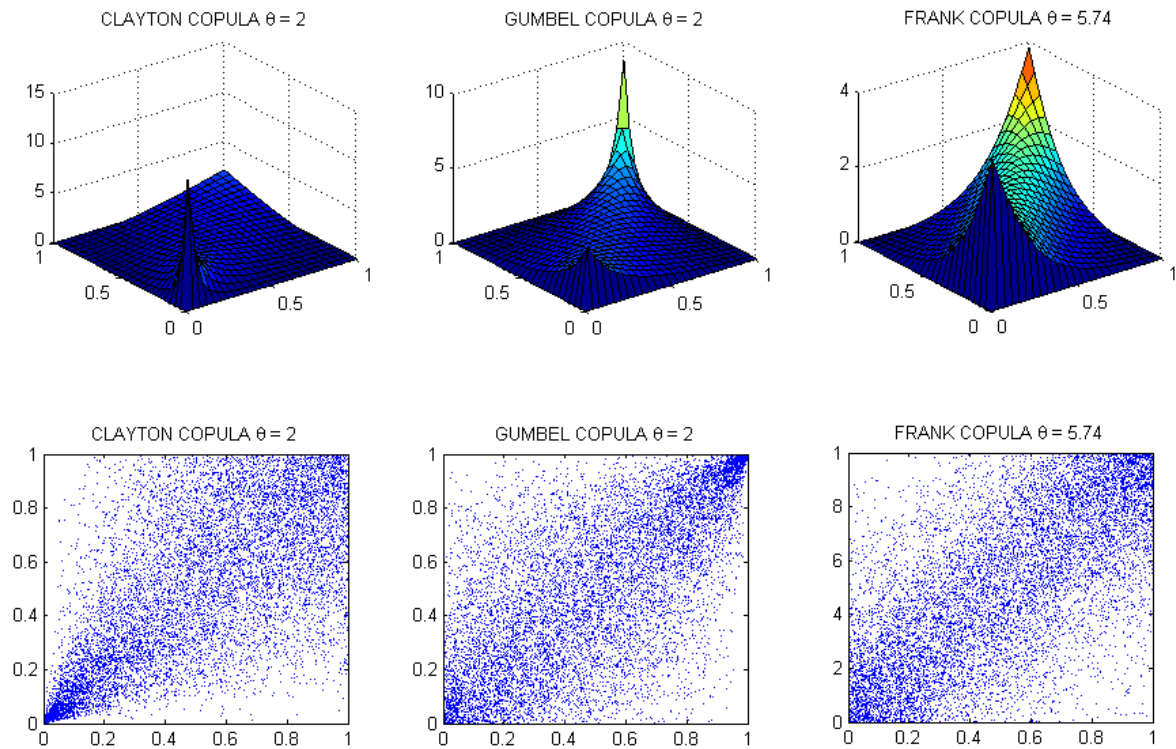
Figure 2.3 The three upper graphs show the density of the (bivariate) Clayton Copula, the Gumbel Copula and the Frank copula respectively. The lower scatter plots show a simulation sample of *n*=10,000 of the three corresponding copula specifications. In all graphs, the copula parameter values are set to match Kendall's tau fixed to $\tau = 0.5$.

2.4 Estimating Copula-based multivariate density models

An important assumption we make in this thesis, is that the parameters of the marginals can be separated from each other and from those of the copula function.[9] This implies that the estimation of the marginal distributions and the copula parameters can proceed separately. In the context of this thesis, this assumption implies that the calibration of the parameters of a copula based multivariate density model can be separated in multiple stages. The multistage approach has the main advantage that it makes the estimation procedure more flexible. However, a relative disadvantage is that the multistage procedure comes at the cost of a loss of efficiency, as compared to the single step procedure in which the parameters of the marginal distributions and their copula are estimated simultaneously. Parameter estimates of a copula based density model may be obtained by either parametric, semi-parametric or nonparametric estimation. In the remainder of this section we briefly discuss each of these three estimation approaches, and provide several popular single stage and multistage estimation procedures.

2.4.1 Parametric estimation

The advantage of a fully parametric copula-based multivariate density model is that maximum likelihood (ML) can be applied to obtain efficient estimates of the (conditional) copula and their (conditional) marginals. The parameters of a d-dimensional multivariate density model based on a conditional copula C are obtained by maximizing the following likelihood:

$$\sum_{t=1}^{n} \sum_{j=1}^{d} \log f_j(y_{j,t}|\mathcal{F}_{t-1}) + \sum_{t=1}^{n} \log c\big(F_1(y_{1,t}|\mathcal{F}_{t-1}), F_2(y_{2,t}|\mathcal{F}_{t-1}), \dots, F_d(y_{d,t}|\mathcal{F}_{t-1})|\mathcal{F}_{t-1}\big) \qquad (2.20)$$

Where $F_j(y_{j,t}|\mathcal{F}_{t-1})$ and $f_j(y_{j,t}|\mathcal{F}_{t-1}), j = 1, \dots, d$, are the conditional marginal distributions and densities respectively, and $c(y_{j,t}|\mathcal{F}_{t-1})$ is the conditional copula density, defined as:

$$c(u_1, \dots, u_d|\mathcal{F}_{t-1}) = \frac{\partial^d}{\partial u_1 \dots \partial u_d} C(u_1, \dots, u_d|\mathcal{F}_{t-1}) \qquad (2.21)$$

The most straightforward way to estimate a fully parameterized copula-based multivariate density model, consists of a single step in which all parameters in the above expression are estimated simultaneously. Under the assumption that the parameters of the marginals can be separated from each other and from those of the copula function, multistage procedures can be introduced. A widely used method is the Inference Functions for Margins (IFM) procedure (Joe, 1997), given in figure 2.4. This approach basically divides the likelihood of (2.20) in a "marginal part" (the left part of the expression) and a "copula part" (the right part of the expression) and estimates the corresponding parameters in separate stages. The advantage of this approach is that it is computationally relatively easy and tractable, but at the cost of efficiency compared to the single-step approach. One could also use the two-step approach to decide on the most appropriate copula family and then estimate all parameters (marginal and the selected copula) in a final fully parametric round of estimation (McNeil et al, 2005).

---

[9] This assumption is justified by our argument that in principle, the dependence structure of the copula has nothing to do with the individual dependence structure of the marginals, see section 2.1.

Figure 2.4: Stepwise IFM method (see Joe, 1997) for parametric estimation of both the marginals and the copula

The main drawback of a fully parametric approach is that the results depend on correctly specified distributions of both the marginals and the copula specification, which turns out to be rather difficult in practice.

2.4.2 Semi-parametric estimation

If one is highly uncertain about the functional form of the marginals, one could consider non-parametric estimation of the marginals. This method is rank-based and makes use of the empirical cdf to obtain an estimate of $F_j(.)$[10]:

$$F_{j,n}^*(y) = \frac{1}{n+1} \sum_{t=1}^{n} I\big(Y_{j,t} \leq y\big) \tag{2.22}$$

If the copula is estimated parametrically, this estimation approach belongs to the class of *semi-parametric* estimation methods. A two step estimation procedure (Genest et al, 1995) is given in figure 2.5 and is usually called the *Canonical Maximum Likelihood*. Its main advantage is that it does not assume a structure of the marginals (Alexander, 2008), i.e. the procedure is robust to misspecification of the marginal distribution. Working with ranks has several advantages; they are always defined and are invariant by monotone increasing transformations of the marginals (Genest and Favre, 2007). However, a major disadvantage of relying solely on ranks is that one could ignore some valuable information about the variables of interest.

A compromise between fully parameterized and non-parameterized estimation of the marginals is also possible. Chen and Fan (2006) introduce the semi-parametric copula-based multivariate dynamic (SCOMDY) model which assumes a parameterized conditional mean and conditional variance of the random variables under consideration. The procedure consists of three steps which are summarized in figure 2.6. First, the conditional mean and variance are estimated parametrically using Quasi Maximum Likelihood. The (standardized) innovations that result from the first stage can now be used to estimate the marginals non-parametrically. In the third step, Maximum Likelihood can be used to estimate the copula parameters.

Figure 2.5: Stepwise approach for the Canonical Maximum Likelihood estimation procedure (Genest et al, 1995)

---

[10] Note that the division is by n+1 instead of n to ensure that the marginals lie strictly in the unit cube, see McNeil et al (2005)

As any conditional mean and conditional variance specification can be used in the first step, the structure of the SCOMDY model is very flexible. For example, one can choose a certain (asymmetric) ARMA-GARCH model to capture the possible nonlinear, asymmetric dependence structures of the marginals.

Semi-Parametric Copula-based Multivariate Dynamic (SCOMDY) (Chen and Fan, 2006)

Stage 1: Univariate Quasi Maximum Likelihood (QML): estimate the parameters of the conditional mean and conditional variance of the variables of interest.

Stage 2: Apply (2.22) to the residuals from Stage 1 to obtain the marginal distributions $F_j(.)$.

Stage 3: ML: Estimate parameters of the copula (right part of (2.20))

Figure 2.6: Stepwise SCOMDY procedure (see Chen and Fan, 2006) for semi-parametric estimation of the marginals and parametric estimation of the copula

2.4.3 Non-parametric estimation

If the dependence structure of both the marginals and their copula specification are unknown and highly uncertain, the copula C may be estimated non-parametrically by the *empirical copula*. This method is based on a transformation of a random sample of $(X_1, Y_1), ..., (X_n, Y_n)$ to the pairs of ranks $(R_1, S_1), ..., (R_n, S_n)$. De Heuvels (1979) formally defined the empirical copula as:

$$C_n(u, v) = \frac{1}{n+1} \sum_{i=1}^{n} I\left(\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v \right), \quad u, v \in [0,1] \tag{2.23}$$

Where C_n is an asymptotically unbiased estimate of C. In the empirical copula (2.23) both the marginals and the copula are estimated non-parametrically, which has the obvious advantage that it is free of parameter uncertainty. Moreover, the approach is robust to misspecification of the marginals. However, relying solely on ranks also has its disadvantages. As highlighted in the last subsection, nonparametric estimation could ignore valuable information about either the marginal distributions or their copula structure. Furthermore, Alexander (2008) points out that in practice the empirical copula can be very "spiky" in the sense that small changes in the sample can cause a substantial change in the estimated parameters. Another disadvantage is that a loss of efficiency occurs compared to the single stage parametric ML approach.

3 Methods

The main goal of this thesis is to develop a new goodness-of-fit test that compares the predictive ability of copula based multivariate dependence models in the left tail. That is, we basically focus on the (out-of-sample) *goodness-of-fit step*, that shows which model best captures the basic data characteristics in the left tail. To this end, we introduce scoring rules and equal predictive ability tests based on KLIC scores, measuring the distance between the true probability density and the candidate density. In this section, we first discuss the estimation and prediction procedures that precede goodness-of-fit testing in Section 3.1, and elaborate on our newly developed goodness-of-fit tests in Section 3.2.

3.1 Estimation Framework and Forecast Methods

As described in the last section, a wide range of estimation procedures is available to fit a copula model to data. Our newly developed goodness-of-fit test has the advantage that it is valid under any of the parametric, semi-parametric and nonparametric estimation procedures of copula models as described in section 2.4. This is due to the fact that the test is valid under general conditions by adopting the forecasting framework of Giacomini and White (2006). This specific framework compares the accuracy of *forecast methods* rather than *forecast models*. A forecast method is defined to be the set of choices the forecaster makes at the time of prediction, including the density model itself, the parameter estimation, and the estimation window. In the framework of Giacomini and White (2006) the (rolling) estimation window is required to be finite, and the model parameters are estimated in a moving window of fixed size. The advantage of comparing forecast *methods* rather than forecast *models*, is that in the former parameter estimation is treated as an integral part of the density forecast. Consequently, parameter estimation uncertainty is not an issue when comparing competing (copula based) density forecasts, as it is part of the respective competing forecasting method. Comparing forecast methods has the further advantage that it allows the comparison of both nested and non-nested copula models. (Diks et al, 2009, 2010)

Adopting this framework of forecast methods simplifies our research considerably. We can create an environment in which we can keep all things equal, except for the different selected competing copula based density forecast models. To this end, we define an in-sample estimation window of a fixed size of R past observations in which we estimate both the parameters of the marginal distributions and the copula parameters. After the in-sample *estimation step*, we create P one-step-ahead forecasts of the selected copula based density forecast models and apply our test of equal predictive accuracy. In this way we isolate the performance of the selected density forecasting models so we can focus on the *goodness-of-fit step* and need not to worry about estimation uncertainty.

3.2 Scoring rules and equal predictive accuracy tests

In this section we develop scoring rules that *directly* compare the *out-of-sample tail behavior* of two competing copula specifications for multivariate density forecasts. These scoring rules can be used as the basis for a formal goodness-of-fit test to evaluate the forecasting ability of alternative copula specifications in tails.

In the literature, several likelihood-based scoring rules have appeared to measure the distance between the proposed estimated density forecast and the true (but unobserved) density. We will focus on scoring rules based on the Kullback-Leibler Information Criterion (KLIC), which measures the distance between the density forecast and the true (but unknown) density. KLIC-type scoring rules have a clear intuition and are particularly useful for testing relative predictive accuracy, i.e. comparing one predictive density *directly* with another. We combine the KLIC-based scoring rules introduced in Diks et al (2009) to compare *multivariate* density forecasts for the *entire* density with the techniques proposed in Diks et al (2010) for evaluating *univariate* density forecasts in tails.

In the next subsection, the KLIC-based scoring rule to compare forecasts of entire univariate distributions will be further elaborated. This basic scoring rule is extended two different ways in the forthcoming subsections. First, the KLIC-based scoring rule will be extended to compare entire copula-based *multivariate* densities, following Diks et al (2009). Second, we extend the univariate scoring rule to focus on specific regions of the distribution, such as the left tail, following Diks et al (2010). Finally, we will come up with a newly developed scoring rule to evaluate multivariate density forecasts in a specific region, based on Kullback-Leibler divergence.

3.2.1 Comparing univariate density forecasts

The univariate Kullback-Leibler Information Criterion (KLIC) for a certain predictive density $\hat{f}_t$ is given as:

$$\text{KLIC}(\hat{f}_t) = \int p_t(y_{t+1}) \log \frac{p_t(y_{t+1})}{\hat{f}_t(y_{t+1})} dy_{t+1} = \mathbb{E}_t\left[\log p_t(Y_{t+1}) - \log \hat{f}_t(Y_{t+1})\right] \tag{3.1}$$

where p_t is the true conditional density. For the purpose of goodness-of-fit testing, the goal is to minimize $\text{KLIC}(\hat{f}_t)$. Intuitively, the smaller the KLIC score, the closer is $\hat{f}_t$ to the true conditional density p_t. Therefore, a higher value of $\mathbb{E}_t(\log \hat{f}_t(Y_{t+1}))$ is preferred. In practice however, p_t is unknown, hence $\mathbb{E}_t(\log \hat{f}_t(Y_{t+1}))$ is unknown. In order to circumvent this problem, we introduce the so called *KLIC-score* S_{t+1}:

$$S_{t+1} = \log \hat{f}_t(Y_{t+1}) \tag{3.2}$$

Which has conditional mean $\mathbb{E}_t(S_{t+1}) = \mathbb{E}_t(\log \hat{f}_t(Y_{t+1}))$. The KLIC-score has the advantage that it is observable, given a certain predictive density. Furthermore, it has a clear intuition as it boils down to a straightforward logarithmic scoring rule. S_{t+1} is high if an observation falls in a region with high predictive density, and attains a low value if it falls in a region with low predictive density. Obviously, the highest KLIC-score is preferred.

Two competing predictive densities $\hat{f}_{A,t}$ and $\hat{f}_{B,t}$ can be statistically compared using a Diebold Mariano type test statistic[11] with the null hypothesis

$$H_0: \mathrm{E}(\log \hat{f}_{A,t}(Y_{t+1})) = \mathrm{E}(\log \hat{f}_{B,t}(Y_{t+1})) \qquad \text{for all } t = \mathrm{R}, \mathrm{R} + 1, \dots, \mathrm{R} + \mathrm{P} - 1 \qquad (3.3)$$

with R the length of the rolling in-sample period, and P the number of observations in the forecasting period. We can define the score differences as:

$$d_{t+1} = \log \hat{f}_{A,t}(Y_{t+1}) - \log \hat{f}_{B,t}(Y_{t+1}) \qquad (3.4)$$

Next we can use the sample average of the score difference, $\bar{d}_{R,P} = \frac{1}{P}\sum_{t=R}^{T-1} d_{t+1}$, to construct a Diebold Mariano (DM) type test statistic:

$$t_{R,P} = \frac{\bar{d}_{R,P}}{\sqrt{\hat{\sigma}_{R,P}^2/P}} \qquad (3.5)$$

Where $\hat{\sigma}_{R,P}^2$ is a heteroskedasticity and autocorrelation-consistent (HAC) variance estimator of the asymptotic variance $\sigma_{R,P}^2 = Var(\sqrt{n}\bar{d}_{R,P})$, which satisfies $\hat{\sigma}_{R,P}^2 - \sigma_{R,P}^2 \xrightarrow{p} 0$ (Diks et al, 2010). Giacomini and White (2006) show that the test statistic $t_{R,P}$ is asymptotically standard normally distributed under the null-hypothesis of equal KLIC-scores[12].

3.2.2 Comparing multivariate density forecasts

In order to extend the basic KLIC score of (3.2) to a higher dimension d, we define a d-dimensional vector of interest $\boldsymbol{Y}_t = (Y_{1,t}, Y_{2,t}, \dots, Y_{d,t})'$ with corresponding marginal distributions $F_1, F_2, \dots, F_d$. We consider a conditional copula based model C_t as in (2.4) with log-likelihood:

$$\sum_{j=1}^d \log \hat{f}_j(y_{j,t}) + \log c_t(F_1(y_{1,t}), F_2(y_{2,t}), \dots, F_d(y_{d,t})) \qquad (3.6)$$

Where c_t is the conditional copula density. Analogue to (3.1), the KLIC of a multivariate one-step-ahead predictive density $\hat{f}_t$ for $\boldsymbol{Y}_{t+1}$ is given as:

$$\mathrm{KLIC}(\hat{f}_t) = \mathbb{E}_t\left[\log p_t(\boldsymbol{Y}_{t+1}) - \left\{\sum_{j=1}^d \log \hat{f}_{j,t}(Y_{j,t+1}) + \log \hat{c}_t(\widehat{\boldsymbol{U}}_{t+1})\right\}\right] \qquad (3.7)$$

With $\widehat{\boldsymbol{U}}_{t+1} = (F_{1,t}(y_{1,t+1}), F_{2,t}(y_{2,t+1}), \dots, F_{d,t}(y_{d,t+1}))'$ the conditional multivariate PIT of $\hat{c}_t$. Again, the goal is to minimize $\mathrm{KLIC}(\hat{f}_t)$, so a higher value of $\mathbb{E}_t\left[\sum_{j=1}^d \log \hat{f}_{j,t}(Y_{j,t+1}) + \log \hat{c}_t(\widehat{\boldsymbol{U}}_{t+1})\right]$ is to be preferred. Along the same reasoning as the univariate case, we introduce KLIC-score S_{t+1}:

$$S_{t+1} = \sum_{j=1}^d \log \hat{f}_{j,t}(Y_{j,t+1}) + \log \hat{c}_t(\widehat{\boldsymbol{U}}_{t+1}) \qquad (3.8)$$

---

[11] Note that this is a general result, provided that the densities under consideration are normalized to have unit total probability: by taking differences of the $\mathrm{KLIC}(\hat{f}_t)$ in (3.1) the term $\mathbb{E}_t[\log p_t(Y_{t+1})]$ drops out.

[12] This statement holds if (i) $\{Z_t\}$ is ϕ-mixing of size $- q/(2q - 2)$ with $q \geq 2$, or α-mixing of size $-q/(q-2)$ with $q \geq 2$; (ii) $\mathrm{E}|d_{t+1}|^{2q} < \infty$ for all t; and (iii) $\sigma_{R,P}^2 = Var(\sqrt{n}\bar{d}_{R,P}) > 0$ for all n sufficiently large. See Theorem 4 of Giacomini and White (2006) for a proof and further details.

Given d predictive marginal densities $\hat{f}_{j,t}$ for $j = 1 \dots d$ and a conditional copula specification $\hat{c}_t$, S_{t+1} is observable. Two competing multivariate density models with different (well-defined) conditional copula specifications $\hat{c}_{A,t}$ and $\hat{c}_{B,t}$, can now be compared using a DM-type test statistic as in (3.5). Furthermore, we introduce *identical* predictive marginal densities $\hat{f}_{j,t}$ for both models, because we want to isolate the predictive performance of the different copula specifications. Now we can test the null-hypothesis of equal KLIC-scores:

$$H_0: \text{E}(S_{A,t+1}) = \text{E}(S_{B,t+1}) \tag{3.9}$$

Since the identical predictive marginal densities cancel out, the null hypothesis boils down to:

$$H_0: \text{E}\left(\log \hat{c}_{A,t}(\hat{\boldsymbol{U}}_{t+1})\right) = \text{E}\left(\log \hat{c}_{B,t}(\hat{\boldsymbol{U}}_{t+1})\right) \tag{3.10}$$

We can define the score differences as:

$$d_{t+1} = \log \hat{c}_{A,t}(\hat{\boldsymbol{U}}_{t+1}) - \log \hat{c}_{B,t}(\hat{\boldsymbol{U}}_{t+1}) \qquad \text{for } t+1 = R+1, \dots, T = R+P \tag{3.11}$$

If we compute the sample average of the score difference as $\bar{d}_{R,P} = \frac{1}{P}\sum_{t=R}^{T-1} d_{t+1}$, we obtain a test-statistic similar to (3.5):

$$Q_{R,P} = \frac{\bar{d}_{R,P}}{\sqrt{\hat{\sigma}_{R,P}^2 / P}} \tag{3.12}$$

For $\hat{\sigma}_{R,P}^2$ we again use a conventional HAC-estimator of $\sigma_{R,P}^2$. $Q_{R,P}$ can be compared to the standard normal distribution under the null-hypothesis of equal KLIC-scores.

A comparison of the univariate and multivariate scenario is given in table 3.1.

	Univariate Density models	Multivariate density models
$\boldsymbol{H_0: \text{E}(S_{t+1}^A) = \text{E}(S_{t+1}^B)}$	$\text{E}(\log \hat{f}_{A,t}(Y_{t+1}))$ $= \text{E}(\log \hat{f}_{B,t}(Y_{t+1}))$	$\text{E}\left(\log \hat{c}_{A,t}(\hat{\boldsymbol{U}}_{t+1})\right)$ $= \text{E}\left(\log \hat{c}_{B,t}(\hat{\boldsymbol{U}}_{t+1})\right)$[13]
Score differences: $\boldsymbol{d_{t+1}}$	$\log \hat{f}_{A,t}(Y_{t+1}) - \log \hat{f}_{B,t}(Y_{t+1})$	$\log \hat{c}_{A,t}(\hat{\boldsymbol{U}}_{t+1}) - \log \hat{c}_{B,t}(\hat{\boldsymbol{U}}_{t+1})$
Test statistic	$t_{R,P} = \dfrac{\bar{d}_{R,P}}{\sqrt{\hat{\sigma}_{R,P}^2 / P}}$	$Q_{R,P} = \dfrac{\bar{d}_{R,P}}{\sqrt{\hat{\sigma}_{R,P}^2 / P}}$

Table 3.1: The ingredients for Diebold Mariano based goodness-of-fit tests in (3.5) and (3.12), to compare the predictive ability of density forecast Model A and Model B based on Kullback-Leibler divergence. The left column shows the univariate scenario, whereas the right column shows the multivariate version, under the assumption that model A and B only differ in their copula specification.

---

[13] Under the assumption that model A and B only differ in their copula specification.

3.2.3 Comparing univariate density forecasts in tails

In many practical applications a particular region of the density is of most interest. In finance for example, major interest goes out to left tail behavior, for instance to calculate important risk measures such as Value-at-Risk (VaR) or Expected Shortfall (ES). Therefore, it is relevant to consider goodness-of-fit tests that evaluate the forecasting performance specifically in a certain region of interest. An intuitive approach would be to multiply the scoring rule of (3.2) with a weight function that is high in relevant regions, and low in irrelevant regions. In order to test which model shows best performance in the region of interest, one could consider the Diebold-Mariano test statistic of (3.5). Amisano and Giacomini (2007) proposed such a *weighted logarithmic* (*wl*) scoring rule:

$$S^{wl}(\hat{f}_t; y_{t+1}) = w(y_{t+1}) \log \hat{f}_t(y_{t+1}) \tag{3.13}$$

For the purpose of *left tail* evaluation, one could introduce a threshold weight function $w(y_{t+1}) = I(y_{t+1} \leq r)$ for a certain value of *r*, where $I(A) = 1$ if A occurs and zero otherwise. Two competing predictive densities, $\hat{f}_{A,t}$ and $\hat{f}_{B,t}$, can be compared using a Diebold Mariano type test statistic with the weighted score difference, which is defined as:

$$d_{t+1}^{wl} = w(y_{t+1})\big[\log \hat{f}_{A,t}(y_{t+1}) - \log \hat{f}_{B,t}(y_{t+1})\big] \text{ and } \bar{d}_{R,P}^{wl} = \frac{1}{P}\sum_{t=R}^{T-1} d_{t+1}^{wl} \tag{3.14}$$

For $t = t + 1 = R + 1, \dots, T = R + P$. We can now test the null-hypothesis $H_0: E[d_{t+1}^{wl}] = 0$ with test statistic:

$$t_{R,P} = \frac{\bar{d}_{R,P}^{wl}}{\sqrt{\hat{\sigma}_{R,P}^2/P}} \sim N(0,1) \tag{3.15}$$

However, Diks et al (2010) report several limitations of this weighted logarithmic (*wl*) scoring rule. Their main criticisms are based on the observation that the *wl* scoring rule as defined in (3.13) ignores the *total tail probability* $\hat{F}(r)$. An implication is that S^{wl} may be biased towards density forecasts with higher $\hat{F}(r)$, i.e. fat-tailed densities. This may have adverse consequences, as a fat-tailed density forecast may be favored over thin-tailed forecasts, even if the latter is the true distribution from which the data are drawn. This observation makes the *wl* scoring rule *improper*, following the terminology of Gneiting and Ranjan (2008). They define a scoring rule to be proper if under the true conditional density p_t, no density forecast $\hat{f}_t$ can receive a better average score than the actual conditional density p_t.

Diks et al (2010) provide a clear illustration of the "improperness" of the *wl* scoring rule. They compare the accuracy of $\hat{f}_{A,t}$, a standard normal distribution, and $\hat{f}_{B,t}$ a fat-tailed Student-*t* distribution with v degrees of freedom. They point out that for the case $v = 5$, the relative log-likelihood score $\log \hat{f}_{A,t}(y_{t+1}) - \log \hat{f}_{B,t}(y_{t+1})$ is strictly negative in the left tail $(-\infty, y^*)$, with $y^* \approx 2.5$, see the left panel of figure 3.1. This implies that if the threshold $r \leq y^*$ is used, the average weighted likelihood score difference $\bar{d}_{R,P}^{wl}$ can never be positive, and will be strictly negative whenever there are observations in the left tail. This has far-reaching consequences for the outcome of the test statistic in (3.15): even if the actual conditional density p_t corresponds to the standard normal distribution, the test statistic will always be negative and the fat-tailed distribution $\hat{f}_{B,t}$ will be favored over the (correct) standard normal distribution $\hat{f}_{A,t}$. This example clearly shows that the *wl* scoring rule is *improper.*

Diks et al (2010) introduce two *proper* scoring rules to compare univariate predictive densities in tails, based on *conditional likelihood* (*cl*) and *censored likelihood* (*csl*).

Conditional on the fact that Y_{t+1} is a tail event, we obtain the *conditional likelihood* scoring rule:

$$S^{cl}(\hat{f}_t; y_{t+1}) = I(y_{t+1} \leq r) \log\left[\frac{\hat{f}_t(y_{t+1})}{\hat{F}(r)}\right] \tag{3.16}$$

This *cl* scoring rule is proper as the probability $\hat{f}_t(y_{t+1})$ is normalized with the total tail probability $\hat{F}(r)$ for the region of interest. Moreover, it allows for different tail behavior in the sense that it defines the distribution of f conditional on the outcome of $I(y_{t+1} \leq r)$. A disadvantage of S^{cl} is that it only evaluates the predictive density $\hat{f}_t(y_{t+1})$ *relative* to $\hat{F}(r)$ for $y_{t+1} \leq r$. This implies that the information about $\hat{F}(r)$ is *lost* and the focus is *only* on the shape for $y_{t+1} \leq r$. A consequence is that the *cl* scoring rule does not differentiate between density forecasts with similar tail shape, even if the total tail probability is totally different.

The *censored likelihood* scoring rule combines the information of the total tail probability $\hat{F}(r)$, and the actual value of $\hat{f}_t(y_{t+1})$ given that Y_{t+1} is a tail event. The *csl* scoring rule is given as:

$$S^{csl}(\hat{f}_t; y_{t+1}) = I(y_{t+1} \leq r) \log \hat{f}_t(y_{t+1}) + I(y_{t+1} > r) \log[1 - \hat{F}(r)] \tag{3.17}$$

Similar to S^{cl}, the *csl* scoring rule is proper (in the sense that it corrects for total tail probability) and takes into account tail behaviour. Intuitively, S^{csl} corrects S^{wl} with information from the total tail probability $\hat{F}(r)$. It "shifts" the *wl* scoring rule to the right, as the correction term $I(y_{t+1} > r) \log[1 - \hat{F}(r)]$ is always positive. Compared to *cl* scoring rule, the *csl* scoring rule uses more information, as it assigns nonzero scores to observations outside the region of interest ($y_{t+1} > r$), whereas S^{cl} fixes these scores to zero. The *csl* scoring rule however ignores the shape of $\hat{f}_t(y_{t+1})$ for values that fall outside the region of interest ($y_{t+1} > r$).
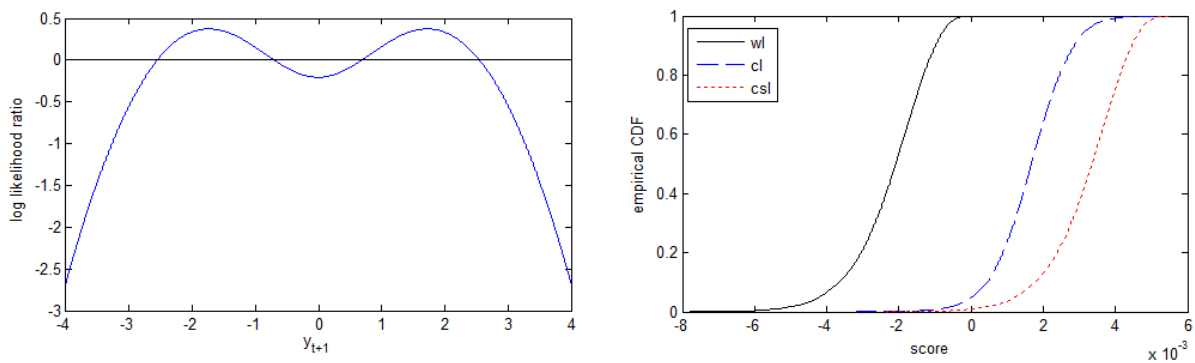

Figure 3.1: Both figures compare the pdf of the standard normal distribution (Model A), with the (standardized) Student-t(5) distribution (ModelB). The left panel shows the relative log-likelihood scores $\log \hat{f}_{A,t}(y_{t+1}) - \log \hat{f}_{B,t}(y_{t+1})$. The right panel shows the emprical CDFs of the mean relative scores for the weighted logarithmic (*wl*) scoring rule in (3.14), the conditional likelihood (*cl*) scoring rule in (3.18) and the censored likelihood (*csl*) scoring rule in (3.19) for a sample size of $n = 2000$ independent observations from a standard normal distribution. We use the threshold weight function $I(y_{t+1} \leq 2.5)$ and the scoring rule is defined as the score for the (correct) standard normal distribution minus the (incorrect) student-t distribution. The graph is based on 10,000 replications. (See Diks et al, 2009)

The proposed proper scoring rules can be interpreted in terms of Kullback-Leibler divergence, such that the relative scores of competing models can be computed:

$$d_{t+1}^{cl} = S^{cl}(\hat{f}_{A,t}; y_{t+1}) - S^{cl}(\hat{f}_{B,t}; y_{t+1}) \tag{3.18}$$

$$d_{t+1}^{csl} = S^{csl}(\hat{f}_{A,t}; y_{t+1}) - S^{csl}(\hat{f}_{B,t}; y_{t+1}) \tag{3.19}$$

These relative scores can be used as the basis for the Diebold Mariano type test statistics as in (3.15) to test the null-hypothesis of equal predictive ability:

$$t_{R,P}^* = \frac{\bar{d}_{R,P}^*}{\sqrt{\hat{\sigma}_{R,P}^2/P}} \sim N(0,1) \tag{3.20}$$

Diks et al (2010) illustrate the properties of these scoring rules using the same example as before, comparing $\hat{f}_{A,t}$, the (correct) standard normal predictive density, with $\hat{f}_{B,t}$, the fat tailed standardized-t distribution with $v = 5$. Using the results from the left panel of figure 3.1, we set the threshold value of weight function $I(y_{t+1} \leq r)$ to $r = -2.5$. The mean relative scores of the weighted logarithmic scoring rule $\bar{d}_{R,P}^{wl}$ in (3.14), the conditional likelihood scoring rule $\bar{d}_{R,P}^{cl}$ in (3.18) and the censored likelihood scoring rule $\bar{d}_{R,P}^{csl}$ in (3.19), are compared using the empirical CDFs for series of $n = 2,000$ independent observations from a standard normal distribution, based on 10,000 replications.

The right panel of figure (3.1) clearly shows that the (improper) *wl* scoring rule shows a bias towards the incorrect Student-t distribution, with almost exclusively negative mean relative scores, indicating a lower score to the (correct) standard normal distribution. The proper scoring rules do not suffer from this bias and show mostly positive scores. It should be noted that the mean relative scores based on censored likelihood show the best results in the sense that they show more positive scores, and stochastically dominate the both the weighted likelihood and conditional likelihood scores (in this experiment).

Diks et al (2010) also show that these scoring rules can be generalized to more general weight functions $w_t(y)$, which may focus on different regions of interest and are allowed to be dynamic (e.g. using a time-varying threshold r_t). Also in the general context, different predictive densities can be compared using Kullback-Leibler divergence.

3.2.4 Comparing multivariate density forecasts in tails

The theory presented in section 3.2.3 can be extended to the multivariate case. Instead of focusing on the tail behavior of a univariate density forecast $\hat{f}_t(y_{t+1})$, we are now interested in the tail behavior of the multivariate density forecast $\hat{f}_t$ for $\boldsymbol{Y}_{t+1}$, where $\boldsymbol{Y}_t = (Y_{1,t}, \dots, Y_{d,t})'$ is the vector of interest with corresponding conditional marginal distributions $F_{1,t}, \dots, F_{d,t}, j = 1, \dots, d$. Therefore, it is convenient to work with the KLIC-score of multivariate density forecasts as in (3.7). We reintroduce the conditional copula C_t with corresponding conditional density c_t, and $\boldsymbol{U}_{t+1} = (F_{1,t}(Y_{1,t}), \dots, F_{d,t}(Y_{d,t}))'$ its multivariate conditional PIT.

