Semiparametric estimation of the tail of the error distribution in multivariate regression

By Mervyn J Silvapulle, Gunky Kim, and Paramsothy Silvapulle

Department of Econometrics and Business Statistics, Monash University, P.O. Box 197, Caulfield East, Melbourne, Australia 3145.

mervyn.silvapulle@buseco.monash.edu.au, gunky.kim@buseco.monash.edu.au, param.silvapulle@buseco.monash.edu.au

Abstract

This paper develops a new semiparametric method for modelling multivariate data, particularly when it is desired to assign different levels of importance to various regions in the domain of the joint distribution. The main idea is to estimate the unknown parameter by maximizing a Cramer-von-Mises type distance between a particular type of empirical distribution and the true distribution, which is specified in a semiparametric way that allows a flexible form for the multivariate distribution. The Cramer-von-Mises function incorporates a weight function to assign differential weights to different parts of the domain of the distribution. The main focus of this paper is estimating the lower tail of the joint distribution accurately. In a simulation study, the performance of the new semiparametric method is found to be better than that of the inference function method. The proposed estimation method is used to estimate the joint distribution of DM-USD and Yen USD exchange rates.

JEL Classifications: C13, C14 and C32

Some key words: Copula; minimum distance; Cramer - von Mises; pseudolikelihood; GARCH; tail probability

1 INTRODUCTION

The class of semiparametric methods continue to attract considerable interest as a suitable compromise between fully parametric and fully nonparametric methods for modeling multivariate data. The former, particularly that based on multivariate normal distribution, is usually very restrictive because it is capable of representing only a limited range of shapes, but the method is usually easier to implement. By contrast, the latter, which includes multivariate kernel and spline estimates, is very flexible but usually difficult to implement and may require unrealistically large samples. In this paper, we develop a new semiparametric method for modeling multivariate data when it is desired to assign different levels of importance to different regions in the domain of the distribution. In terms of flexibility and ease of implementation, the method developed in this paper lies between the two extremes of fully parametric and fully nonparametric approaches.

To estimate the unknown parameter in the model, we maximize a Cramèr-von-Mises type distance between a particular type of empirical distribution and the true distribution, which is specified in a semiparametric way allowing a flexible form for the multivariate distribution. Further, we also use a weight function to assign differential weights to different parts of the domain of the distribution. This is an attractive feature of this new method compared to the likelihood based ones and those based on estimating equations.

To provide a brief overview of the type of situations that we have in mind, let us consider an example. Let Y_1 and Y_2 denote the returns of two shares, say a bank and a mining company respectively. Let x denote the return on a market index such as the Dow Jones Index. Suppose that $Y_1 = \mathbf{x}_1^T\boldsymbol{\beta}_1 + \varepsilon_1$ and $Y_2 = \mathbf{x}_2^T\boldsymbol{\beta}_2 + \varepsilon_2$, where $\mathbf{x}_1 = \mathbf{x}_2 = (1, x)^T$. Thus, after accounting for the overall market movements, ε_1 and ε_2 represent the risks that are not under the control of the investor. In this setting, examples of quantities that are of interest include the following, where a_1 and a_2 are given numbers: (i) The probability that the returns from the investments would fall below certain specified levels for a given value of x, for example, $\text{pr}(Y_1 \leq a_1 \text{ and } Y_2 \leq a_2 \mid \mathbf{x})$; (ii) The probability that the return from one investment falls below a specified level given that the other has already fallen below a certain level, for example, $\text{pr}(Y_1 \leq a_1 \mid Y_2 \leq a_2, \mathbf{x})$, and (iii) the quantile c, defined by $\text{pr}\{b_1Y_1 + (1-b_1)Y_2 \leq c \mid \mathbf{x}\} = \alpha$, where α is a given number, for example $\alpha = 0.05$ or 0.4. In quantitative finance, the quantity c in the last example is usually referred to as the *Value at Risk* of a portfolio in which b_1 and $(1-b_1)$ are the proportions of investment in the first and second assets respectively.

There are similar quantities of interest in other contexts, where one is essentially interested in estimating the joint distribution of ε_1 and ε_2 in the region where both of them are negative. In risk management where the variable of interest is the return from an investment, investors and hence portfolio managers pay more attention to the lower tail, and quantities such as *semi-variance*, which is essentially the variance of the negative values. Note that in contrast to the traditional objective of regression modeling, wherein the main interest is the regression parameter with the error distribution being a nuisance parameter/function, their roles are reversed in the foregoing example. In this type of inferential settings, wherein the error distribution itself is the object of interest, we cannot afford to assume that the error distribution is restricted to be multivariate normal. The quantities of particular interest are functions of the joint distribution of $(\varepsilon_1, \varepsilon_2)$.

Thus, the methodological part of the problem requires to estimate the joint distribution of $(\varepsilon_1, \varepsilon_2)$ using a procedure that assigns more importance to its joint negative tail without imposing strong parametric forms for the joint distribution of $\boldsymbol{\varepsilon} := (\varepsilon_1, \varepsilon_2)$. This paper develops a new flexible method to suit this objective. To this end, we specify the joint distribution of $\boldsymbol{\varepsilon}$ using copulas because of their attractive features in this type of settings (see (Joe, 1997) and (Cherubini *et al.*, 2004)). It would be helpful to briefly recall some of these relevant results.

The joint cumulative distribution function $H(\mathbf{x})$ of a random vector $\mathbf{X}$ with continuous marginals $F_j(x_j) = \mathrm{pr}(X_j \leq x_j)$ has the unique representation $H(\mathbf{x}) = C\{F_1(x_1), \ldots, F_k(x_k)\}$, where $C(\mathbf{u})$ is the joint cumulative distribution of $\mathbf{U} := (U_1, \ldots, U_k)$ and U_j is the random variable $F_j(X_j)$ which is distributed uniformly on $[0, 1]$, $j = 1, \ldots, k$ (Sklar (1959)). The function C is called the *copula* of $\mathbf{X}$. This suggests that the joint distribution of $\mathbf{X}$ can be specified in two convenient stages independently: first, specify distributions $F_1, \ldots, F_k$ for the margins, and in the second stage specify a joint distribution C on the unit cube. Two of the reasons for increased interest on copulas include the flexibility they offer because they can represent practically any shape for the joint distribution, and their ability to separate the intrinsic measures of association between the components of $\mathbf{X}$ and the functional forms of the marginal distributions.

Usually the copula belongs to a parametric family, and in fact, the marginal distributions themselves may be from parametric families. In this case, the joint cumulative distribution $H(\mathbf{x})$ of $\mathbf{X}$ takes the form $H(\mathbf{x}; \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k, \boldsymbol{\theta}) = C\{F_1(x_1; \boldsymbol{\alpha}_1), \ldots, F_k(x_k; \boldsymbol{\alpha}_k); \boldsymbol{\theta}\}$, and $\boldsymbol{\theta}$ is called the dependence parameter or association parameter.

This helps to separate the marginal parameters from the intrinsic association which is captured by $\boldsymbol{\theta}$. An attractive feature of this approach is that the copula C and the association parameter $\boldsymbol{\theta}$ are invariant under continuous and monotonically increasing transformations of the marginal variables. Hence copulas have an advantage when the interest centers on intrinsic association among the marginals (Wang and Ding (2000), Oakes and Wang (2003)).

