2001-17

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

ESTIMATION OF COPULA MODELS FOR TIME SERIES OF POSSIBLY DIFFERENT LENGTHS

BY

ANDREW J. PATTON

DISCUSSION PAPER 2001-17
NOVEMBER 2001

Estimation of Copula Models for Time Series of Possibly Different Lengths

Andrew J. Patton*
University of California, San Diego

12 November, 2001

Abstract

The theory of conditional copulas provides a means of constructing flexible multivariate density models, allowing for time-varying conditional densities of each individual variable, and for time-varying conditional dependence between the variables. Further, the use of copulas in constructing these models often allows for the partitioning of the parameter vector into elements relating only to a marginal distribution, and elements relating to the copula. This paper presents a two-stage (or multi-stage) maximum likelihood estimator for the case that such a partition is possible. We extend the existing statistics literature on the estimation of copula models to consider data that exhibit temporal dependence and heterogeneity. The estimator is flexible enough that the case that unequal amounts of data are available on each variable is easily handled. We investigate the small sample properties of the estimator in a Monte Carlo study, and find that it performs well in comparisons with the standard (one-stage) maximum likelihood estimator. Finally, we present an application of the estimator to a model of the joint distribution of daily Japanese yen - U.S. dollar and euro - U.S. dollar exchange rates. We find some evidence that a copula that captures asymmetric dependence performs better than those that assume symmetric dependence.

Keywords: copulas, maximum likelihood, two-stage estimation, exchange rates, missing data.
J.E.L. Codes: C13, C32, C51, F31.

*This work is preliminary, and comments or questions are welcome. The author would like to thank Graham Elliott, Rob Engle, Raffaella Giacomini, Giuseppe Ragusa, Kevin Sheppard and Allan Timmermann for useful comments and suggestions. The simulations presented in this paper were made possible by Vince Crawford and the UCSD Economics Experimental and Computational Laboratory. Financial support from the UCSD Project in Econometric Analysis Fellowship is gratefully acknowledged. Contact information: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093-0508, U.S.A. Email: apatton@econ.ucsd.edu.

1 Introduction

This paper presents a two-stage (or multi-stage) maximum likelihood estimator for multivariate density models of time series data, and we allow for cases where the amount of data on each variable differs. We focus on the case that the multivariate density model is constructed using the theory of copulas, and that the parameter vector is such that it may be partitioned into elements that relate only to a marginal distribution and elements that relate only to the copula. If such a partition is not possible, the familiar one-stage maximum likelihood estimator is the natural estimator to employ. When this partitioning is possible however, great computational savings may be achieved by employing a two-stage estimator.

Evidence of non-normality of the distribution of many interesting economic variables grows each year. One of the first papers to report such evidence was Mandelbrot (1963), who found that the returns on financial assets exhibit too much kurtosis to be adequately described as Gaussian, and numerous studies[1] have since reported further evidence. The implication of these papers is that the multivariate normal distribution is simply not a good model for the joint distribution of many interesting economic variables. This leads us to the problem of finding more appropriate multivariate models. Copula theory is perfectly suited to help us in this quest.

The theory of copulas dates back to Sklar (1959), but its application in statistical modelling is a more recent phenomenon. Sklar (1959) showed that we may decompose a joint distribution into its k marginal distributions, and a *copula*, which describes the dependence between the variables. One of the uses of this theorem to the researcher is in the construction of flexible multivariate distributions[2]: we may combine k marginal distributions of any form (normal, Student's t, exponential, log-normal, etcetera) with any copula to form a valid multivariate distribution. Most existing multivariate distributions are simple extensions of univariate distributions, and often have the restrictive property that all of the marginal distributions are of the same type (all marginal distributions of a multivariate normal are normal, all marginal distributions of a multivariate Student's t_6 are univariate Student's t_6, and so on). If the individual variables of interest were known to be best fitted by different univariate distributions, the choice of a suitable joint distribution was difficult. Copula theory resolves this difficulty.

The application of copula theory to the analysis of economic problems is a new and fast-growing field. Some examples of work in this field (though this list will surely be out-of-date within a month

[1]See, *inter alia*, Fama (1965), Bollerslev (1987), Richardson and Smith (1993), Erb, *et al.* (1994), Bae, *et al.* (2000), Campbell, *et al.* (2000), and most recently, Ang and Chen (2001) and Longin and Solnik (2001).