In order to focus on the left "tail" of the copula, we introduce the threshold weight function $w(\boldsymbol{u}_{t+1}) = \mathrm{I}(\boldsymbol{u}_{t+1} \le \boldsymbol{r})$ where $\boldsymbol{u}_{t+1} = (u_{1,t+1}, \dots, u_{d,t+1})'$, $j = 1,2 \dots d$ and $\boldsymbol{r}$ a $d \times 1$ vector of threshold values $r \in [0,1]$. In the two-dimensional case, one could visualize this weight function as focusing on a "block" of size $r \times r$ in the lower left corner of scatter plot, see the area highlighted with the blue scatter points in the left corner of the graphs in figure 3.2. Using this multivariate threshold weight function, we can extend the three scoring rules as introduced in the last subsection to a higher dimension. For the weighted likelihood scoring rule in (3.13) we obtain:

$$S^{wl}(\hat{f}_t; \hat{c}_t; \boldsymbol{u}_{t+1}) = \mathrm{I}(u_{t+1} \le r) \left[\sum_{j=1}^{d} \log\left(\hat{f}_{j,t}(y_{t+1})\right) + \log\left(\hat{c}_t(\widehat{\boldsymbol{U}}_{t+1})\right) \right] \tag{3.21}$$

For the conditional likelihood (*cl*) scoring rule in (3.19):

$$S^{cl}(\hat{f}_t, \hat{c}_t; \boldsymbol{u}_{t+1}) = \mathrm{I}(u_{t+1} \le r) \left[\sum_{j=1}^{d} \log\left(\frac{\hat{f}_{j,t}(y_{t+1})}{\hat{C}_t\left(\hat{F}_{1,t}(r), \dots, \hat{F}_{d,t}(r)\right)} \right) + \log\left(\frac{\hat{c}_t(\widehat{\boldsymbol{U}}_{t+1})}{\hat{C}_t\left(\hat{F}_{1,t}(r), \dots, \hat{F}_{d,t}(r)\right)} \right) \right] \tag{3.22}$$

For the censored likelihood (*csl*) scoring rule in (3.20):

$$S^{csl}(\hat{f}_t, \hat{c}_t; \boldsymbol{u}_{t+1}) = \mathrm{I}(u_{t+1} \le r) \left[\sum_{j=1}^{d} \log\left(\hat{f}_{j,t}(y_{t+1})\right) + \log\left(\hat{c}_t(\widehat{\boldsymbol{U}}_{t+1})\right) \right]$$
$$+ \mathrm{I}(u_{t+1} > r) \left[\log\left(1 - \hat{C}_t\left(\hat{F}_{1,t}(r), \dots, \hat{F}_{d,t}(r)\right) \right) \right] \tag{3.23}$$
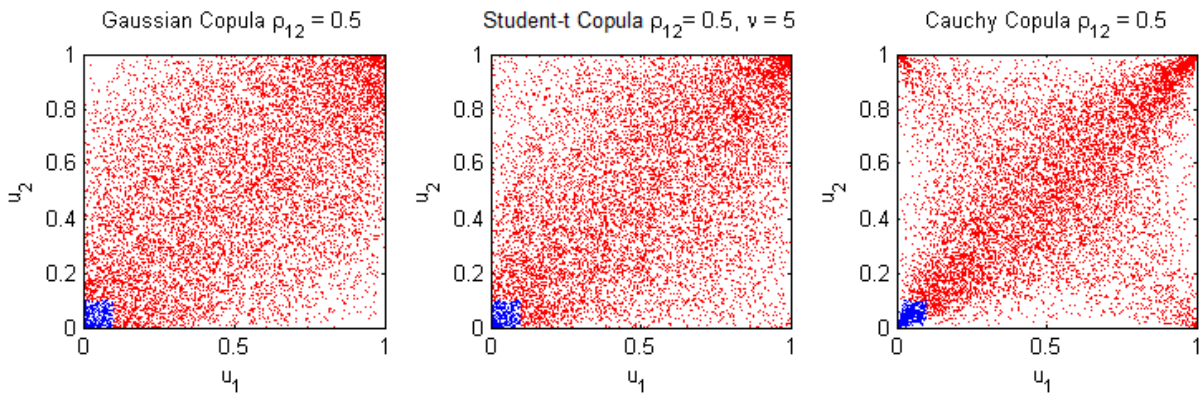

Figure 3.2: For three elliptical copulas (from left to right the Gaussian, Student-t(5) and Cauchy Copula), the region of interest is highlighted with the blue scatter points for weight function I($u_1, u_2 \le r$) and r=0.1.

Let us reintroduce the two competing multivariate density models from section 3.2.2, Model A and Model B, that only differ in their copula specification (that is, the corresponding marginal distributions of both models are identical). The relative difference in scoring rules can now be computed as:

$$d_{t+1}^* = S^*\left(\hat{f}_t, \hat{c}_{A,t}; \boldsymbol{u}_{t+1}\right) - S^*\left(\hat{f}_t, \hat{c}_{B,t}; \boldsymbol{u}_{t+1}\right) \tag{3.24}$$

Where (*) stands for the *wl*, *cl* or *csl* scoring rules as in (3.22), (3.23) and (3.24). The score differences between Model A and Model B for the three scoring rules can be expressed as:

$$d_{t+1}^{wl} = \mathrm{I}(\boldsymbol{u}_{t+1} \le r)\left[\log\left(\hat{c}_{A,t}(\widehat{\boldsymbol{U}}_{t+1})\right) - \log\left(\hat{c}_{B,t}(\widehat{\boldsymbol{U}}_{t+1})\right)\right] \tag{3.25}$$

$$d_{t+1}^{cl} = \mathrm{I}(\boldsymbol{u}_{t+1} \le r)\left[\log\left(\frac{\hat{c}_{A,t}(\widehat{\boldsymbol{U}}_{t+1})}{\hat{C}_{A,t}\left(\hat{F}_{1,t}(r), \ldots, \hat{F}_{d,t}(r)\right)}\right) - \log\left(\frac{\hat{c}_{B,t}(\widehat{\boldsymbol{U}}_{t+1})}{\hat{C}_{B,t}\left(\hat{F}_{1,t}(r), \ldots, \hat{F}_{d,t}(r)\right)}\right)\right] \tag{3.26}$$

$$d_{t+1}^{csl} = \mathrm{I}(\boldsymbol{u}_{t+1} \le r)\left[\log\hat{c}_{A,t}(\widehat{\boldsymbol{U}}_{t+1}) - \log\hat{c}_{B,t}(\widehat{\boldsymbol{U}}_{t+1})\right] +$$

$$\mathrm{I}(\boldsymbol{u}_{t+1} > r)\left[\log\left(1 - \hat{C}_{A,t}\left(\hat{F}_{1,t}(r), \ldots, \hat{F}_{d,t}(r)\right)\right) - \log\left(1 - \hat{C}_{B,t}\left(\hat{F}_{1,t}(r), \ldots, \hat{F}_{d,t}(r)\right)\right)\right] \tag{3.27}$$

As before, these relative scores can be used as the basis for the Diebold Mariano type test statistic for multivariate density forecasts as in (3.12). We test $H_0: \mathrm{E}\left(S_{A,t+1}\right) = \mathrm{E}\left(S_{B,t+1}\right)$ of equal predictive ability by using the test statistic:

$$Q_{R,P}^* = \frac{\bar{d}_{R,P}^*}{\sqrt{\hat{\sigma}_{R,P}^2/P}} \qquad \text{for } t + 1 = R + 1, \ldots, P + R \tag{3.28}$$

Where $\hat{\sigma}_{R,P}^2$ is a conventional HAC-estimator of the sample variance and (*) symbolizes *wl*, *cl* and *csl*. The obtained test statistics can be compared with the standard normal distribution.

The question rises whether the newly developed scoring rules exhibit similar characteristics as their univariate counterparts. Since our particular interest goes out to the question whether the proposed scoring rules are *proper,* we extend the example presented section 3.2.3 to the bivariate case. We replace Model A from section 3.2.3 (the standard normal distribution) for the independent bivariate Gaussian copula ($C_{A,t}^{Ga}$) and the (fattailed) Student-*t* copula is set as the competing model ($C_{B,t}^{Ga}$). As a first graphical check to get a sense of the differences between the two models, we provide the relative log-likelihood $\log\hat{c}_{A,t}^{Ga}(y_{t+1}) - \log\hat{c}_{B,t}^{T}(y_{t+1})$ in the three-dimensional plot in the left panel of figure 3.3. It is striking to see that in the relative log-likelihood scores are *strictly* negative in the corners of the figure, i.e. in the "tails" of the copula distribution. This implies that if the threshold value r is chosen to be small enough (r < ca. 0.1), the average weighted likelihood score $\bar{d}_{R,P}^{wl}$ (from equation 3.25) can never be positive and is strictly negative whenever there are observations in the region of interest. Note that this observation is similar to the univariate case, and along the same line of reasoning we conclude that weighted likelihood scoring rule is *improper*.

Analogue to section 3.2.3, we compare the behavior of the mean relative scores of the *wl* scoring rule, the *cl* scoring rule and the *csl* scoring rule in a simulation of $n = 2,000$ independent samples of Model A ($C_{A,t}^{Ga}$) and Model B ($C_{B,t}^{Ga}$), with 10,0000 replications. In the right graph of fiure 3.3 we display the simulation results of $\bar{d}_{t+1}^{wl}$, $\bar{d}_{t+1}^{cl}$ and $\bar{d}_{t+1}^{csl}$ from (3.25), (3.26) and (3.27) with the threshold value fixed at $r = 0.1$. The results are very similar to those of the univariate case. As expected from the left graph of figure 3.3, the *wl* scoring rule turns out to be *improper,* because *all* mean relative scores are negative, indicating a clear bias towards the (incorrect) fat-tailed model. The proper *cl* and *csl* scoring rules do not suffer from this bias and show (mostly) positive scores. Again, the *csl* scoring rule shows the best performance with the highest (and strictly positive) mean relative scores and stochastically dominates the other scoring rules.
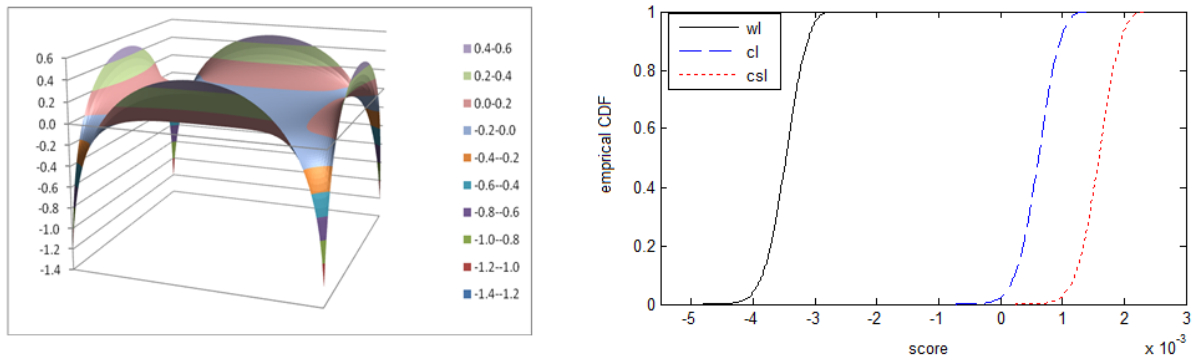
Figure 3.3: Bivariate extension of figure 3.1. We compare the density of the independent Gaussian Copula (Model A), with the independent Student-t(5) Copula (Model B). The left panel shows the relative log-likelihood scores log $\hat{c}_{A,t}^{Ga}(y_{t+1}) - $ log $\hat{c}_{B,t}^{T}(y_{t+1})$. The right panel shows the empricial CDFs of the mean relative scores for the weighted logarithmic (*wl*) scoring rule (3.25), the conditional likelihood (*cl*) scoring rule (3.26) and the censored likelihood (*csl*) scoring rule (3.27) for a sample size of $n = 2000$ independent observations from a independent Gaussian Copula. We use the threshold weight function I($y_{t+1} \leq 0.1$) and the scoring rule is defined as the score for the (correct) Gaussian Copula minus the (incorrect) Student-t Copula. The graph is based on 10,000 replications.

4 Monte Carlo Simulations

In this section we assess the finite sample properties of the predictive ability tests based on the weighted logarithmic (*wl*) scoring rule in (3.21), the conditional likelihood (*cl*) scoring rule in (3.22), and the censored likelihood (*csl*) scoring rule in (3.23). Specifically, we focus on the size and power properties of the test of equal predictive ability of two competing multivariate density forecasts based on the Diebold-Mariano test statistic as given in (3.28).

We test the null hypotheses that the two competing predictive densities have equal expected scores, versus the alternative hypothesis that one of the distributions has better predictive ability, or:

$$\left.\begin{array}{l} H_0\colon \mathrm{E}[d^*_{t+1}] = 0 \\ H_a\colon \mathrm{E}[d^*_{t+1}] < 0 \end{array}\right\} \text{ for } t = R, R+1, \dots, R+P-1 \tag{4.1}$$

Where * denotes either *wl*, *cl* or *csl*, such that the score differences as given in resp. (3.25), (3.26) and (3.27) are obtained. As before, R denotes the length of the rolling in-sample window and P denotes the number of one-step-ahead density forecasts. The one-sided alternative hypothesis is chosen to test the properness of a scoring rule, i.e. to test whether a scoring rule on average favors a correctly specified density forecast over an incorrect one.

4.1 Size

This section contains a bivariate extension of the Monte Carlo simulations of Diks et al (2010) in which they assess the size properties of univariate scoring rules in tails.

In their article, Diks et al (2010) design a case in which the competing density forecasts under consideration are both "equally (in)correct", independent of the value of the threshold r. Key insight is to focus on the central part of the distribution by taking the threshold weight function $w(y) = \mathrm{I}(-r \leq y \leq r)$, and choosing the data generating process (DGP) a symmetric conditional density with mean zero. Two competing densities are defined: $f_1(x)$ which is the DGP shifted to the right with a conditional mean fixed at a certain positive constant $\lambda > 0$, while $f_2(x)$ corresponds to the DGP shifted to the left with conditional mean fixed at $-\lambda$. Diks et al (2010) emphasize that this test-scenario is suitable for assessing the finite sample *size* properties of the proposed Diebold Mariano type test statistic as in this scenario the two competing predictive densities are "equally incorrect". This follows from the fact that the simulation setup is entirely symmetric, and the scoring rules considered are invariant under a reflection of zero under both the DGP and the two competing predictive densities. Note that the two competing density forecasts have equal predictive accuracy for *all* values of r, due to the focus on the central part of the distribution.

The simulation experiment of Diks et al (2010) can be extended to a multivariate case. For simplicity, we focus on 2 dimensions and the Gaussian copula. Let $f(x)$ and $g(y)$ denote i.i.d. standard normal distributed marginal distributions, where we use the notation $F(x) = u$ and $G(y) = v$, with $u, v \in [0,1]$ for the corresponding cdfs. Furthermore, we take an independent Gaussian copula as DGP: $C_{DGP}^{Ga}(u, v)$. In the bivariate case the correlation parameter $\rho_{12} = \rho_{21}$ is the only parameter of the Gaussian copula (which is thus set to zero in the DGP). At this point, we introduce two competing Gaussian copulas $C_A^{Ga}(u, v)$ and $C_B^{Ga}(u, v)$ with their respective correlation parameters fixed at $\rho_{12} = -0.7$ and $\rho_{12} = 0.7$ respectively, to assure equal distance from the DGP. In order to focus on a central, symmetric part of the copula distribution, we introduce the weight function $w(u, v) = I(r \leq u \leq 1 - r, r \leq v \leq 1 - r)$, with $0 \leq r < 0.5$[14]. A graphical illustration of the situation is given in figure 4.2, for a sample of $n = 1000$ observations and $r = 0.4$. For this combination of DGP and predictive copula specifications, the relative scores d_{t+1}^{*} of the *wl, cl* and *csl* scoring rules based on $w(u, v)$ are identical. This is due to the fact that the two competing copula specifications have equal distance from the DGP and their total probability mass in the region of interest is equal (indicated by the "block" of blue dots in figure 4.1) due to the symmetric property of the Gaussian copula.
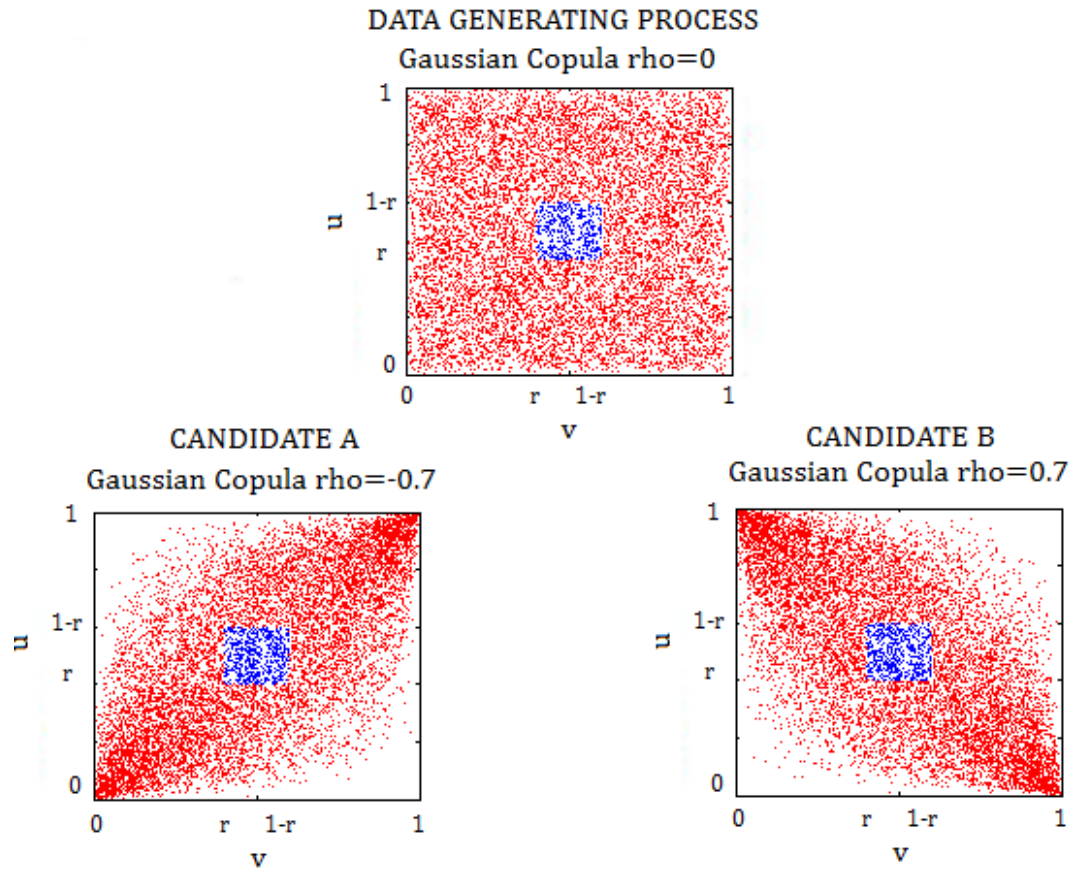

Figure 4.1 Simulation design with equally incorrect predictive copula densities independent of the choice of r. The DGP is set as an independent Gaussian Copula (upper scatter plot), while the two competing predictive Gaussian copulas have their correlation coefficients fixed at resp. $\rho_{12} = -0.7$ (left graph) and $\rho_{12} = 0.7$ (right graph). The focus is on the central part of the distribution with weight function $w(u, v) = I(r \leq u \leq 1 - r, r \leq v \leq 1 - r)$, with $r = 0.4$, and is highlighted with the blue dots in the three scatterplots.

---

[14] Note that the probability mass of this weight function corresponds to a "block" in the center of the copula distribution and is always positive due to the quasi-monotonicity of copulas, see requirement (iii) of lemma 2.1.

We use the test as defined in (3.28) with the null hypothesis of equal predictive accuracy of the two competing densities against the alternative that the $C_A^{Ga}(u,v)$ has better predictive ability. We compute the one-sided average rejection rates (at nominal significance levels of 1%, 5% and 10%) as a function of the threshold value r. The sample size of the simulation experiments is chosen to be dependent on the threshold level r, such that the expected number of observations within the region of interest is constant at a value m. We do this by setting the sample size equal to $m/P(r \leq u \leq 1-r, r \leq v \leq 1-r)$. In this way, a fair comparison of rejection rates can be made for different threshold values. The rejection rates of the tests are expected to be very close to the nominal significance levels, for every choice $0 \leq r < 0.5$.

Figure 4.2 depicts the results of the size experiments for $m = 50$ and $m = 100$, based on 10,000 replications. The rejection rates of the simulations follow our expectations, as they are rather close to the nominal significance levels for all values of r. For $m = 50$ and $m = 100$ the size properties of the predictive accuracy test appears to be satisfactory. Results for $m < 20$ however, show less stable results. Unreported results for $m = 5$ and $m = 10$ show a slight overestimation of the rejection rates for small values of r, which slowly converges to the nominal significance level for $r \rightarrow 1$. Apparently, the asymptotic distribution of our Diebold Mariano type test statistic is not a suitable approximation of the real (small-sample) distribution. For sample sizes with m smaller than 20, one could better use bootstrapping to get a good approximation of the true distribution. This may be an interesting topic for further research.
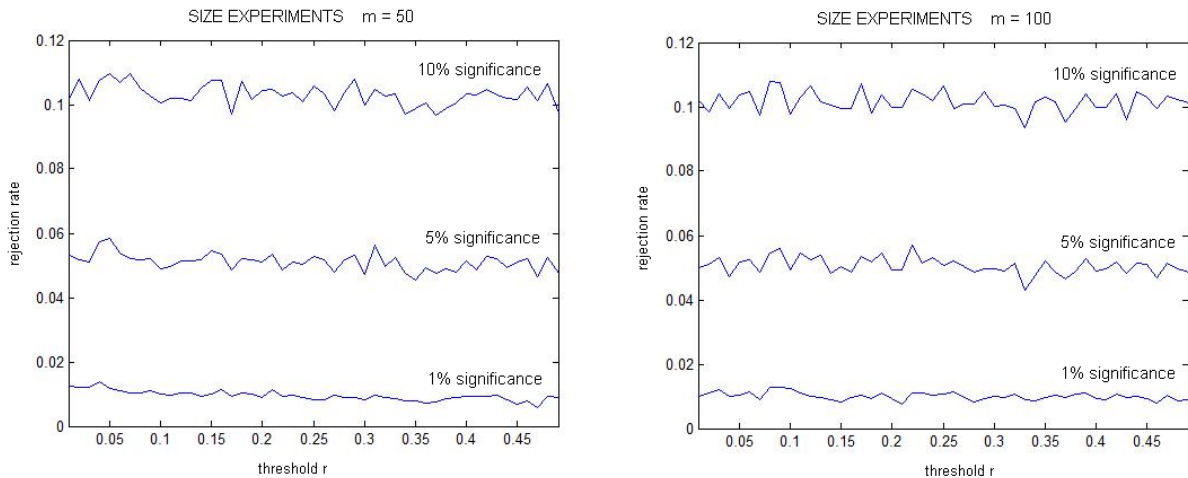

Figure 4.2: One-sided rejection rates to assess the size properties of the equal predictive ability test as defined in (3.28) when using the weighted logarithmic (*wl*), the conditional likelihood (*cl*) and the censored likelihood (*csl*) scoring rules under the weight function I($r \leq u \leq 1-r, r \leq v \leq 1-r$), with $0 < r < 0.5$, based on 10,000 replications. The sample size is chosen to be $m/P(r \leq u \leq 1-r, r \leq v \leq 1-r)$, and we provide results for m=50 (left) and m=100 (right). The DGP is an independent Gaussian copula, whereas the two competing Gaussian copulas have correlation coefficients of resp. $\rho_{12} = -0.7$ and $\rho_{12} = 0.7$. The graphs show rejection rates against the alternative that the copula with $\rho_{12} = -0.7$ has better predictive ability.

4.2 Power

To assess the power properties of our test statistics based on multivariate scoring rules, we perform several simulation experiments where one of the competing density forecasts is correct, i.e. corresponds exactly with the underlying DGP.

The simulation experiment for two dimensions is set up as follows. As before, we introduce two independently drawn standard normal marginal distributions $f(x)$ and $g(y)$, and an independent Gaussian Copula as DGP, i.e. the correlation parameter ρ_{12} is fixed to zero. Furthermore, we introduce two competing copula specifications. The first is equal to the DGP, that is an independent Gaussian copula, $C_A^{Ga}(u,v)$, the second is an independent students-t copula with degrees of freedom fixed to five, $C_B^T(u,v)$. We focus on the left tail of the copula specification by choosing weight function $w(u,v) = I(u, v \leq r)$ where $0 < r \leq 1$. Again we compute one-sided rejection rates against the null hypothesis of equal scores as given in (3.28). We compute two different alternative hypotheses in order to asses:

- "True Power": Alternative Hypothesis is: correct copula performs better
- "Spurious Power" : Alternative Hypothesis is: incorrect copula performs better

All things kept equal, we repeat this experiment for a Student-t Copula as DGP. Obviously, now $C_B^T(u,v)$ is the correct copula specification, whereas $C_A^{Ga}(u,v)$ is incorrectly specified. In all, we thus provide four different experiments: we assess *true* and *spurious* power for both a Gaussian copula as DGP and a Student-t copula as DGP. We use the sample size $m/P(u \leq r, v \leq 1-r)$, for a given threshold value r and a fixed constant m. Figure 4.3 and 4.4 depict the results of the four power experiments for resp. $m = 5$ and $m = 40$ expected observations in the region of interest, based on 1,000 replications[15]. Several conclusions can be drawn.

First, the power of the *wl* scoring rule appears to depend strongly on the choice of the DGP and the value of the threshold r. In case of the Gaussian copula as DGP, the *wl* scoring rule performs poorly for both sample sizes. For the small sample size in figure 4.3, the true power is close to zero for all threshold values r and the spurious power is high, in particular for values of the threshold values $r < 0.2$. The situation does not improve much for the large sample size in figure 4.4. The true power improves slightly in the middle region of the copula ($0.4 < r < 0.6$), but on the flipside the spurious power in the tails of the copula soars (to even ca. 100% for $r < 0.2$). This implies that the substantial spurious power of the *wl* scoring rule is *not* a small sample problem. If the DGP is set to the Student-t(5) copula however, the *wl* scoring rule performs very well. Loosely speaking, compared to the results of the Gaussian DGP the roles for true power and spurious power are reversed: now the spurious power is approximately zero for both sample sizes, while the true power is high (in particular for low threshold values) and increases when the sample size becomes larger. These conflicting results of the *wl* scoring rule are problematic if the correct copula specification is unknown, which often is the case in practice. If the null hypothesis of equal predictive ability is rejected based on the *wl* scoring rule, this could be due to *true power* as in the upper right graphs of figure 4.3 and 4.4, or to the *spurious* power exhibited in the lower left figures. This implies that (for this simulation design) the *wl* scoring rule is unreliable, in particular for small values of r where spurious power can be close to 100%.

---

[15] The number of replications is reduced by the amount of cases when zero observations fall in the region of interest (this might occur as m denotes the *expected* number of observations in the region of interest). These (rare) cases are deleted because in test statistic (3.29) this implies $\bar{d}_{R,P}^* = \hat{\sigma}_{R,P}^2 = 0$, i.e. division by zero.
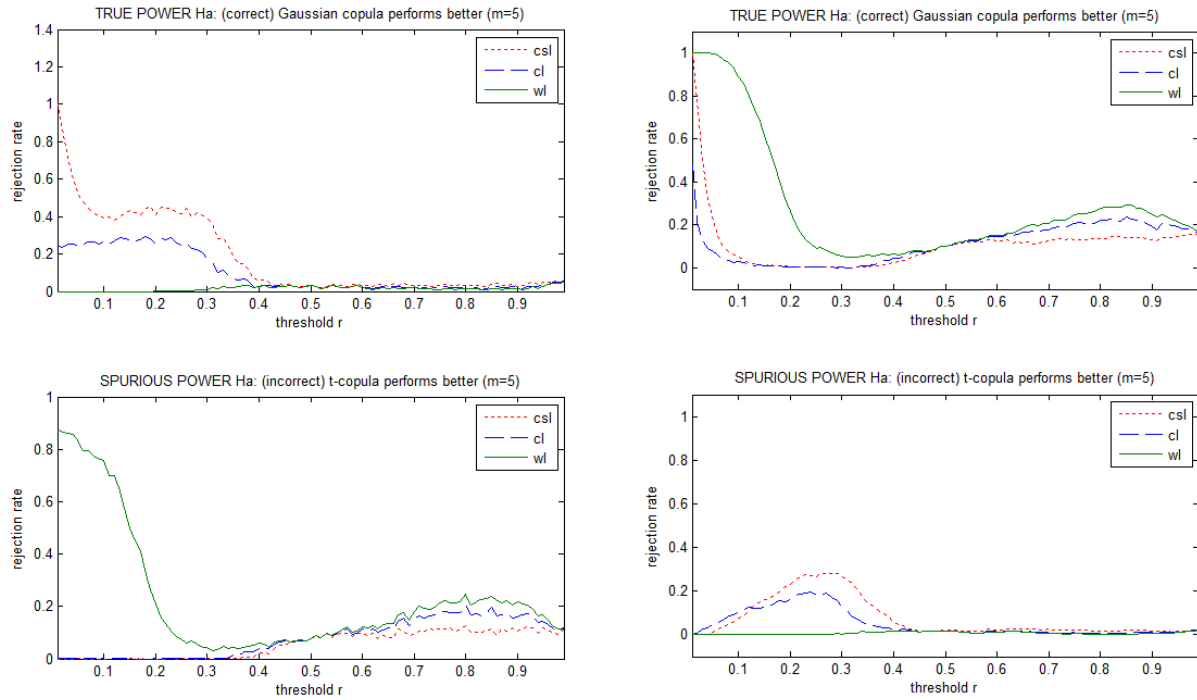
Figure 4.3: One sided rejection rates at significance level 5% to assess the power properties of the equal predictive ability test as defined in (3.28) when using the weighted logarithmic (*wl*), the conditional likelihood (*cl*) and the censored likelihood (*csl*) scoring rules under the weight function $w(u,v) = \mathrm{I}(u, v \leq r)$ where $0 < r < 1$, based on 1,000 replications and m=5 expected observations in the region of interest. For the graphs in the left (right) columns, the DGP is the independent Gaussian Copula (independent Student-*t*(5) copula). The graphs in the top (bottom) panels show average rejection rates against superior predictive ability of the correct (incorrect) copula specification, i.e. true (spurious) power.
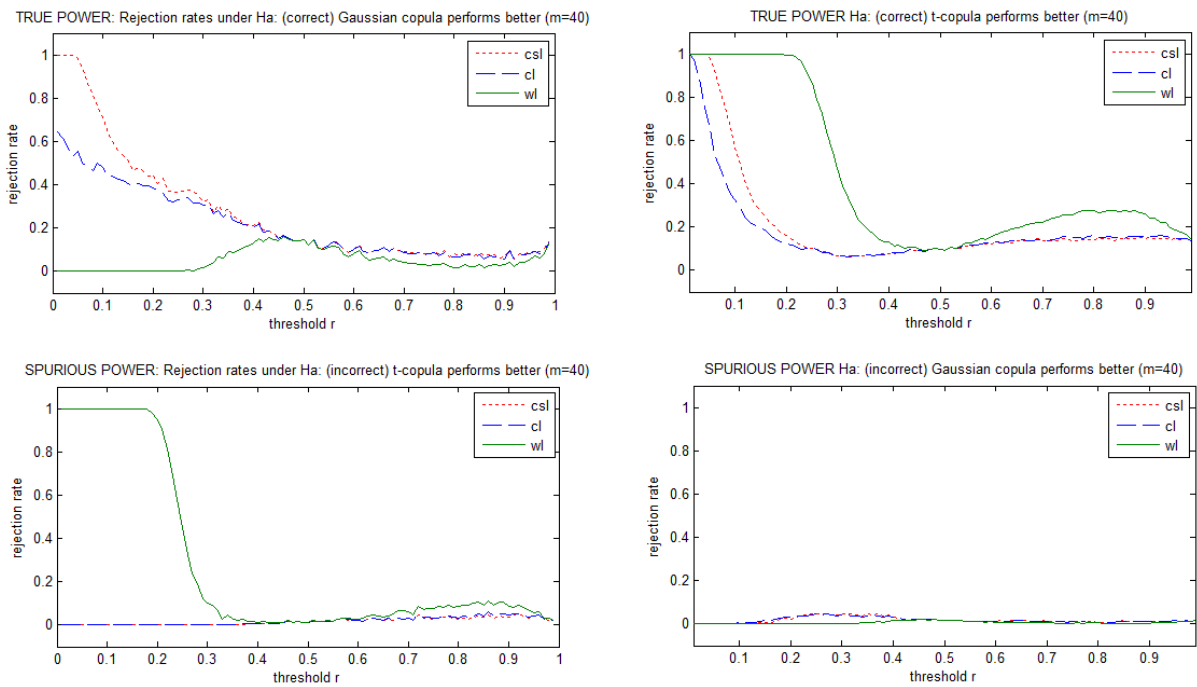
Figure 4.4: This figure is similar to Figure 4.4, but now with m=40 expected observations in the left tail.

34

Second, similar patterns of *spurious* power of the *wl* scoring rule are not apparent for the *proper* conditional (*cl*) and censored likelihood (*csl*) scoring rules. The proper scoring rules show limited spurious power with a maximum of ca. 30% in the lower right chart of figure 4.4 ($m = 5$ with the Student-t copula as DGP). In stark contrast to the *wl* scoring rule, this appears to be a small sample problem, since for $m = 40$ the maximum spurious power quickly decreases to approximately 5% for both the *cl* and *csl* scoring rule, which we deem acceptable.

Third, we observe that in general the *csl* scoring rule performs better than the *cl* scoring rule for both sample sizes. The *true* power of the *csl* scoring rule turns out to be at least as high as the *cl* scoring rule, in particular for $r < 0.5$. This can be explained by the fact that the *csl* scoring rule uses more information by censoring the observations *outside* the region of interest, as discussed in section 3.2.3. In particular for purposes requiring small values of r, such as risk management, the *csl* scoring rule might therefore be considered as favorable to the *cl* scoring rule.

Fourth, in case of the (fat-tailed) Student-t Copula as DGP, the *cl* and *csl* scoring rules perform somewhat disappointing in terms of true power. For all threshold values considered, the true power of the *wl* scoring rule performs at least as good as the *cl* and *csl* scoring rule, for both sample sizes. However, the true power for threshold values of $r < 0.1$, i.e. the left tail, is increasing for larger sample size, which is an encouraging result. In particular the performance of the *csl* scoring rule is promising, as the true power in the left tail is close to 100% for $m = 40$.

Fifth, for large values of *r*, that is $r \rightarrow 1$, the *wl, cl* and *csl* scoring rules behave similarly. This pattern is as expected, because analytically, the scoring rules are equal for $r = 1$.[16]
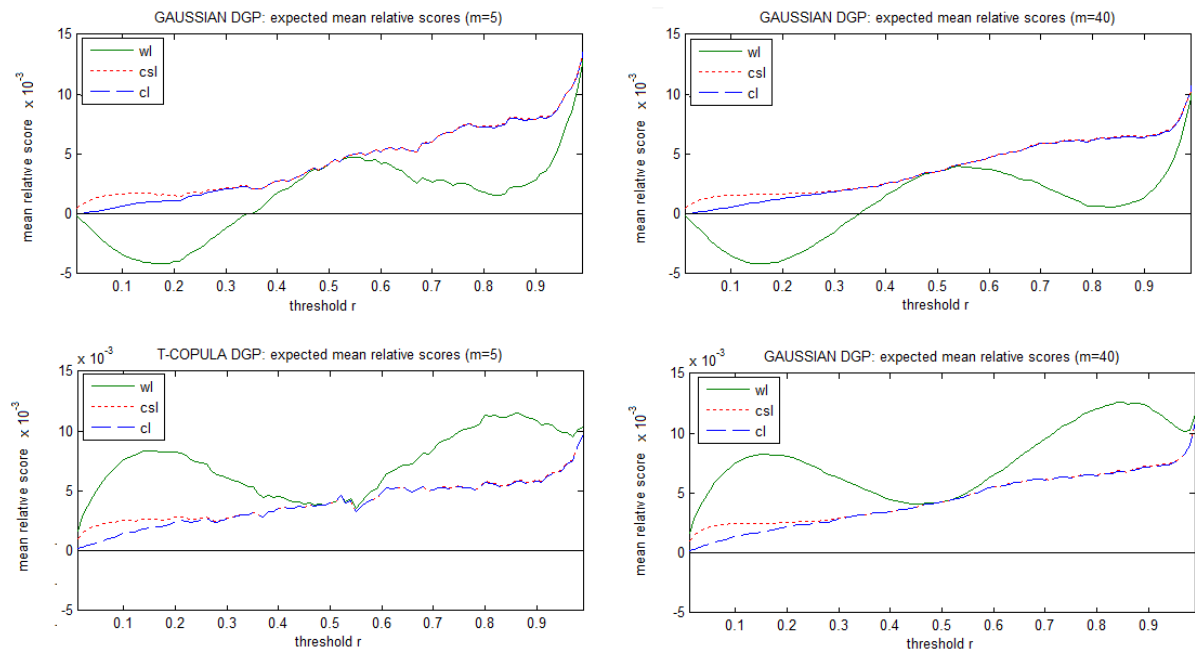

Figure 4.5: Proxy of the mean relative scores for sample sizes m=5 (left graphs) and m=40 (right graphs), when using the weighted logarithmic (*wl*), the conditional likelihood (*cl*) and the censored likelihood (*csl*) scoring rules under the weight function $w(u, v) = I(u, v \leq r)$ where $0 < r < 1$, based on 10,000 replications. The upper (lower) graphs show the results for the Gaussian copula (Student-t(5) copula) as DGP.