Now, we provide a brief outline of the approach developed in this paper. The estimation is carried out in two stages. In the first stage, we estimate the regression parameters for each margin separately. Let $\tilde{\varepsilon}_{1j}, \ldots, \tilde{\varepsilon}_{nj}$, denote the residuals for the jth component and let $\tilde{\boldsymbol{\varepsilon}}_i = (\tilde{\varepsilon}_{i1}, \ldots, \tilde{\varepsilon}_{ik})^T$, $i = 1, \ldots, n$. These are then used to construct an *empirical copula*, $\tilde{\mathbb{C}}_n(\mathbf{u})$, to estimate the true unknown copula in a way that has some resemblance to the traditional empirical distribution function. Now, we estimate $\boldsymbol{\theta}$ by $\tilde{\boldsymbol{\theta}}$ defined by

$$\tilde{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} J_n(\boldsymbol{\theta}) \quad \text{where } J_n(\boldsymbol{\theta}) = \int \{\tilde{\mathbb{C}}_n(\mathbf{u}) - C(\mathbf{u};\boldsymbol{\theta})\}^2 W(\mathbf{u}, \tilde{\boldsymbol{\gamma}})\, d\mathbf{u}, \tag{1}$$

$W(\mathbf{u}, \tilde{\boldsymbol{\gamma}})$ is some weight function that enables us to assign different levels of importance to different parts in the domain of $C(\mathbf{u};\boldsymbol{\theta})$, and $\tilde{\boldsymbol{\gamma}}$ is a statistic. Thus, the weight function may depend on the data, for example, $W(\mathbf{u}, \tilde{\boldsymbol{\gamma}}) = I(u_1 \leq \hat{\gamma}_1, u_2 \leq \hat{\gamma}_2)$, where $(\gamma_1, \gamma_2) = (F_1(0), F_2(0))$ so that a positive weight is assigned only to the region where ε_1 and ε_2 are estimated to be negative.

Let us note that the existing results for estimation using Cramèr-von-Mises distance between the empirical distribution and a member of the family of parametric distributions, are not applicable to our approach based on (1) because $\tilde{\mathbb{C}}_n(\mathbf{u})$ is not the usual empirical distribution function, but the empirical copula based on estimated parameters and functions. Even for independently and identically distributed observations, the empirical copula is not a sum of independently and identically distributed random variables. Consequently, we encounter some mathematical challenges.

In the foregoing example, we considered a simple form of the linear model. In financial risk management, usually high-frequency financial time series data are used. For these data, GARCH-type models are used frequently in practice. In fact, this class of models have become the work-horse of financial econometrics. In the main body of this paper, we consider linear and nonlinear GARCH models.

We show that the resulting estimator $\tilde{\boldsymbol{\theta}}$ is asymptotically normal and obtain an estimator for its asymptotic covariance matrix. In the asymptotic results, the efficiency of the estimators of $\boldsymbol{\beta}$ does not affect the efficiency of $\tilde{\boldsymbol{\theta}}$, but does affect the efficiency of estimators of events such as $\mathrm{pr}(\boldsymbol{\epsilon} \leq \mathbf{a})$.

Once $\tilde{\boldsymbol{\theta}}$ has been computed, the joint distribution of $\boldsymbol{\varepsilon}$ would be estimated by $C(\tilde{\mathbb{F}}_n(\boldsymbol{\varepsilon}); \tilde{\boldsymbol{\theta}})$, where $\tilde{\mathbb{F}}_n(\mathbf{t}) = \big(\tilde{F}_{n1}(t_1), \ldots, \tilde{F}_{nk}(t_k)\big)$ and $\tilde{F}_{nj}(t_j) = n^{-1}\Sigma_i I(\tilde{\varepsilon}_{ij} \leq t_j)$, the empirical distribution of the residuals obtained in the first stage of estimation. This estimated joint distribution of the error term is a fundamental quantity of interest for statistical inference. If in fact the true copula $C_0(\mathbf{u})$ of $\boldsymbol{\varepsilon}$ is not a member of the family $C(\mathbf{u}; \boldsymbol{\theta})$ then $\tilde{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}_0$, the point at which $\int \{C_0(\mathbf{u}) - C(\mathbf{u}; \boldsymbol{\theta})\}^2 W(\mathbf{u}, \boldsymbol{\gamma}_0)\, d\mathbf{u}$ reaches its minimum. Thus, even if our choice of the parametric form $C(\mathbf{u}; \boldsymbol{\theta})$ is incorrect, the distribution being estimated is still the 'best' in terms of being 'closest' to the true copula with respect to the distance $d(C_1, C_2) = \int \{C_1(\mathbf{u}) - C_2(\mathbf{u})\}^2 W(\mathbf{u}, \boldsymbol{\gamma}_0)\, d\mathbf{u}$.

Thus, for the purpose of estimating probabilities of various events of interest, for example, as in the investment example discussed earlier, a method based on a distance measure, such as that proposed in this paper, is intuitively appealing. Even though we allow the marginal distributions to be unknown and a nonparametric method is adopted for this aspect, our estimation method does not require complicated smoothing methods such as a kernel or spline estimates because the estimation method requires only an estimate of the cumulative distribution function for which we use an empirical distribution function.

The plan of the paper is as follows: Section 2 states the main results and an indication of the main arguments for the proofs. Section 3 provides discussion of a simulation study, and section 4 discusses an example. Section 5 concludes.

2 The Main Results

Since the results for multivariate nonlinear time-series models are significantly more complicated, we shall first provide the results for the linear regression model in detail. Then we shall extend to the more general case.

2.1 The linear regression model

We assume that the true model for $\mathbf{Y} := (Y_1, \ldots, Y_k)$ is $Y_j = \mathbf{x}_j^T \boldsymbol{\beta}_j + \varepsilon_j$, where $\mathbf{x}_j$ is a nonstochastic vector of covariates associated with Y_j, the jth component of $\mathbf{Y}$. Although we assume that $\mathbf{x}_j$ is nonstochastic, most of the results would hold when $\mathbf{x}_j$ is stochastic, with appropriate modifications. Suppose that there are n independent observations $(Y_{i1}, \mathbf{x}_{i1}, \ldots, Y_{ik}, \mathbf{x}_{ik})$, $(i = 1, \ldots, n)$. Thus, we have $Y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}_j + \varepsilon_{ij}$, $(i = 1, \ldots, n, j = 1, \ldots, k)$. Let the probability density and cumulative distribution functions of ε_j be denoted by f_j and F_j respectively. Let $C_0(\mathbf{u})$ denote the true copula of $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_k)^T$ where $\mathbf{u} = (u_1, \ldots, u_k)^T$. Let $C(\mathbf{u}; \theta)$ and $c(\mathbf{u}; \theta)$ denote the assumed parametric form of the copula of $\boldsymbol{\varepsilon}$ and the corresponding density function, respectively. For the time being we shall assume that θ is scalar for simplicity. The results for the case when it is a vector will be presented later; this extension is quite straight forward except for complicated notation. Let $\mathbf{F} = (F_1, \ldots, F_k)^T$ and $\mathbf{F}(\boldsymbol{\varepsilon}) = (F_1(\varepsilon_1), \ldots, F_k(\varepsilon_k))^T$. Let the empirical distribution functions for the univariate and multivariate cases be defined as $F_{nj}(t) = (n+1)^{-1} \sum_i I(\varepsilon_{ij} \leq t)$ and $\mathbb{F}_n(\boldsymbol{\varepsilon}) = (n+1)^{-1} \sum I(\boldsymbol{\varepsilon}_i \leq \boldsymbol{\varepsilon})$ respectively. Now, the *empirical copula* $\mathbb{C}_n$ is defined by

$$\mathbb{C}_n(\mathbf{u}) = \mathbb{F}_n(F_{n1}^{-1}(u_1), \ldots, F_{nk}^{-1}(u_k)), \tag{2}$$

where $F_{nj}^{-1}(u_j)$ is defined as $\inf\{t : F_{nj}(t) \geq u_j\}$, for example, see Fermanian *et al.* (2004) or Tsukahara (2005).