[2]As recently as Farebrother (1992), for example, it was a considerable challenge in econometric theory to construct an asymmetric bivariate density with common marginal densities. Employing copula theory renders the task almost trivial: simply select any asymmetric copula and use it to link any two marginal distributions of the same type. Suitable copulas include the Clayton and the Gumbel copulas, see Joe (1997) or Nelsen (1999) for more on these copulas.

of writing) includes Rosenberg (1999) and (2000), Bouyé, *et al.,* (2000), Li (2000), Scaillet (2000), Embrechts, *et al.,* (2001), Patton (2001a,b) and Rockinger and Jondeau (2001). The application of copula theory to economics, however, adds a new dimension of difficulty for the researcher: time series dependence.

There is a quite large body of work on the estimation theory underlying the numerous applications of copula theory that have appeared in the statistics literature[3], see Oakes (1982), Genest and Rivest (1993), Genest, *et al.,* (1995), Shih and Louis (1995), Joe and Xu (1996), Xu (1996), Capéraà, *et al.,* (1997) and Glidden (2000). This theory, however, was developed for applications where the data could be assumed to be independent and identically distributed (*i.i.d.*), an assumption that is rejected for almost every economic time series. The first contribution of this paper is to extend the theory on the estimation of parametric copula models to allow for data that may exhibit both temporal dependence and heterogeneity, employing the two-stage maximum likelihood framework of Newey and McFadden (1994) and White (1994). Thus our estimator can be used in the estimation of time-varying conditional density models, that allow for time-varying conditional marginal distributions and a time-varying conditional copula. A nonparametric copula estimator for time series data has recently been proposed by Scaillet (2000), however, as for all nonparametric procedures, the marginal distributions and copula must be assumed to be constant in this framework.

A further contribution is that we also present results for this estimator in the case that the amount of data available on each variable differs. These results may be interpreted as an extension of some of the results presented by Anderson (1957), Little and Rubin (1987) and Stambaugh (1997). The case of unequal amounts of data arises in a number of interesting applications, such as the analysis of: developed markets and emerging markets, which may have only recently begun maintaining data sets; market returns and the returns on a recently floated company; market returns and the returns on a company that went bankrupt; any pair of assets where one is denominated in euros and the other is not. The latter example is the one examined in this paper. Stambaugh (1997) showed the importance of making use of all available data in a simple asset allocation example. It should be pointed out that the theory presented in this paper is only applicable in the case that the starting date(s) (or ending date(s), as applicable) of the truncated series do not contain any information for the parameters of interest that is not contained in the observed data. The 'missing-data mechanism', in the terminology of Little and Rubin (1987), must be ignorable. Examples in economics where the missing-data mechanism is *not* ignorable are to be found in Brown, *et al.* (1995), Goetzmann and Jorion (1999), and Kofman and Sharpe (2000), *inter alia*.

.

[3]In biostatistics and epidemiology, see Clayton (1978), Denuit and Genest (1999) and Fine and Jiang (2000), for example, and in other fields of applied statistics, see Cook and Johnson (1981) and Oakes (1989), amongst others.

The small sample properties of the estimator are investigated and compared with existing techniques in a Monte Carlo study. We simulate processes with characteristics common to many financial time series at the daily frequency and compare the mean squared error (MSE) of the estimator with that of the standard maximum likelihood estimator. The two-stage estimator is found to have lower small sample MSE than the one-stage estimator in a number of situations, and only moderately greater MSE in the remaining cases.

Finally, we present an application of the estimator to a model of the joint distribution of daily Japanese yen - U.S. dollar and euro - U.S. dollar exchange rates. These rates are the two most frequently traded exchange rates, and suffer from the problem that we have much less data available on the euro than we do on the yen. We consider three different copula models, the Gaussian, Plackett and Clayton copulas, and find evidence that a copula that captures asymmetric dependence performs better than those that assume symmetric dependence.

The remainder of the paper is organised as follows. In Section 2 we provide a brief introduction to copula theory. In Section 3 we present the two-stage estimator and discuss the consistent estimation of its asymptotic covariance matrix. We also discuss a modification of the two-stage estimator that achieves full efficiency. In Section 4 we present the results of a Monte Carlo study of the small sample properties of the estimator and in Section 5 we apply the estimator to a model of the joint distribution of daily Japanese yen - U.S. dollar and euro - U.S. dollar exchange rates. Finally, in Section 6 we conclude and present some of the many challenges that remain for future research. The assumptions required for the maximum likelihood estimator and all proofs are contained in the appendix.