---

[16] If threshold r attains its maximum value of 1, the total tail probability $\hat{C}_t\left(\hat{F}_{1,t}(r), \dots, \hat{F}_{d,t}(r)\right)$ equals 1 and I($u_{t+1} > r$) is zero, i.e. the scoring rules given in (3.21), (3.22) and (3.23) are equal.

To understand the behavior of the power curves and their relating conclusions more fully, we proceed by examining the expected mean relative scores of the *wl*, *cl* and *csl* scoring rules. Recall that the mean relative score is a crucial ingredient for the power plots due to its close relation with test statistic in (3.28). To obtain a "proxy" of the mean relative scores we simulate 1,000 independent samples from the competing copula specifications A and B. We set Model A equal to the correct copula specification (equal to the DGP) and calculate the mean relative scores following (3.25), (3.26) and (3.27). The simulated values of $E[\bar{d}_{t+1}^{wl}, r], E[\bar{d}_{t+1}^{cl}, r]$ and $E[\bar{d}_{t+1}^{csl}, r]$ are depicted in figure 4.5 for both the Gaussian DGP and the Student-t(5) DGP and sample sizes with $m = 5$ and $m = 40$. Several observations can be made.

First, we observe that figure 4.5 is in line with the power plots in figure 4.3 and 4.4. The areas with negative expected mean relative scores of the *wl* scoring rule correspond to the regions where spurious power is high. Furthermore, the pattern of stronger power for the *csl* scoring rule as compared to the *cl* scoring rule is also apparent in figure 4.5. In particular for small values of r, the curve of the expected mean relative score of the *csl* scoring rule exceeds the curve corresponding to the *cl* scoring rule. These results should be expected, since the mean relative score serves as numerator for the test statistic in (3.28).

Second, the curves of the two graphs at the left hand side in figure 4.5, corresponding to the small sample size ($m = 5$) appear to be less stable than the graphs at the right hand side ($m = 40$). To gain better insight in the uncertainty of the mean relative scores, we provide the 10% and 90% percentiles for the Gaussian DGP in figure 4.6 and the Student-t(5) copula as DGP in figure 4.7. By comparing the left and right graphs of both figures, we observe that increasing the sample size tightens the confidence interval of the mean relative scores considerably, for all three scoring rules considered. At this point, it should be emphasized that decreasing pattern of the *true* power plots for the *cl* and *csl* scoring rules in figure 4.3 and 4.4 is due to the decreasing sample size, and *not* due to lower mean relative scores (in fact, the expected mean relative score are *increasing* for larger threshold values *r*). Indeed, if the threshold value r goes up, a smaller sample size is required to assure *m* expected observations in the region of interest. In general, the (absolute) test statistic in (3.28) decreases with sample size[17]. In order to obtain reliable test statistics, it is therefore crucial to take a large enough sample size.

Because we are mostly interested in the left tail of the copula distribution, we analyze the lower graphs of figure 4.6 and figure 4.7 in more detail, since these graphs "zoom" in the tail of the copula distributions with $r \in [0.01, 0.1]$. It is striking to see that the overlap of the (80%) confidence intervals of the three scoring rules is very limited in case of the large sample size ($m = 40$). For the Gaussian DGP in figure 4.6, the confidence interval of the *csl* scoring rule clearly shows the highest mean relative scores, and thus outperforms the other scoring rules. The *cl* scoring rule ranks second, and the *wl* scoring rule performs the worst with an entirely negative confidence interval. Note that the bad performance of the *wl* scoring rule is in line with its excess *spurious* power for the Gaussian DGP, which can be understood from the example and related discussion concerning figure 3.3. In case of the Student-t(5) copula as DGP in figure 4.7, the *wl* scoring rule ranks best, followed by the *csl* and *cl* scoring rules. These rankings exactly correspond to the rankings that are apparent from figure 4.3 and 4.4. Again, this suggests that the power plots and the figures depicting the expected mean relative scores are complementary.

---

[17] NB: This effect is direct through a smaller sample size P and indirect through increasing uncertainty $\hat{\sigma}_{R,P}^2$.
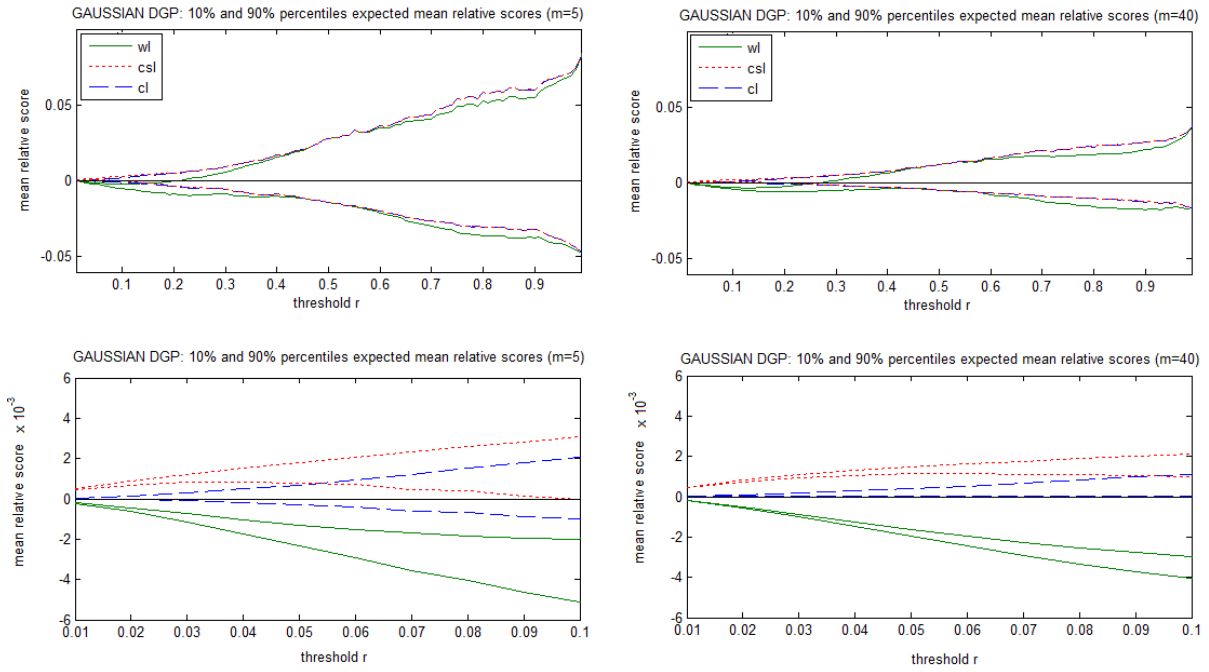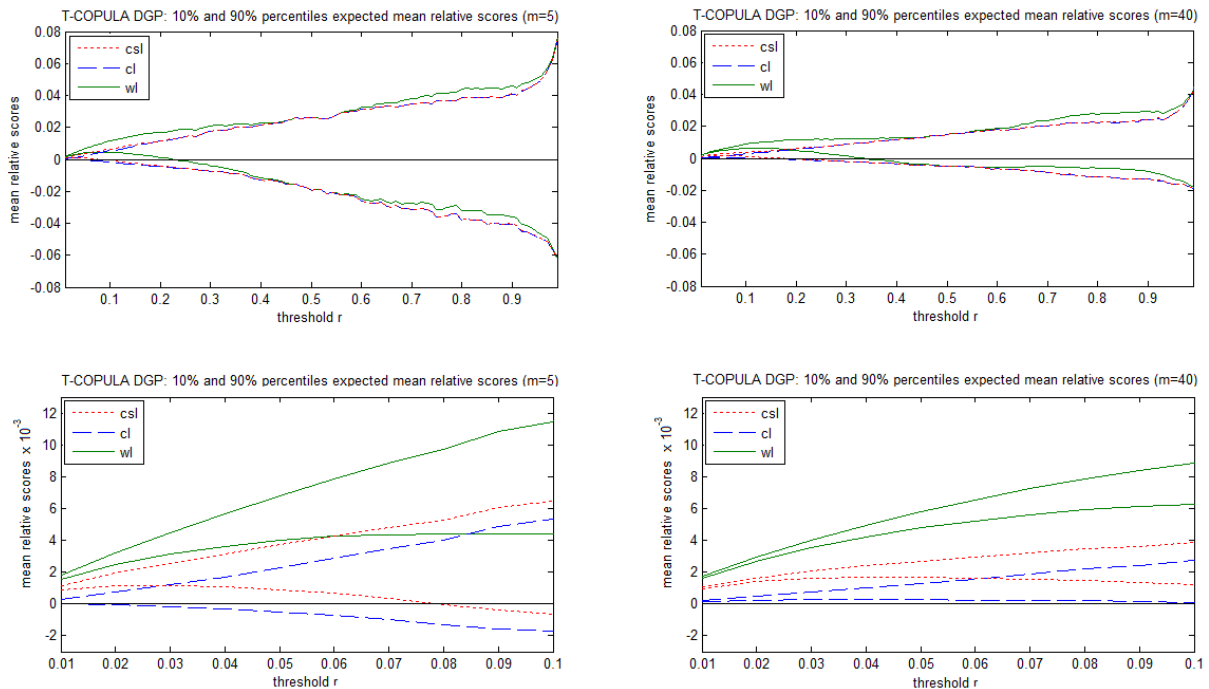
Figure 4.6: 10% and 90% percentiles for the upper two graphs of figure 4.5, i.e. the mean expected relative scores of comparing an independent Gaussian Copula with the independent Student-t(5) Copula, with the independent Gaussian Copula as DGP. The two graphs in the left (right) panel correspond to the m=5 (m=40) expected observations in the region of interest. The upper two graphs show the results for the full range of r, that is $r \in [0.01, 1]$, the lower two graphs "zoom" in the left tail of the copula with $r \in [0.01, 0.1]$.

Figure 4.7: 10% and 90% percentiles for the lower two graphs of figure 4.5, i.e. the mean expected relative scores of comparing an independent Student-t(5) copula with the independent Gaussian copula, with the independent Student-t(5) copula as DGP. The two graphs in the left (right) panel correspond to the m=5 (m=40) expected observations in the region of interest. The upper two graphs show the results for the full range of r, that is $r \in [0.01, 1]$, the lower two graphs "zoom" in the left tail of the copula with $r \in [0.01, 0.1]$.

To examine how the power changes for varying values of the correlation parameter ρ_{12} the results of the above simulation experiments for $\rho_{12} = 0.3, 0.6$ and 0.9 are given in Appendix 5. Before we summarize the main findings of these results, we examine the relative log-likelihood scores $(\log c_{A,t}^{Ga}(y_{t+1}) - \log c_{B,t}^{T}(y_{t+1}))$ for increasing ρ_{12} in more detail (figure 4.8). These graphs suggest that for increasing correlation, it becomes relatively harder to differentiate between the (log-)pfds of the competing copula distributions in the center of the copula distributions and, in particular, in their left and right tails. The results in Appendix 5 confirm this observation: in general, both *spurious* and *true* power decrease for all three scoring rules considered, i.e. for increasing correlation, the *wl*, *cl* and *csl* scoring rules have more difficulties in differentiating between the two competing models. The expected mean relative scores of the *cl* and *csl* scoring rules exhibit a stable positive pattern, but their confidence intervals are getting wider. This pattern, particularly present in the left tail, causes the drop in true power for both proper scoring rules. Moreover, we experience that for increasing correlation the *cl* scoring rule is converging to the *csl* scoring rule and even performs somewhat better in the (extreme) case of $\rho_{12} = 0.9$. For the *wl* scoring rule the confidence interval also widens, but the entire interval experiences a "shift" towards those of the proper scoring rules. However, only in the case of $\rho_{12} = 0.9$, we experience overlap between the confidence intervals of the *wl* scoring rule and either of the proper scoring rules. Based on the results for $\rho_{12} = 0.6$ and $m = 40$, we emphasize that for reasonable correlation and large enough sample size the proper scoring rules are still highly preferable to the improper *wl* scoring rule.

So far, we have only considered independent symmetric copulas for our simulation experiments. The question rises whether similar power results follow if one of the two competing copulas is *asymmetric.* To this end, we compare the predictive accuracy of the symmetric Gaussian copula with the asymmetric Clayton copula, which shows positive dependence in the left tail, as discussed in section 3.2.3. Basically, we use the exact same set up as described in the beginning of this section, but substitute the Student-*t*(5) copula for the Clayton copula. In order to make a fair comparison between the two copulas we fix the measure of concordance to $\tau = 0.2$ which corresponds to $\rho = \sin\left(\frac{\pi}{2}\tau\right) = 0.31$ for the Gaussian copula parameter and $\theta = \frac{2\tau}{1-\tau} = 0.50$ for the Clayton parameter. The corresponding results are given in figure 4.9 to figure 4.13.
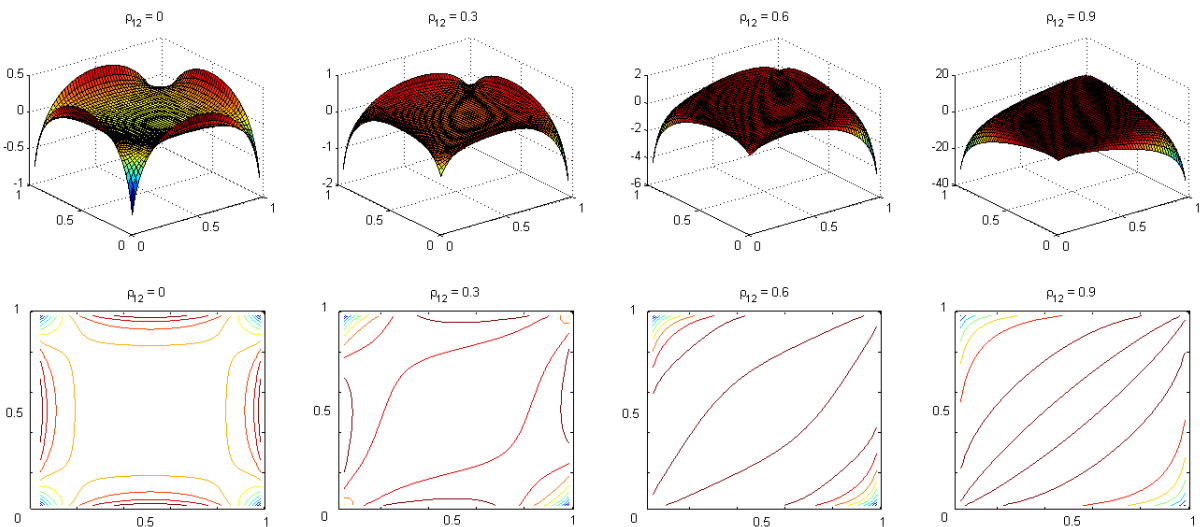

Figure 4.8 The upper panel shows the relative log-likelihood scores $\log c_{A,t}^{Ga}(y_{t+1}) - \log c_{B,t}^{T}(y_{t+1})$ for increasing values of parameter$\rho_{12} = 0, 0.3, 0.6$ and 0.9 (l-r). The lower panel shows their corresponding contourplots.
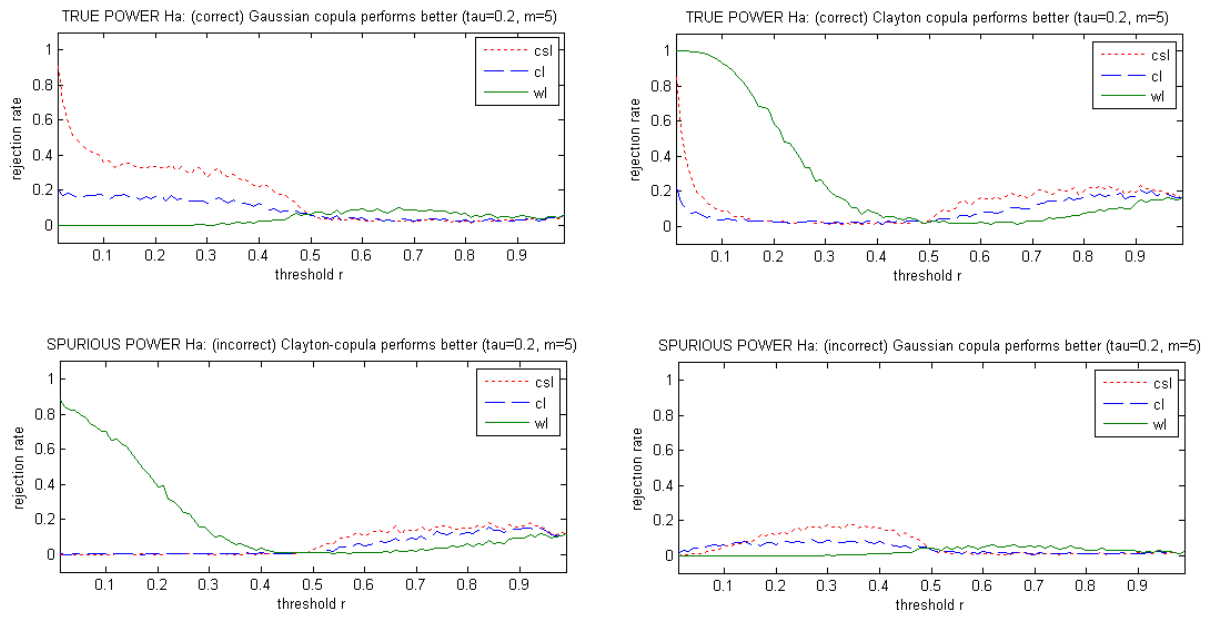
Figure 4.9 One sided rejection rates at significance level 5% to assess the power properties of the equal predictive ability test as defined in (3.28) when using the weighted logarithmic (wl), the conditional likelihood (cl) and the censored likelihood (csl) scoring rules under the weight function $w(u, v) = I(u, v \leq r)$ where $0 < r < 1$, based on 1,000 replications and m=5 expected observations in the region of interest. For the graphs in the left (right) columns, the DGP is the Gaussian Copula (Clayton Copula), both with parameter fixed at the measure of concordance $\tau = 0.2$. The graphs in the top (bottom) panels show average rejection rates against superior predictive ability of the correct (incorrect) copula specification, i.e. true (spurious) power.
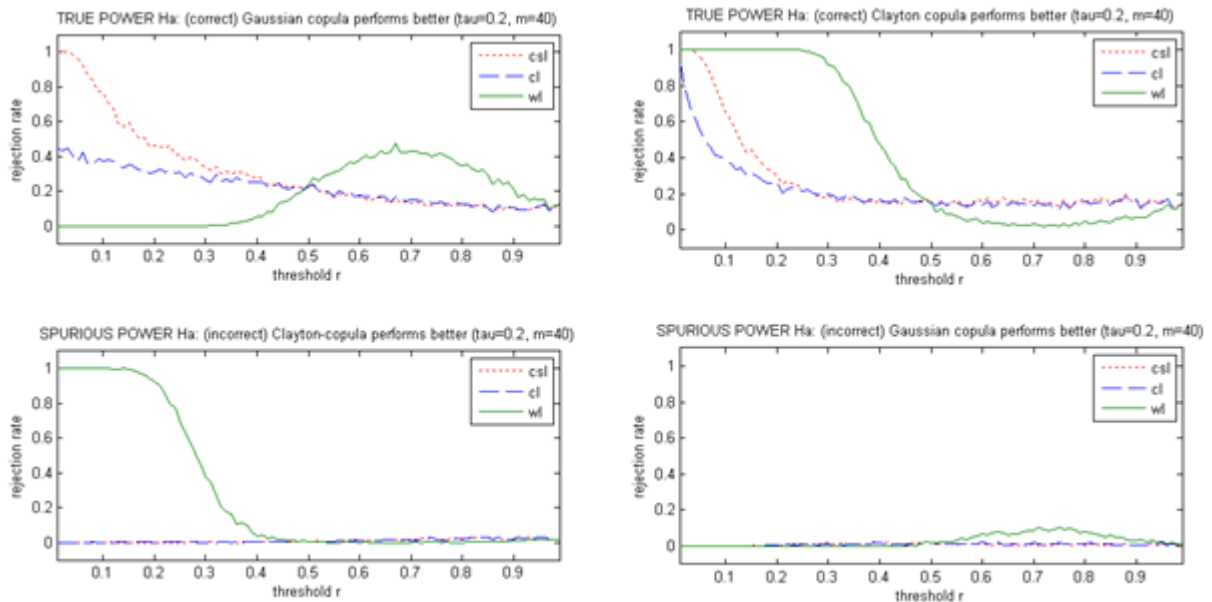

Figure 4.10 This figure is similar to Figure 4.8, but now with m=40 expected observations in the left tail.
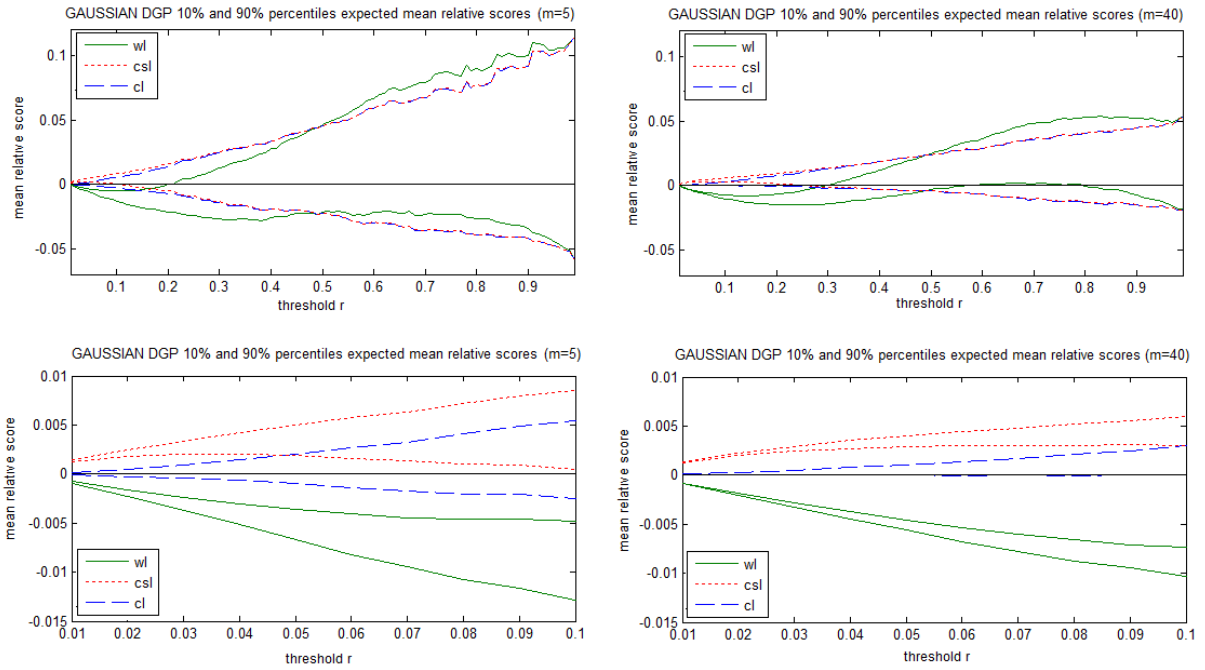
Figure 4.11: 10% and 90% percentiles for the upper two graphs of figure 4.8, i.e. the mean expected relative scores of comparing a Gaussian Copula with a Clayton Copula, with the Gaussian Copula as DGP and the measure of concordance fixed to $\tau = 0.2$. The two graphs in the left (right) panel correspond to the m=5 (m=40) expected observations in the region of interest. The upper two graphs show the results for the full range of r, that is $r \in [0.01, 1]$, the lower two graphs "zoom" in the left tail of the copula with $r \in [0.01, 0.1]$.

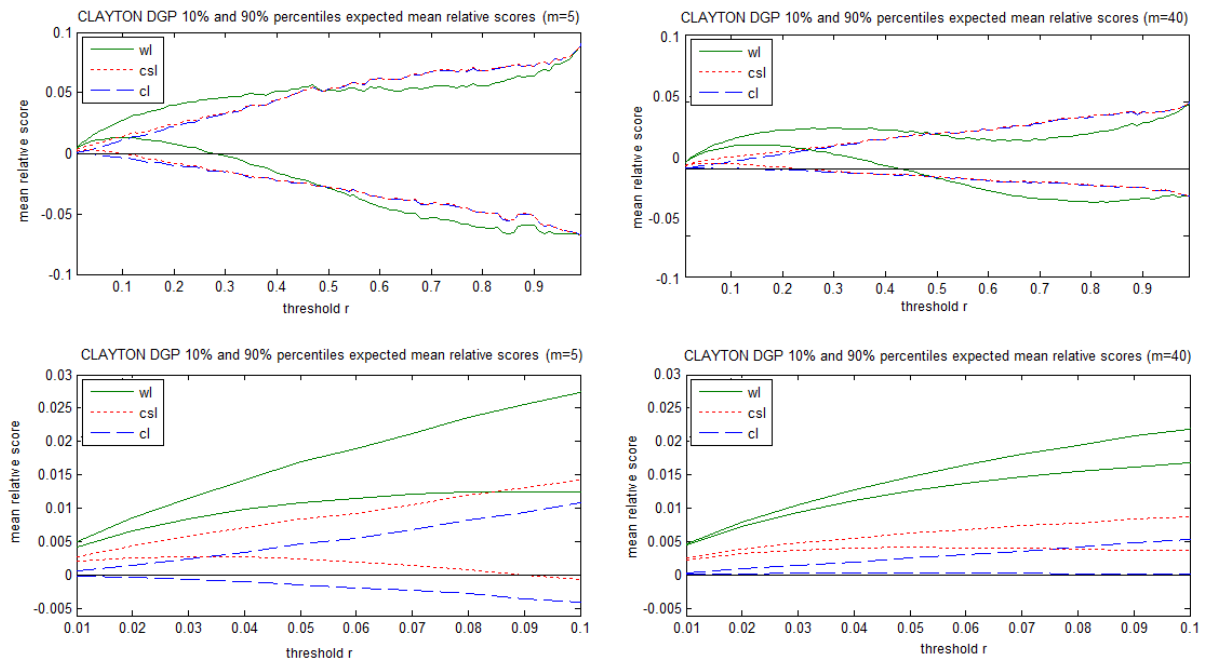**Figure 4.1210% and 90% percentiles for the lower two graphs of figure 4.8, i.e. the mean expected relative scores of comparing a Clayton Copula with a Gaussian Copula, with the Clayton Copula as DGP and the measure of concordance fixed to $\tau = 0.2$. The two graphs in the left (right) panel correspond to the m=5 (m=40) expected observations in the region of interest. The upper two graphs show the results for the full range of r, that is $r \in [0.01, 1]$, the lower two graphs "zoom" in the left tail of the copula with $r \in [0.01, 0.1]$.**
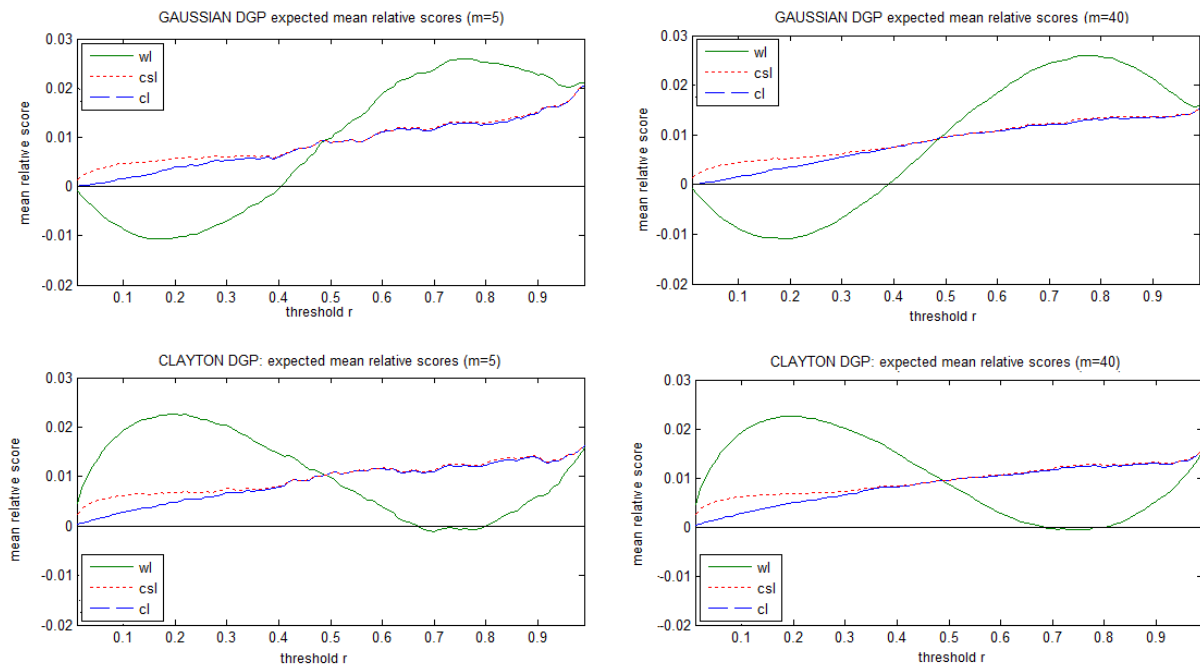
Figure 4.13: Proxy of the mean relative scores for sample sizes m=5 (left graphs) and m=40 (right graphs), when using the weighted logarithmic (*wl*), the conditional likelihood (*cl*) and the censored likelihood (*csl*) scoring rules under the weight function $w(u, v) = I(u, v \leq r)$ where $0 < r < 1$, based on 10,000 replications. The upper (lower) graphs show the results for the Gaussian copula (Clayton copula) as DGP.

The above results concerning the competing Gaussian and Clayton copula are in line with the conclusions drawn from the comparison of the (independent) Gaussian and Student-t(5) copula. Again, the *wl* scoring rule shows substantial spurious power, favoring the copula model with the largest probability mass in the region of interest. The *cl* and *csl* scoring rule show limited spurious power, which virtually disappears when increasing the sample size, i.e. their spurious power is a small sample problem. In terms of expected mean relative scores in figure 4.13, the two proper scoring rules again show a stable performance with entirely positive expected mean relative scores. Similar to earlier results, the *csl* scoring rule performs slightly better, in particular for small values of *r*. At last, also the "zoomed" confidence intervals are similar, the rankings between the different scoring rules (and their related true power plots) are equal. This result would be expected, since the Student-t(5) copula and Clayton copula both contain substantial lower tail dependence.

Note that the main difference between the (independent) symmetric and asymmetric simulation results is visible by comparing their corresponding expected mean relative scores in figure 4.5 and 4.13 respectively. While the curve of the *wl* scoring does not cross the curves of the proper scoring rules, there is a clear intersection in figure 4.13 near $r = 0.5$. This intersection suggests that, relatively to the proper scoring rules, the *wl* favors the Clayton copula for small values $r < ca. 0.5$ (the left tail), and the Gaussian copula for $r > ca. 0.5$. This result is plausible, as the Clayton copula contains positive lower tail dependence and very weak upper tail dependence.

5 Empirical Application

In this section we examine the practical relevance of our goodness-of-fit test for comparing alternative multivariate density models in the left tail. We start this section with a brief introduction of our dataset, which consists of MSCI Total Return Indices of four different countries. The corresponding data characteristics are thoroughly investigated in section 5.2, with particular interest for the stylized facts of financial asset returns. We use this information to decide on a suitable model for the marginal distributions in section 5.3 and analyze whether the proposed model works properly. In section 5.4 we present our main results in the form of tests statistics that are obtained by comparing the six (bivariate) copula specifications as introduced in section 2.3. We end this section with a short discussion of our results.

5.1 Data

The empirical data sample used in this thesis consists of daily MSCI Total Return Indices of four countries (USA, Mexico, Argentina and Brazil) in the period from January 2 1992 to July 27 2010 (source: Datastream). To assure continuous data, we remove the days in which the exchange was closed in at least one of the countries and transform the MSCI Total Return Indices to log-returns hereafter. The total sample size for each country consists of 4,241 observations, which allows us to use an in-sample period of R=2,000 observations.

MSCI Total Return Indices measure the market performance and include the performance of both the price and the income from dividend payments[18]. Total return indices have the advantage that they are available over a large span of time, and they are highly frequently dealt throughout the world.

We consider these four specific countries from the Americas for two main reasons. First, they consist of a nice mixture of developed (USA) and developing (Brazil, Argentina, and Mexico) economies. This allows us to consider dependence structures of both between and within these different groups. Second, they are situated in similar time zones, such that the information asymmetry between the countries is limited. For example, if one would compare the total return indices of a European country with a (Latin-) American country, the time difference could considerably influence the dependence structure between the two indices, as equities are traded with different information.

---

[18] Source: www.mscibarra.com

5.2 Stylized facts of asset returns

5.2.1 Univariate stylized facts

When working with financial asset returns, it is informative to check whether the individual samples exhibit the commonly acknowledged univariate stylized facts in empirical asset returns. In this section we analyze our data for the presence of these univariate stylized facts, given by:

(i)   The distributions of returns are not normal
(ii)   (almost) No significant autocorrelations in returns
(iii)   Small but very slowly declining autocorrelations in squared returns

Stylized fact (i) can be checked by investigating the descriptive statistics of our data in table 5.1. We compare the results for our data with the characteristics of the normal distribution, which is symmetric (zero skewness) and has a kurtosis of three. For all four countries considered, the descriptive statistics in table 5.1 show excess kurtosis and nonzero skewness, implying fat-tailed and asymmetric distributions. The Jarque-Bera test statistics confirm these observations, as the null-hypotheses of normally distributed returns are convincingly rejected for all countries, with p-values of 0.0000.

Figure 5.1 can be used to detect whether stylized fact (ii) and (iii) are present in the data. The results in the upper panel of figure 5.1 suggest that autocorrelations in (total) returns hardly exceed the 5% significance bound for all countries considered, which is evidence for stylized fact (ii). The autocorrelations of squared returns in the lower panel of figure 5.2 however are significant and slowly decline for increasing lag size. This is evidence for the third stylized fact, suggesting the presence of *volatility clustering*, i.e. periods of large returns alternate with periods of small returns.