The main parametric methods of estimating θ are maxmimum likelihood[ML] and *inference function for margins*[IFM] (for example, see Joe (1997)). These estimators are asymptotically normal under some standard regularity conditions which include that the marginal distributions be correctly specified. For the case when the copula is correctly specified, Genest *et al.* (1995a) introduced a more appealing method (see also Oakes (1994)).

Now, we introduce the following semiparametric estimator of the copula parameter θ, when the marginal distributions are unknown: (a) Let $\tilde{\boldsymbol{\beta}}_j$ be an estimator of $\boldsymbol{\beta}_j$ such that $n^{1/2}(\tilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j) = O_p(1)$. (b) Compute the residuals $\tilde{\varepsilon}_{ij} = y_{ij} - \mathbf{x}_{ij}^T \tilde{\boldsymbol{\beta}}_j$, for $i = 1, \ldots, n$. (c) Estimate $F_j(t)$ by $\tilde{F}_{nj}(t)$, the empirical distribution of the residuals defined by $\tilde{F}_{nj}(t) = (n+1)^{-1} \Sigma_{i=1}^n I(\tilde{\varepsilon}_{ij} \leq t)$; thus $\tilde{F}_{nj}$ is the empirical distribution of $\{\tilde{\varepsilon}_{1j}, \ldots, \tilde{\varepsilon}_{nj}\}$. (d) Compute the copula by $\tilde{\mathbb{C}}_n$, the *empirical copula for residuals* defined by

$$\tilde{\mathbb{C}}_n(\mathbf{u}) = \tilde{\mathbb{F}}_n(\tilde{F}_{n1}^{-1}(u_1), \ldots, \tilde{F}_{nk}^{-1}(u_k)) \tag{3}$$

where $\tilde{\mathbb{F}}_n(\boldsymbol{\varepsilon}) = (n+1)^{-1} \sum I(\tilde{\boldsymbol{\varepsilon}}_i \leq \boldsymbol{\varepsilon})$. (e) Estimate θ by $\tilde{\theta}$, defined by $\tilde{\theta} = \text{argmin}_\theta J_n(\theta)$ where

$$J_n(\theta) = \int \{\tilde{\mathbb{C}}_n(\mathbf{u}) - C(\mathbf{u}; \theta)\}^2 W(\mathbf{u}, \tilde{\boldsymbol{\gamma}}) \, d\mathbf{u} \tag{4}$$

and $W(\mathbf{u}, \tilde{\boldsymbol{\gamma}})$ is a given nonnegative weight function. This two-stage procedure is similar to the Pseudo-likelihood approach of Genest *et al.* (1995b) which was also indicated in broad terms by Oakes (1994), both for the case when $(Y_{i1}, \ldots, Y_{ik})$ are independent and identically distributed for $i = 1, \ldots, n$. It turns out that $\tilde{\mathbb{C}}_n$ converges in probability to the true copula $C_0(\mathbf{u})$. Consequently, $J_n(\theta)$ also converges in probability, uniformly in $\boldsymbol{\theta}$, to $J_0(\theta)$ defined by

$$J_0(\theta) = \int \{C_0(\mathbf{u}) - C(\mathbf{u}; \theta)\}^2 W(\mathbf{u}, \boldsymbol{\gamma}_0) \, d\mathbf{u} \tag{5}$$

where $\boldsymbol{\gamma}_0$ is the probability limit of $\tilde{\boldsymbol{\gamma}}$. Therefore, $\tilde{\theta}$ is likely to be a reasonable estimator of θ_0, which we define as the point at which $J_0(\theta)$ reaches its minimum. If $C_0(\cdot)$ is a member of the family $\{C(\cdot; \theta) : \theta \in \Theta\}$ then θ_0 is the true value.

We will show that $\tilde{\theta}$ is consistent and asymptotically normal, and obtain a closed form expression for the asymptotic variance. By substituting sample estimates for the asymptotic variance formulae, we shall obtain a consistent estimator of the large sample variance of $\tilde{\boldsymbol{\theta}}$.

Now, let us indicate the essentials for obtaining the asymptotic distribution of $\sqrt{n}(\tilde{\theta} - \theta_0)$. Under quite general conditions, $\tilde{\theta}$ is a consistent estimator of θ_0. By a one-term Taylor expansion, we have

$$\sqrt{n}(\tilde{\theta} - \theta_0) = \{\ddot{J}_n(\bar{\theta})\}^{-1} \sqrt{n} \dot{J}_n(\theta_0) \tag{6}$$

where $\dot{J}$ and $\ddot{J}$ are the first and second derivatives of J_n, and $\bar{\theta} \in (\theta_0, \tilde{\theta})$. Convergence of $\{\tilde{\mathbb{C}}_n, \tilde{\gamma}, \bar{\theta}\}$ in probability ensures that $\ddot{J}_n(\bar{\theta}) \xrightarrow{p} \ddot{J}_0(\theta_0)$. Therefore, it remains to obtain the asymptotic distribution of $\sqrt{n}\dot{J}_n(\theta_0)$.

Note that $\tilde{\mathbb{C}}_n(\mathbf{u})$ in (3) is different from the $\mathbb{C}_n(\mathbf{u})$ in (2), the former is the empirical copula of the residuals $\tilde{\boldsymbol{\varepsilon}}_1, \ldots, \tilde{\boldsymbol{\varepsilon}}_n$. Since residuals are expected to be close to the true errors, we would expect $\tilde{\mathbb{C}}_n(\cdot)$ and $\mathbb{C}_n(\cdot)$ to be close. It turns out, as stated in the next lemma, that they are very close, and this plays a crucial role in the derivations.