1.1 Notation

We have two (scalar) random variables of interest, X and Y, and possibly some conditioning variables $\mathbf{W}$. The variables' conditional distribution is: $(X_t, Y_t)\,|\mathcal{F}_{t-1} \sim H_t \equiv C_t\,(F_t, G_t)$, where H_t is some bivariate distribution function, the marginal distributions of X_t and Y_t are F_t and G_t, and the copula is C_t. (The notation '$H \equiv C\,(F, G)$' will become clear in the next section.) We will assume that all distributions are continuous, though this assumption may be relaxed at the expense of further complication. The information set is defined as $\mathcal{F}_t \equiv \sigma\,(X_t, Y_t, \mathbf{W}_{t+1}, X_{t-1}, Y_{t-1}, \mathbf{W}_t, ...)\,.$ As usual, we will denote random variables in upper case, X_t, and realisations of random variables in lower case, x_t. We will often need to refer to the history of the random variables, which will be denoted $Z^t \equiv \left(X_t, Y_t, \mathbf{W}_{t+1}', X_{t-1}, Y_{t-1}, \mathbf{W}_t', ...\right)'$. Throughout this paper we will denote the distribution (or *c.d.f.*) of a random variable using an upper case letter, and the corresponding density (or *p.d.f.*) using the lower case letter. We will denote the extended real line as $\bar{\mathbb{R}} \equiv \mathbb{R} \cup \{\pm\infty\}$. Convergence in probability is denoted as $\xrightarrow{p}$, and convergence in distribution as $\xrightarrow{\mathcal{D}}$. We denote a $k \times k$ identity matrix as I_k.

It should be pointed out that although we focus on the bivariate case in this paper, both the theory of copulas and the estimation methods presented here extend quite naturally to the general multivariate case.

2 An introduction to copula theory

The introduction to copula theory presented below follows closely that of Patton (2001a). We will firstly present the probability integral transformation, and will then introduce the copula via standard theory on the distribution of transformations of random variables. Following that, the more general theory of conditional copulas is presented. A very readable and thorough introduction to the theory of copulas may be found in Nelsen (1999).

The first analysis of the distribution of the probability integral transformation is quite old, dating back to Fisher (1932). For a more recent reference see, for example, Casella and Berger (1990). Let $U_t \equiv F_t(X_t)$ and $V_t \equiv G_t(Y_t)$. We then say that U_t and V_t are the 'probability integral transforms of X_t and Y_t'. The distribution of the probability integral transform is given in Theorem 1 below.

Theorem 1 (Fisher, 1932) *If F_t and G_t are continuous distribution functions, then $U_t \equiv F_t(X_t) \sim Unif(0,1)$ and $V_t \equiv G_t(Y_t) \sim Unif(0,1)$.*

With this result in hand, we may introduce the copula using basic statistical theory.

2.1 The copula and transformations of random variables

In this section we will suppress the dependence of the random variables and their distributions on t, for the sake of simplicity. Let $U \equiv F(X)$ and $V \equiv G(Y)$, as above. We will now attempt to find the joint density of U and V according to basic results in mathematical statistics on the distribution of transformations of random variables. We will denote the joint density of U and V as c, which turns out to be the 'copula density'.

Since F and G are strictly increasing and continuous, we have that $X = F^{-1}(U)$ and $Y = G^{-1}(V)$, and $\frac{\partial X}{\partial U} = \left(\frac{\partial U}{\partial X}\right)^{-1} = \left(\frac{\partial F(X)}{\partial X}\right)^{-1} = f(X)^{-1}$ and $\frac{\partial Y}{\partial V} = \left(\frac{\partial V}{\partial Y}\right)^{-1} = \left(\frac{\partial G(Y)}{\partial Y}\right)^{-1} = g(Y)^{-1}$. Note that $\frac{\partial X}{\partial V} = \frac{\partial Y}{\partial U} = 0$. Then,

$$\begin{aligned} c(u,v) &= h(X(u), Y(v)) \cdot \begin{vmatrix} \frac{\partial X}{\partial U} & \frac{\partial X}{\partial V} \\ \frac{\partial Y}{\partial U} & \frac{\partial Y}{\partial V} \end{vmatrix} \\ &= h\left(F^{-1}(u), G^{-1}(v)\right) \cdot \frac{\partial X}{\partial U} \cdot \frac{\partial Y}{\partial V} \\ c(u,v) &= \frac{h\left(F^{-1}(u), G^{-1}(v)\right)}{f\left(F^{-1}(u)\right) \cdot g\left(G^{-1}(v)\right)} \qquad (1) \end{aligned}$$

Equation (1) shows that the copula density of X and Y is equal to the ratio of the joint density, h, to the product of the marginal densities, f and g. From this expression we can obtain a first result on the properties of copulas: if X and Y are independent, then the copula density takes the value 1 everywhere, since in that case the joint density is equal to the product of the marginal densities. Since we know that the marginal densities of U and V are uniform, by Theorem 1 above, we thus have that if X and Y are independent the joint distribution of U and V is the bivariate Uniform$(0,1)$ distribution.