	USA	Mexico	Argentina	Brazil
Mean	-0.0000	0.0004	0.0005	0.0004
Median	0.0006	0.0008	0.0010	0.0012
Maximum	0.1034	0.1191	0.1634	0.2310
Minimum	-0.0950	-0.1304	-0.1844	-0.1407
Std. Dev.	0.0129	0.0169	0.0246	0.0239
Skewness	-0.2395	0.0578	-0.2923	0.0896
Kurtosis	10.0805	7.6018	9.8247	9.3101
Jarque-Bera	8899.5	3744.4	8290.8	7041.8
Probability	0.0000	0.0000	0.0000	0.0000
Observations	4241	4241	4241	4241

Table 5.1: Descriptive statistics of the daily log MSCI Total Return series in the period 2-1-1992 to 26-7-2010
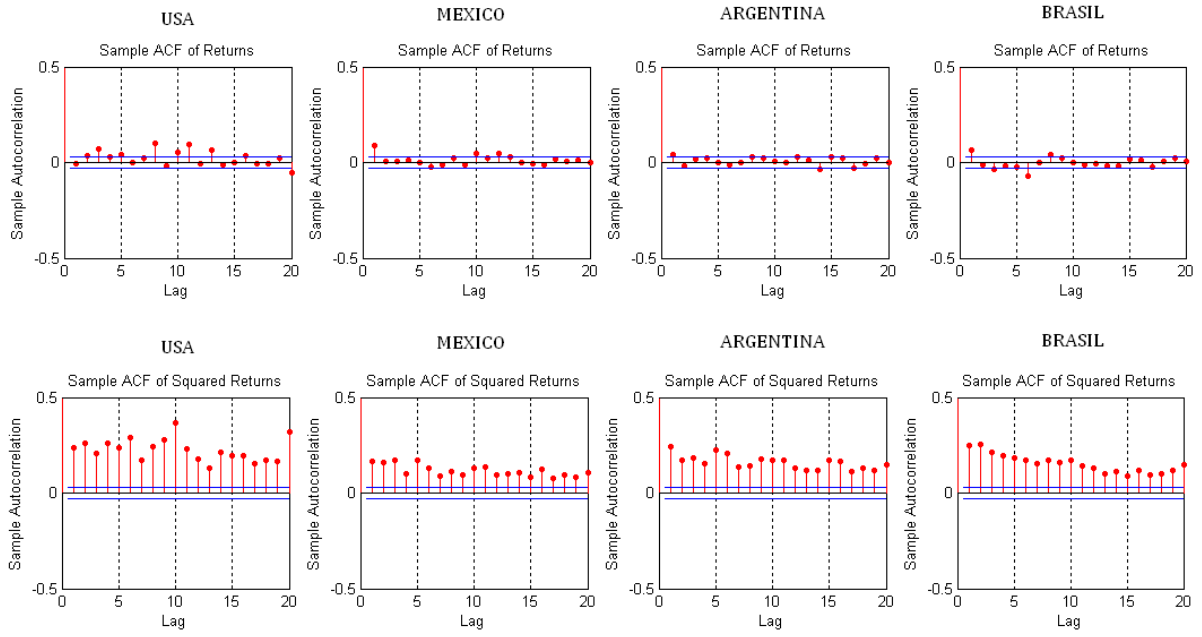
Figure 5.1: The upper panel exhibits the sample autocorrelation of the total returns of the full dataset, that is the period from 2-1-1992 to 26-7-2010. The lower panel shows the sample autocorrelation of the corresponding squared total returns. The horizontal blue lines display the 5% significance levels.

44

5.2.2 Multivariate stylized facts

The joint distribution of multivariate empirical asset returns is typically not normally distributed, partly due to the univariate stylized facts of asset returns, as discussed in the last subsection. In this section we focus on the (bivariate) *dependence characteristics* of our dataset, and evaluate whether the joint behavior of our data deviates from what would be expected under normality. In short, we focus on:

(i)   Tail dependence
(ii)   Asymmetric dependence
(iii)   Time varying dependence

In the remainder of this section, we will consider several graphical tools to detect whether the three above mentioned dependence characteristics are present in our data.

We start off with a brief description of the Chi-plot introduced by Fisher and Switzer (1985), a useful nonparametric tool to visualize dependence. The Chi-plot complements the rank scatterplot and relies on the empirical marginal distributions F_i and G_i and their empirical joint distribution H_i. Under the null hypothesis of independence, one should have $H_i \approx F_i \times G_i$. The Chi-plot is obtained by plotting the measures

$$\chi_i = \frac{H_i - F_i G_i}{\sqrt{F_i(1 - F_i)G_i(1 - G_i)}} \tag{5.1}$$

$$\lambda_i = 4 \, \text{sign}\left(\tilde{F}_i \tilde{G}_i\right) \max\left(\tilde{F}_i^2, \tilde{G}_i^2\right) \tag{5.2}$$

on the horizontal axis and the vertical axis respectively, where $\tilde{F}_i = F_i - 1/2$, $\tilde{G}_i = G_i - 1/2$ for $1 \leq i \leq n$. The measure on the vertical axis, λ_i, is bounded by the interval [-1,1] and can be interpreted as a (signed) measure of distance between the pair (X_i, Y_i) and the center of the dataset. The intuition of $\chi_i \in$ [-1,1] on the vertical axis is straightforward, as it measures the departure from bivariate independence. Fisher and Switzer (1985) propose confidence bounds of $\pm 1.78 / \sqrt{n}$ for a 95% confidence interval to detect whether positive or negative values of χ_i are significant departures from the null hypothesis of independence. Appendix 7 provides examples of Chi-plots of different copula specifications, and how they change for varying parameter values. For more details, see Fisher and Switzer (1985, 2001).

Figure 5.2 provides a matrix with the rank scatter plots and corresponding Chi-plots of the six country pairs of our dataset (full sample). Note that from the rank scatter plots it is hard to judge by human eye whether (asymmetric) tail dependence is present. It is therefore more informative to analyze their corresponding Chi-plots. As becomes clear from the Chi-plots, all country pairs clearly contain positive association that does not go back to zero for $|\lambda| \to 1$, what would be expected under normality, but it persists. This observation suggests tail dependence, since λ is a measure of distance between the pair (X, Y) and the center of the dataset. As a rough check we compare the results in figure 5.2 with the Chi-plots of different copulas in Appendix 7. As expected, the Chi-plots generated from our data correspond better to the Student-t Chi-plots than Gaussian Chi-plots. Note that for $\tau = 0.2$ the shape of the asymmetric Gumbel Chi-plot (with $\theta = 1.25$) in Appendix 7 also shows similarities to the Chi-plots in figure 5.2. However, a quick comparison also tells us that the asymmetry in tail dependence is not likely to be as extreme as the Gumbel and Clayton copulas with large values of Kendall's tau ($\tau > 0.35$).
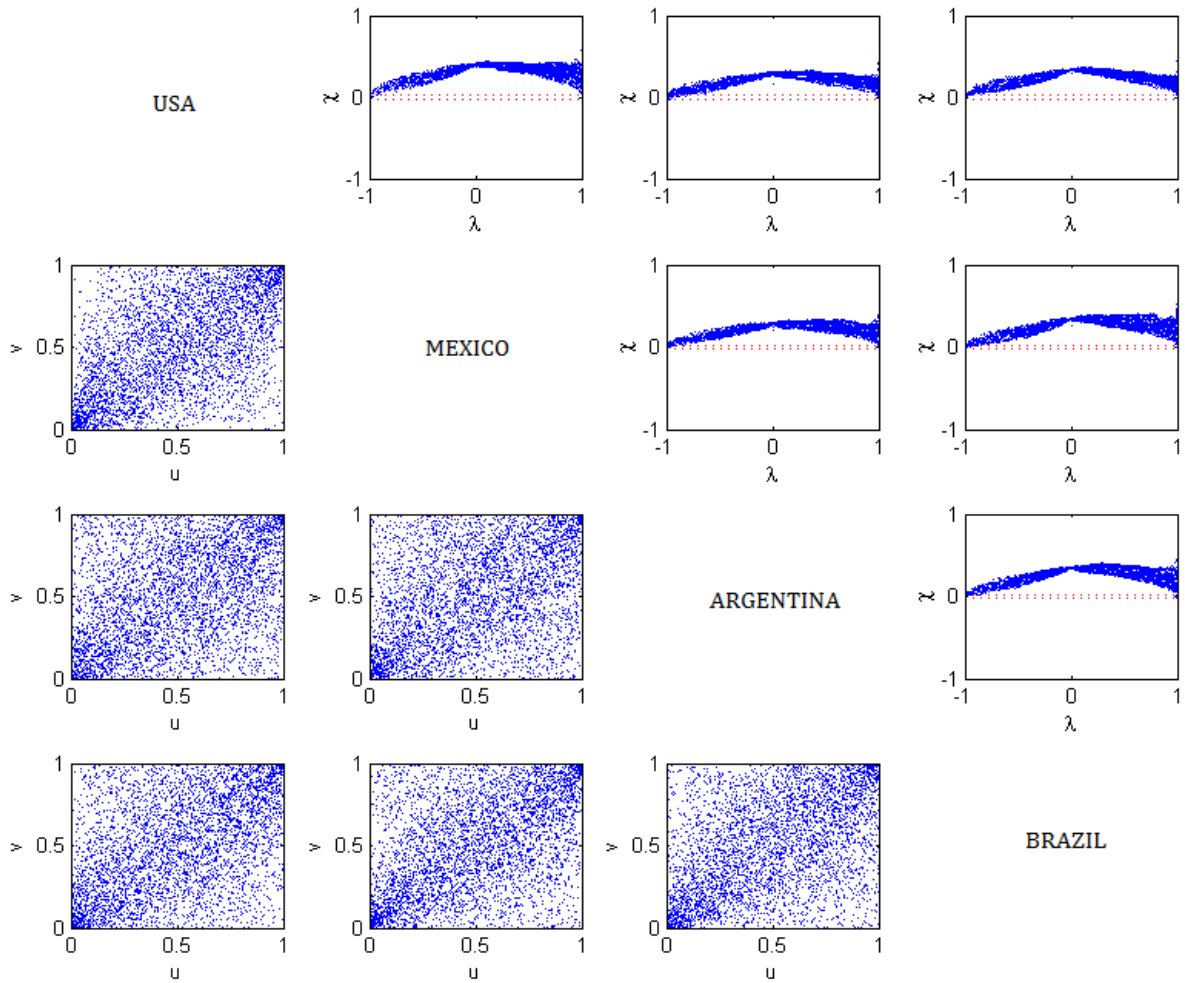
Figure 5.2 Rank scatter plots and corresponding chi-plots of all country pairs using the full dataset, that is the period from 2-1-1992 to 26-7-2010. The horizontal red dotted lines in the chi-plots display the 5% significance control limits.

To analyze the tail dependence from a different perspective, figure 5.3 compares the *quantile dependence* of the country pairs in our dataset with their normally distributed counterparts. In order to create the "tent-shaped" curve of the normally distributed quantile dependence, we used the empirical correlations derived from our dataset. In figure 5.3, we observe that in general, the dependence in the lowest quantiles exceeds the dependence corresponding to the normally distributed distribution, suggesting *negative* tail dependence. In the highest quantiles however, the dependence structures of the normal and empirical data are rather similar. These results suggest asymmetric (tail) dependence, i.e. correlations in "bear" markets are higher than in "bull" markets. This clearly is a deviation from the symmetric multivariate normal distribution.

Figure 5.4 provides a rough check of time-varying dependence, showing the 250-day unconditional correlation of the six country pairs for the full sample period. We observe highly nonlinear patterns for all six pairs that considerably vary over time. Again this clearly is a departure from the multivariate normal distribution, which assumes constant correlation.
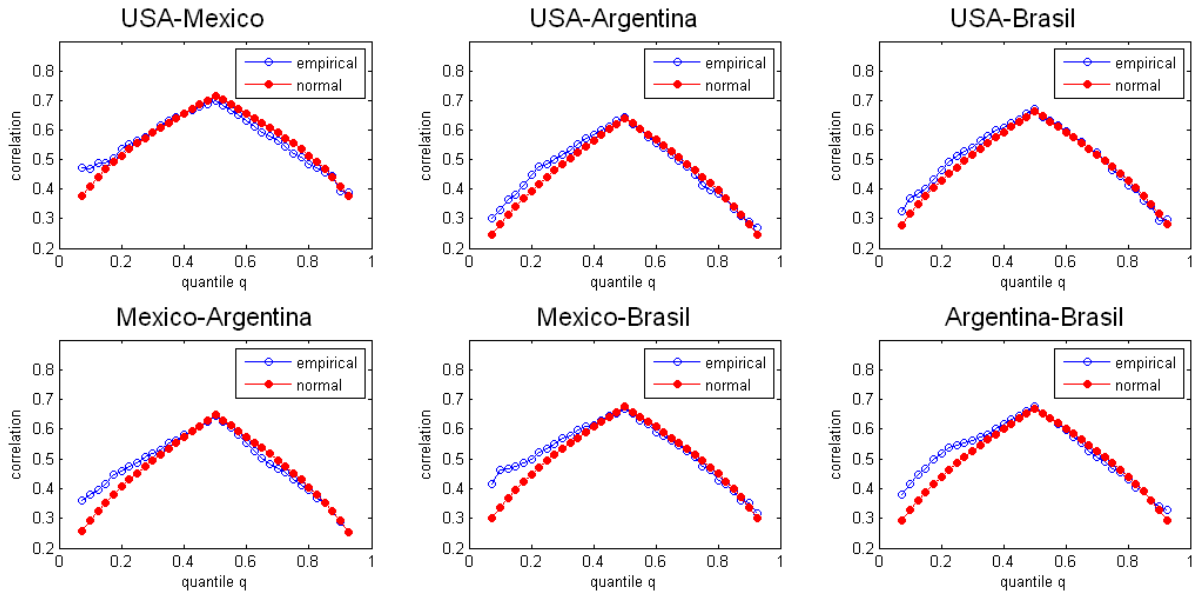
Figure 5.3: Quantile dependence of all country pairs using the full dataset, that is the period from 2-1-1992 to 26-7-2010. The blue (red) curve displays the empirical (normally distributed) data.
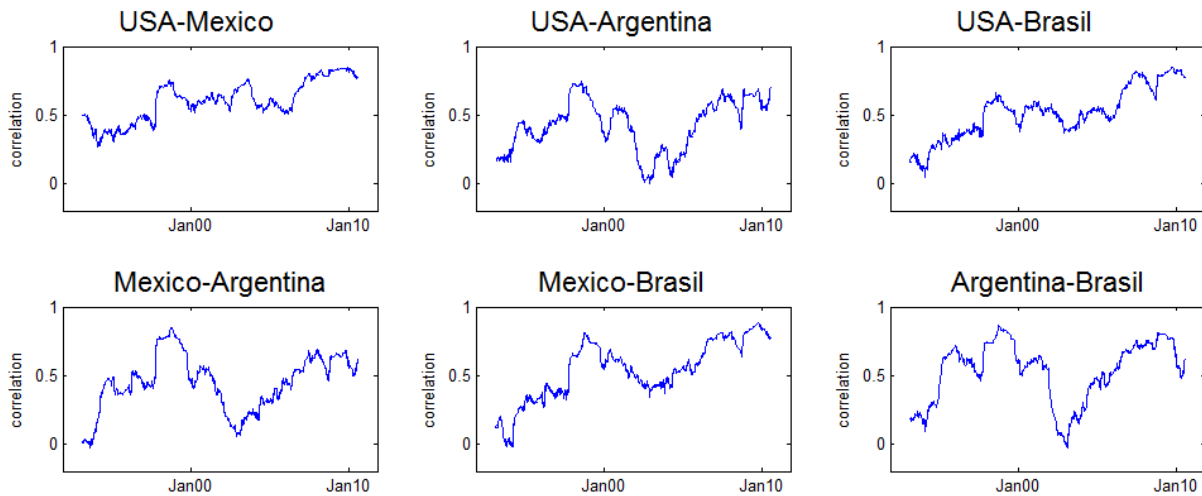

Figure 5.4: 250-day historical unconditional correlation of all country pairs using the full dataset, that is the period from 2-1-1992 to 26-7-2010.

So far, we have found evidence for all three dependence characteristics mentioned at the beginning of this section. The question rises whether (asymmetric) *tail* dependence changes over time. To this end, we introduce the concept of exceedance correlations from Ang and Chen (2002), and provide the figures of two non-overlapping time periods. Specifically, we select the first in-sample period of $R = 2,000$ observations (in figure 5.5) and the last 2,000 observations of the sample (in figure 5.6).

As can be observed in the relating figures, the specific shape of the exceedance correlation plots vary for every country pair and deviate quite strongly from the "tent-shaped" curve as shown by their normally distributed counterparts. By comparing figure 5.5 and figure 5.6, one can find evidence for all three dependence characteristics as given in at the beginning of the section. First, the exceedance correlations in (at least one of) the tails are typically higher than the middle part of the distribution. Second, this pattern is often asymmetric, in the sense that one of the tails shows more exceedance correlations than the other. Third, the exceedance correlation plots seem to change quite strongly over time, as the shapes of the curves in figure 5.5 are rather different than the corresponding curves in figure 5.6.
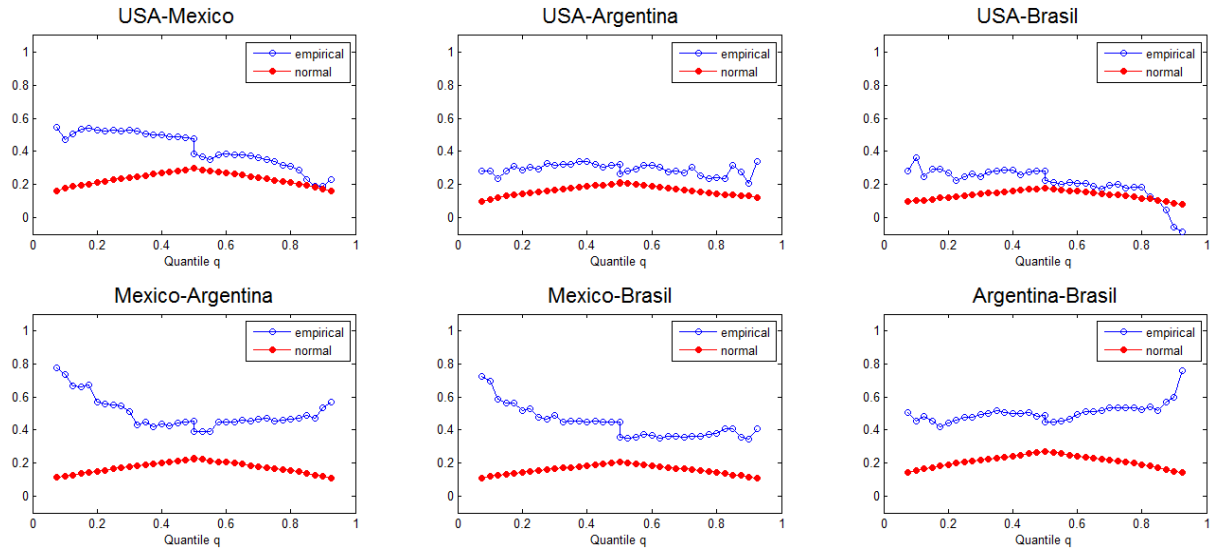

Figure 5.5: Exceedance correlations of all country pairs using the first in-sample period, i.e. the first 2,000 observations from 2-1-1992 to 6-10-2000. The blue (red) curves display the empirical (normally distributed) data.
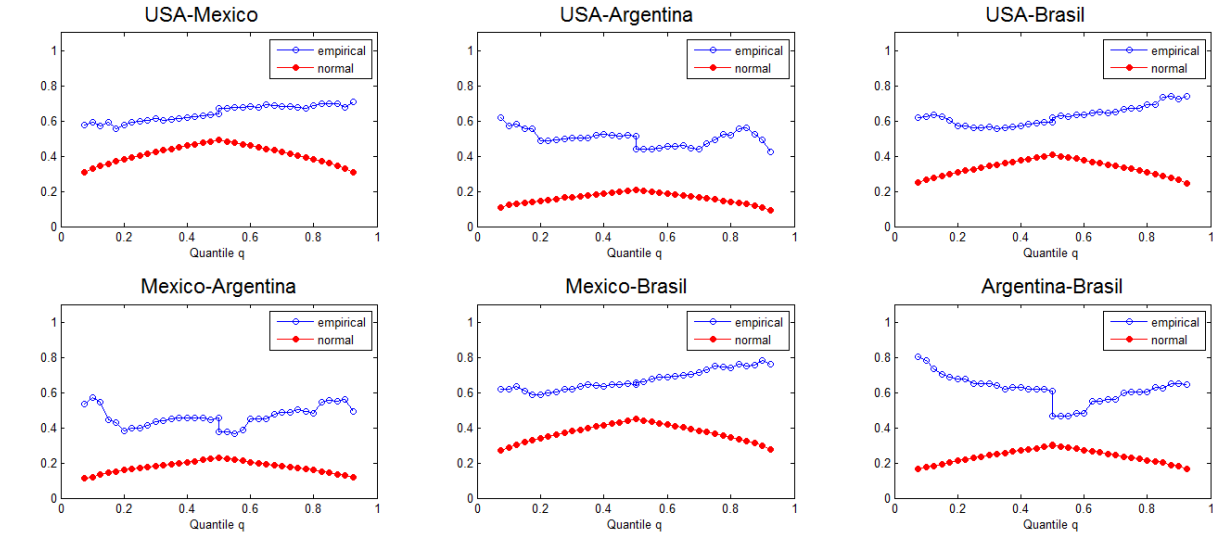

Figure 5.6 Exceedance correlations of all country pairs using the last 2,000 observations of the data sample from 5-11-2001 to 26-7-2010. The blue (red) curves display the empirical (normally distributed) data.

5.3 Modeling the marginals

In order to capture the volatility effects that are present in our data (see section 5.2.1), we model the marginal distributions of our dataset using a model from the class of well-developed GARCH-type conditional volatility models. This implies that we adopt the SCOMDY framework from section 2.4.2 to implement the multivariate density models.

The first stage of a SCOMDY model consists of estimating the conditional mean and conditional variance to obtain the standardized residuals. We consider the asymmetric GJR-model to model the conditional volatility, following Glosten, Jagannathan and Runkle (1993). Contrary to general symmetric GARCH specifications, this model is able to treat negative shocks differently than positive shocks. Intuitively this makes sense, because in practice, "bad news" generally has larger impact than "good news". We consider an AR(5) model for the conditional mean return combined with a GJR(1,1) model for the conditional variance, given by

$$y_{j,t} = \mu_{j,t} + \varepsilon_{j,t} = \mu_{j,t} + \sqrt{h_{j,t}}\eta_{j,t}, \quad \text{for } j = 1,2,3,4 \tag{5.3}$$

where the standardized residuals are i.i.d. with mean zero and variance one and j denotes the country indices of USA (1), Mexico (2), Argentina (3) and Brazil (4). The conditional mean $\mu_{j,t}$ and the conditional variance $h_{j,t}$ in (5.1) for $j = 1,2,3,4$ are given by

$$\mu_{j,t} = \delta_{j,0} + \sum_{i=1}^{5} \delta_{j,i} \, y_{j,t-i}, \tag{5.4}$$

$$h_{j,t} = \omega_j + \alpha_j \varepsilon_{j,t-1}^2 + \gamma_j \mathrm{I}[\varepsilon_{j,t-1} < 0]\varepsilon_{j,t-1}^2 + \beta_j h_{j,t-1}, \tag{5.5}$$

with $\mathrm{I}[\varepsilon_{j,t-1} < 0]$ equal to one if $\varepsilon_{j,t-1} < 0$ and zero otherwise.

In the second stage of the SCOMDY model, we use the standardized residuals $\varepsilon_{j,t}$ resulting from (5.1) to the marginal distributions $F_1, \dots, F_4$ by using the empirical CDF. In the third stage, these marginal distributions are used to estimate the competing copula models parametrically.

Recall that the analysis regarding to the stylized facts of asset returns in section 5.2.1 and 5.2.2 is based on the (raw) returns $y_{j,t}$. The marginal distributions $F_1, \dots, F_4$ of the SCOMDY model however, are based on $\varepsilon_{j,t}$, i.e. the standardized residuals obtained from (5.3). Since $F_1, \dots, F_4$ are used to calibrate the competing copula sepcifications, it is worthwhile to revisit the stylized facts of asset returns for $\varepsilon_{j,t}$ and compare them with the results based on $y_{j,t}$. We present the (squared) autocorrelations (figure 5.7), quantile dependence (figure 5.8), Chi-plots (figure 5.9), and exceedance correlation plots (figure 5.10) based on the standardized residuals for the first in-sample period of $R = 2,000$ observations ranging from 2-1-1992 to 6-10-2000. Note that the analyses in section 5.2.1. and 5.2.2 are based on the full sample of 4241 observations, while the model in (5.3) only considers the in-sample period. In appendix 6 we therefore provide a "fair" comparison of the figures, and provide the corresponding figures of $y_{j,t}$ and $\varepsilon_{j,t}$ based on the first 2,000 observations of the sample.

The results suggest that the AR(5)-GJR(1,1) model filters the volatility effects fairly well for the first in-sample-period of 2000 observations. Compared to figure 5.1, the autocorrelations in squared returns are substantially lower, suggesting that in most cases, the influence of *volatility clustering* is limited.

The multivariate stylized facts of asset returns however, are still present in our data. In general, the quantile dependence plots are closer to their normally distributed counterparts than in section 5.2.2, but still there is quite some evidence for (asymmetric) tail dependence, i.e. deviation from normality. The curves of the country pairs USA-Argentina and USA-Brasil however, are hard to distinguish from the curves corresponding to the normal distrubution (figure 5.9). The Chi-plot exhibits similar characteristics as in the last subsection. Tail dependence appears to be present and again the Chi-plots show most similarity with the shape of the Student-t copula or the Gumbel copula with relatively low values of θ. The exceedance correlation plots are quite similar to those in the last subsection; again there generally is evidence for (asymmetric) tail dependence.
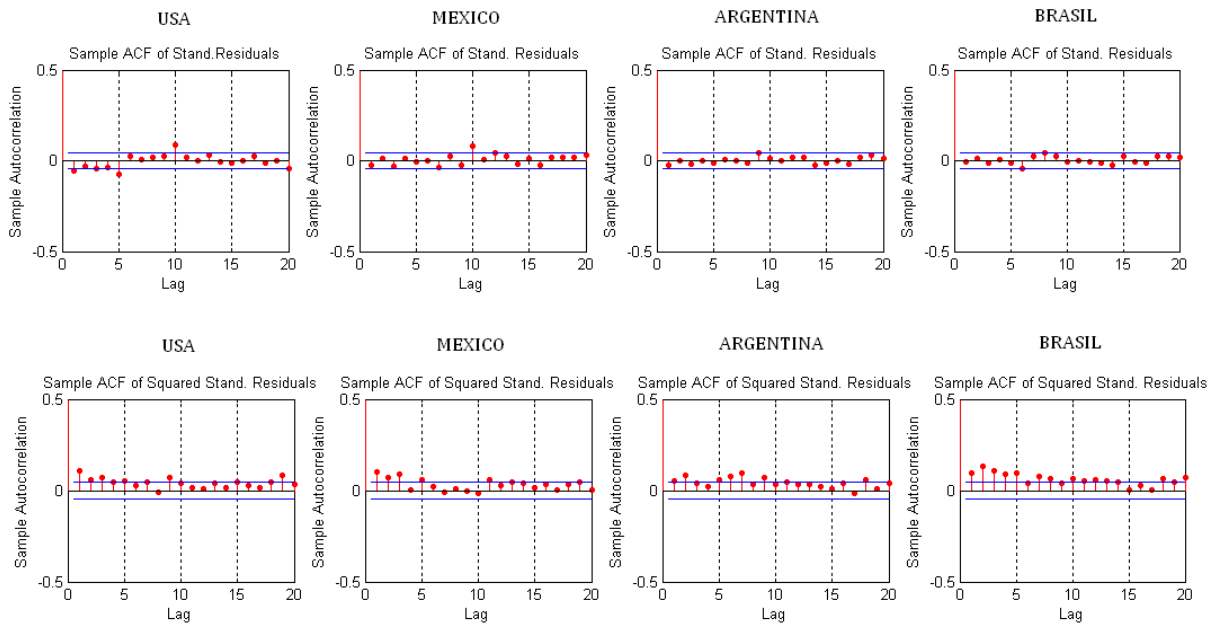

Figure 5.7: The upper panel exhibits the sample autocorrelation of the standardized residuals of the first in-sample period from 2-1-1992 to 6-10-2000. The lower panel shows the sample autocorrelation of the corresponding squared standardized residuals. The horizontal blue lines display the 5% significance levels.
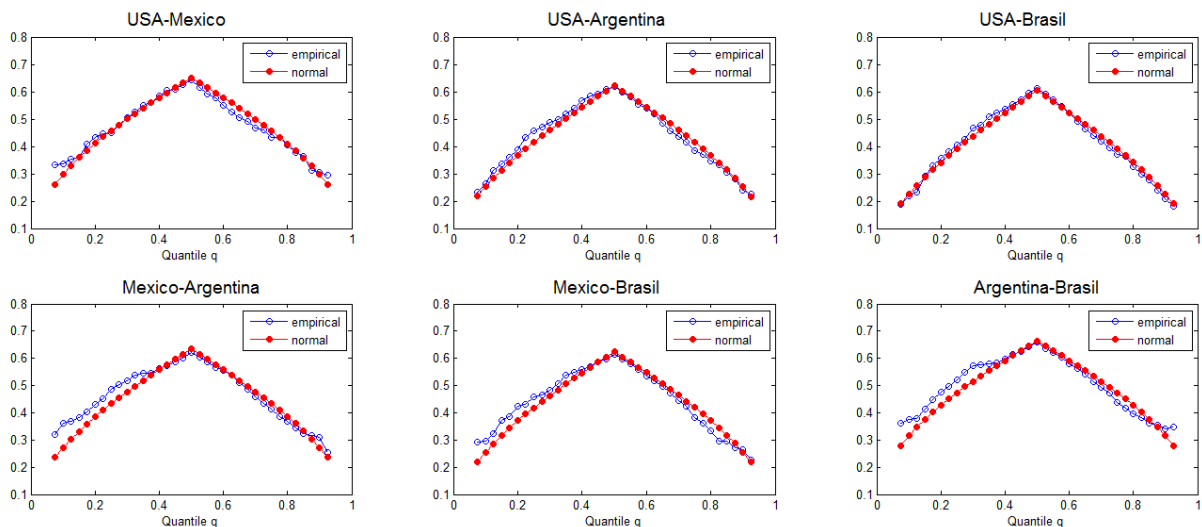

Figure 5.8: Quantile dependence of all country pairs using the standardized residuals of the first in-sample period from 2-1-1992 to 6-10-2000. The blue (red) curve displays the empirical (normally distributed) data.

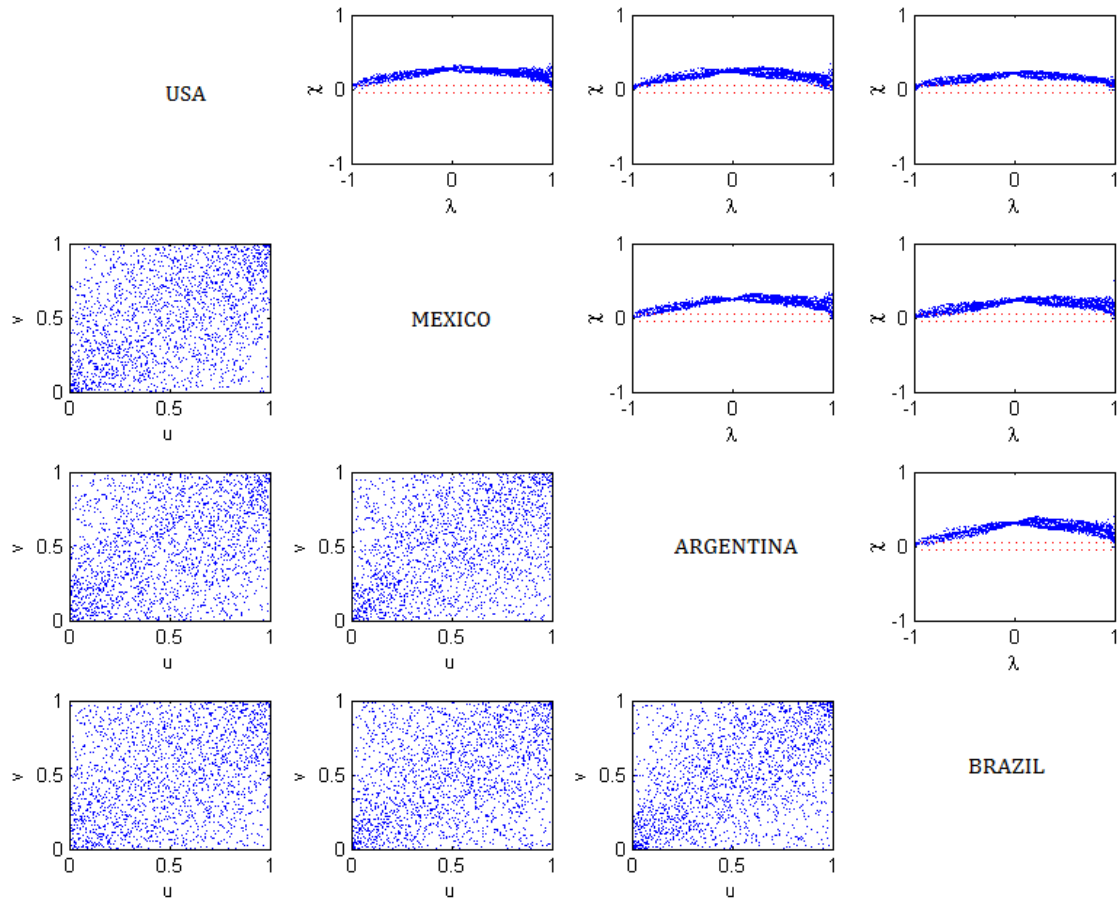RANK SCATTERPLOTS AND CORRESPONDING CHI-PLOTS STAND. RESIDUALS

Figure 5.9: Rank scatter plots and corresponding chi-plots of all country pairs using the standardized residuals of the first in-sample period from 2-1-1992 to 6-10-2000. The horizontal red dotted lines in the chi-plots display the 5% significance control limits.
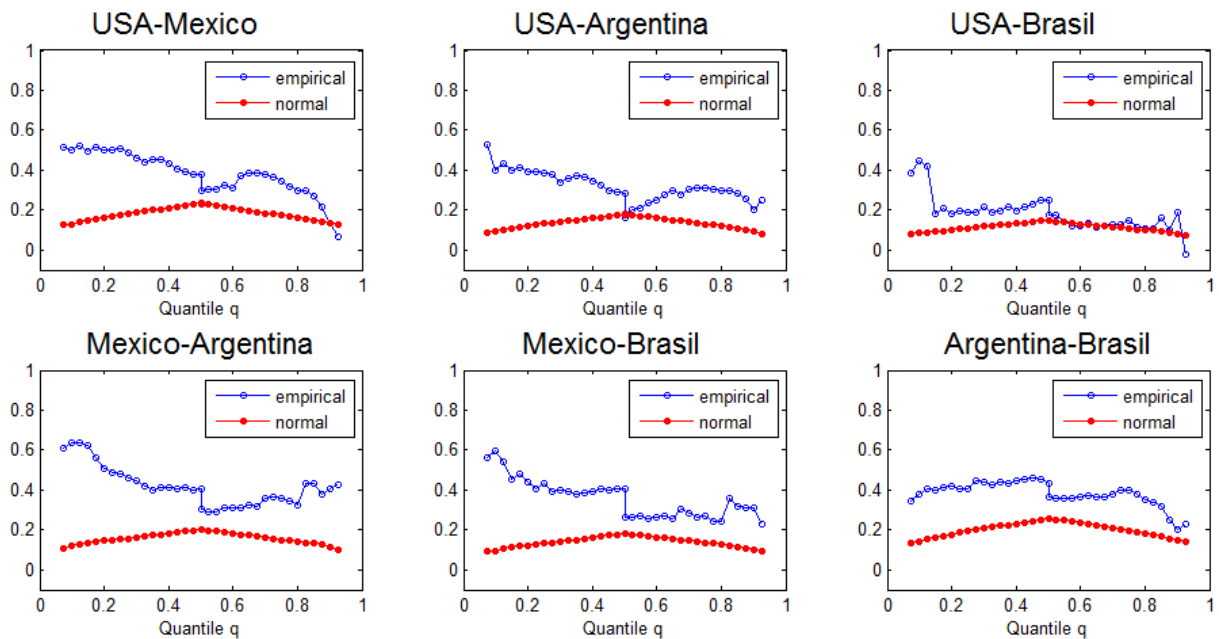

Figure 5.10: Exceedance correlations of all country pairs using the standardized residuals of the first in-sample period from 2-1-1992 to 6-10-2000. The blue (red) curves display the empirical (normally distributed) data.