To indicate the main steps in deriving the asymptotic distribution of $\sqrt{n}(\tilde{\theta} - \theta_0)$, we first assume that $\boldsymbol{\gamma}$ is a scalar and rewrite (6) as

$$\sqrt{n}(\tilde{\theta} - \theta_0) = P1 + P2 + P3 + o_p(1) \tag{7}$$

where

$$P1 = \{\ddot{J}_n(\bar{\theta})\}^{-1}\sqrt{n}\int\{\tilde{\mathbb{C}}_n(\mathbf{u}) - \mathbb{C}_n(\mathbf{u})\}\dot{C}(\mathbf{u};\theta_0)W(\mathbf{u},\tilde{\gamma})\ d\mathbf{u} \tag{8}$$

$$P2 = \{\ddot{J}_n(\bar{\theta})\}^{-1}\sqrt{n}\int\{\mathbb{C}_n(\mathbf{u}) - C(\mathbf{u};\theta_0)\}\dot{C}(\mathbf{u};\theta_0)W(\mathbf{u},\gamma_0)\ d\mathbf{u} \tag{9}$$

$$P3 = \{\ddot{J}_n(\bar{\theta})\}^{-1}A_n\sqrt{n}(\tilde{\gamma} - \gamma_0) \tag{10}$$

$$A_n = \int\{\mathbb{C}_n(\mathbf{u}) - C(\mathbf{u};\theta_0)\}\dot{C}(\mathbf{u};\theta_0)\dot{W}(\mathbf{u},\gamma_0)\ d\mathbf{u}, \tag{11}$$

where $\dot{W}(\mathbf{u},\gamma) = (\partial/\partial\gamma)W(\mathbf{u},\gamma)$.

Lemma 1.

$$\sqrt{n}\int\{\tilde{\mathbb{C}}_n(\mathbf{u}) - \mathbb{C}_n(\mathbf{u})\}\dot{C}(\mathbf{u};\theta)W(\mathbf{u},\tilde{\gamma})\ d\mathbf{u} = o_p(1), \quad \theta \in \Theta. \tag{12}$$

Since $\mathbb{C}_n$ is the empirical copula process based on independently and identically distributed random variables, we can use existing results for empirical copula, for example see Tsukahara (2005), to write

$$P2 = n^{-1/2}\sum\zeta_i + o_p(1) \tag{13}$$

where $\zeta_1, \ldots, \zeta_n$ are *iid* and ζ_i is a function of $\boldsymbol{\varepsilon}_i$; explicit expressions for ζ_i are given below. We make the mild assumption that the preliminary statistic $\tilde{\gamma}$ used in the weight function can be expressed as

$$\sqrt{n}(\tilde{\gamma} - \gamma_0) = n^{-1/2}\sum\xi_i^* + o_p(1) \tag{14}$$

for some independent random variables $\xi_1^*, \ldots, \xi_n^*$ where ξ_i^* is a function of $\boldsymbol{\varepsilon}_i$. Therefore, we have

$$P3 = n^{-1/2}\sum\xi_i + o_p(1) \tag{15}$$

where $\xi_i = \ddot{J}_0(\boldsymbol{\theta}_0)^{-1}A\xi_i^*$ and $A = plimA_n$. Now, combining (7)-(15), we have

$$\sqrt{n}(\tilde{\theta} - \theta_0) = n^{-1/2}\sum(\xi_i + \zeta_i) + o_p(1). \tag{16}$$

Therefore, $\sqrt{n}(\tilde{\theta} - \theta_0)$ converges to a normal distribution. These heuristic arguments hold when θ and γ are vectors as well. The relevant results are stated below briefly.

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^T$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^T$. Now, $P1, P2$ and $P3$ in (9, 10, 11), $\dot{C}$, $\dot{J}_n$ are vectors, and A in (11) and $\ddot{J}_n$ are matrices. Let $\sqrt{n}(\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = n^{-1/2}\sum \boldsymbol{\xi}_i^* + o_p(1)$ and $\boldsymbol{\xi}_i = \ddot{J}_0(\boldsymbol{\theta}_0)^{-1}A\boldsymbol{\xi}_i^*$. Let

$$\boldsymbol{\zeta}_i(\boldsymbol{\theta}) = \int \mathbf{b}(\mathbf{u}, \boldsymbol{\theta})\{d_{1i}(\mathbf{u}, \boldsymbol{\theta}) - d_{2i}(\mathbf{u}, \boldsymbol{\theta})\}\ d\mathbf{u}, \tag{17}$$

where

$$d_{1i}(\mathbf{u}, \boldsymbol{\theta}) = I(\mathbb{F}(\boldsymbol{\varepsilon}_i) \leq \mathbf{u}) - C(\mathbf{u}; \theta) \tag{18}$$

$$d_{2i}(\mathbf{u}, \boldsymbol{\theta}) = \sum_{p=1}^{k}(\partial/\partial u_p)C(\mathbf{u}; \theta)\{I(F(\varepsilon_{ij}) \leq u_p) - u_p\} \tag{19}$$

$$\mathbf{b}(\mathbf{u}, \boldsymbol{\theta}) = \mathcal{K}(\boldsymbol{\theta})^{-1}\dot{C}(\mathbf{u}, \boldsymbol{\theta}) \tag{20}$$

$$\mathcal{K}(\boldsymbol{\theta}) = \int \dot{C}(\mathbf{u}, \boldsymbol{\theta})\dot{C}(\mathbf{u}, \boldsymbol{\theta})^T W(\mathbf{u}, \boldsymbol{\gamma}_0)\ d\mathbf{u}. \tag{21}$$

This $\boldsymbol{\zeta}_i$ corresponds to the the ζ_i in (13). Next, we state the main result.

Proposition 1. *Under some regularity conditions, we have* $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, V)$, *where* $V = \lim n^{-1}\sum var(\boldsymbol{\xi}_i + \boldsymbol{\zeta}_i)$.

For statistical inference on $\boldsymbol{\theta}$, we also need to be able to estimate the asymptotic covariance matrix V. In most practical situations, this can be done in a straight forward way. First, note that if the weight function W in the definition of J_n does not depend on a preliminary statistic, then $\boldsymbol{\xi}_i = \mathbf{0}$ and hence $V = \text{var}(\boldsymbol{\zeta})$. Since $\boldsymbol{\zeta}_i$ is a function of $(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\varepsilon}_i)$, a consistent estimator of V is the sample variance of $\tilde{\boldsymbol{\zeta}}_1, \ldots, \tilde{\boldsymbol{\zeta}}_n$, where $\tilde{\boldsymbol{\zeta}}_i = \boldsymbol{\zeta}_i(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\varepsilon}}_i)$.