We can also use equation (1) to derive an expression for h as a function of x and y instead:

$$\begin{aligned} h\left(F^{-1}(u), G^{-1}(v)\right) &= f\left(F^{-1}(u)\right) \cdot g\left(G^{-1}(v)\right) \cdot c(u,v) \\ h(x,y) &= f(x) \cdot g(y) \cdot c(F(x), G(y)) \end{aligned} \quad (2)$$

Equation (2) is the 'density version' of Sklar's (1959) theorem: the joint density, h, can be decomposed into product of the marginal densities, f and g, and the copula density, c. Sklar's theorem holds under more general conditions than the ones we imposed for this illustration, and below we discuss the general proof.

2.2 The theory of the conditional copula

For an introduction to the general theory of copulas the reader is referred to Nelsen (1999) or Chapter 6 of Schweizer and Sklar (1983). We will start with a few very basic, but very important, definitions based on those in Nelsen (1999). The second condition below refers to the 'H_t-volume' of a rectangle $[x_1, x_2] \times [y_1, y_2]$ in $\bar{\mathbb{R}}^2$, denoted by V_{H_t}. This is simply the probability of observing a point in the region $[x_1, x_2] \times [y_1, y_2]$. It is expressed in the following way as it generalises more easily to the multivariate case.

Definition 1 (Conditional bivariate distribution function) *A conditional bivariate distribution function is a right continuous function $H_t : \bar{R}^2 \rightarrow [0,1]$ with the properties:*

1. $H_t(x, -\infty|\mathcal{F}_{t-1}) = H_t(-\infty, y|\mathcal{F}_{t-1}) = 0$, *and* $H_t(\infty, \infty|\mathcal{F}_{t-1}) = 1$
2. $V_{H_t}([x_1, x_2] \times [y_1, y_2]) \equiv H_t(x_2, y_2|\mathcal{F}_{t-1}) - H_t(x_1, y_2|\mathcal{F}_{t-1}) - H_t(x_2, y_1|\mathcal{F}_{t-1}) + H_t(x_1, y_1|\mathcal{F}_{t-1}) \geq 0$ *for all* $x_1, x_2, y_1, y_2 \in \bar{\mathbb{R}}$, *and* $x_1 \leq x_2$, $y_1 \leq y_2$.

   where $\mathcal{F}_{t-1}$ is some conditioning set.

The first condition simply provides the upper and lower bounds on the distribution function. The second condition ensures that the probability of observing a point in the region $[x_1, x_2] \times [y_1, y_2]$ is non-negative[4]. We now define the conditional copula.

[4] If we set $x_2 = x_1 + \varepsilon$ and $y_2 = y_1 + \varepsilon$ and let $\varepsilon \rightarrow 0^+$, then it becomes clear that this definition is just the generalisation of the condition that if the bivariate density exists, it must be non-negative on the domain of H_t.

Definition 2 (Conditional copula) *A two-dimensional conditional copula is a function* $C_t : [0,1] \times [0,1] \to [0,1]$ *with the following properties:*

1. $C_t(u,0|\mathcal{F}_{t-1}) = C_t(0,v|\mathcal{F}_{t-1}) = 0$*, and* $C_t(u,1|\mathcal{F}_{t-1}) = u$ *and* $C_t(1,v|\mathcal{F}_{t-1}) = v$*, for every* u, v *in* $[0,1]$

2. $V_{C_t}\left([u_1,u_2] \times [v_1,v_2] \,|\mathcal{F}_{t-1}\right) \equiv C_t\left(u_2,v_2|\mathcal{F}_{t-1}\right) - C_t\left(u_1,v_2|\mathcal{F}_{t-1}\right) - C_t\left(u_2,v_1|\mathcal{F}_{t-1}\right) + C_t\left(u_1,v_1|\mathcal{F}_{t-1}\right) \geq 0$ *for all* $u_1, u_2, v_1, v_2 \in [0,1]$*, such that* $u_1 \leq u_2$ *and* $v_1 \leq v_2$*.*

   where $\mathcal{F}_{t-1}$ *is some conditioning set.*

The first condition of Definition 2 provides the lower bound on the distribution function, and ensures that the marginal distributions, $C_t\left(u,1|\mathcal{F}_{t-1}\right)$ and $C_t\left(1,v|\mathcal{F}_{t-1}\right)$, are uniform. The condition that V_{C_t} is non-negative has the same interpretation as the second condition of Definition 1: it simply ensures that the probability of observing a point in the region $[u_1,u_2] \times [v_1,v_2]$ is non-negative.