5.4 Results

The results of our test statistics in (3.28) for the six country pairs from our empirical application are displayed in table 5.2 to table 5.7. In the remainder of this section we briefly describe the used *forecast methods*[19] and explain how to interpret the given tables.

We use the full dataset as described in section 5.1 with a fixed rolling in-sample window of $R = 2,000$ observations and consider $P = 2214$ one-step-ahead forecasts. The out-of-sample period consists of the period from 6-10-2000 to 26-7-2010. Furthermore, we use the SCOMDY procedure to estimate the marginal distributions and their copula, as discussed in the last subsection. Recall that we isolate the performance of the copula specifications by comparing *forecasts methods* that only differ in their respective copula models, as discussed in section 3.1. In this application, we consider the six different parametric copula specifications from section 2.3, which are divided in two groups: the Elliptical Copulas (the Gaussian, Student-t and Cauchy copula) and the Archimedean copulas (the Clayton, Gumbel and Frank copula). Note that in the remainder we use the survival (or rotated) Gumbel copula, because we concentrate on the *left tail*, and the ordinary Gumbel copula exhibits *upper* tail dependence (see section 2).

The results are provided in six different tables, each representing a different country pair. Every table exhibits four threshold values: $r = 0.01$, $r = 0.05$, $r = 0.10$ and $r = 1$, with threshold weight function $w(u, v) = I(u, v \leq r)$. For all threshold values, we display the results of fifteen pairs of competing copula specifications in a matrix, where copula A is given in the columns, and copula B is given in the rows. Positive (negative) values of the test statistic thus indicate better performance of the corresponding copula in the column (row). We separately report the test statistics based on the weighted likelihood (*wl*) scoring rule, conditional likelihood (*cl*) scoring rule and censored likelihood (*csl*) scoring rule. We do not distinguish between these scoring rules in case of $r = 1$, as in this case the scoring rules are equal.

As additional information we also included several useful figures in parentheses. Under every threshold value r we include the figure m in parentheses, which corresponds to the number of tail observations for the given threshold value. Furthermore, in every column we include the total tail probabilities of the six copula specifications corresponding to the given threshold value r, based on the *average* estimated parameter values over the entire out-of-sample period. [20]

---

[19] As discussed in section 3.1 we follow the framework of Giacomini and White (2006), who define a forecast method as the set of choices the forecaster makes at the time of prediction, including the forecasting models, the parameter estimation, and the estimation window

[20] In fact, these are "proxies" of the total tail probabilities, since they are calculated once based on the *average* parameter values over the entire out-of-sample period.

Table 5.2: USA vs. Mexico: Pair-wise tests of equal predictive accuracy of copulas in tails.

r = 0.01 (m=11)		Gaussian (0.0018)	Student-t (0.0025)	Cauchy (0.0044)	Clayton (0.0050)	Gumbel (0.0047)	Frank (0.0004)
Gaussian	csl		-1,97**	-1,25	-1,37	-1,40	2,57**
	cl		-1,21	-0,56	-1,66*	-1,54	-0,48
	wl		-3,26***	-2,88***	-3,25***	-3,22***	3,22***
Student-t	csl			-0,75	-1,02	-1,04	2,45**
	cl			-0,38	-1,55	-1,37	-0,22
	wl			-2,40**	-3,21***	-3,14***	3,23***
Cauchy	csl				-0,41	-0,40	2,12**
	cl				-0,27	-0,27	0,16
	wl				-1,15	-0,88	3,23***
Clayton	csl					0,36	2,09**
	cl					0,21	0,38
	wl					2,20**	3,27***
Gumbel	csl						2,11**
	cl						0,36
	wl						3,26***

r = 0.05 (m=63)		Gaussian (0.0151)	Student-t (0.0171)	Cauchy (0.0222)	Clayton (0.0253)	Gumbel (0.0245)	Frank (0.0089)
Gaussian	csl		-3,89***	-1,06	-2,65***	-2,62***	4,17***
	cl		-2,65***	0,19	-1,74*	-1,58	1,44
	wl		-6,63***	-3,61***	-6,80***	-6,59***	6,54***
Student-t	csl			0,00	-2,05**	-1,95*	4,15***
	cl			0,77	-1,19	-0,96	1,76*
	wl			-2,19**	-6,78***	-6,43***	6,62***
Cauchy	csl				-2,07**	-2,10**	2,99***
	cl				-1,86*	-1,89*	0,56
	wl				-3,72***	-3,44***	5,82***
Clayton	csl					1,43	3,57***
	cl					1,07	1,73*
	wl					4,88***	6,87***
Gumbel	csl						3,59***
	cl						1,67*
	wl						6,81***

r = 0.10 (m=115)		Gaussian (0.0381)	Student-t (0.0406)	Cauchy (0.0449)	Clayton (0.0520)	Gumbel (0.0512)	Frank (0.0304)
Gaussian	csl		-4,20***	-0,75	-2,49**	-2,59***	4,14***
	cl		-3,89***	-0,31	-2,37**	-2,45	3,32***
	wl		-7,28***	-2,60***	-7,47***	-7,40***	7,15***
Student-t	csl			0,26	-1,72*	-1,80*	4,25***
	cl			0,49	-1,58	-1,65*	3,57***
	wl			-1,10	-7,44***	-7,30***	7,34***
Cauchy	csl				-1,83*	-1,98**	2,66
	cl				-1,87*	-2,00**	1,80*
	wl				-4,81***	-4,73***	5,48***
Clayton	csl					0,00	3,51***
	cl					-0,03	3,14***
	wl					3,13***	7,71***
Gumbel	csl						3,59***
	cl						3,22***
	wl						7,72***

r = 1.00 (m=2241)		Gaussian (1.00)	Student-t (1.00)	Cauchy (1.00)	Clayton (1.00)	Gumbel (1.00)	Frank (1.00)
Gaussian	all		-5,84***	8,62***	8,65***	2,42**	7,11***
Student-t	all			10,84***	10,07***	5,64***	8,17***
Cauchy	all				-3,40***	-8,76***	-5,17***
Clayton	all					-13,04***	-3,70***
Gumbel	all						3,45***

Note: The table represents the test statistic in (3.28) based on the *csl, cl* and *wl* scoring rules using $w(u,v) = I(u,v \leq r)$, for r= 0.01, r=0.05, r=0.10 and r=1. Positive (Negative) values indicate better performance of the corresponding copula in the column (row). The asterisks *, ** and *** indicate significance at (two-sided) 10%, 5% and 1% significance respectively. We consider P =2241 one-step-ahead forecasts based on a rolling in-sample window of length R=2000. The number of tail observations *m* and total tail probability C(r,r) based on average parameter values of the copulas, are given in parentheses.

Table 5.3: USA vs. Argentina: Pair-wise tests of equal predictive accuracy of copulas in tails.

r = 0.01 (m=8)		Gaussian (0.0008)	Student-t (0.0018)	Cauchy (0.0039)	Clayton (0.0031)	Gumbel (0.0032)	Frank (0.0003)
Gaussian	csl		-1.81*	-1.37	-1.48	-1.48	2.29**
	cl		-0.32	-0.92	-0.25	-0.59	-0.21
	wl		-2.63***	-2.68***	-2.72***	-2.71***	2.67***
Student-t	csl			-0.85	-0.67	-0.85	2.07**
	cl			-1.03	0.06	-1.05	0.03
	wl			-2.66***	-2.76***	-2.78***	2.67***
Cauchy	csl				0.85	0.68	1.73*
	cl				1.15	0.91	0.49
	wl				2.12**	1.99**	2.71***
Clayton	csl					-1.35	1.84*
	cl					-1.90*	0.02
	wl					-2.39**	2.71***
Gumbel	csl						1.82*
	cl						0.20
	wl						2.70***
r = 0.05 (m=38)		Gaussian (0.0090)	Student-t (0.0123)	Cauchy (0.0195)	Clayton (0.0170)	Gumbel (0.0173)	Frank (0.0062)
Gaussian	csl		-2.58**	-0.62	-1.90*	-1.82*	3.14***
	cl		-1.40	0.36	-1.25	-1.12	1.60
	wl		-4.97***	-3.98***	-5.44***	-5.25***	4.84***
Student-t	csl			0.97	-0.58	-0.57	2.92***
	cl			1.15	-0.62	-0.48	1.54
	wl			-2.53**	-5.84***	-5.43***	4.97***
Cauchy	csl				-1.40	-1.70*	1.60
	cl				-1.29	-1.56	0.29
	wl				0.05	-0.14	4.60***
Clayton	csl					-0.21	2.45**
	cl					-0.21	1.44
	wl					-0.98	5.24***
Gumbel	csl						2.40**
	cl						1.39
	wl						5.16***
r = 0.10 (m=78)		Gaussian (0.0257)	Student-t (0.0303)	Cauchy (0.0396)	Clayton (0.0374)	Gumbel (0.0375)	Frank (0.0222)
Gaussian	csl		-2.76***	-0.20	-2.01**	-1.95*	3.07***
	cl		-2.63***	0.19	-2.21**	-2.05**	2.46**
	wl		-6.03***	-4.03***	-6.94***	-6.63***	5.13***
Student-t	csl			1.37	-0.60	-0.67	2.99***
	cl			1.40	-0.90	-0.87	2.60***
	wl			-2.05**	-7.79***	-7.09***	5.69***
Cauchy	csl				-1.69*	-2.05**	1.24
	cl				-1.62	-1.96**	0.74
	wl				-0.79	-1.03	4.88***
Clayton	csl					-0.38	2.50**
	cl					-0.37	2.37**
	wl					-0.60	6.31***
Gumbel	csl						2.47**
	cl						2.33**
	wl						6.20***
r = 1.00 (m=2241)		Gaussian (1.00)	Student-t (1.00)	Cauchy (1.00)	Clayton (1.00)	Gumbel (1.00)	Frank (1.00)
Gaussian	all		-5.66***	7.82***	0.37	-3.12***	-0.28
Student-t	all			11.25***	4.45***	2.90***	4.76***
Cauchy	all				-7.98***	-9.61***	-7.85***
Clayton	all					-6.02***	-0.52
Gumbel	all						2.35**

Note: The table represents the test statistic in (3.28) based on the *csl, cl* and *wl* scoring rules using $w(u,v) = \mathbf{I}(u, v \leq r)$, for r= 0.01, r=0.05, r=0.10 and r=1. Positive (Negative) values indicate better performance of the corresponding copula in the column (row). The asterisks *, ** and *** indicate significance at (two-sided) 10%, 5% and 1% significance respectively. We consider P =2241 one-step-ahead forecasts based on a rolling in-sample window of length R=2000. The number of tail observations *m* and total tail probability C(r,r) based on average parameter values of the copulas, are given in parentheses.

Table 5.4: USA vs. Brazil: Pair-wise tests of equal predictive accuracy of copulas in tails.

r = 0.01 (m=9)		Gaussian (0.0013)	Student-t (0.0017)	Cauchy (0.0041)	Clayton (0.0040)	Gumbel (0.0040)	Frank (0.0003)
Gaussian	csl		-1,95*	-0,40	-1,10	-1,05	2,39**
	cl		-0,07	0,80	0,21	0,24	0,42
	wl		-2,91***	-2,14**	-2,91***	-2,85***	2,88***
Student-t	csl			0,41	-0,48	-0,38	2,30**
	cl			0,83	0,31	0,32	0,37
	wl			-1,30	-2,76***	-2,58***	2,90***
Cauchy	csl				-1,09	-1,18	1,69*
	cl				-0,96	-1,04	-0,61
	wl				-0,90	-0,95	2,86***
Clayton	csl					0,52	1,89*
	cl					0,28	0,23
	wl					0,58	2,94***
Gumbel	csl						1,88*
	cl						0,20
	wl						2,94***
r = 0.05 (m=50)		**Gaussian (0.0121)**	**Student-t (0.0135)**	**Cauchy (0.0206)**	**Clayton (0.0212)**	**Gumbel (0.210)**	**Frank (0.0075)**
Gaussian	csl		-3,57***	-0,28	-2,12**	-2,02**	3,89***
	cl		-2,14**	0,57	-1,31	-1,16	2,13**
	wl		-5,62***	-3,06***	-5,95***	-5,77***	5,80***
Student-t	csl			0,86	-1,22	-1,05	3,87***
	cl			1,10	-0,69	-0,50	2,22**
	wl			-1,77*	-5,93***	-5,53***	5,85***
Cauchy	csl				-2,12**	-2,23**	2,26**
	cl				-1,79*	-1,92*	0,52
	wl				-2,26**	-2,26**	5,09***
Clayton	csl					0,97	3,15***
	cl					0,50	1,96*
	wl					1,91*	6,08***
Gumbel	csl						3,13***
	cl						1,91*
	wl						6,04***
r = 0.10 (m=100)		**Gaussian (0.0322)**	**Student-t (0.0416)**	**Cauchy (0.0416)**	**Clayton (0.0448)**	**Gumbel (0.0445)**	**Frank (0.0263)**
Gaussian	csl		-3,80***	0,86	-2,27**	-2,03**	4,12***
	cl		-3,03***	1,32	-1,98**	-1,56	3,48***
	wl		-6,31***	-1,63	-7,21***	-6,79***	6,59***
Student-t	csl			2,02**	-1,36	-0,97	4,13***
	cl			2,13**	-1,21	-0,61	3,50***
	wl			-0,17	-7,44***	-6,68***	6,67***
Cauchy	csl				-3,27***	-3,40***	1,44
	cl				-3,29***	-3,41***	0,50
	wl				-4,21***	-4,32***	4,48***
Clayton	csl					1,80*	3,37***
	cl					1,88*	3,03***
	wl					2,87***	7,21***
Gumbel	csl						3,28***
	cl						2,86***
	wl						7,09***
r = 1.00 (m=2241)		**Gaussian (1.00)**	**Student-t (1.00)**	**Cauchy (1.00)**	**Clayton (1.00)**	**Gumbel (1.00)**	**Frank (1.00)**
Gaussian	all		-6,13***	9,56***	6,76***	1,56	5,14***
Student-t	all			11,38***	8,35***	4,88***	6,43***
Cauchy	all				-6,28***	-9,95***	-7,27***
Clayton	all					-10,89***	-3,17***
Gumbel	all						2,66***

Note: The table represents the test statistic in (3.28) based on the *csl, cl* and *wl* scoring rules using $w(u,v) = \mathbf{I}(u, v \leq r)$, for r= 0.01, r=0.05, r=0.10 and r=1. Positive (Negative) values indicate better performance of the corresponding copula in the column (row). The asterisks *, ** and *** indicate significance at (two-sided) 10%, 5% and 1% significance respectively. We consider P =2241 one-step-ahead forecasts based on a rolling in-sample window of length R=2000. The number of tail observations *m* and total tail probability C(r,r) based on average parameter values of the copulas, are given in parentheses.

Table 5.5: Mexico vs. Argentina: Pair-wise tests of equal predictive accuracy of copulas in tails.

r = 0.01 (m=6)		Gaussian (0.0009)	Student-t (0.0019)	Cauchy (0.0040)	Clayton (0.0034)	Gumbel (0.0035)	Frank (0.0003)
Gaussian	csl		-1,10	-0,82	-0,87	-0,90	1,91*
	cl		-0,68	-0,83	-0,92	-1,05	0,70
	wl		-2,33**	-2,30**	-2,40**	-2,39**	2,39**
Student-t	csl			-0,64	-0,65	-0,74	1,62
	cl			-0,78	-1,20	-1,27	0,69
	wl			-2,21**	-2,43**	-2,40**	2,37**
Cauchy	csl				0,48	0,30	1,27
	cl				0,60	0,44	0,87
	wl				1,42	1,32	2,36**
Clayton	csl					-1,15	1,35
	cl					-1,22	0,82
	wl					-1,77*	2,40**
Gumbel	csl						1,36
	cl						0,92
	wl						2,39**

r = 0.05 (m=38)		Gaussian (0.0099)	Student-t (0.0129)	Cauchy (0.0199)	Clayton (0.0185)	Gumbel (0.0184)	Frank (0.0065)
Gaussian	csl		-1,84*	-1,73*	-1,35	-1,59	2,38**
	cl		-0,78	-1,67*	-0,70	-1,25	0,11
	wl		-5,10***	-4,92***	-5,41***	-5,36***	4,53***
Student-t	csl			-1,59	-0,94	-1,38	2,21**
	cl			-1,76*	-0,59	-1,61	0,35
	wl			-4,57***	-5,56***	-5,44***	4,81***
Cauchy	csl				1,78*	1,48	2,07**
	cl				1,87*	1,56	1,30
	wl				2,78***	2,67***	5,10***
Clayton	csl					-3,41	1,83*
	cl					-3,37	0,41
	wl					-3,28***	5,12***
Gumbel	csl						1,96*
	cl						0,73
	wl						5,13***

r = 0.10 (m=83)		Gaussian (0.0276)	Student-t (0.0317)	Cauchy (0.0403)	Clayton (0.0399)	Gumbel (0.0396)	Frank (0.0231)
Gaussian	csl		-2,37**	-1,01	-1,60	-1,79*	2,27**
	cl		-1,93*	-0,59	-1,43	-1,69*	1,25
	wl		-6,52***	-4,29***	-7,08***	-6,78***	5,05***
Student-t	csl			-0,50	-0,97	-1,36	2,37**
	cl			-0,21	-0,89	-1,45	1,53
	wl			-3,24***	-7,38***	-6,84***	5,76***
Cauchy	csl				0,12	-0,20	1,54
	cl				-0,03	-0,37	0,90
	wl				0,30	0,11	5,03***
Clayton	csl					-2,21**	1,94*
	cl					-2,22**	1,39
	wl					-1,54	6,42***
Gumbel	csl						2,06**
	cl						1,58
	wl						6,33***

r = 1.00 (m=2241)		Gaussian (1.00)	Student-t (1.00)	Cauchy (1.00)	Clayton (1.00)	Gumbel (1.00)	Frank (1.00)
Gaussian	all		-3,15***	9,77***	0,90	-1,98**	1,60
Student-t	all			11,97***	2,51**	0,38	3,51***
Cauchy	all				-9,74***	-11,35***	-8,92***
Clayton	all					-5,05***	0,08
Gumbel	all						2,52**

Note: The table represents the test statistic in (3.28) based on the *csl, cl* and *wl* scoring rules using $w(u,v) = I(u, v \leq r)$, for r= 0.01, r=0.05, r=0.10 and r=1. Positive (Negative) values indicate better performance of the corresponding copula in the column (row). The asterisks *, ** and *** indicate significance at (two-sided) 10%, 5% and 1% significance respectively. We consider P =2241 one-step-ahead forecasts based on a rolling in-sample window of length R=2000. The number of tail observations *m* and total tail probability C(r,r) based on average parameter values of the copulas, are given in parentheses.

Table 5.6: Mexico vs. Brasil: Pair-wise tests of equal predictive accuracy of copulas in tails.

r = 0.01 (m=10)		Gaussian (0.0015)	Student-t (0.0025)	Cauchy (0.0042)	Clayton (0.0044)	Gumbel (0.0044)	Frank (0.0004)
Gaussian	csl		-1,56	-0,53	-1,03	-1,07	2,55**
	cl		0,00	0,50	0,10	0,06	0,11
	wl		-2,87***	-2,02**	-2,82***	-2,80***	3,09***
Student-t	csl			0,25	-0,48	-0,52	2,31**
	cl			0,61	0,16	0,09	0,08
	wl			-1,08	-2,65***	-2,61***	3,08***
Cauchy	csl				-0,95	-1,00	1,77*
	cl				-0,87	-0,91	-0,43
	wl				-1,26	-1,10	2,86***
Clayton	csl					-0,19	1,92*
	cl					-0,54	-0,02
	wl					2,00**	3,05***
Gumbel	csl						1,95*
	cl						0,01
	wl						3,05***

r = 0.05 (m=53)		Gaussian (0.0132)	Student-t (0.0162)	Cauchy (0.0258)	Clayton (0.0235)	Gumbel (0.227)	Frank (0.0079)
Gaussian	csl		-2,91***	-0,56	-1,81*	-1,89*	3,76***
	cl		-1,81*	0,47	-0,93	-0,95	1,62
	wl		-5,83***	-3,31***	-6,05***	-5,92***	5,91***
Student-t	csl			0,67	-0,87	-0,96	3,60***
	cl			1,10	-0,10	-0,16	1,78*
	wl			-1,53	-5,99***	-5,73***	6,05***
Cauchy	csl				-1,61	-1,77*	2,42**
	cl				-1,59	-1,71*	0,41
	wl				-2,80***	-2,56**	5,35***
Clayton	csl					-0,32	2,94***
	cl					-0,34	1,49
	wl					4,07***	6,25***
Gumbel	csl						3,01***
	cl						1,51
	wl						6,22***

r = 0.10 (m=110)		Gaussian (0.0344)	Student-t (0.0381)	Cauchy (0.0430)	Clayton (0.0488)	Gumbel (0.0478)	Frank (0.0276)
Gaussian	csl		-3,90***	-0,73	-2,51**	-2,64***	3,85***
	cl		-3,60***	-0,22	-2,30**	-2,38**	2,68***
	wl		-7,26***	-3,10***	-7,81***	-7,66***	6,64***
Student-t	csl			0,66	-1,40	-1,56	4,02***
	cl			0,84	-1,05	-1,15	3,13***
	wl			-0,99	-7,92***	-7,65***	7,10***
Cauchy	csl				-1,83*	-2,02**	2,48**
	cl				-1,70*	-1,84*	1,45
	wl				-4,25***	-4,07***	5,58***
Clayton	csl					-0,57	3,37***
	cl					-0,44	2,79***
	wl					3,99***	7,71***
Gumbel	csl						3,47***
	cl						2,86***
	wl						7,67***

r = 1.00 (m=2241)		Gaussian (1.00)	Student-t (1.00)	Cauchy (1.00)	Clayton (1.00)	Gumbel (1.00)	Frank (1.00)
Gaussian	all		-6,61***	5,94***	6,07***	-0,63	6,86***
Student-t	all			9,20***	8,14***	4,22***	8,56***
Cauchy	all				-2,44**	-6,73***	-2,98***
Clayton	all					-11,51***	-1,26
Gumbel	all						5,60***

Note: The table represents the test statistic in (3.28) based on the *csl, cl* and *wl* scoring rules using $w(u,v) = \mathbf{I}(u, v \leq r)$, for r= 0.01, r=0.05, r=0.10 and r=1. Positive (Negative) values indicate better performance of the corresponding copula in the column (row). The asterisks *, ** and *** indicate significance at (two-sided) 10%, 5% and 1% significance respectively. We consider P =2241 one-step-ahead forecasts based on a rolling in-sample window of length R=2000. The number of tail observations *m* and total tail probability C(r,r) based on average parameter values of the copulas, are given in parentheses.

Table 5.7: Argentina vs. Brasil: Pair-wise tests of equal predictive accuracy of copulas in tails.

r = 0.01 (m=7)		Gaussian (0.0013)	Student-t (0.0025)	Cauchy (0.0042)	Clayton (0.0043)	Gumbel (0.0041)	Frank (0.0003)
Gaussian	csl		-1,30	-0,80	-0,86	-0,90	2,02**
	cl		-1,33	-0,70	-1,40	-1,32	1,09
	wl		-2,55**	-2,37**	-2,58***	-2,56**	2,54**
Student-t	csl			-0,26	-0,32	-0,40	1,78*
	cl			-0,30	-1,37	-1,12	1,19
	wl			-1,94*	-2,59***	-2,53**	2,55**
Cauchy	csl				0,03	-0,09	1,44
	cl				-0,09	-0,15	1,01
	wl				-0,05	0,06	2,52**
Clayton	csl					-0,80	1,46
	cl					-0,28	1,28
	wl					0,76	2,57**
Gumbel	csl						1,49
	cl						1,27
	wl						2,56**

r = 0.05 (m=40)		Gaussian (0.0122)	Student-t (0.0158)	Cauchy (0.0211)	Clayton (0.0222)	Gumbel (0.216)	Frank (0.0076)
Gaussian	csl		-1,99**	-0,62	-1,07	-1,19	2,61***
	cl		-1,69*	-0,26	-1,33	-1,38	1,33
	wl		-5,15***	-3,73***	-5,43***	-5,31***	4,75***
Student-t	csl			0,27	-0,13	-0,34	2,43**
	cl			0,33	-0,78	-0,90	1,50
	wl			-2,34**	-5,60***	-5,34***	4,98***
Cauchy	csl				-0,48	-0,71	1,64
	cl				-0,73	-0,89	0,83
	wl				-1,23	-1,08	4,67***
Clayton	csl					-1,51	1,86*
	cl					-0,88	1,43
	wl					2,20**	5,26***
Gumbel	csl						1,94*
	cl						1,47
	wl						5,22***

r = 0.10 (m=99)		Gaussian (0.0323)	Student-t (0.0371)	Cauchy (0.0426)	Clayton (0.0465)	Gumbel (0.0456)	Frank (0.0266)
Gaussian	csl		-2,68***	0,04	-1,54	-1,64	2,68***
	cl		-2,03**	0,70	-1,04	-1,10	1,69*
	wl		-6,49***	-2,52**	-7,02***	-6,78***	5,33***
Student-t	csl			1,14	-0,44	-0,63	2,79***
	cl			1,48	0,07	-0,12	1,90*
	wl			-0,59	-7,35***	-6,87***	5,99***
Cauchy	csl				-1,63	-1,86*	1,09
	cl				-1,87*	-2,08**	0,10
	wl				-3,12***	-3,10***	4,18***
Clayton	csl					-0,98	2,18**
	cl					-0,72	1,48
	wl					2,44**	6,62***
Gumbel	csl						2,25**
	cl						1,53
	wl						6,54***

r = 1.00 (m=2241)		Gaussian (1.00)	Student-t (1.00)	Cauchy (1.00)	Clayton (1.00)	Gumbel (1.00)	Frank (1.00)
Gaussian	all		-3,89***	8,33***	1,25	-2,76***	0,72
Student-t	all			12,38***	4,00***	1,54	4,56***
Cauchy	all				-7,94***	-10,46***	-7,90***
Clayton	all					-6,67***	-0,63
Gumbel	all						2,89***

Note: The table represents the test statistic in (3.28) based on the *csl, cl* and *wl* scoring rules using $w(u,v) = \mathbf{I}(u, v \leq r)$, for r= 0.01, r=0.05, r=0.10 and r=1. Positive (Negative) values indicate better performance of the corresponding copula in the column (row). The asterisks *, ** and *** indicate significance at (two-sided) 10%, 5% and 1% significance respectively. We consider P =2241 one-step-ahead forecasts based on a rolling in-sample window of length R=2000. The number of tail observations *m* and total tail probability C(r,r) based on average parameter values of the copulas, are given in parentheses.

5.5 Discussion of results

In this section, we discuss the results of section 5.4. First we examine whether our results for threshold value $r = 1$ (i.e. considering the full copula distribution) correspond to our expectations. Hereafter, we elaborate on the results when focusing on the left tail of the distribution ($r = 0.1$, $r = 0.05$ and $r = 0.01$) and explore the differences between the test statistics based on the weighted likelihood (*wl*), conditional likelihood (*cl*) and censored likelihood (*csl*) scoring rules.

When considering the full copula distribution (i.e. setting $r = 1$), most test statistics in section 5.4 are significant based on the 1% significance level, indicating that the null hypothesis of equal predictive ability is generally convincingly rejected. The results suggest superior forecast ability of the Student-t copula. The Gumbel copula generally ranks second, closely followed by the Gaussian copula. In fact, the Gaussian copula outperforms the Gumbel copula for two country pairs. The Frank copula typically ranks fourth, followed by the Clayton copula. The Cauchy copula clearly performs the worst, for all country pairs considered. The results are in line with our expectations based on the extensive data analysis in section 5.2 and 5.3. In particular the Chi-plot seems to be a valuable graphical tool to detect the entire dependence structure, as our expectations regarding to the Chi-plots in section 5.3 correspond closely to the best performing models in section 5.4. Moreover, we note that our results are in line with the results of Diks et al (2009), who also find superior predictive accuracy of the Student-t copula and relatively bad performance of the Clayton copula. They apply the same test on daily exchange rate returns of several major currencies against the US dollar.[21]

In the remainder of this section, we focus on the left tail of the distribution, and consider the results for threshold values $r = 0.1$, $r = 0.05$ and $r = 0.01$. Most of the test statistics based on the *wl* scoring rule are significant, such that a clear performance ranking can be made for the competing models. For the test statistics based on the (*proper*) *cl* and *csl* scoring rules however, such a ranking is harder to make. This is due to the fact that, in particular for $r = 0.01$, there often is not enough evidence to reject the null hypothesis of equal predictive accuracy of two competing models. In the following, we therefore separately analyze the results of the *wl* scoring rule and the *proper* scoring rules, and compare the results hereafter.

The observation that most test statistics based on the *weighted likelihood (wl)* scoring rule are significant is in line with the high power of the *wl* scoring rule in our Monte Carlo experiments. Recall from section 4.2 however, that we found excess *spurious power* for the *wl* scoring rule due to its bias towards fat-tailed models. Since the exact distribution of our data is unknown, it is unclear whether these significant rejections of the null-hypothesis of equal performance are reliable (i.e. in favor to the (most) correct distribution). To investigate whether the test statistics favor the model with the highest probability mass in the left tail, it is convenient to order the six competing models based on their total tail probabilities (displayed in parentheses in table 5.2 to 5.7). It is striking to see that these rankings correspond closely to the rankings based on the (significant) test statistics, which indeed suggests a bias towards fat-tailed models. In both rankings, the Clayton copula generally performs best, closely followed by the Gumbel and Cauchy copula. The Student-t copula typically ranks fourth, the Gaussian copula fifth, and the

---

[21] Furthermore, we note that Diks et al (2009) use copula specifications of three rather than two dimensions. Still, the results are roughly similar; the main difference is that in our results the Gumbel copula performs slightly better than the Gaussian copula, while in Diks et al (2009) these roles are turned.

Frank copula clearly performs the worst. To gain more insight in these specific rankings, we analyze the performance of the three elliptical copulas in more detail. *A priori*, we might expect that the Student-*t* copula shows the best forecasting performance among the elliptical copulas considered, due to its flexibility in the tails of the distribution[22]. As expected, the Student-*t* copula significantly outperforms the Gaussian copula in *all* cases (based on a 5% significance interval). Notably, the Cauchy copula is in *all* cases favored to the Student-*t* copula, although not always to a significant extent. Despite the relative disadvantage of the Student-t copula for estimating an additional parameter, it is very unlikely that in *all* cases considered the Cauchy copula is indeed more appropriate than the Student-*t* copula for the purpose of forecasting left tail events. Therefore, we conclude that the results of the *wl* scoring rule contain clear evidence for a bias towards fat-tailed models. Note that in risk management applications, this bias towards fat-tailed forecasting models generally results in an *overestimation* of risk, i.e. the selected models are too conservative. This could have adverse economic consequences for risk managers, such as sub-optimal asset allocation and "over-hedging" (Diks et al, 2010).

The conditional likelihood (*cl*) and censored likelihood (*csl*) scoring rules show less significant results as compared to those based on the *wl* scoring rule.[23] Both scoring rules generally point towards the same conclusion, although the test statistics based on the *csl* scoring rule generally show more powerful results, in particular for $r = 0.01$. These observations correspond to our Monte Carlo power experiments in section 4.2. As noted earlier, clear copula rankings based on significant results are often hard to make. Still, we mention two major differences in performance between the proper scoring rules and the *wl* scoring rule. First, the Gumbel copula generally perceives a better ranking for the proper scoring rules. In fact, for two country pairs (Mexico-Brazil and Argentina-Brazil) the Gumbel copula shows the best performance for all considered threshold values (although not to a significant extent), whereas the *wl* scoring rule significantly favors the ("fatter" tailed) Clayton copula. Second, we observe that and Student-t copula mostly performs better than the Cauchy copula, except for the case of $r = 0.01$.[24] These results indeed suggest that the *cl* and *csl* scoring rules are not biased towards fat-tailed models.

In the remainder we discuss two important factors that influence the power of our tests. The first factor is the sample size, or the closely related expected number of left tail observations (m). In our Monte Carlo simulations we demonstrated that the power of our test statistics increases with sample size. We did not test our dataset for different sample sizes, but the relatively weak power of our *proper* scoring rules could be simply due to the fact that m is too small, in particular for $r = 0.01$. For larger r (and thus more left tail observations) we observe more significant results for the proper scoring rules, which is a promising result. To what extent this power increase is due to the larger region of interest (larger r) or due to the higher number of left tail observations (larger m) could be identified by further research.

Secondly, we consider the estimated parameter values, because we experienced in our Monte Carlo simulations that a different measure of association can influence the power of our test considerably. We included figures of the estimated parameters over the entire out-of-sample

---

[22] The Gaussian copula ($v = 0$) and Cauchy copula ($v = \infty$) can be interpreted as special cases of the Student-*t* copula with resp. minimum and maximum tail dependence (for a discussion see section 2.3.1).
[23] Note that this is also apparent in the power plots of section 4.2: "aggregating" the *spurious* and *true* power plots results in higher (or equal) power for the wl scoring rule than for the cl/csl scoring rules.
[24] Note that the country pair Mexico-Argentina (showing high tail dependence in section 5.3) is an exception as the Cauchy copula outperforms the Student-t copula for all threshold values in the left tail.

period in Appendix 8. The country pairs with the most extreme parameter values are depicted in figure 5.11 and 5.12: the pair USA-Mexico has the highest average ρ_{12} of 0.59 (figure 5.11), while the pair USA-Argentina shows the lowest average ρ_{12} of 0.38 (figure 5.12). From appendix 8, several observations can be made. First, we observe similar upward sloping parameter patterns for the pairs USA-Mexico, USA-Brazil and Mexico-Brazil, suggesting that since 2000 both Mexico and Brazil are getting more associated to USA, the only developed country in the dataset. All country pairs containing Argentina show parameter patterns similar to those in figure 5.12, where the 2001-2002 currency crisis in Argentina seems to be of considerable. The parameter values typically show a peak just before the burst of the crisis and experience a slight decrease in the after-crisis period for about five to six years. Hereafter, the parameter values generally move up again, similar to the other country pairs. Second, we clearly see that the association between the country pairs substantially varies over time. Our rolling estimation window appears to capture the stylized fact of time-varying dependence as discussed in section 5.2. The relatively high correlation parameters of both the Student-t and Gaussian copula might indeed cause less powerful test statistics in our empirical application, in line with our results in section 4.2. However, the varying parameter values generally complicate the analysis of comparing the power results of section 5.4 to our Monte Carlo experiments.
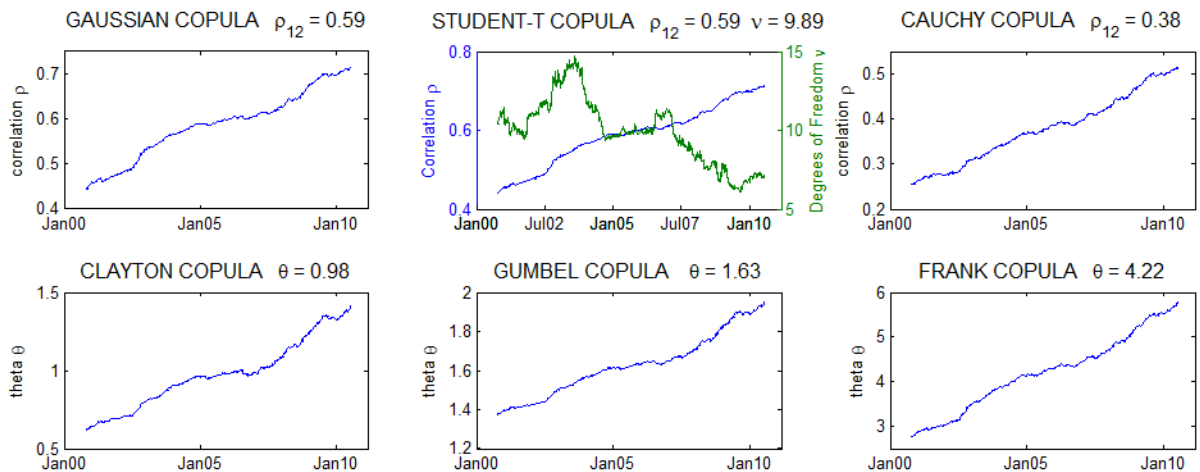
Figure 5.11: Estimated parameter values of the country-pair USA-Mexico, for the six different copulas for R=2000 and the out-of-sample period from 6-10-2000 to 26-7-2010, including the average parameter values.
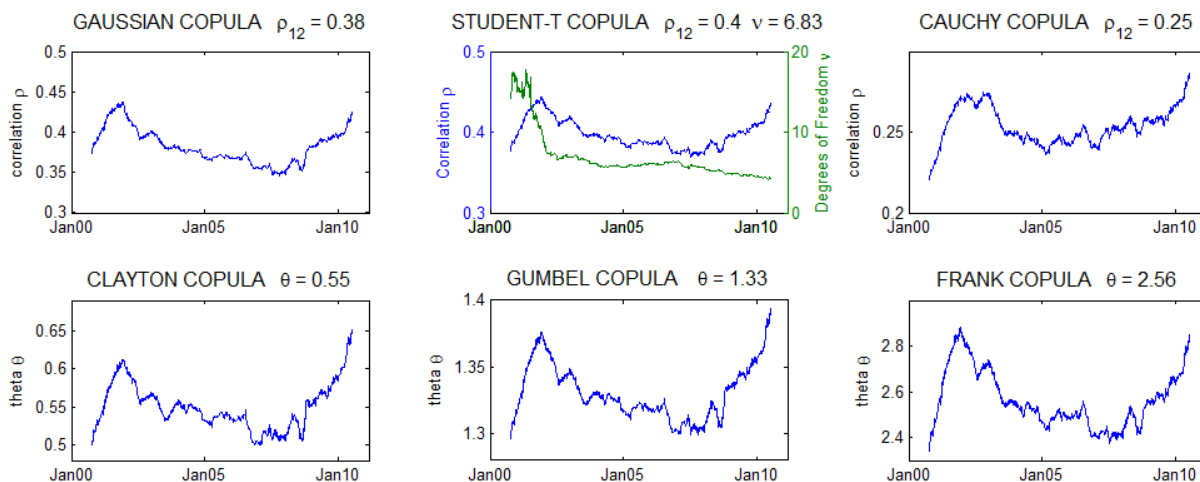
Figure 5.12: Estimated parameter values of the country-pair USA-Argentina, for the six different copulas for R=2000 and the out-of-sample period from 6-10-2000 to 26-7-2010, including the average parameter values.