A slight modification of this method can be applied when the weight function W depends on $\tilde{\boldsymbol{\gamma}}$. To illustrate this let us assume that $n^{-1/2}(\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = n^{-1/2}\sum D_i\boldsymbol{\tau}_i^* + o_p(1)$ where $\{D_i : i = 1, 2, \ldots\}$ are some matrices of constants and $\{\boldsymbol{\tau}_i^* : i = 1, 2, \ldots\}$ are *iid.* As an example, this condition is satisfied when $\tilde{\boldsymbol{\beta}}$ is the least squares or an M-estimator and the weight function assigns a positive weight only to the region where the residuals are negative. Now, we have $P_3 = n^{-1/2}\sum E_i\boldsymbol{\tau}_i + o_p(1)$ where $\{E_i : i = 1, 2, \ldots\}$ are some matrices of constants and $\{\boldsymbol{\tau}_i : i = 1, 2, \ldots\}$ are *iid*. Therefore, we have

$$V = \mathrm{var}(\boldsymbol{\zeta}) + (n^{-1}\Sigma D_i)\mathrm{var}(\boldsymbol{\tau})(n^{-1}\Sigma D_i)^T + (n^{-1}\Sigma D_i)\mathrm{cov}(\boldsymbol{\zeta}, \boldsymbol{\tau}).$$

Now, $\mathrm{var}(\boldsymbol{\zeta}), \mathrm{var}(\boldsymbol{\tau})$, and $\mathrm{cov}(\boldsymbol{\zeta}, \boldsymbol{\tau})$ can be estimated as above by sample moments and this leads to the consistent estimator

$$\tilde{V} = \tilde{\mathrm{var}}(\boldsymbol{\zeta}) + (n^{-1}\Sigma D_i)\tilde{\mathrm{var}}(\boldsymbol{\tau})(n^{-1}\Sigma D_i)^T + (n^{-1}\Sigma D_i)\tilde{\mathrm{cov}}(\boldsymbol{\zeta}, \boldsymbol{\tau}). \qquad (22)$$

Now, we consider the more general case of nonlinear time series models with heteroscedastic errors.

2.2 Nonlinear time series models with GARCH errors

Let $\{\mathbf{y}_i : i = 0, \pm 1, \pm 2, \ldots\}$ denote a vector time series, where $\mathbf{y}_i = (y_{1i}, \ldots, y_{ki})^T$, and let the data generating process be

$$y_{ij} = \mu_{ij}(\boldsymbol{\alpha}_{1j}) + \sqrt{h_{ij}(\boldsymbol{\alpha}_j)}\varepsilon_{ij}, \qquad (23)$$

where $\{(\varepsilon_{i1}, \ldots, \varepsilon_{ik}) : i = 1, \ldots, n\}$ are n independently and identically distributed random variables, $\boldsymbol{\alpha}_j = (\boldsymbol{\alpha}_{1j}^T, \boldsymbol{\alpha}_{2j}^T)^T$, and μ_{ij} and h_{ij} may depend on past observations and covariates. Here, we shall adopt the notation in Koul and Ling (2006). Throughout, we shall assume that the functional forms of $\mu_{ij}(\boldsymbol{\alpha}_j)$ and $h_{ij}(\boldsymbol{\alpha}_j)$ are known and that they are twice continuously differentiable functions of $\boldsymbol{\alpha}_{1j}$ and $\boldsymbol{\alpha}_j$ respectively. Further, we shall assume that the series $\{y_{ij}\}$ is strictly stationary and ergodic, and that y_{0j} is independent of all previous observations, for every j. Let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_k^T)^T$ and $\boldsymbol{\alpha}^0$ denote the true parameter value of $\boldsymbol{\alpha}$.

Let F_j be the cumulative distribution function of $\{\varepsilon_{ij}\}$ and let f_j denote the corresponding density function. Let $C_0(\mathbf{u})$ and $c_0(\mathbf{u})$ denote the copula of $\boldsymbol{\varepsilon}$ and the corresponding density function, respectively, where $\mathbf{u} = (u_1, \ldots, u_k)^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_k)^T$. Let $\mathbb{F} = (F_1, \ldots, F_k)$, and $\mathbb{F}(\boldsymbol{\varepsilon}) = (F_1(\varepsilon_1), \ldots, F_k(\varepsilon_k))$.

We assume that the time series model (23) can be estimated for each margin separately, and let $\tilde{\boldsymbol{\alpha}}_j$ denote a $\sqrt{n}$−consistent estimator of $\boldsymbol{\alpha}_j^0$, $(j = 1, \ldots, k)$. (see

Koul and Ling (2006)). Now, define the residuals corresponding to $\{\varepsilon_{ij}\}$ by

$$\tilde{\varepsilon}_{ij} = [y_{ij} - \mu_{ij}(\tilde{\boldsymbol{\alpha}}_{j1})]/\sqrt{h_{ij}(\tilde{\boldsymbol{\alpha}}_j)}. \tag{24}$$

Let $\tilde{F}_{nj}(t) = (n+1)^{-1}\Sigma_{i=1}^{n} I(\tilde{\varepsilon}_{ij} \leq t)$ and $\tilde{\mathbb{F}}_n(\boldsymbol{\varepsilon}) = (\tilde{F}_{n1}(\varepsilon_1), \ldots, \tilde{F}_{nk}(\varepsilon_k))^T$; thus, $\tilde{F}_{nj}(t)$ is the *edf* of the residuals $\{\tilde{\varepsilon}_{ij}\}$.

Now we define the minimum distance estimator $\tilde{\boldsymbol{\theta}}$ as the minimizer of $J_n(\boldsymbol{\theta})$ as in the previous subsection except that $\tilde{\mathbb{C}}$ now is the empirical copula of the residuals $\tilde{\boldsymbol{\varepsilon}}_1, \ldots, \tilde{\boldsymbol{\varepsilon}}_n$. Let $\mathbb{C}_n$ denote the empirical copula of the unobserved errors $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$. Now, the heuristic arguments presented in the previous subsection for the regression model are just as applicable to the present time-series models as well. The technical details required to provide a rigorous proof of the asymptotic normality of the estimator of $\boldsymbol{\theta}$ in the next result are considerably more complicated, however.

Proposition 2. *Under some regularity conditions, we have* $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, V)$ *where* V *as in Theorem 1. A consistent estimator of* V *is (22) with* $\tilde{\epsilon}_{ji}$ *in (24).*

This theorem says that the MD method provides an asymptotically normal estimator of the copula parameter, which in turn can be used for estimating various quantities of interest such as the probabilities of joint events. The MD estimator is consistent when the parametric family is correctly specified. To illustrate an application of this estimator, we may think of the random error components as the risky parts that are not in our control. In this context, the risk manager may be interested to estimate the probability that both these fall below their respective 30% quantiles. In other words, we wish to estimate $Pr[\epsilon_{1,T+1} \leq F_1^{-1}(0.3)$ and $\epsilon_{2,T+1} \leq F_2^{-1}(0.3) \mid \mathcal{F}_T]$ where F_1 and F_2 are the distribution functions of $\epsilon_{1,T+1}$ and $\epsilon_{1,T+1}$ respectively. An estimate of this is $C[\tilde{F}_1^{-1}(0.3), \tilde{F}_2^{-1}(0.3); \tilde{\boldsymbol{\theta}}]$ where $\tilde{F}_j$ is the empirical of the residuals of the jth component $(j = 1, 2)$.

Instead of stating the event in terms on the error terms we may wish to state it in terms of the observed variables. Suppose that we wish to estimate $pr[Y_{1,T+1} \leq b_1$ and $Y_{2,T+1} \leq b_2 \mid \mathcal{F}_T]$ for some given b_1 and b_2. This may be estimated by $C(a_1, a_2; \hat{\boldsymbol{\theta}})$ where $a_1 = b_1 - \hat{\mu}_1$ and $a_2 = b_2 - \hat{\mu}_2$. Bootstrap methods maybe used for computing standard errors.

3 Simulation Study

A simulation study was carried out to compare the IFM method with the MD method. Since our purpose of introducing the MD method was broader than estimating the unknown parameter, there is no unique criterion for such a comparison. In this simulation we will compare how well the probability of the lower tail of a bivariate distribution is estimated by this method, particularly when the parametric family of copulas is misspecified.