By drawing on the above conditions for the conditional copula, and extending its domain to $\bar{\mathbb{R}}^2$, we may alternatively define a conditional copula as the conditional bivariate distribution of a pair of random variables (U_t, V_t) having margins that are $Unif\,(0,1)$. The extension of the domain to $\bar{\mathbb{R}}^2$ is accomplished as follows:

$$\text{Let} \quad C_t^*\left(u,v|\mathcal{F}_{t-1}\right) = \begin{cases} 0 & \text{for } u < 0 \text{ or } v < 0, \\ C_t\left(u,v|\mathcal{F}_{t-1}\right) & \text{for } (u,v) \in [0,1] \times [0,1], \\ u & \text{for } u \in [0,1], v > 1, \\ v & \text{for } u > 1, v \in [0,1], \\ 1 & \text{for } u > 1, v > 1. \end{cases} \tag{3}$$

The link between the probability integral transformation and the theory of copulas now becomes clear: the copula is the joint distribution function of the probability integral transforms of each of the variables X_t and Y_t with respect to their marginal distributions, F_t and G_t. We now move on to an extension of the the key result in the theory of copulas: Sklar's (1959) theorem for conditional distributions:

Theorem 2 (Sklar's Theorem for Continuous Conditional Distributions) *Let* H_t *be a conditional bivariate distribution function with continuous margins* F_t *and* G_t*, and let* $\mathcal{F}_{t-1}$ *be some conditioning set. Then there exists a unique conditional copula* $C_t : [0,1] \times [0,1] \to [0,1]$ *such that*

$$H_t(x,y|\mathcal{F}_{t-1}) = C_t(F_t(x|\mathcal{F}_{t-1}), G_t(y|\mathcal{F}_{t-1})|\mathcal{F}_{t-1}), \quad \forall x, y \in \bar{\mathbb{R}} \tag{4}$$

Conversely, if C_t *is a conditional copula and* F_t *and* G_t *are the conditional distribution functions of two random variables* X_t *and* Y_t*, then the function* H_t *defined by equation (4) is a bivariate conditional distribution function with margins* F_t *and* G_t*.*

The density function equivalent of (4) is useful for maximum likelihood analysis, and is obtained quite easily, provided that F_t and G_t are differentiable, and H_t and C_t are twice differentiable.

$$\begin{aligned}
h_t(x,y|\mathcal{F}_{t-1}) &\equiv \frac{\partial^2 H_t(x,y|\mathcal{F}_{t-1})}{\partial x \partial y} \\
&= \frac{\partial F_t(x|\mathcal{F}_{t-1})}{\partial x} \cdot \frac{\partial G_t(y|\mathcal{F}_{t-1})}{\partial y} \cdot \frac{\partial^2 C_t(F_t(x|\mathcal{F}_{t-1}), G_t(y|\mathcal{F}_{t-1})|\mathcal{F}_{t-1})}{\partial (F_t(x|\mathcal{F}_{t-1})\,\partial(G_t(y|\mathcal{F}_{t-1}))} \\
&\equiv f_t(x|\mathcal{F}_{t-1}) \cdot g_t(y|\mathcal{F}_{t-1}) \cdot c_t(u,v|\mathcal{F}_{t-1}), \quad \forall (x,y) \in \bar{\mathbb{R}}^2 \qquad (5)
\end{aligned}$$

where $u \equiv F_t(x|\mathcal{F}_{t-1})$, and $v \equiv G_t(y|\mathcal{F}_{t-1})$. The expression in equation (5) is precisely the same as that in equation (2), which we obtained using the theory on the distribution of transformations of random variables. Taking logs of both sides we obtain:

$$\mathcal{L}_{XY} = \mathcal{L}_X + \mathcal{L}_Y + \mathcal{L}_C \qquad (6)$$

and so the joint log-likelihood is equal to the sum of the marginal log-likelihoods and the copula log-likelihood.