6 Conclusion

In this thesis we developed a testing framework for evaluating and comparing the accuracy of copula-based multivariate density models to forecast joint extreme events. We have shown that we can use likelihood based scoring rules to evaluate multivariate density forecasts in the left tail. Specifically, we have introduced three KLIC-based scoring rules based on *weighted* (full*)* likelihood, *conditional* likelihood and *censored* likelihood respectively. Based on these scoring rules, we have extensively tested the finite sample properties of our testing framework using Monte Carlo simulation experiments. Moreover, the practical usefulness of the testing framework is examined in an empirical application using daily MSCI Total Return Indices of USA, Mexico, Argentina and Brazil. Our main findings are summarized below.

The *weighted likelihood* (*wl*) scoring rule appears to be biased towards models with the highest probability mass in the left tail. In our Monte Carlo simulations we found strong evidence of excess *spurious power*, i.e. cases in which the *wl* scoring rule significantly favors the *incorrect* (fat-tailed) density. Notably, this spurious power is *increasing* for larger sample sizes. Also in our empirical application the *wl* scoring rule is biased, since the copula models with the highest total tail probability significantly outperform the competing models in practically all cases. In fact, the entire rankings based on the performance of our test statistic and the rankings based on the total tail probability are (almost) identical. In financial risk management applications, favoring fat-tailed density models could have adverse economic consequences. It leads to a bias towards the most conservative models, resulting in suboptimal asset allocation. For the purpose of forecasting joint extreme events, we therefore recommend not to rely on test statistics based on *wl* scoring rule.

The scoring rules based on conditional likelihood (*cl*) and censored likelihood (*csl*) are *proper*[25] in the sense that they do not suffer from this bias towards fat-tailed densities. Both scoring rules use the total tail probability to offset the bias of the *wl* scoring rule. The *csl* scoring rule corrects the *wl* scoring rule by censoring the observations outside the region of interest, while the *cl* scoring rule normalizes the *wl* scoring rule with the total tail probability within the region of interest. Our Monte Carlo simulations demonstrate that the *spurious power* of both the *cl* and *csl* scoring rules is limited. This (limited) *spurious power* virtually disappears for increasing sample size, i.e. for spurious power is a small sample problem (in stark contrast to the *wl* scoring rule). Moreover, we found that the *true power* increases with sample size, i.e. the scoring rules significantly favor the *correct* density more often. Larger sample sizes thus considerably strengthen the reliability of our test based on the *cl* and *csl* scoring rule.

Among the two proper scoring rules the *csl* scoring rule generally performs better than the *cl* scoring rule, in particular in the left tail. This is due to the fact that the *csl* scoring rule uses more information as it also uses the observations outside the region of interest. In our Monte Carlo simulations we often found relatively higher true power for the *csl* scoring rule in the left tail. In our practical applications we observe that the null-hypothesis of equal predictive performance is rejected relatively more often based on the *csl* scoring rule, which is a sign of relatively more powerful tests. In financial risk management applications where the left tail is of interest, the *csl* scoring rule should therefore be preferred over the *cl* scoring rule.

---

[25] This term is adopted from Winkler and Murphy (1968), who define a scoring rule to be proper if a correctly specified model always receives a higher average score than an incorrectly specified model.

At last we provide several critical notes regarding to the *proper cl* and *csl* scoring rules. Although the proper scoring rules do not suffer from excess *spurious power* in our Monte Carlo simulations, their *true* power turned out to be somewhat disappointing for small sample sizes. Moreover, we found in our Monte Carlo simulations that the power for differentiating between the Gaussian and Student-t copula is *decreasing* with correlation. Further research should investigate whether similar decreasing power patterns are also present for increasing parameters of different competing copula models. Most likely, these relations had impact on the results of our empirical application, as the proper scoring rules often could not make a clear judgment between the competing copulas, in particular for the smallest regions of interest. At this point, it should be emphasized however that in *all* cases considered the power increases with sample size. Although our *proper* scoring rules need some further research on changing parameter values and increasing dimensions, we can safely state that our flexible testing framework provides us with a valuable tool to evaluate and compare forecasts of joint extreme events.

7 Further Research

In the conclusion we pointed in the direction of two interesting topics for further research regarding to our *proper* scoring rules. First, it should be examined how the power changes for increasing parameter values in case of Archimedean copulas, by adopting similar Monte Carlo experiments as in Appendix 5. Another possibility is to include parameter uncertainty in our Monte Carlo experiments, such as in Diks et al (2009). They simulate from a GARCH model to obtain the standardized residuals that can be used as inputs for Step 2 of the SCOMDY model (see section 2.4.2). Such a setup has the advantage that a comparison with the empirical application is easier to make. However, this procudere turns out to be very time intensive. Second, a straightforward contribution to this thesis would be to increase the dimension of the models, because we only considered bivariate density forecasts. The question rises whether the power generally increases with dimension, such as in the results of Berg (2009).

Regarding to the relatively weak power results of the *proper* scoring rules in our empirical applications, it would be interesting to investigate whether more significant results are obtained when using a larger sample size, or choosing a less correlated dataset. More powerful results may also be obtained by considering more advanced copula specifications, such as skewed copula specifications, mixtures of elliptical copulas, or extreme value copulas.

Furthermore, we only consider statistical evaluation criteria in our empirical example. Additional insights could be obtained by considering economic criteria, by for instance calculating Value-at-Risk (VaR) and Expected Shortfall (ES) measures based on the results of the test statistics (see Diks et al, 2010). Alternatively, one could set up a dynamic hedging strategy based on the selected models, and measure the degree of "over-hedging".

In the Monte Carlo experiments assessing the size properties of our test, we found that for sample sizes with less than 20 observations in the region of interest, the asymptotic distribution of our Diebold Mariano type test statistic is not a suitable approximation of the true distribution. To solve this problem of small sample sizes, bootstrapping should be used. Another (rigorous) solution would be to adopt a Bayesian inference approach, as it contains the main advantage that it provides exact results even for small datasets, i.e. it does not rely on asymptotics. Moreover, it has the advantage that it is possible to include *a priori* knowledge. Geweke and Amisano (2010) propose a framework to compare different Bayesian likelihood based prediction models with alternative predictive distributions. However, this approach is methodologically very different from the theory present in this thesis (which is frequentist). Adopting a Bayesian inference approach therefore corresponds to the construction of an entirely new testing framework.

At last, our test statistics could be extended to allow for more general weight functions. A possible direction for further research is to allow for a time-varying region of interest. Moreover, note that in this thesis we only focus on a "block" within the left tail of the copula distribution. For application areas other than risk management, it could be useful to consider different "shapes" and/or different locations as region of interest.

8 References

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

Alexander, C. (2008). *Market Risk Analysis, Volume II: Practical Financial Econometrics*. John Wiley & Sons, Lt, Hoboken, New Jersey.

Ang, A., and J. Chen. (2002). Asymmetric Correlations of Equity Portfolios. *Journal of Financial Economics* 63, 443-494.

Berg, D. (2009). Copula goodness-of-fit testing: An overview and power comparison. *European Journal of Finance*; Forthcoming, DOI:10.1080/13518470802697428.

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

Diks, C., Panchenko, V. and Van Dijk, D. (2009). Out-of-sample comparison of copula specifications in multivariate density forecasts. *Tinbergen Institute Discussion Paper TI 2008-105/4*

Diks, C., Panchenko, V. and Van Dijk, D. (2010). Partial Likelihood-Based Scoring Rules for Evaluating Density Forecasts in Tails. *Tinbergen Institute Discussion Paper TI 2008-050/4*

Embrechts, P., Klüppelberg, C. and Mikosch T. (1997) *Modelling extremal events for Insurance and finance*. Springer Verlag, Berlin.

Embrechts, P., McNeil, A., Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond* (ed. M.A.H. Dempster), pp. 176-223. Cambridge University Press, Cambridge.

Embrechts, P. (2008). Copulas: A personal view *Journal of Risk and Insurance* 76, 639-650.

Fang, K.-T., Kotz, S., and Ng, K.-W. (1990). *Symmetric Multivariate and Related Distributions*. Chapman & Hall, New York.

Fisher, N. I., and Switzer, P. (1985). Chi-plots for assessing dependence. *Biometrika*, 72(2), 253–265.

Fisher, N. I., and Switzer, P. (2001). Graphical assessment of dependence: Is a picture worth 100 tests? *American Statistics*, 55(3), 233–239.

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

Genest C., Rémillard, B. and Beaudoin D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44, 199-213.

Geweke, J., Amisano, G. (2010). Comparing and evaluating predictive distributions of asset returns. *International Journal of Forecasting*, vol. 26, pp. 216-230

Glosten, L.R., Jagannathan, R., Runkle, D.E. (1993). On the Relation between Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, vol. 48, pp. 1779-1801.

Granger, C.W.J., 2002. Time series concept for conditional distributions. Manuscript, UCSD.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Monographs in Statistics and Probability 73, Chapman and Hall, London.

McNeil, A.J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press, Princeton.

Mikosch, T. (2006). Copulas: Tales and Facts. *Extremes* 9, 3-20. Discussion, 21-53. Rejoinder 55-62.

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

Patton, A. (2004). On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation. *Journal of Financial Econometrics*, Vol. 2, No. 1, pp. 130—168

Patton, A. (2006). Copula–based models for financial time series. In *Handbook of Financial Time Series* To be published in Andersen, T.G., Davies, R.A., Kreiss, J.–P. and Mikosch, T. (Eds.) (2007), Springer, Berlin.

Scarsini, M. (1984). On measures of concordance. *Stochastica*, 8, 201–218.

Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9(4):879–885.

Winkler, R.L. and Murphy, A.H. (1968). Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory*, 4, 210-230.

Appendix 1 Explosive development of Copula Theory
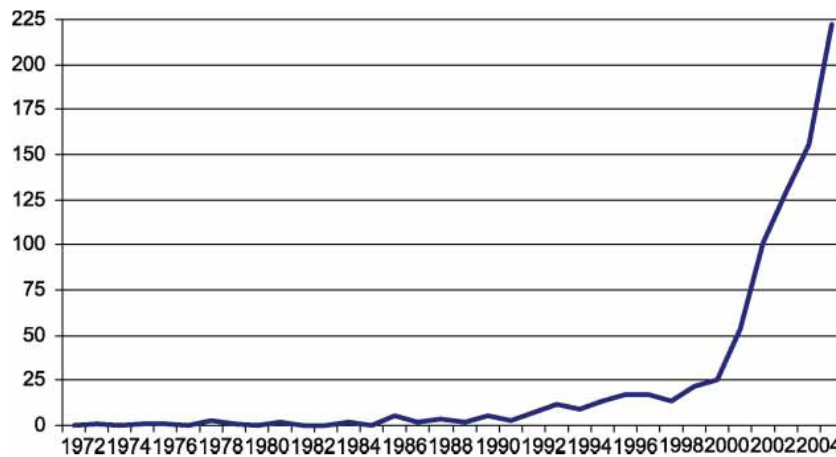
Genest et al (2009) made a survey about the explosive development of copula theory the last few decades. They clearly observed a "boom" in copula research starting the end of the nineties, see figure A1.1. They also found that copulas are used in numerous applications areas in diverse disciplines ranging from health sciences, biology, hydrology, environmental science to finance, see figure A1.2. Not surprisingly, the latter experienced the most attention due to the booming banking and insurance business in the last few decades. The risk management methodology acknowledged an explosive growth, mostly driven by new regulatory guidelines and new financial products in the eighties and nineties (Embrechts 2008). Due to their flexibility, copula modelling was considered to be a invaluable tool for handling the presence of several phenomena in empirical asset prices, such as the *stylized facts* of assets prices as discussed in section 5.2.
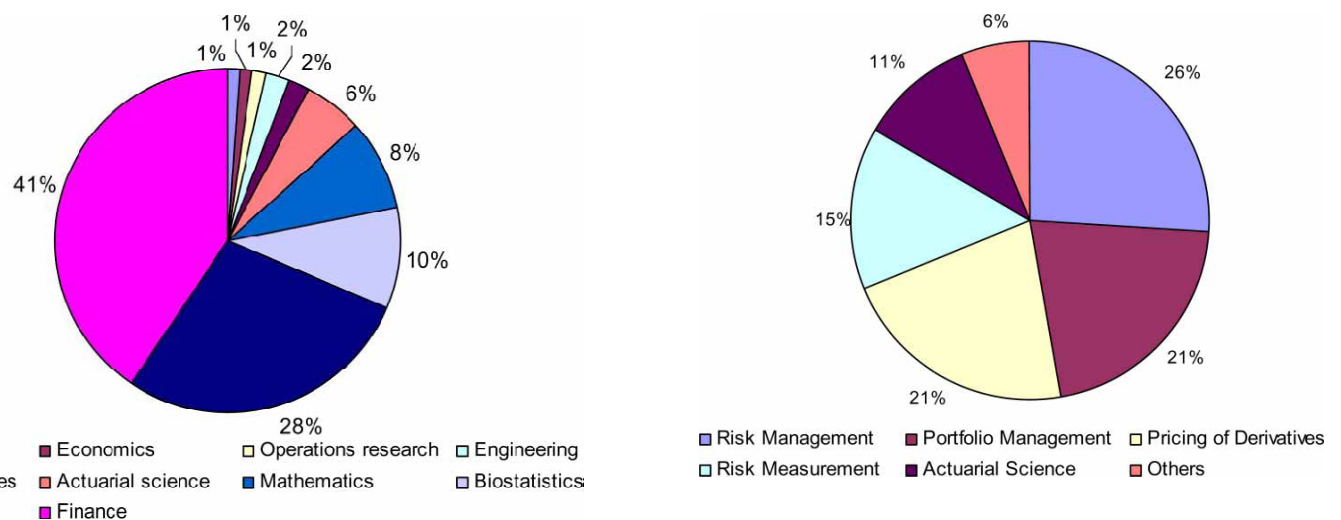
Appendix 2 Overview of goodness-of-fit tests for Copulas

Here we discuss the main findings two recent studies concerning tests for appropriate copula specification performed by Berg (2009) and Genest, Rémillard, and Beaudoin (2009). Both articles present an overview of existing goodness-of-fit tests. We categorized the tests in four major groups. Further details can be found in the relating articles.

First, we note the Chi square tests based on discretized data by gridding the probability space. However, this "binned approach" is not recommended because of 3 reasons. It is unfeasible for high-dimensional data due to the curse of dimensionality. Moreover, the grouping of the data is arbitrary so the results of the tests depend on the number of categories selected. At last, due to dependence between ranks the distribution is not exactly the same as in the classical context.

The second group consists of copula specific tests, such as procedures for testing the dependence structure of the Normal copula and the Clayton family.

The third group consists of general tests involving tuning parameters. These tests are applicable to *any* copula family but their implementation involves a tuning parameter or kernels, weight functions and associated smoothing parameters.

The fourth class is the largest; it consists of the dimension reduction approaches, applicable to any copula structure and requiring no strategic choice for their use (also called 'blanket tests' by Genest et al, 2009). The tests currently available are based on:
   a) Empirical copula, using a rank-based version of the Cramer-von Mises or Kolmogorov-Smirnov statistic. Approximate *p*-values can only be obtained by a parametric bootstrap procedure.
   b) Kendall's transform, which is a specific probability integral transformation of the data. Again the *p*-values can only be obtained via specially adapted Monte Carlo methods (bootstrapping).
   c) Rosenblatt's transform, which is based on the Anderson-Darling test statistic together with a parametric bootstrap procedure. Note: this Rosenblatt transform is non-unique, and is therefore criticized as being somewhat arbitrary.

Regarding to the dimension reduction approaches, Genest et al (2009) found that in general, the greater the sample size, the better. They further conclude that the tests relying on Rosenblatt's transform are at an advantage, since they require a single bootstrap to approximate the null distribution to extract *p*-values (the empirical copula and Kendall's transform require double bootstrapping). The Cramer von Mises (CvM) test statistics are almost invariable more powerful than Kolmogorov-Smirnov functionals. They found that the CvM test statistics based on the empirical copula and corrected Rosenblatt transform yield the best results. The CvM test statistic based on Kendall's transform is also recommendable, especially when the null hypothesis is *Archimedean* (since K is then available in closed form).

Berg (2009) found increasing power with sample size, dimension and dependence. The best performance is the proposed approach of Berg that averages several empirical copula approaches. Special tests for elliptical copulas or the Clayton copula may however be preferable, especially for high dimensions and large sample sizes.

Appendix 3 Pitfalls of Linear correlation

If we are aware of the restrictions of these traditional models, why are they still widely used? The main reason is that in the specific case of elliptical distributions, such as the multivariate normal and students-t distribution, linear correlation can be used as dependence measure. Linear correlation, or Pearson's correlation coefficient r, between two random variables X and Y is given by:

$$r(X,Y) = cor(X,Y) = \frac{E(XY) - E(X)E(Y)}{\sqrt{var(X)var(Y)}} \qquad (A3.1)$$

Embrechts et al (2002) recall the advantages of working with linear correlation, as it is straightforward to work with, easy to manipulate under linear transformations and it is a natural measure of dependence in multivariate normal distributions. Due to these reasons, linear correlation is widely adopted in financial theory, such as CAPM and APT. However, one should keep in mind that it can only be used when working with elliptical distributions, which requires symmetric marginal distributions and linear dependence. One should be aware of the pitfalls of using linear correlation (see Embrechts et al (2002) for a detailed description):

1. r only measures *linear* association. This implies that it is not invariant under nonlinear strictly increasing functions, such as the log-transform.
2. $X \perp Y \Longrightarrow r = 0$, but the reverse only holds in the specific case of the multivariate normal distribution! Moreover, it holds in general that r can be close to zero even if X and Y are strongly dependent[26], which clearly may lead to misleading results.
3. r does not always exist, e.g. the bivariate Cauchy distribution has infinite second moments, such that the Pearson's correlation coefficient in (A3.1) cannot be computed.

Note that these drawbacks all can lead to misleading results, which makes linear correlation an undesirable dependence measure.

Copula models avoid these restrictions, and have therefore become an important tool to describe dependence between random variables. The most basic copula is the independence copula, $\Pi(u,v) = uv$, which is a straightforward consequence of Sklar's theorem given in Theorem 2.1:

$$X \perp Y \Leftrightarrow C = \Pi \qquad \text{with } \Pi(u,v) = uv \qquad (A3.2)$$

If however $C \neq \Pi$, then X and Y are dependent. To what extent C can deviate from Π can be quantified by a certain measure of association/dependence that is based on the copula of (X,Y) only. Such a measure always exists, because any copula is a bounded function on a bounded set by definition. Scarsini (1984) suggested a measure of concordance based on copulas, which is defined as follows:

---

[26] i.e. if $r \approx 0$, this does not necessarily mean that X and Y are close to independence.

Definition A3.1 *A measure of concordance is a function $\kappa(X,Y)$ satisfying the following axioms:*

(i) κ is defined for any random pair (X,Y) with continuous margins.

(ii) $\kappa(X,Y) = \kappa(Y,X)$.

(iii) If the copula C of (X,Y) and the copula D of (W,Z) satisfy $C \prec_{PQD} D$ then $\kappa(X,Y) \leq \kappa(W,Z)$.

(iv) The range of κ is [-1,1].

(v) $X \perp Y$ implies $\kappa(X,Y) = 0$.

(vi) $\kappa(-X,Y) = -\kappa(X,Y)$.

(vii) If $(X_n, Y_n) \rightsquigarrow (X,Y)$ then $\kappa(X_n, Y_n) \to \kappa(X,Y)$ as $n \to \infty$.

For details, see Scarsini (1984). It can be shown that for a random pair (X,Y) with its copula C, all 7 axioms are satisfied, using among others that copulas are invariant under increasing transformations and are bounded functions on a bounded set by definition. The only drawback of working with measures of concordance is that the reverse of axiom (v) does not hold, i.e. $\kappa(X,Y) \nRightarrow X \perp Y$. Measures that satisfy $\delta(X,Y) \Leftrightarrow X \perp Y$ are called *measures of dependence*. The definition differs slightly from the above definition by changing the last three axioms (see for example Schweizer and Wolff, 1981), but they are more complicated to work with.

One of the most common copula-based measure of concordance is Kendall's τ. In 1938, Kendall introduced the following measure of concordance for the random pair (X,Y):

$$\tau(X,Y) = P\{(X - X^*)(Y - Y^*) > 0\} - P\{\{(X - X^*)(Y - Y^*) < 0\} \tag{A3.3}$$

Where (X,Y) is a random pair and (X^*,Y^*) its independent copy. In other words, it substracts the probability of that two pairs (X,X^*) and (Y,Y^*) are *concordant* (that is: $(X - X^*)(Y - Y^*) > 0$) to the probability that the two pairs are *discordant* (that is: $(X - X^*)(Y - Y^*) > 0$). It can be shown that this notation can be rewritten as:

$$\tau(X,Y) = 4 \int_0^1 \int_0^1 C(u,v) dC(u,v) - 1 \tag{A3.4}$$

To obtain the relationship of Kendall's tau with the Gaussian copula C_Σ^{Ga} we derive the integral:

$$\int_0^1 \int_0^1 C_\Sigma^{Ga}(u,v) dC_\Sigma^{Ga}(u,v) = \frac{1}{4} + \{\arcsin(r)\}/2\pi \tag{A3.5}$$

Consequently, $\tau\left(C_\Sigma^{Ga}\right) = \frac{2}{\pi}\arcsin(r)$, which holds for all elliptical copulas (McNeill et al 2005, Proposition 5.37).

Nelsen (2006) provides a proof for the following relation of Kendall's tau and Archimedean copulas, in terms of their density generator $\psi_\theta(x)$

$$\tau = 1 + 4 \int_0^1 \frac{\psi_\theta(t)}{\psi_{\theta'}(t)} dt \tag{A3.6}$$

To emphasize that Kendall's tau is not the only possible measures of concordance, we note that there also exist other measures of concordance such as Spearman's ρ, Van den Waarden's coefficient ω, Gini's γ or Blomqvuist's β.

Appendix 4 Copula examples

Copula densities and simulation scatter plots of the six copulas of interest with $\tau = 0.2$.
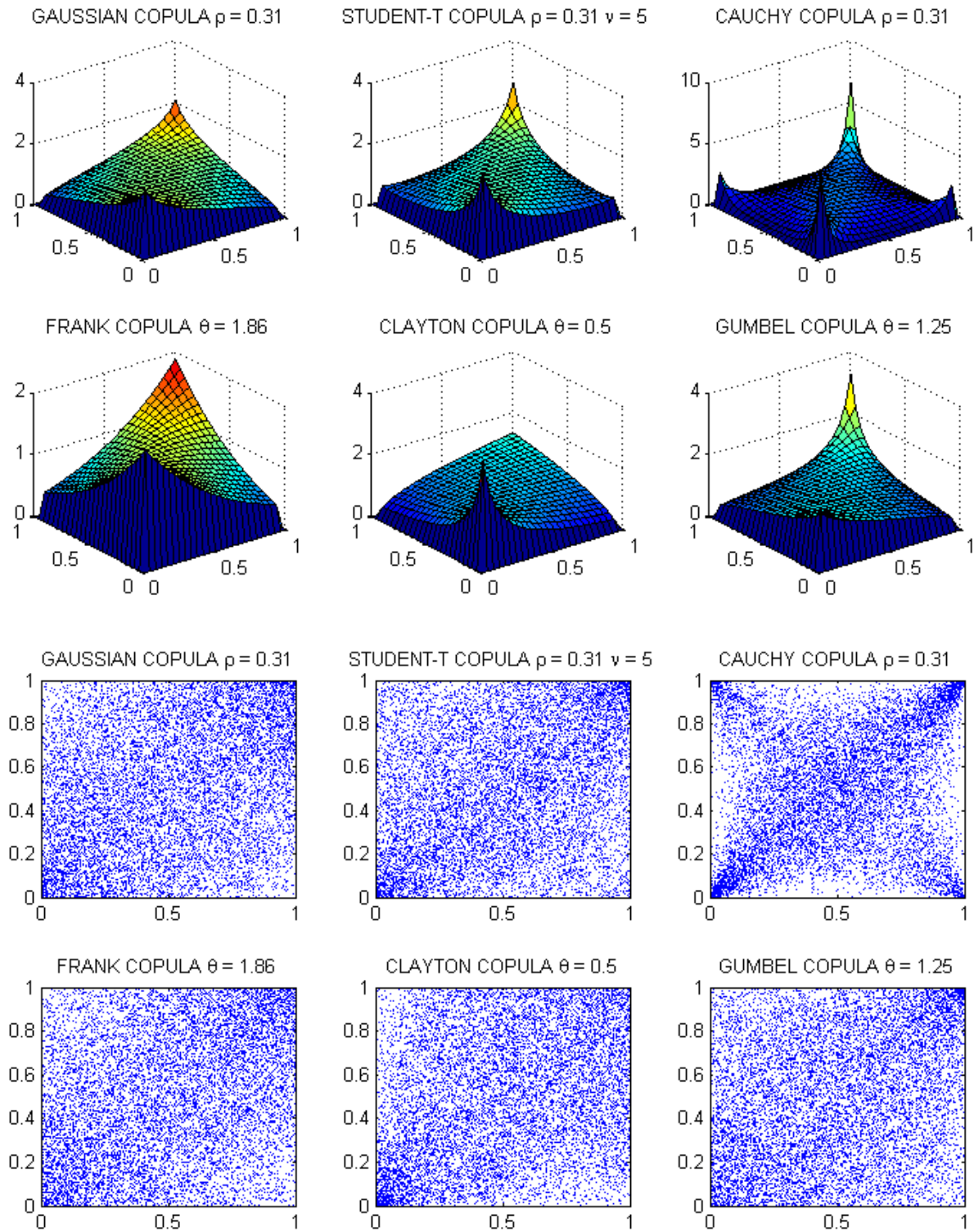
Figure A4.1 The six upper graphs show from left to right the density of the (bivariate) Gaussian copula, the Student-t Copula, the Cauchy Copula, the Clayton Copula, the Gumbel Copula and the Frank copula respectively. The lower scatterplots show a simulation sample of N=10,000 of the six corresponding copula specifications. In all graphs, the parameter values are set to match Kendalls tau fixed to $\tau = 0.2$.

Copula densities and simulation scatter plots of the six copulas of interest with $\tau = 0.5$.
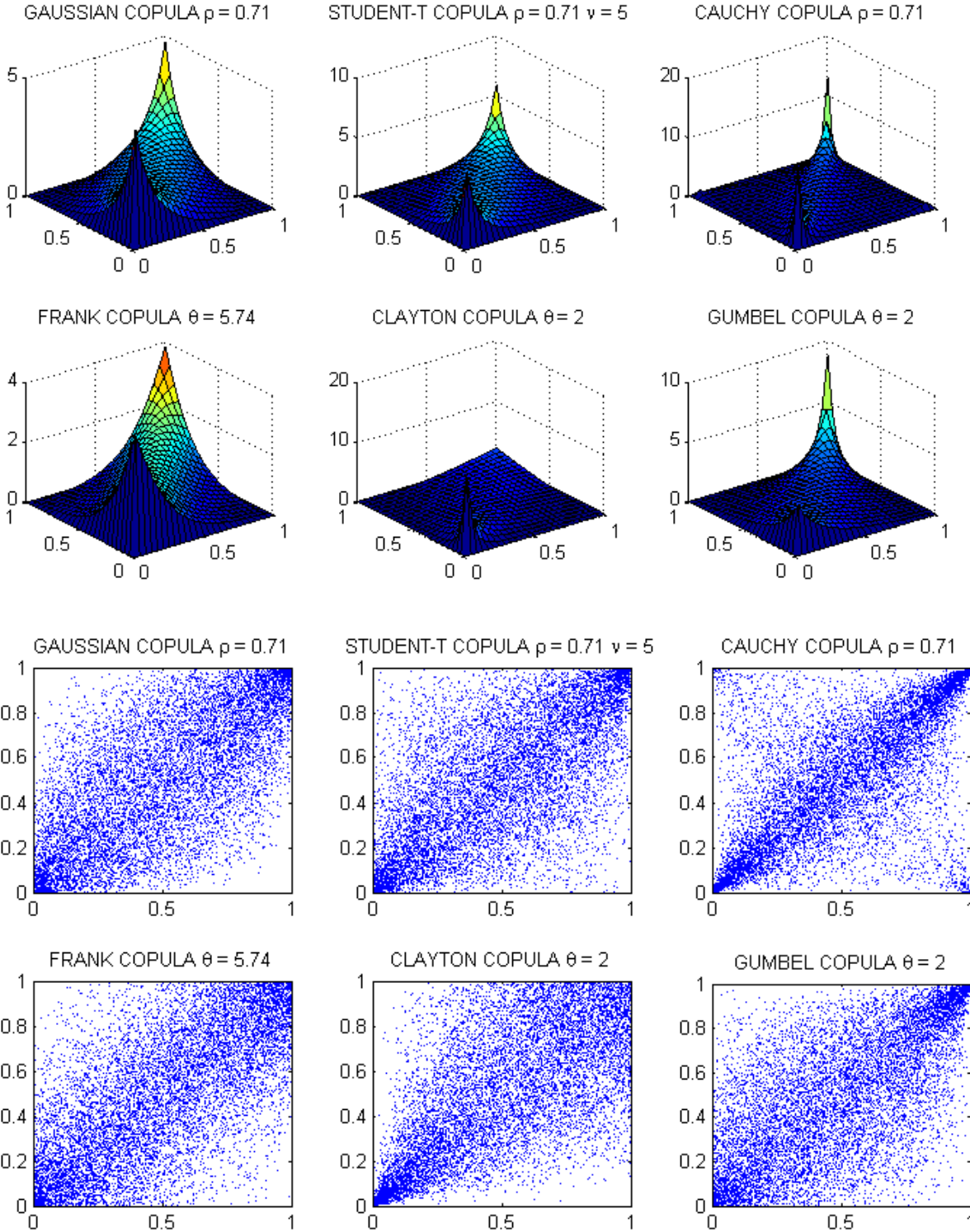
Figure A4.2 The six upper graphs show from left to right the density of the (bivariate) Gaussian copula, the Student-t Copula, the Cauchy Copula, the Clayton Copula, the Gumbel Copula and the Frank copula respectively. The lower scatterplots show a simulation sample of N=10,000 of the six corresponding copula specifications. In all graphs, the parameter values are set to match Kendalls tau fixed to $\tau = 0.5$.

Appendix 5 Additional Power Experiments: Increasing Correlation

Results for the Power plots with correlation fixed to $\rho_{12} = 0.3$:
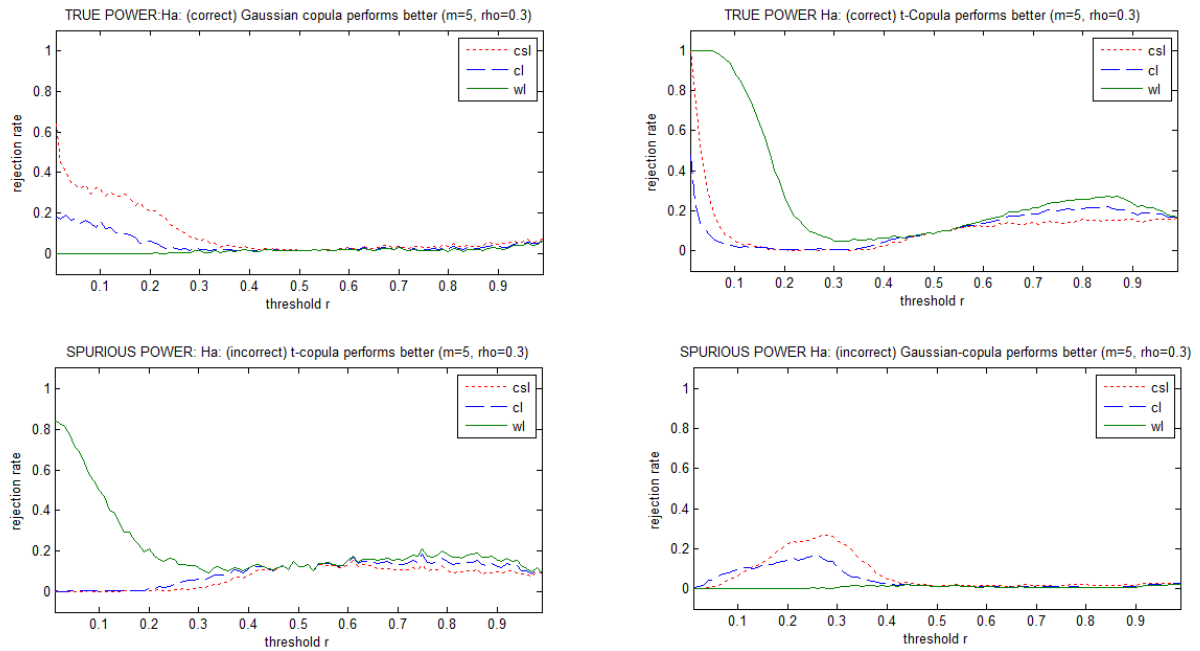
Figure A5.1: One sided rejection rates at significance level 5% to assess the power properties of the equal predictive ability test as defined in (3.28) when using the weighted logarithmic (wl), the conditional likelihood (cl) and the censored likelihood (csl) scoring rules under the weight function $w(u,v) = \mathrm{I}(u,v \le r)$ where $0 < r < 1$, based on 1,000 replications and m=5 expected observations in the region of interest. For the graphs in the left (right) columns, the DGP is the Gaussian Copula (Student-t(5) copula) with rho=0.3. The graphs in the top (bottom) panels show average rejection rates against superior predictive ability of the correct (incorrect) copula specification, i.e. true (spurious) power.