Design of the study: There are two parts to this simulation study. In the first, we considered bivariate independently and identically distributed observations. In the second, we considered a bivariate GARCH model. These are described below in turn.

To describe the iid setting, let $\mathbf{Y}_t = (Y_{1t}, Y_{2t})^T$ denote a bivariate random variable. Assume that $Y_1, \ldots, Y_n$ are *iid* with *cdf* $C_0(F_1(y_1), F_2(y_2)$, where F_1 and F_2 are the marginal distributions and C_0 is the copula of $\mathbf{Y}$. The following pairs of distributions were studied for (F_1, F_2): (i) N-N: $Y_1, Y_2 \sim N(0, 1)$; (ii) T-T: $Y_1, Y_2 \sim t_r$; (iii) T-ST: $Y_1 \sim t_r$ and $Y_2 \sim$ skew t_r with skewness $= 0.5$; (iv) T-C: $Y_1 \sim t_r$ and $Y_2 \sim \chi^2_5$.

For the copulas, the followings families were studied: Frank, Clayton, Gumbel, Joe, and symmetrized Joe-Clayton. The value of the copula parameter was chosen so that the Kendall's τ is 0.5.

The IFM method assumed that the marginal distributions were normal with unknown means and standard deviations. Therefore, when the marginal distributions are N-N, they are correctly specified for the IFM. In this case, if the copula is also correctly specified then the IFM would be consistent. When the error distribution is T-T, or T-ST, or T-C, the marginal distributions are misspecified for the IFM. Therefore, in these three cases, IFM estimates are unlikely to be consistent.

The objective of the simulation study was to evaluate how well the methods perform for estimating $C_0(a, a)$ for a range of values of a near 0.5; we chose $a = 0.3, 0.4, 0.5, 0.6$. Since the quantity being estimated, namely $C_0(a, a) = \text{pr}[Y_1 \leq F_1^{-1}(a)$ and $Y_2 \leq F_2^{-1}(a)]$, is a probability associated with the lower tail, we would like to fit the data well for the lower tail region. To this end, we chose a weight function that assigns a positive weight to the lower tail, $\{u \leq 0.75 \text{ and } v \leq 0.75\}$, in the domain $[0,1]^2$ of the copula. More specifically we chose $w(u,v) = I\{u \leq 0.75 \text{ and } v \leq 0.75\}[\{C(u,v;\tilde{\theta})+0.01\}\{1.01-C(u,v;\tilde{\theta}\}]^{-1}$ where $\tilde{\theta}$ is a $n^{1/2}$-consistent estimator of θ.

In the second part of the study, we considered the following GARCH model: $Y_{1t} = \sqrt{h_{1t}}\eta_{1t}$ and $Y_{2t} = \sqrt{h_{2t}}\eta_{2t}$ where $h_{1t} = 0.2 + 0.1y_{1,t-1}^2 + 0.15h_{1,t-1}$, $h_{2t} = 0.1 + 0.1y_{1,t-1}^2 + 0.15y_{2,t-2}^2 + 0.1h_{2,t-1}$, and (η_{1t}, η_{2t}) are iid with distribution function $C_0\{F_1(y_1), F_2(y_2)\}$, where F_1, F_2 and C_0 as in the previous part of the design for iid observations. The quantity being estimated is $C_0(a,a) = \text{pr}[\eta_{1,t+1} \leq F_1^{-1}(a)$ and $\eta_{2,t+1} \leq F_2^{-1}(a) \mid \mathcal{F}_t]$ for $a = 0.3, 0.4, 0.5, 0.6$, which corresponds to that in the previous part for the iid setting.

Since the simulations were computing intensive, the MD method was applied only with data generated from the normal distribution for the error terms. We could have evaluated the method with data generated from T-T, T-ST, and T-C. Fortunately, the simulation results to be presented show that this is not necessary.

Our focus is on the realistic setting when the copula and marginal distributions are misspecified. Consequently, the copula parameter itself is not of much interest and every method studied in this paper is inconsistent. The relevant question is, how large is the difference between what would like to estimate and what is being estimated. Therefore, we focus on the asymptotic bias.

Results:

(i). *Marginal distributions and the copula are correctly specified:* These results are in Table 1 under the heading N-N for IFM. The results in these tables under the headings T-T, T-ST and T-C correspond to the cases when the margins are misspecified for IFM since IFM requires teh marginal distributions to be specified by a parametric family, and we chose the normal distribution for this purpose. These results show that the parameters being estimated are equal to the true parameter values.

(ii). *The copula is correctly specified, but the marginal distributions are incorrectly specified for the IFM:* These results are also in Table 1, but under the headings N-N, T-T and T-ST. They indicate that estimates based on the IFM method are inconsistent, but the bias is small.

When the observations are iid, the empirical copula is the same for different marginal distributions. Therefore, the MD and the semiparametric estimates would remain unchanged for different marginal distributions. This is not the case for the GARCH model. It is reassuring to note that Table 1 shows that the estimates for MD and semiparametric methods under GARCH structure are the same as those under iid observations.

(iii). *The copula is incorrectly specified:* These results are given in Tables 2 - 5. Since the copula is incorrectly specified, the 'pseudo likelihood' used by IFM and the semiparametric methods are incorrect irrespective of whether the margins are correctly specified or not. Here the term 'pseudo likelihood' is used loosely. However, the results for different marginal distributions are useful to assess the sensitivity of IFM to the specification of the marginal distributions. For the semiparametric and MD methods, the exact forms of the marginal distributions are not required because they are estimated by the empirical distributions of the residuals. The results show that the MD method performs better than the others. This was precisely what we expected as was motivated in the previous sections.

In addition to estimating the tail probability $C_0(a, a)$ for different values of a, we also estimated the global measures, Kendall's τ and Spearman's ρ. Although estimation of these global measures were not our prime goal, it is encouraging to note that the MD estimates are closer to the true value than the others, in almost all cases.

4 An example

Let us consider an example to indicate the nature of the topic studied in this paper. The following example is a slightly modified version of that studied by Patton (2006). Some of the following discussions are similar to those for the illustrative example in Kim *et al.* (2008).

Let $E_{1,t}$ and $E_{2,t}$ denote the DM-USDollar and Yen-USDollar exchange rates respectively. Let $Y_{1t} = logE_{1,t} - logE_{1,t-1}$ and $Y_{2t} = logE_{2,t} - logE_{2,t-1}$. Thus, Y_{1t} and Y_{2t} can be seen as measures of returns of the two investments. Consider the model

$$
\begin{aligned}
Y_{1t} &= \mu_1 + \epsilon_{1t}, \qquad \epsilon_{1t} = \sqrt{h_{1t}}\eta_{1t}, \quad h_{1t} = \alpha_{11} + \alpha_{12}\epsilon_{1,t-1}^2 + \alpha_{13}h_{1,t-1}, \\
Y_{2t} &= \mu_2 + \epsilon_{2t}, \qquad \epsilon_{2t} = \sqrt{h_{2t}}\eta_{2t}, \quad h_{2t} = \alpha_{21} + \alpha_{22}\epsilon_{2,t-1}^2 + \alpha_{23}h_{2,t-1},
\end{aligned} \tag{25}
$$

where $\{\eta_{1t}, \eta_{2t}\}$ are *iid*. Let $C_0(u_1, u_2)$ denote its copula.