Figure A5.2: This figure is similar to Figure A5.1, but now with m=40 expected observations in the left tail.

Results for the Power plots with correlation fixed to $\rho_{12} = 0.6$:
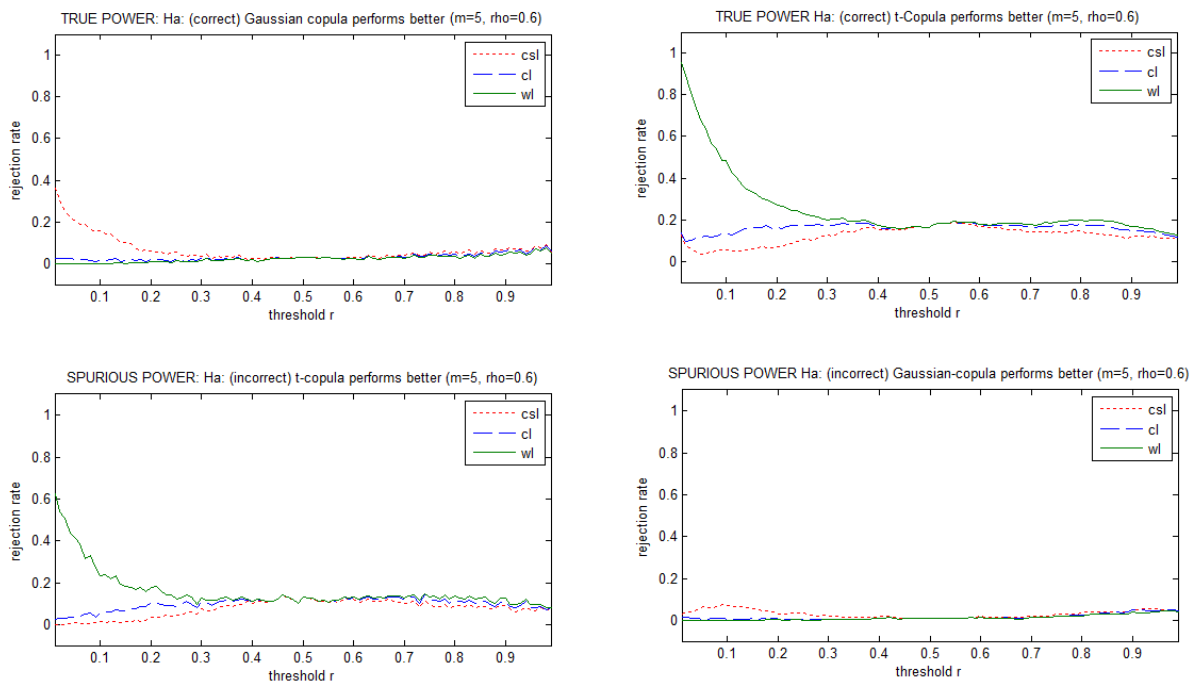
Figure A5.3: One sided rejection rates at significance level 5% to assess the power properties of the equal predictive ability test as defined in (3.28) when using the weighted logarithmic (wl), the conditional likelihood (cl) and the censored likelihood (csl) scoring rules under the weight function $w(u, v) = \mathrm{I}(u, v \le r)$ where $0 < r < 1$, based on 1,000 replications and m=5 expected observations in the region of interest. For the graphs in the left (right) columns, the DGP is the Gaussian Copula (Student-t(5) copula) with rho=0.6. The graphs in the top (bottom) panels show average rejection rates against superior predictive ability of the correct (incorrect) copula specification, i.e. true (spurious) power.

Figure A5.4: This figure is similar to Figure A5.3, but now with m=40 expected observations in the left tail.

Results for the Power plots with correlation fixed to $\rho_{12} = 0.9$:
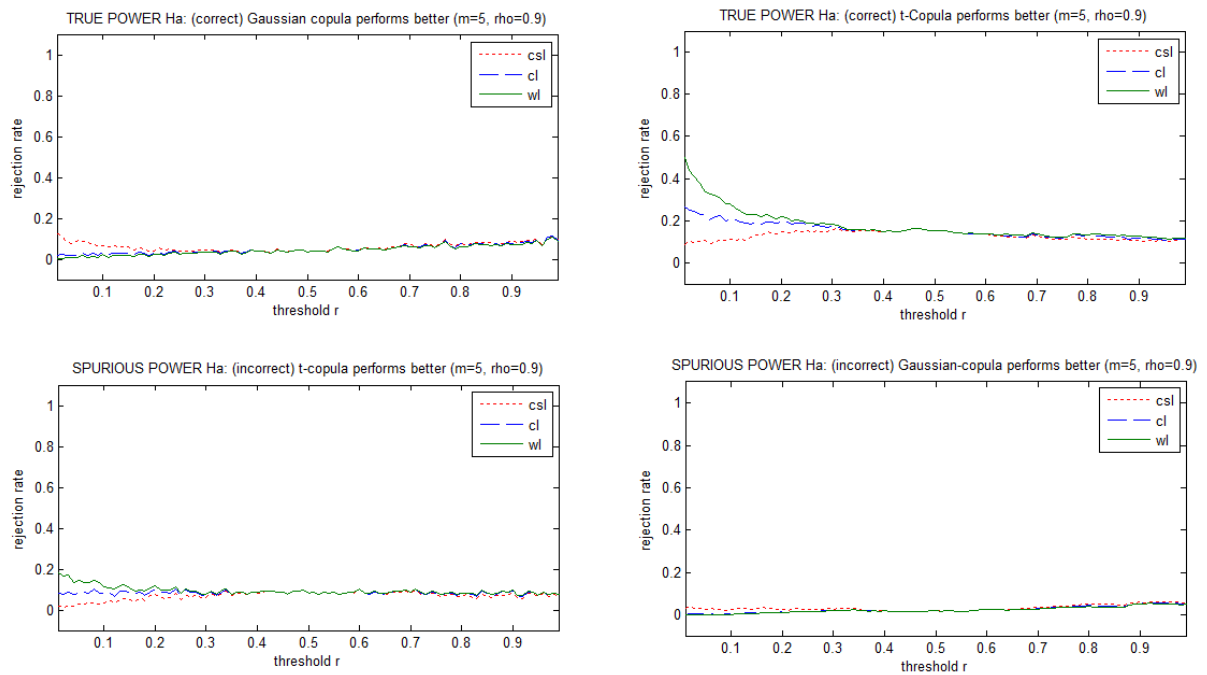
Figure A5.5: One sided rejection rates at significance level 5% to assess the power properties of the equal predictive ability test as defined in (3.28) when using the weighted logarithmic (wl), the conditional likelihood (cl) and the censored likelihood (csl) scoring rules under the weight function $w(u, v) = \mathrm{I}(u, v \leq r)$ where $0 < r < 1$, based on 1,000 replications and m=5 expected observations in the region of interest. For the graphs in the left (right) columns, the DGP is the Gaussian Copula (Student-t(5) copula) with rho=0.9. The graphs in the top (bottom) panels show average rejection rates against superior predictive ability of the correct (incorrect) copula specification, i.e. true (spurious) power

Rho.9 m=40
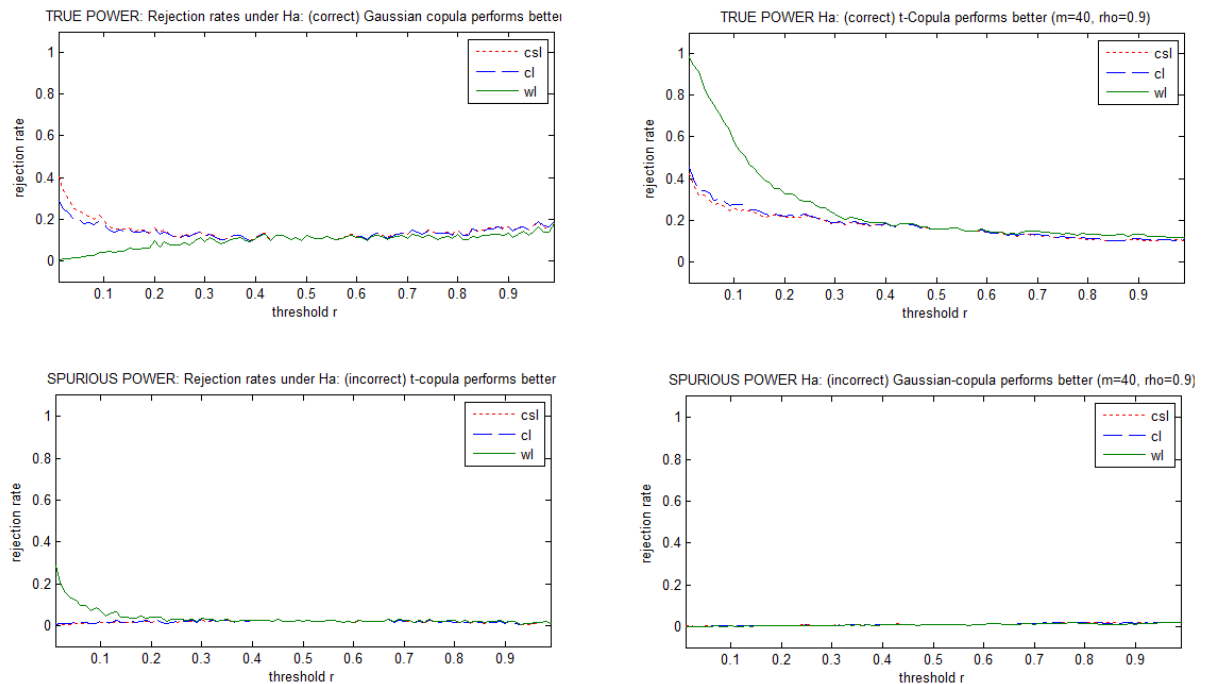
Figure A5.6: figure is similar to Figure A5.5, but now with m=40 expected observations in the left tail.

Expected mean relative scores with Gaussian DGP: the correlation parameter is increasing from $\rho_{12} = 0.3$ in the top panel to $\rho_{12} = 0.6$ in the middle panel and $\rho_{12} = 0.9$ in the lowest panel.
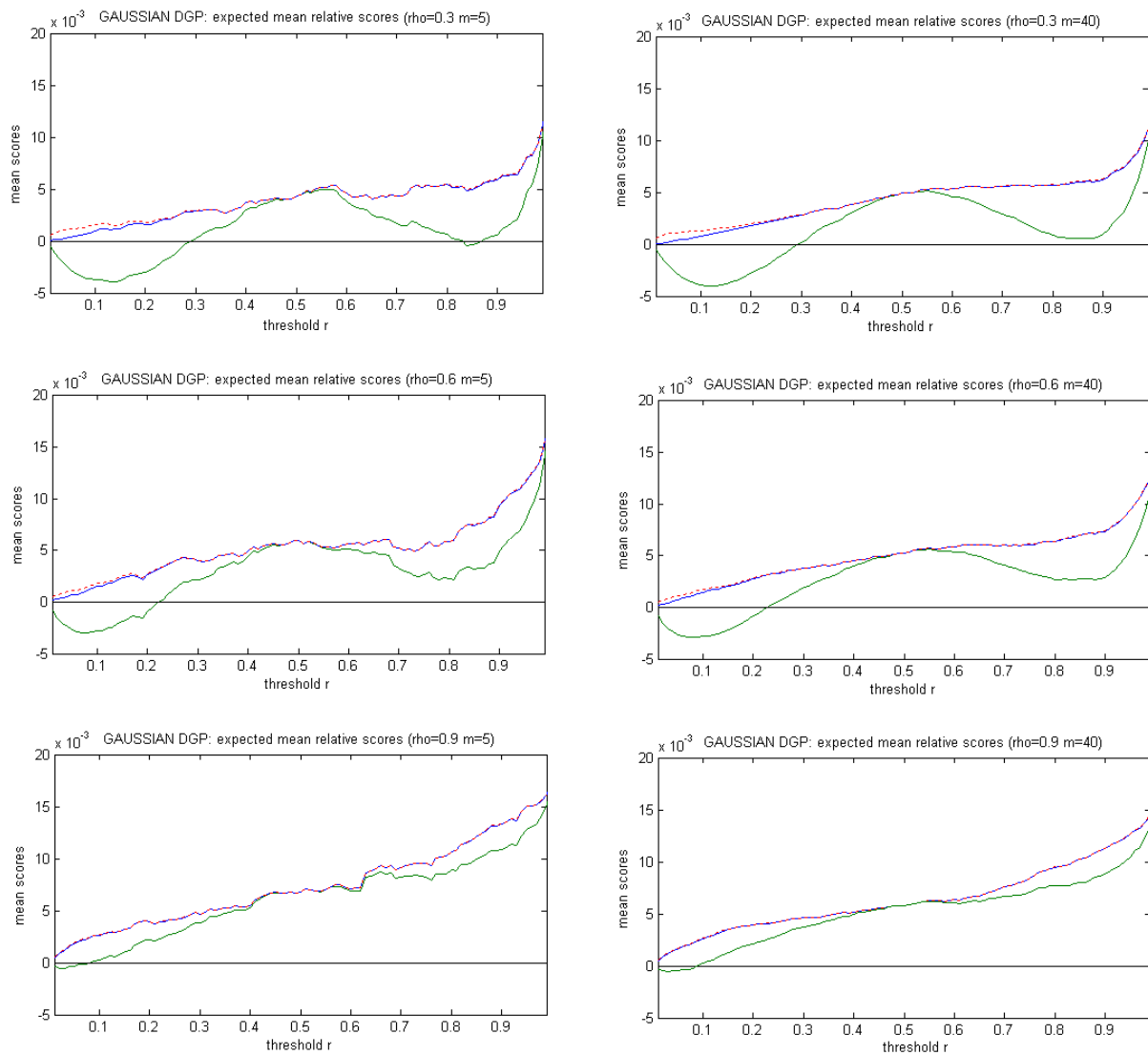

Figure A5.7: Proxy of the mean relative scores for sample sizes m=5 (left graphs) and m=40 (right graphs), when using the weighted logarithmic (*wl*), the conditional likelihood (*cl*) and the censored likelihood (*csl*) scoring rules under the weight function $w(u, v) = I(u, v \leq r)$ where $0 < r < 1$, based on 10,000 replications. The graphs show the results for the Gaussian copula as DGP, with the correlation parameter fixed to rho=0.3 in the upper two graphs, rho=0.6 in the middle two graphs and rho=0.9 in the lower two graphs.

10% and 90% percentiles mean relative scores with Gaussian DGP: the correlation parameter is increasing from $\rho_{12} = 0.3$ in the top panel to $\rho_{12} = 0.6$ in the middle panel and $\rho_{12} = 0.9$ in the lowest panel.
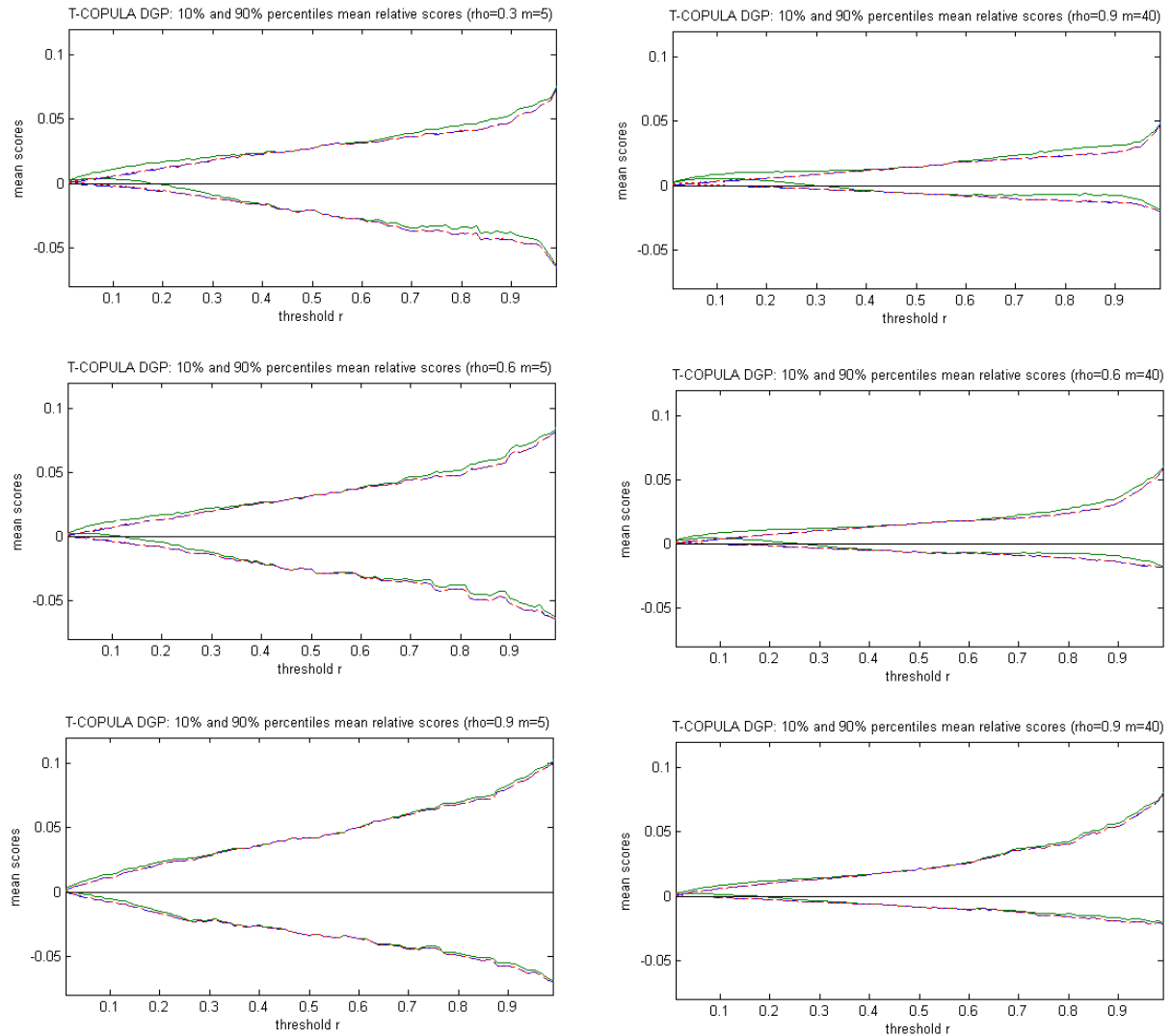

Figure A5.8: 10% and 90% percentiles for the upper two graphs of figure 4.5, i.e. the mean expected relative scores of comparing an Gaussian Copula with the Student-t(5) Copula, with the Gaussian Copula as DGP. The two graphs in the left (right) panel correspond to the m=5 (m=40) expected observations in the region of interest. The graphs show the results for the Gaussian copula as DGP, with the correlation parameter fixed to rho=0.3 in the upper two graphs, rho=0.6 in the middle two graphs and rho=0.9 in the lower two graphs. The graphs show the results for the full range of r, that is $r \in [0.01, 1]$.

10% and 90% percentiles mean relative scores with Gaussian DGP.

This figure is zooming in on the left tail, with the three graphs at the left (right) hand side correspond to the small (large) sample size with m=5 (m=40) expected observations in the region of interest, here the left tail. The correlation parameter is increasing from $\rho_{12} = 0.3$ in the top panel to $\rho_{12} = 0.6$ in the middle panel and $\rho_{12} = 0.9$ in the lowest panel.
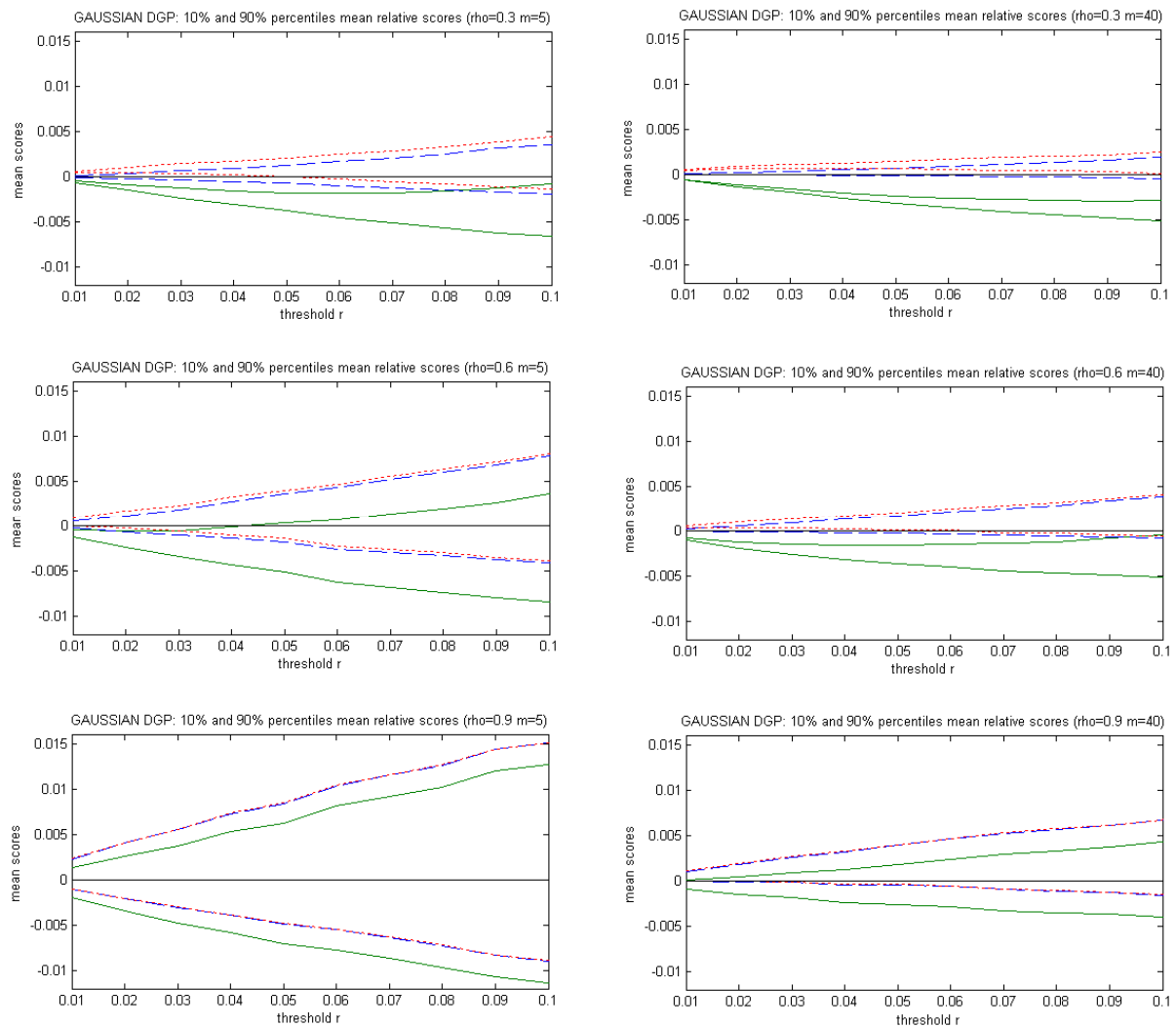
Figure A5.9: 10% and 90% percentiles for the upper two graphs of figure 4.5, i.e. the mean expected relative scores of comparing an Gaussian Copula with the Student-t(5) Copula, with the Gaussian Copula as DGP. The two graphs in the left (right) panel correspond to the m=5 (m=40) expected observations in the region of interest. The graphs show the results for the Gaussian copula as DGP, with the correlation parameter fixed to rho=0.3 in the upper two graphs, rho=0.6 in the middle two graphs and rho=0.9 in the lower two graphs. The graphs "zoom" in the left tail of the copula with $r \in [0.01, 0.1]$.

Expected mean relative scores with Student-t(5) copula as DGP: the correlation parameter is increasing from $\rho_{12} = 0.3$ in the top panel to $\rho_{12} = 0.6$ in the middle panel and $\rho_{12} = 0.9$ in the lowest panel.
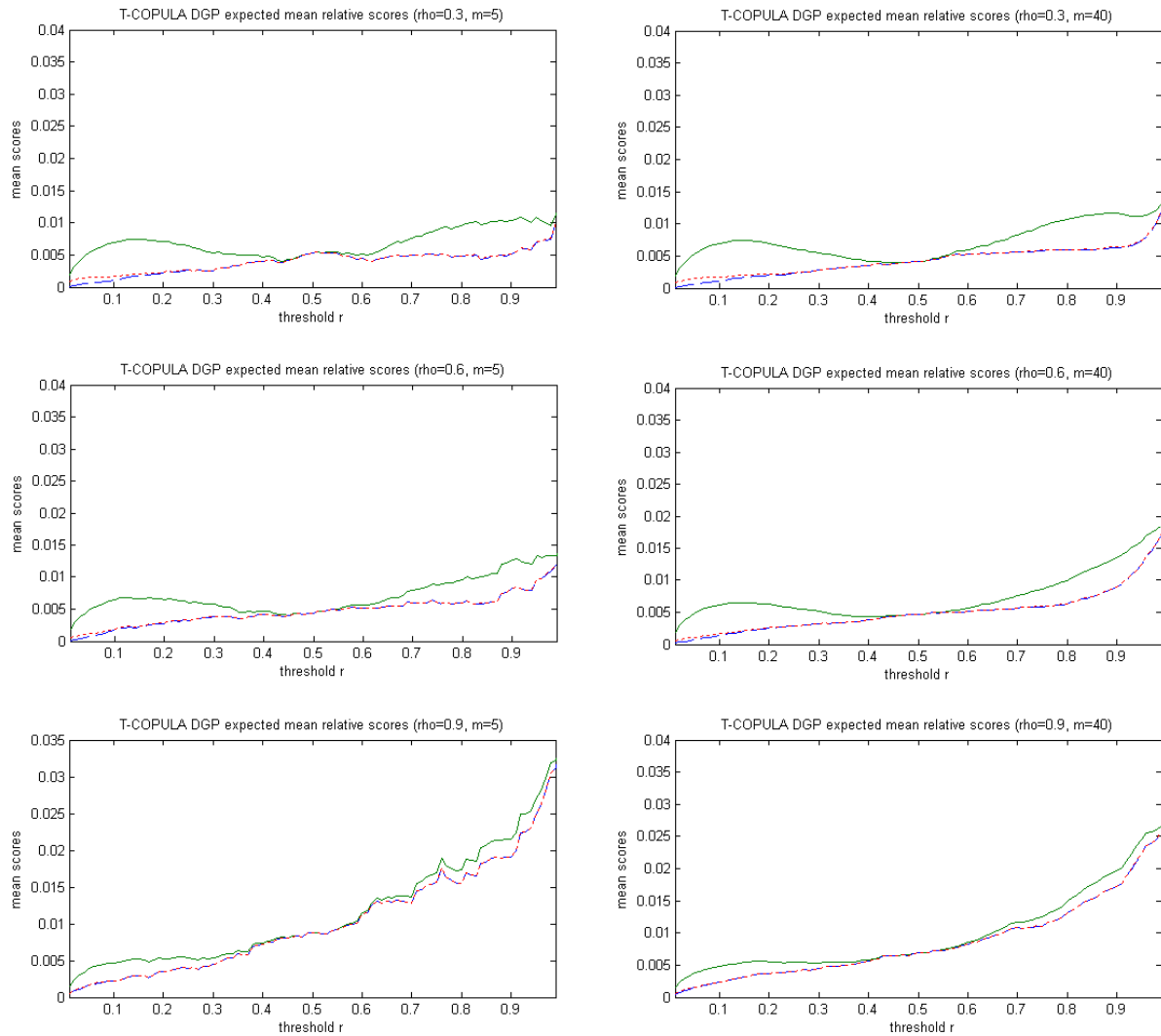

Figure A5.10: Proxy of the mean relative scores for sample sizes m=5 (left graphs) and m=40 (right graphs), when using the weighted logarithmic (*wl*), the conditional likelihood (*cl*) and the censored likelihood (*csl*) scoring rules under the weight function $w(u,v) = \text{I}(u,v \le r)$ where $0 < r < 1$, based on 10,000 replications. The graphs show the results for the Student-t Copula as DGP, with the correlation parameter fixed to rho=0.3 in the upper two graphs, rho=0.6 in the middle two graphs and rho=0.9 in the lower two graphs

10% and 90% percentiles mean relative scores with Student-t(5) copula as DGP.

Expected mean relative scores with Student-t copula as DGP: the correlation parameter is increasing from $\rho_{12} = 0.3$ in the top panel to $\rho_{12} = 0.6$ in the middle panel and $\rho_{12} = 0.9$ in the lowest panel.

Figure A5.11: : 10% and 90% percentiles for the upper two graphs of figure 4.5, i.e. the mean expected relative scores of comparing an Gaussian Copula with the Student-t(5) Copula, with the Student-t(5) Copula as DGP. The two graphs in the left (right) panel correspond to the m=5 (m=40) expected observations in the region of interest. The graphs show the results for the Gaussian copula as DGP, with the correlation parameter fixed to rho=0.3 in the upper two graphs, rho=0.6 in the middle two graphs and rho=0.9 in the lower two graphs. The graphs show the results for the full range of r, that is $r \in [0.01, 1]$.

10% and 90% percentiles mean relative scores with Student-t(5) copula as DGP.

This figure is zooming in on the left tail, with the three graphs at the left (right) hand side correspond to the small (large) sample size with m=5 (m=40) expected observations in the region of interest, here the left tail. The correlation parameter is increasing from $\rho_{12} = 0.3$ in the top panel to $\rho_{12} = 0.6$ in the middle panel and $\rho_{12} = 0.9$ in the lowest panel.
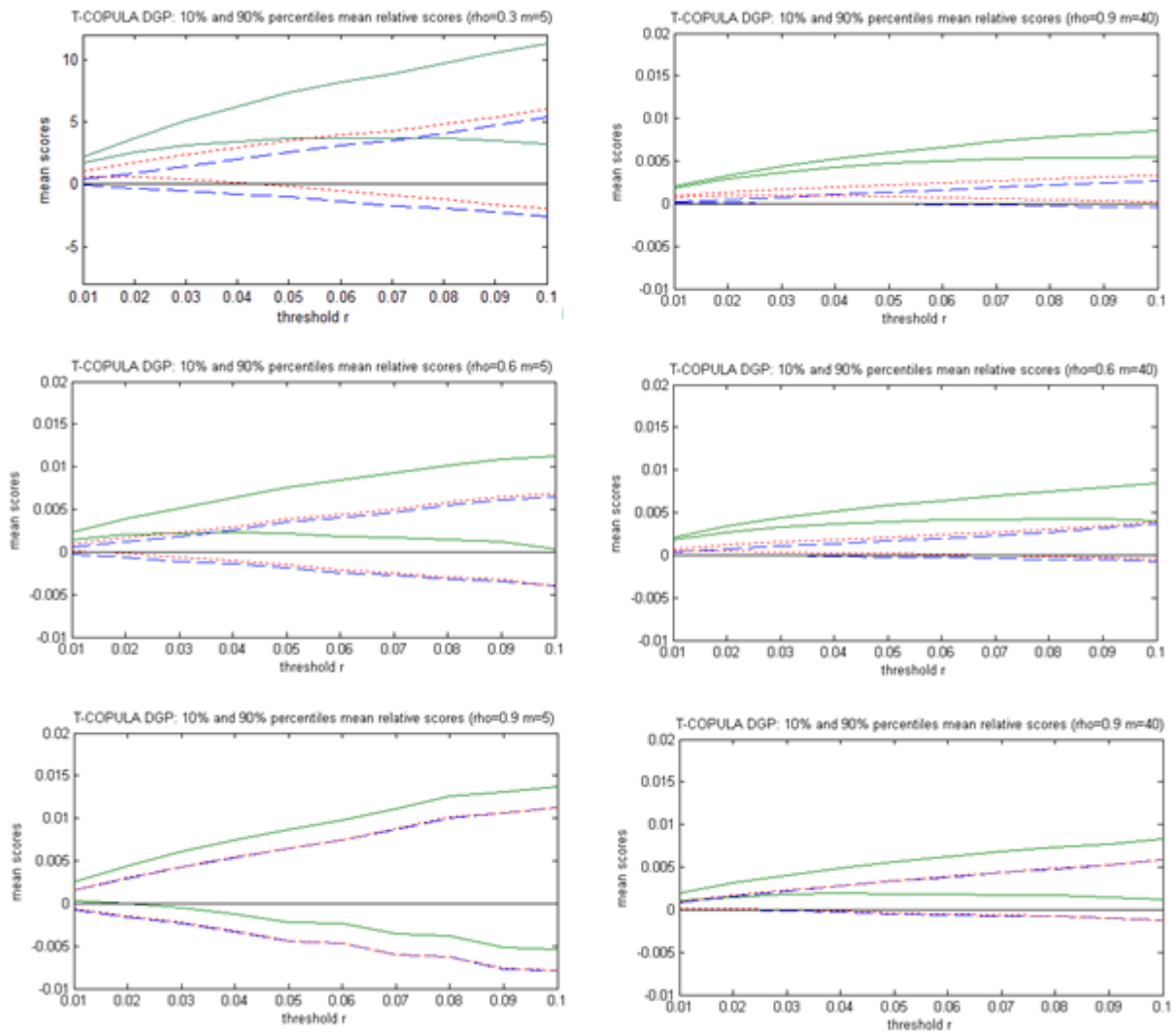
Figure A5.12: 10% and 90% percentiles for the upper two graphs of figure 4.5, i.e. the mean expected relative scores of comparing an Gaussian Copula with the independent Student-t(5) copula, with the Student-t(5) copula as DGP. The two graphs in the left (right) panel correspond to the m=5 (m=40) expected observations in the region of interest. The graphs show the results for the Student-t(5) copula as DGP, with the correlation parameter fixed to rho=0.3 in the upper two graphs, rho=0.6 in the middle two graphs and rho=0.9 in the lower two graphs. The graphs "zoom" in the left tail of the copula with $r \in [0.01, 0.1]$.