We use the data for the period Jan 1991 - Jan 1999, a period prior to the introduction of the Euro. The total number of observations is 2046. Let us suppose that we are interested to estimate probabilities of events such as $P(y_{1n} \leq log(t), y_{2n} \leq log(t))$, and $P(\eta_{1n} \leq t, \eta_{2n} \leq t)$. Since these events correspond to the lower tail, we use the same weight function as that in the simulation study.

We estimated several copulas and evaluated their goodness of fit. The Gaussian copula family appears to fit well. The copulas estimated by the semiparametric and minimum distance methods are $C(\tilde{F}_{1n}(\tilde{\eta}_1), \tilde{F}_{2n}(\tilde{\eta}_2); 0.53)$ and $C(\tilde{F}_{1n}(\tilde{\eta}_1), \tilde{F}_{2n}(\tilde{\eta}_2); 0.55)$ respectively, where $\tilde{F}_{1n}$ and $\tilde{F}_{2n}$ are the empirical distribution functions of the residuals $\{\tilde{\eta}_{1t}\}$ and $\{\tilde{\eta}_{2t}\}$ respectively, and $C(u_1, u_2; \theta)$ is the Gaussian copula. In this particular instance, the semiparametric and MD estimates are close, and hence any estimates based on the methods would also be close. However, in general, this would not be case.

While the SP estimator converges to a point, say $\boldsymbol{\theta}_a$, it is difficult to interpret exactly what this point represent. By contrast, we know that the MD estimator also converges to $\boldsymbol{\theta}_0$ and this is the point such that the true copula C_0 is closest to $C(\cdot; \boldsymbol{\theta}_0)$ in terms of the distance J_0 defined in (5). Therefore, intuitively the MD results are more appealing to the context of the present example. It reassuring that the SP and the MD methods led to results that are close in this particular case. However, in general, they could be very different.

References

Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance.* John Wiley and Sons Ltd, Chichester, U.K.

Fermanian, J.-D., Radulovic, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli,* 10(5), 847–860.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995a). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82, 543–552.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995b). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82, 543–552.

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Chapman and Hall, London.

Kim, G., Silvapulle, M. J., and Silvapulle, P. (2008). Estimating the error distribution in multivariate heteroscedastic time series models. *Journal of Statistical Planning and Inference*, 138, 1442–1458.

Koul, H. L. and Ling, S. (2006). Fitting an error distribution in some heteroscedastic time series model. *Annals of Statistics*, 34, 994–1012.

Oakes, D. (1994). Multivariate survival distributions. *Journal of Nonparametric Statistics*, 3, 343–354.

Oakes, D. and Wang, A. (2003). Copula model generated by Dabrowska's association measure. *Biometrika*, 90(2), 478–481.

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

Sklar, A. (1959). Fonctions de répartition à n dimensionset leurs marges. *Publ. Inst, Statis. Univ. Paris*, 8, 229–231.

Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canad. J. Statist.*, 33, 357–375.

Wang, W. and Ding, A. A. (2000). On assessing the association for bivariate current status data. *Biometrika*, 87, 879–893.

Table 1: Comparison of minimum distance, semiparametric and IFM methods with a correctly specified copula.

Parameter	Population Value	IFM (N-N)	IFM (T-T)	IFM (T-ST)	IFM (T-C)	SP	MD
			True and fitted copula:Frank; iid observations				
$C_0(0.3, 0.3)$	0.1967	0.1967	0.2004	0.1990	0.1968	0.1967	0.1966
$C_0(0.4, 0.4)$	0.2906	0.2905	0.2949	0.2933	0.2908	0.2906	0.2905
$C_0(0.5, 0.5)$	0.3889	0.3887	0.3933	0.3916	0.3890	0.3888	0.3888
$C_0(0.6, 0.6)$	0.4906	0.4905	0.4949	0.4933	0.4908	0.4906	0.4905
τ	0.4999	0.4998	0.5169	0.5105	0.5008	0.5000	0.4999
ρ	0.6945	0.6945	0.7139	0.7067	0.6955	0.6946	0.6945
			True and fitted copula: Frank; GARCH margins				
$C_0(0.3, 0.3)$	0.1967	0.1966	0.2004	0.1989	0.1967	0.1966	0.1967
$C_0(0.4, 0.4)$	0.2906	0.2905	0.2948	0.2932	0.2907	0.2905	0.2906
$C_0(0.5, 0.5)$	0.3889	0.3887	0.3932	0.3915	0.3889	0.3887	0.3888
$C_0(0.6, 0.6)$	0.4906	0.4905	0.4948	0.4932	0.4907	0.4905	0.4906
τ	0.4999	0.4997	0.5167	0.5102	0.5003	0.4996	0.4999
ρ	0.6945	0.6943	0.7136	0.7062	0.6950	0.6942	0.6946

Table 2: Comparison of minimum distance, semiparametric and IFM methods with incorrectly specified copula when the observations are iid.

Parameter	Population Value	IFM (N-N)	IFM (T-T)	IFM (T-ST)	IFM (T-C)	SP	MD
		True copula:Frank; fitted copula family: Clayton					
$C_0(0.3, 0.3)$	0.1967	0.1805	0.1716	0.1902	0.1926	0.1807	0.2079
$C_0(0.4, 0.4)$	0.2906	0.2544	0.2448	0.2648	0.2675	0.2545	0.2844
$C_0(0.5, 0.5)$	0.3889	0.3376	0.3283	0.3478	0.3504	0.3377	0.3673
$C_0(0.6, 0.6)$	0.4906	0.4322	0.4242	0.4412	0.4435	0.4324	0.4587
τ	0.5003	0.3494	0.3140	0.3881	0.3979	0.3500	0.4608
ρ	0.6952	0.4995	0.4528	0.5491	0.5613	0.5003	0.6375
		True copula:Gumbel; fitted copula family: Clayton					
$C_0(0.3, 0.3)$	0.1822	0.1790	0.1719	0.1890	0.1909	0.1792	0.1959
$C_0(0.4, 0.4)$	0.2737	0.2527	0.2451	0.2635	0.2656	0.2530	0.2712
$C_0(0.5, 0.5)$	0.3752	0.3360	0.3287	0.3465	0.3486	0.3362	0.3540
$C_0(0.6, 0.6)$	0.4856	0.4309	0.4245	0.4401	0.4419	0.4311	0.4468
τ	0.4995	0.3433	0.3154	0.3832	0.3910	0.3443	0.4115
ρ	0.6816	0.4916	0.4546	0.5429	0.5527	0.4928	0.5782

Table 3: Comparison of minimum distance, semiparametric and IFM methods with incorrectly specified copula when the observations are iid.