Appendix 6 Comparison Stylized Facts first in-sample period

First we provide the plots of (squared) autocorrelations to examine the univariate stylized facts (i) and (ii) for the first in-sample period from 2-1-1992 to 6-10-2000. Figure A6.3 shows the (squared) autocorrelations for the total returns, figure A6.4 shows the plots for the corresponding standardized residuals obtained by model (5.3).
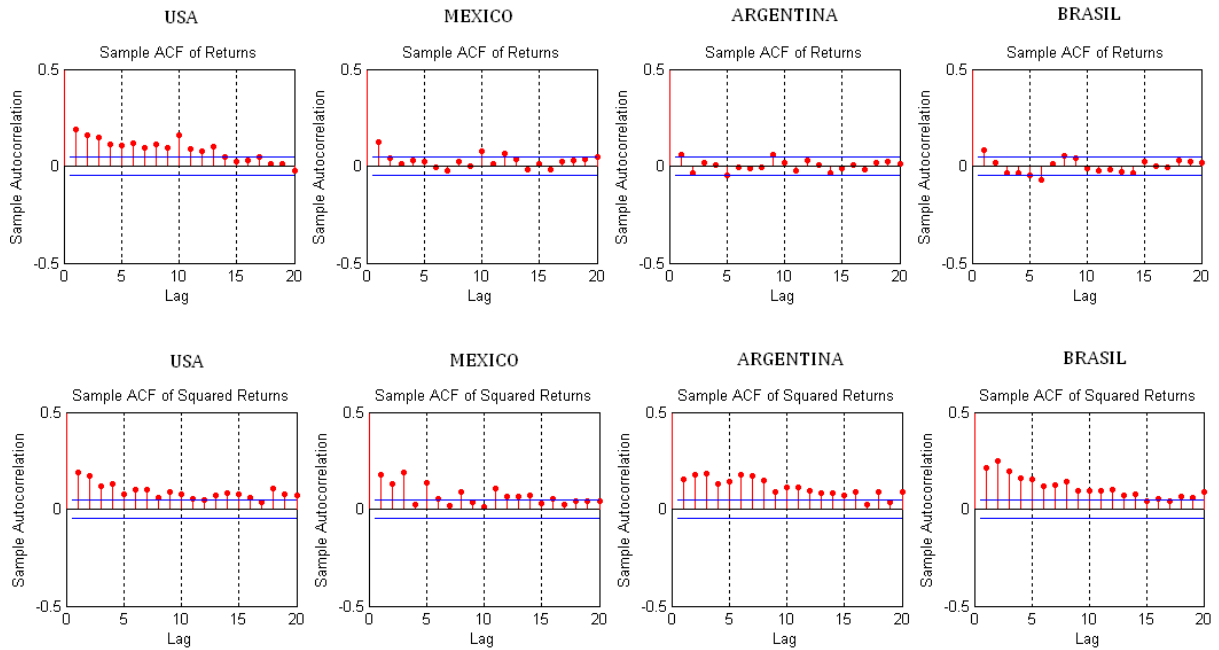

Figure A6.1: The upper panel exhibits the sample autocorrelation of the total returns of the first in-sample period, i.e. the first 2,000 observations from 2-1-1992 to 6-10-2000. The lower panel shows the sample autocorrelation of the corresponding squared total returns. The horizontal blue lines display the 5% significance levels.
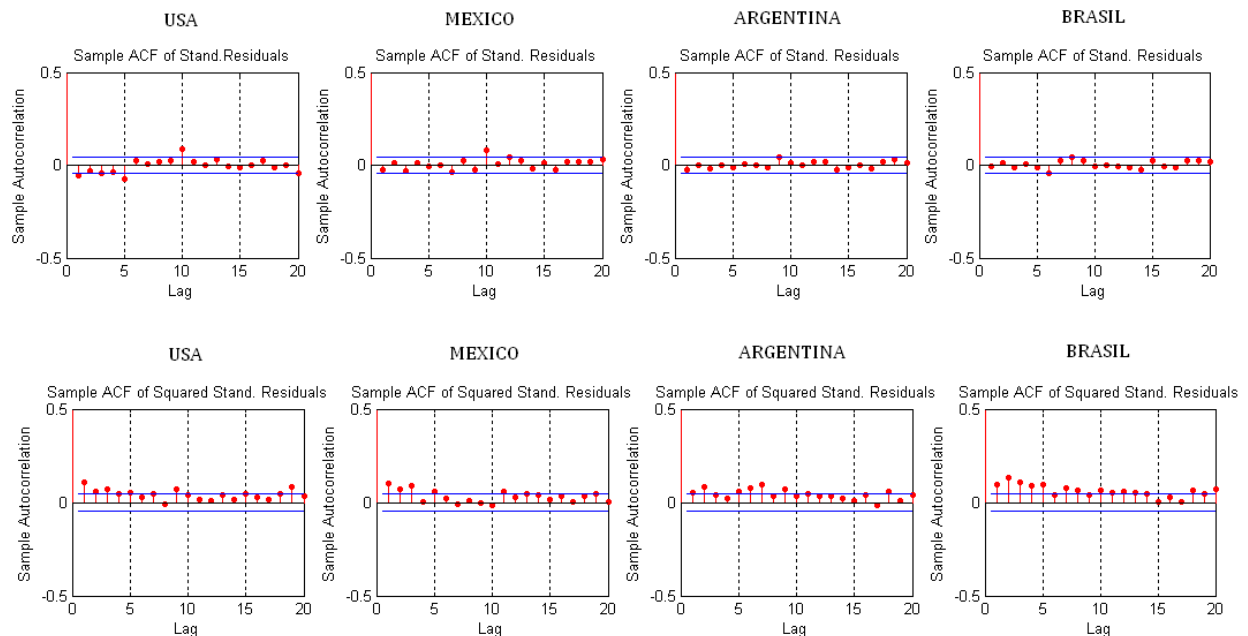
Figure A6.2: The upper panel exhibits the sample autocorrelation of the standardized residuals of the first in-sample period, i.e. the first 2,000 observations from 2-1-1992 to 6-10-2000. The lower panel shows the sample autocorrelation of the corresponding squared standardized residuals. The horizontal blue lines display the 5% significance levels.

Second, we compare several of the multivariate stylized facts. We start off with examining the rank scatterplots and corresponding Chi-plots for the period from 2-1-1992 to 6-10-2000 for all six country pairs. Figure A6.3 shows the scatterplots and corresponding Chi-plots for the total returns, figure A6.4 shows the plots for the corresponding standardized residuals obtained by model (5.3).
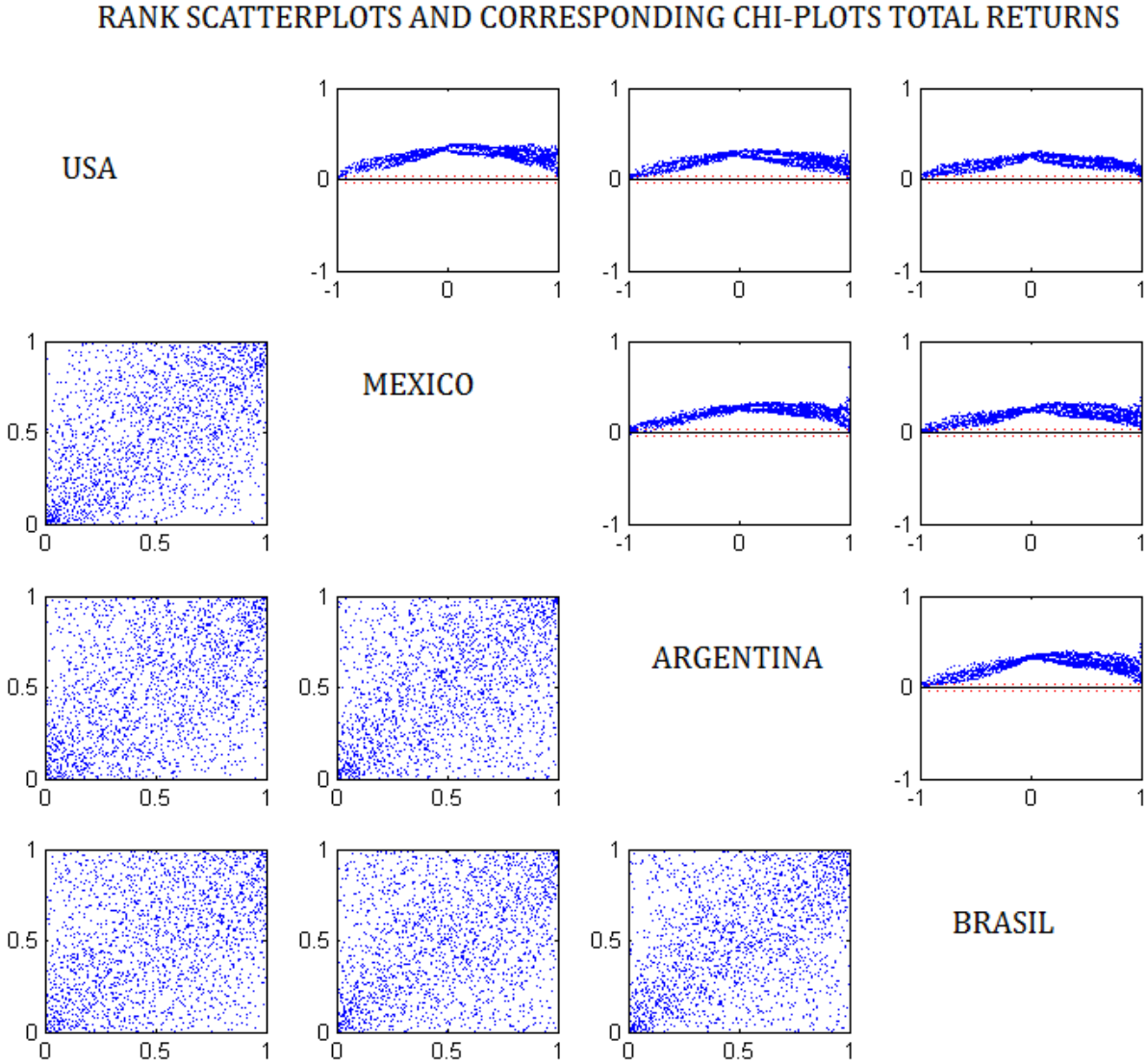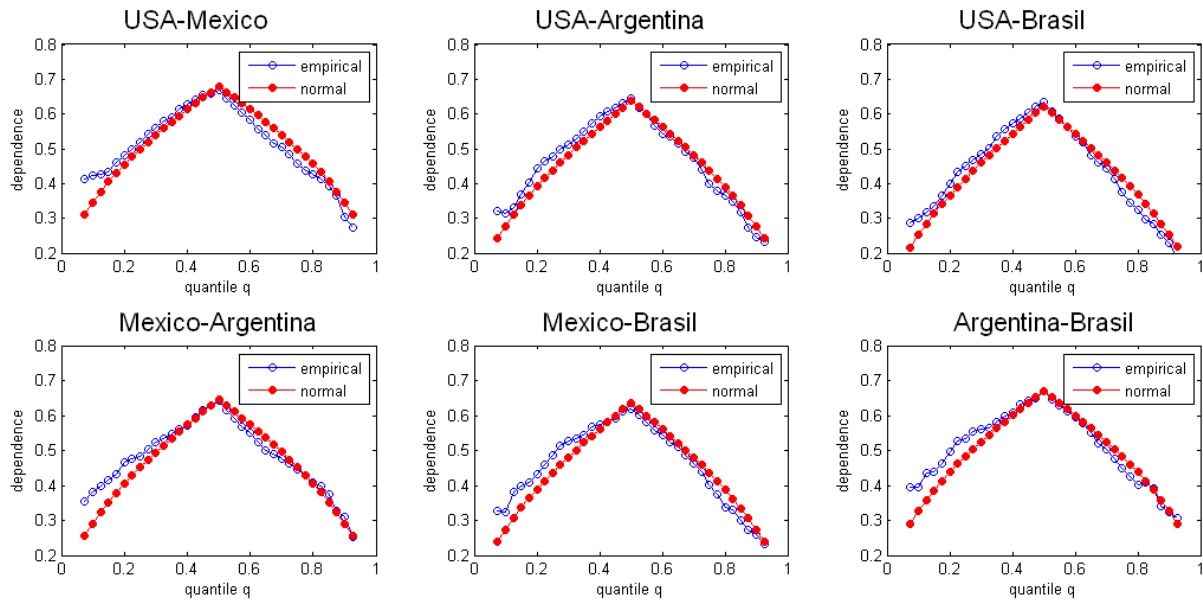

Figure A6.3 Rank scatterplots and corresponding chi-plots of all country pairs using the returns of the first in-sample period, i.e. the first 2,000 observations from 2-1-1992 to 6-10-2000. The horizontal red dotted lines in the chi-plots display the 5% significance control limits.

RANK SCATTERPLOTS AND CORRESPONDING CHI-PLOTS STAND. RESIDUALS
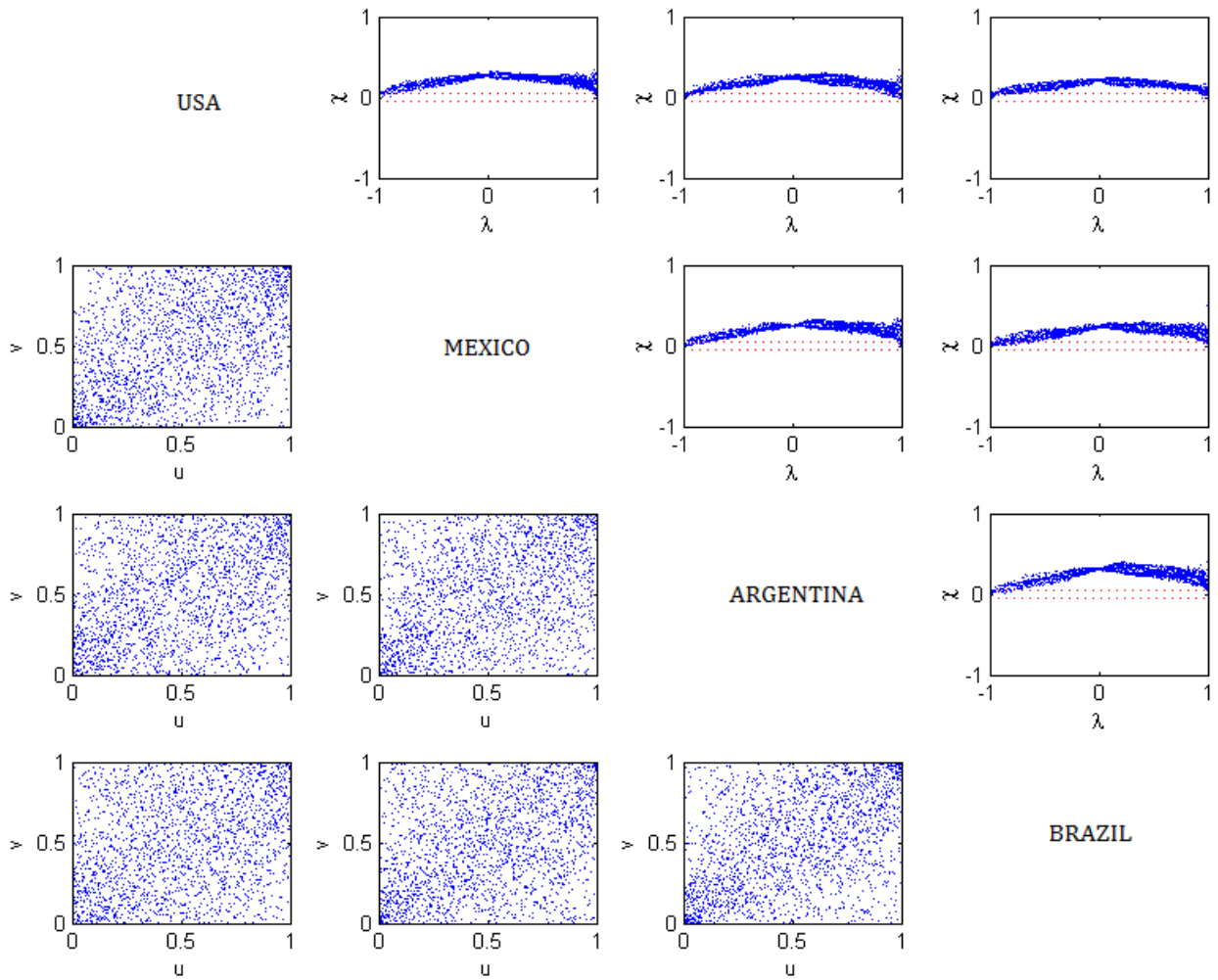
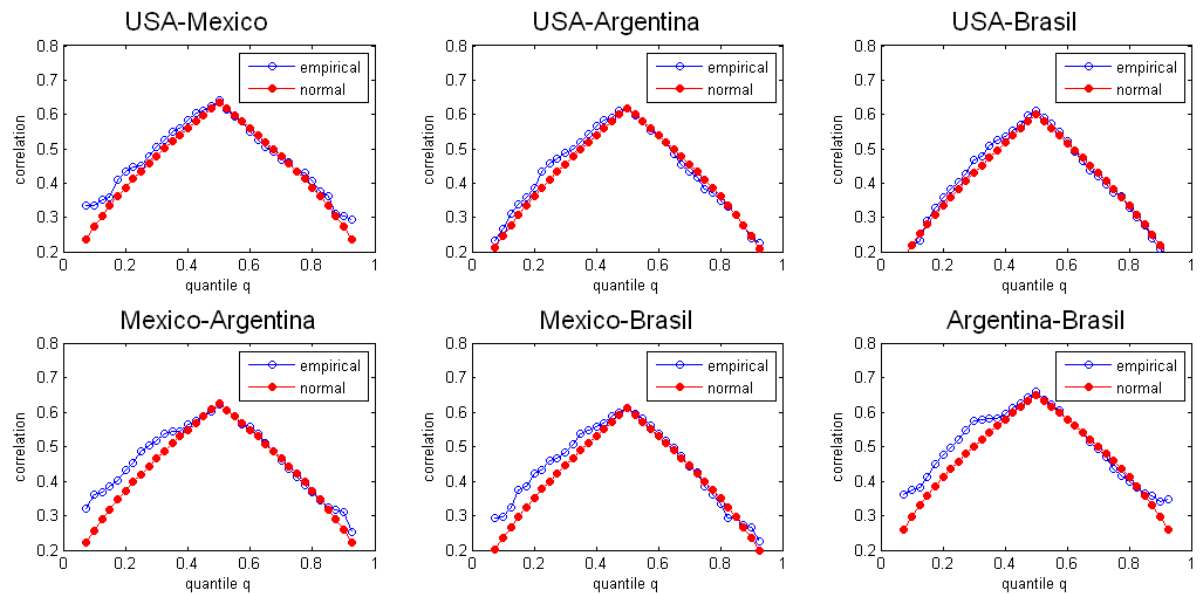**Figure A6.4 Rank scatterplots and corresponding chi-plots of all country pairs using the standardized residuals of the first in-sample period, i.e. the first 2,000 observations from 2-1-1992 to 6-10-2000. The horizontal red dotted lines in the chi-plots display the 5% significance control limits.**

The quantile dependence plots for the period from 2-1-1992 to 6-10-2000 for all six country pairs. Figure A6.5 shows the quantile dependence plots for the total returns, figure A6.6 shows the plots for the corresponding standardized residuals obtained by model (5.3).

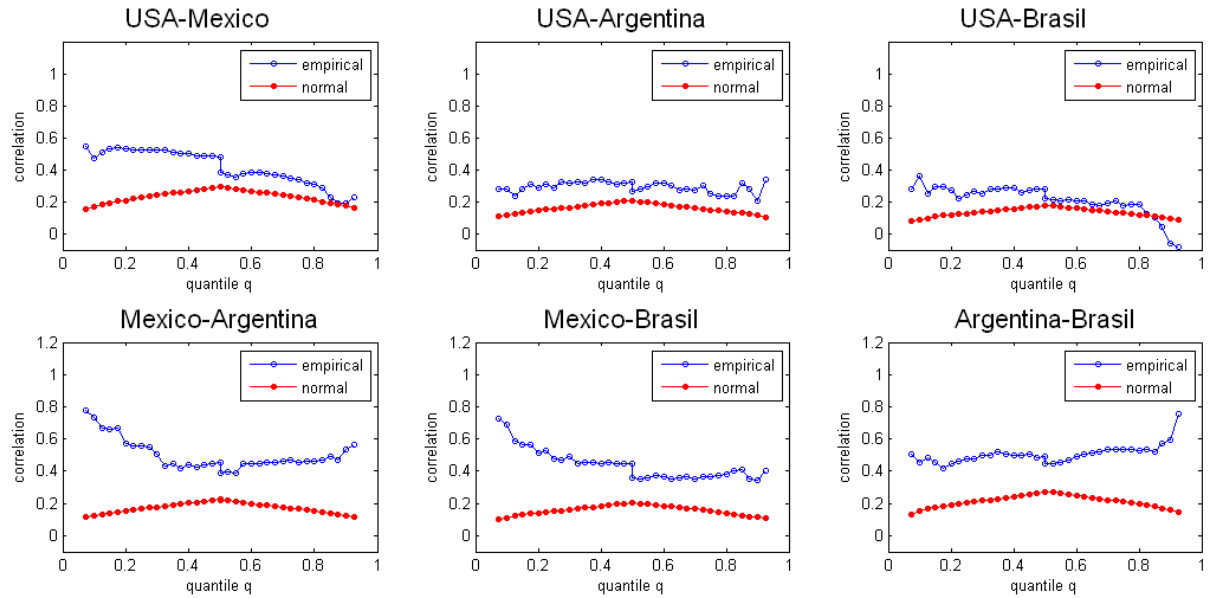**Figure A6.5: Quantile dependence of all country pairs using the returns of the first in-sample period, i.e. the first 2,000 observations from 2-1-1992 to 6-10-2000. The blue (red) curve displays the empirical (normally distributed) data.**
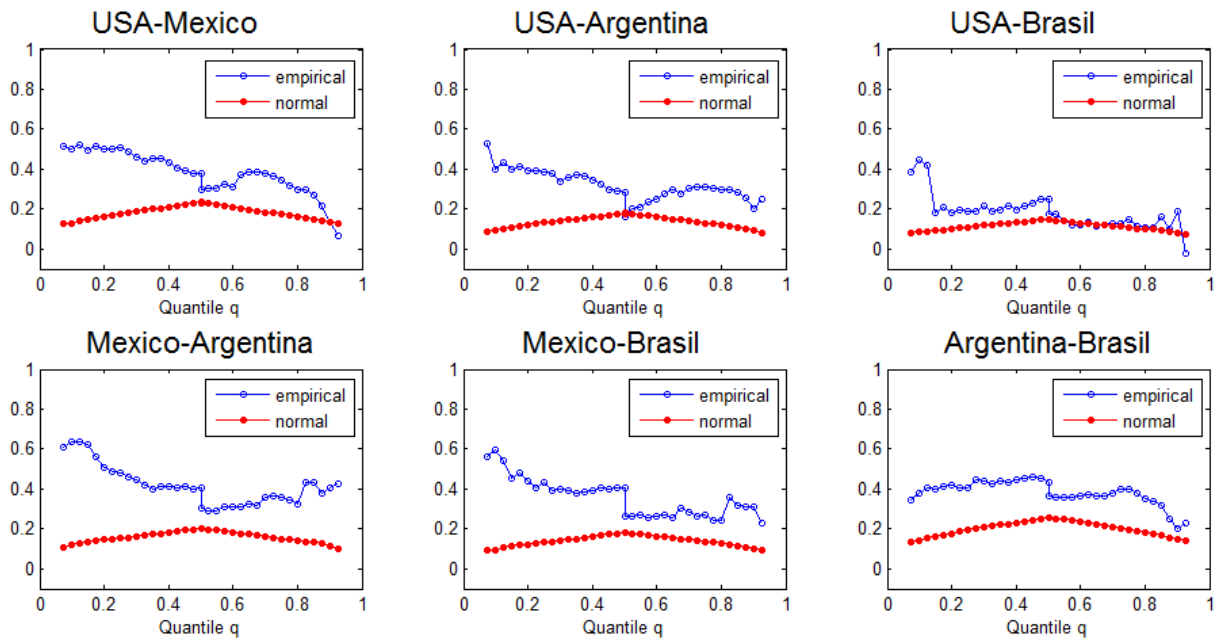

Figure A6.6: Quantile dependence of all country pairs using standardized residuals of the first in-sample period, i.e. the first 2,000 observations from 2-1-1992 to 6-10-2000. The blue (red) curve displays the empirical (normally distributed) data.

The exceedance correlations plots for the period from 2-1-1992 to 6-10-2000 for all six country pairs. Figure A6.7 shows the exceedance correlations plots for the total returns, figure A6.8 shows the plots for the corresponding standardized residuals obtained by model (5.3).

Figure A6.7: Exceedance correlations of all country pairs using the first in-sample period, i.e. the first 2,000 observations from 2-1-1992 to 6-10-2000. The blue (red) curves display the empirical (normally distributed) data.

Figure A6.8: Exceedance correlations of all country pairs using the standardized residuals of the first in-sample period, i.e. the first 2,000 observations from 2-1-1992 to 6-10-2000. The blue (red) curves display the empirical (normally distributed) data.

Appendix 7 Chi Plots

Examples of Chi-plots as in Fisher and Switzer (1985) for 2,000 random samples from the six different copula specifications introduced in section 2.3. Columns 1 to 4 correspond to values of τ of 0, 0.2, .35 and 0.5 respectively.
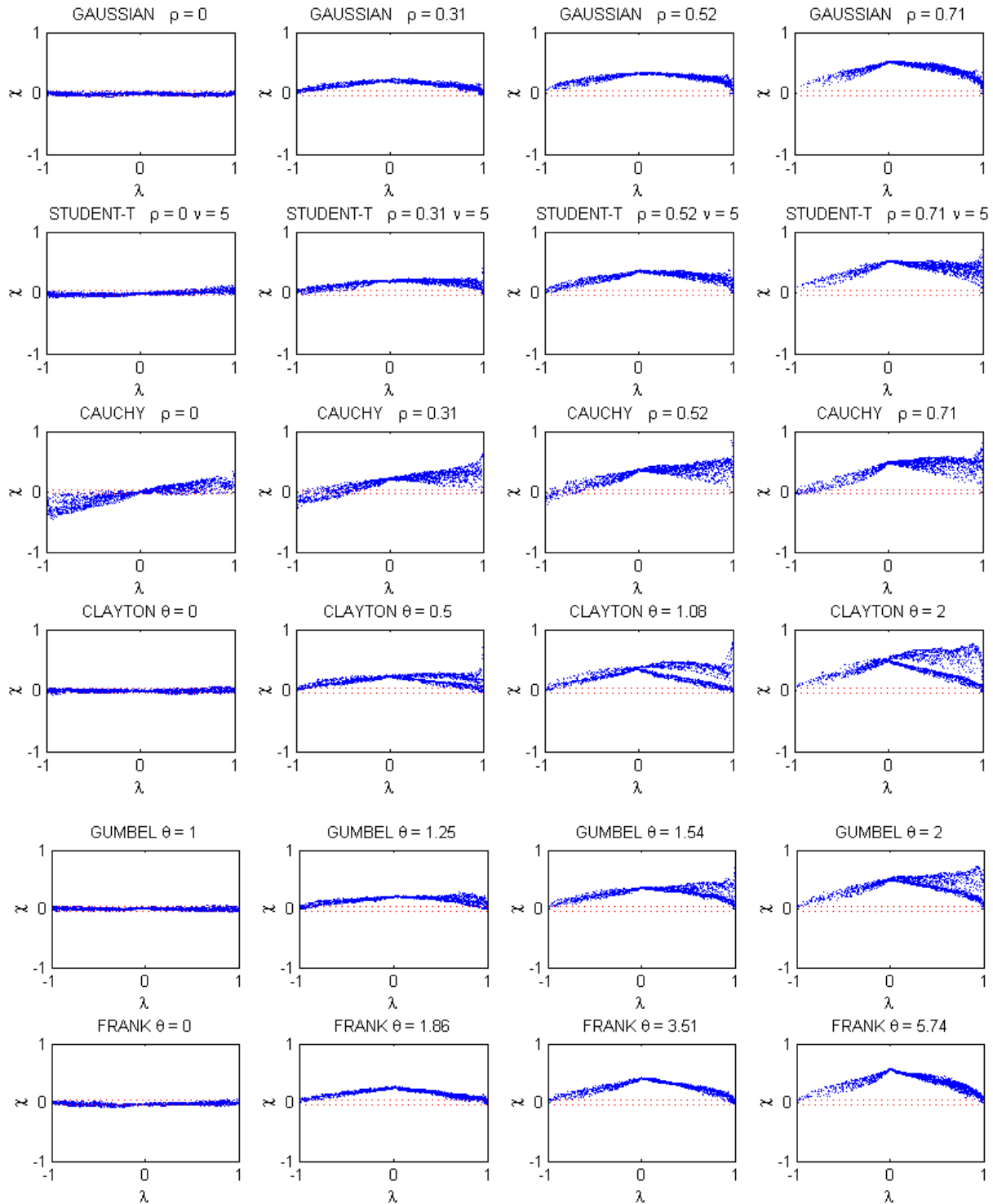

Figure A7.1: Examples of chi-plots for 2,000 simulated random pairs of different copula specifications. Columns 1 to 4 correspond to values of Kendall's τ of 0, 0.2, .35 and 0.5 respectively.The horizontal red dotted lines in the chi-plots display the 5% significance control limits.

Appendix 8 Estimated Parameters Empirical Application

This appendix shows the estimated paramters over the out-of-sample period of the country pairs USA-Brazil, Mexico-Brazil, Argentina-Mexico and Argentina-Brazil.
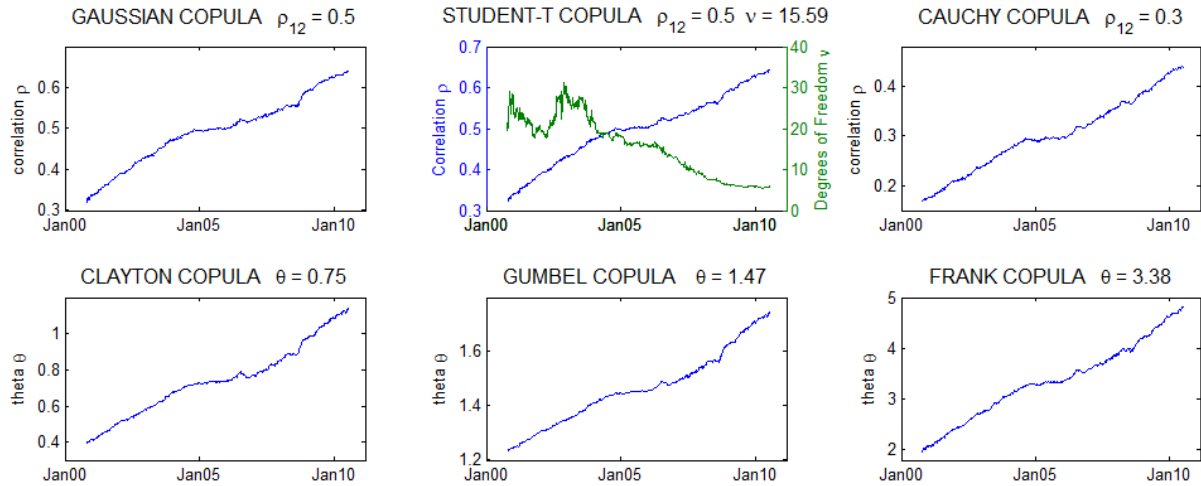

Figure A8.1: Data of the country-pair USA-Brazil: Estimated parameter values for the six different copula specifications from the rolling in-sample window of 2000 observations over the entire out-of-sample period from 6-10-2000 to 26-7-2010.
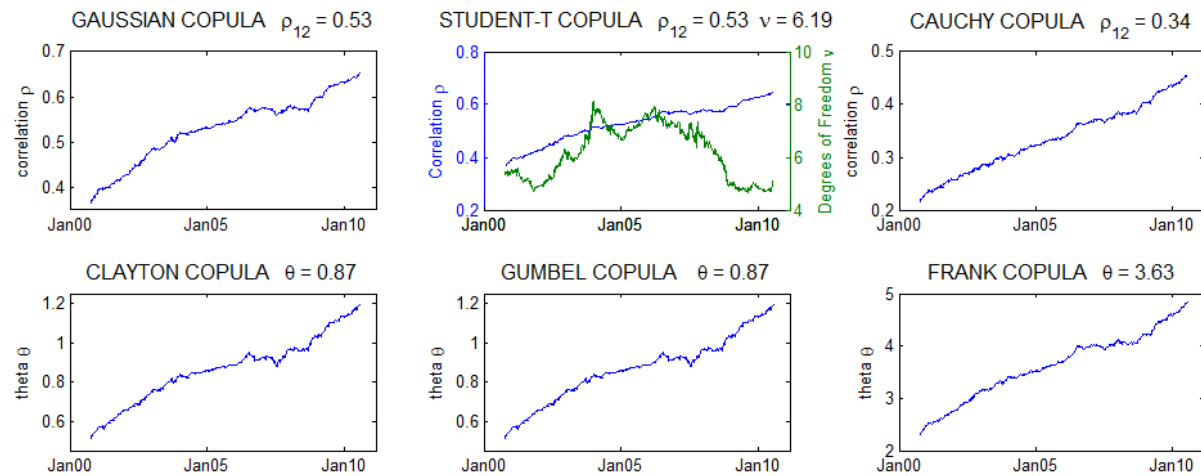

Figure A8.2: Data of the country-pair Mexico Brazil: Estimated parameter values for the six different copula specifications from the rolling in-sample window of 2000 observations over the entire out-of-sample period from 6-10-2000 to 26-7-2010.
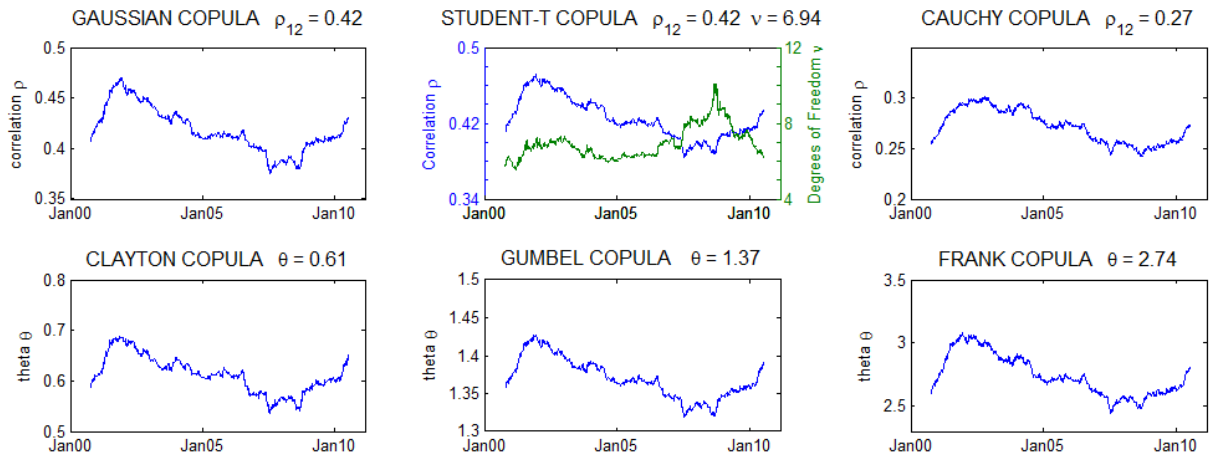
Figure A8.3: Data of the country-pair Mexico Argentina: Estimated parameter values for the six different copula specifications from the rolling in-sample window of 2000 observations over the entire out-of-sample period from 6-10-2000 to 26-7-2010.
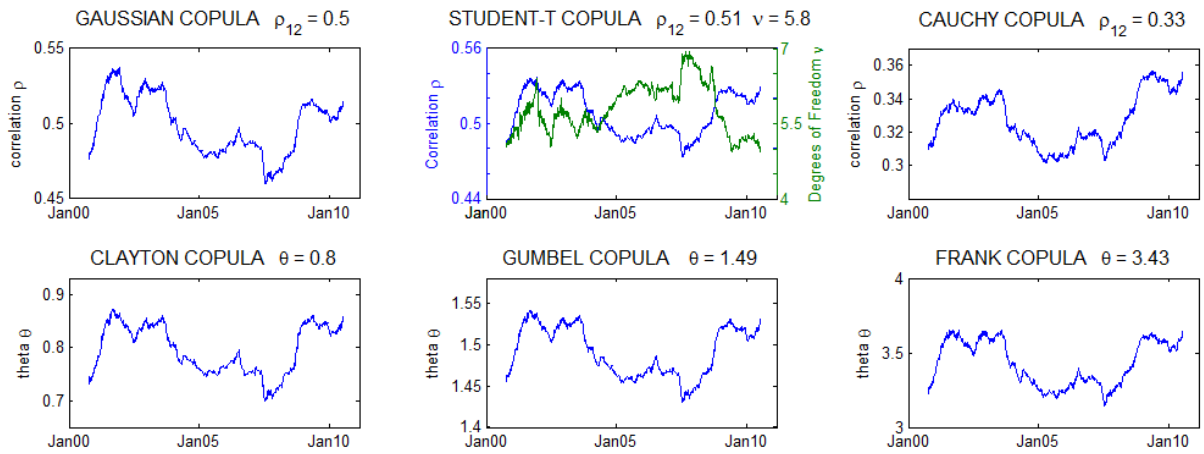

Figure A8.4: Data of the country-pair Argentina Brazil: Estimated parameter values for the six different copula specifications from the rolling in-sample window of 2000 observations over the entire out-of-sample period from 6-10-2000 to 26-7-2010.