Parameter	Population Value	IFM (N-N)	IFM (T-T)	IFM (T-ST)	IFM (T-C)	SP	MD
			True copula: Joe; fitted copula family: Clayton				
$C_0(0.3, 0.3)$	0.1679	0.1651	0.1387	0.1796	0.1895	0.1654	0.1830
$C_0(0.4, 0.4)$	0.2677	0.2379	0.2104	0.2534	0.2641	0.2382	0.2571
$C_0(0.5, 0.5)$	0.3786	0.3217	0.2963	0.3366	0.3470	0.3220	0.3402
$C_0(0.6, 0.6)$	0.4969	0.4186	0.3976	0.4314	0.4405	0.4188	0.4345
τ	0.4998	0.2887	0.1870	0.3459	0.3852	0.2897	0.3594
ρ	0.6798	0.4187	0.2730	0.4949	0.5454	0.4200	0.5124
			True copula: Gumbel; fitted copula family: Frank				
$C_0(0.3, 0.3)$	0.1822	0.1966	0.2018	0.2018	0.2015	0.1966	0.1872
$C_0(0.4, 0.3)$	0.2737	0.2905	0.2964	0.2965	0.2961	0.2905	0.2796
$C_0(0.5, 0.5)$	0.3752	0.3887	0.3949	0.3950	0.3945	0.3887	0.3773
$C_0(0.6, 0.6)$	0.4856	0.4905	0.4964	0.4965	0.4961	0.4905	0.4796
τ	0.4995	0.4996	0.5230	0.5233	0.5216	0.4997	0.4573
ρ	0.6816	0.6942	0.7206	0.7210	0.7191	0.6944	0.6441

Table 4: Comparison of minimum distance, semiparametric and IFM methods with incorrectly specified copula when the margins follow GARCH models.

Parameter	Population Value	IFM (N-N)	IFM (T-T)	IFM (T-ST)	IFM (T-C)	SP	MD
	True copula: Frank; fitted copula family: Clayton						
$C_0(0.3, 0.3)$	0.1967	0.1801	0.1712	0.1894	0.1921	0.1805	0.2076
$C_0(0.4, 0.4)$	0.2906	0.2539	0.2443	0.2640	0.2669	0.2544	0.2841
$C_0(0.5, 0.5)$	0.3889	0.3372	0.3279	0.3470	0.3498	0.3376	0.3670
$C_0(0.6, 0.6)$	0.4906	0.4319	0.4239	0.4405	0.4430	0.4323	0.4584
τ	0.4999	0.3479	0.3124	0.3852	0.3959	0.3495	0.4596
ρ	0.6945	0.4975	0.4507	0.5453	0.5588	0.4996	0.6361
	True copula: Gumbel; fitted copula family: Clayton						
$C_0(0.3, 0.3)$	0.1822	0.1788	0.1717	0.1886	0.1908	0.1791	0.1960
$C_0(0.4, 0.4)$	0.2737	0.2525	0.2449	0.2631	0.2655	0.2528	0.2712
$C_0(0.5, 0.5)$	0.3752	0.3357	0.3285	0.3461	0.3484	0.3361	0.3540
$C_0(0.6, 0.6)$	0.4856	0.4306	0.4244	0.4397	0.4418	0.4309	0.4468
τ	0.4995	0.3424	0.3147	0.3818	0.3905	0.3437	0.4116
ρ	0.6816	0.4903	0.4537	0.5411	0.5520	0.4921	0.5783

Table 5: Comparison of minimum distance, semiparametric and IFM methods with incorrectly specified copula when the margins follow GARCH models

Parameter	Population Value	IFM (N-N)	IFM (T-T)	IFM (T-ST)	IFM (T-C)	SP	MD
True copula: Joe; fitted copula family: Clayton							
$C_0(0.3, 0.3)$	0.1679	0.1649	0.1393	0.1793	0.1893	0.1653	0.1831
$C_0(0.4, 0.4)$	0.2677	0.2377	0.2110	0.2531	0.2639	0.2381	0.2571
$C_0(0.5, 0.5)$	0.3786	0.3216	0.2968	0.3363	0.3469	0.3220	0.3402
$C_0(0.6, 0.6)$	0.4969	0.4185	0.3980	0.4312	0.4404	0.4188	0.4346
τ	0.4998	0.2880	0.1891	0.3447	0.3848	0.2895	0.3596
ρ	0.6798	0.4178	0.2762	0.4934	0.5448	0.4199	0.5126
True copula: Gumbel; fitted copula family: Frank							
$C_0(0.3, 0.3)$	0.1822	0.1965	0.2017	0.2017	0.2013	0.1965	0.1873
$C_0(0.4, 0.4)$	0.2737	0.2904	0.2964	0.2964	0.2960	0.2904	0.2796
$C_0(0.5, 0.5)$	0.3752	0.3886	0.3948	0.3949	0.3944	0.3886	0.3773
$C_0(0.6, 0.6)$	0.4856	0.4904	0.4964	0.4964	0.4960	0.4904	0.4796
τ	0.4995	0.4994	0.5226	0.5229	0.5211	0.4993	0.4574
ρ	0.6816	0.6940	0.7202	0.7205	0.7185	0.6938	0.6443
True copula: Symmetrized Joe-Clayton; fitted copula family: Clayton							
$C_0(0.3, 0.3)$	0.2088	0.2172	0.2177	0.2153	0.2039	0.2172	0.2217
$C_0(0.4, 0.4)$	0.2933	0.2950	0.2956	0.2929	0.2800	0.2951	0.3002
$C_0(0.5, 0.5)$	0.3881	0.3781	0.3787	0.3759	0.3628	0.3782	0.3835
$C_0(0.6, 0.6)$	0.4933	0.4686	0.4692	0.4666	0.4547	0.4687	0.4737
τ	0.5828	0.5004	0.5026	0.4925	0.4444	0.5007	0.5203
ρ	0.7658	0.6828	0.6852	0.6739	0.6180	0.6832	0.7047

Table 6: Semiparametric [SP] and Minimum Distance [MD] estimates of the probabilities of various events in the lower tail for the DM-USD and Yen-USD exchange rates example.

Method	t	0.2	0.3	0.4	0.5
SP	$\mathrm{pr}(\eta_{1n} \leq t, \eta_{2n} \leq t)$	0.0744	0.1475	0.2325	0.3398
MD	$\mathrm{pr}(\eta_{1n} \leq t, \eta_{2n} \leq t)$	0.0751	0.1487	0.2339	0.3414
SP	$\mathrm{pr}(\eta_{1n} \leq t \mid \eta_{2n} \leq t)$	0.2279	0.2266	0.3095	0.4143
MD	$\mathrm{pr}(\eta_{1n} \leq t \mid \eta_{2n} \leq t)$	0.2279	0.3101	0.4148	0.4997
	t	1	0.9	0.7	0.3
SP	$\mathrm{pr}(y_{1n} \leq log(t), y_{2n} \leq log(t))$	0.3328	0.2469	0.1056	0.0052
MD	$\mathrm{pr}(y_{1n} \leq log(t), y_{2n} \leq log(t))$	0.3343	0.2484	0.1067	0.0053
SP	$\mathrm{pr}(y_{1n} \leq log(t) \mid y_{2n} \leq log(t))$	0.4981	0.4250	0.2683	0.0641
MD	$\mathrm{pr}(y_{1n} \leq log(t) \mid y_{2n} \leq log(t))$	0.4982	0.4254	0.2691	0.0649