We can also obtain a corollary to Theorem 2, analogous to that of Nelson's (1999) corollary to Sklar's Theorem, which enables us to extract the conditional copula from any conditional bivariate distribution function, but first we need the definition of the 'quasi-inverse' of a function.

Definition 3 (Quasi-inverse of a distribution function) *The quasi-inverse, $F^{(-1)}$, of a distribution function F is defined as:*

$$F^{(-1)}(u) = \inf\{x : F(x) \geq u\}, \text{ for } u \in [0,1]. \qquad (7)$$

If F is strictly increasing then the above definition returns the usual functional inverse of F, but more importantly it allows us to consider inverses of non-strictly increasing functions.

Corollary 1 *Let H_t be any conditional bivariate distribution with continuous marginal distributions, F_t and G_t, and let $F_t^{(-1)}$ and $G_t^{(-1)}$ denote the (quasi-) inverses of the marginal distributions. Finally, let $\mathcal{F}_{t-1}$ be some conditioning set. Then there exists a unique conditional copula $C_t : [0,1] \times [0,1] \rightarrow [0,1]$ such that*

$$C_t(u,v|\mathcal{F}_{t-1}) = H_t\left(F_t^{(-1)}(u|\mathcal{F}_{t-1}), G_t^{(-1)}(v|\mathcal{F}_{t-1})|\mathcal{F}_{t-1}\right), \quad \forall u,v \in [0,1] \qquad (8)$$

This corollary completes the idea that a bivariate distribution function may be decomposed into three parts. Given any two marginal distributions and any copula we have a joint distribution, and from any given joint distribution we can extract the implied marginal distributions and copula.

3 Two-stage estimation of copula models

Let the conditional distribution $(X_t, Y_t) | \mathcal{F}_{t-1}$ be parameterised as $H_t(\theta_0) \equiv C_t(F_t(\varphi_0), G_t(\gamma_0); \kappa_0)$, so $\mathcal{L}_{XY}(\theta_0) = \mathcal{L}_X(\varphi_0) + \mathcal{L}_Y(\gamma_0) + \mathcal{L}_C(\varphi_0, \gamma_0, \kappa_0)$, where $\varphi_0 \in int(\Phi) \subseteq \mathbb{R}^p$, $\gamma_0 \in int(\Gamma) \subseteq \mathbb{R}^q$, $\kappa_0 \in int(\mathcal{K}) \subseteq \mathbb{R}^r$ and so $\theta_0 \equiv [\varphi_0', \gamma_0', \kappa_0']' \in int(\Theta) \equiv int(\Phi) \times int(\Gamma) \times int(\mathcal{K}) \subseteq \mathbb{R}^{p+q+r} \equiv \mathbb{R}^s$, where $int(A)$ is the interior of the set A. As F, G and C are conditional on $\mathcal{F}_{t-1}$, they will be written as functions of the entire information set and the appropriate parameter: $F_t(Z^t; \varphi)$, $G_t(Z^t; \gamma)$ and $C_t(Z^t; \theta)$, although of course not all of the elements of Z^t will in general be required.

It will not always be the case that the parameter vector θ_0 decomposes so neatly into three components, associated with the first margin, second margin and the copula. Cross-marginal distribution restrictions are one example where this condition would fail to hold. We discuss this scenario in Section 3.3. In the interim sections we will assume that the decomposition is possible.

In this paper we allow for the situation that the amount of data available on X is possibly different to that available on Y, which is also possibly different to the amount of overlapping data on both X and Y. This scenario is depicted in Figure 1. For simplicity, let X be the variable with the most data available. We will denote the number of observations on X, Y and the common sample as n_x, n_y and n_c respectively. All data lengths are assumed to be (fixed) functions of n, and we will set $n_x = n$. We consider cases where $\frac{n_y}{n_x} \to \lambda_y$ and $\frac{n_c}{n_x} \to \lambda_c$, where $0 < \lambda_y \leq 1$ and $0 < \lambda_c \leq 1$. One such situation is that $n_y = n_x - d_1$ and $n_c = n_y$, that is, where the sample on Y started later than the sample on X, and that the difference in the number of observations is constant as $n \to \infty$. Another situation is where $\frac{n_y}{n_x} = d_y$ and $\frac{n_c}{n_x} = d_c$ for all n, that is, that the ratio of the number of observations on Y, and in the common sample, to those available on X is constant. To simplify notation, we assume that all samples (on X, Y and the common sample) start at $t = 1$ and run through until $t = n_x$, n_y and n_c respectively.

There have been other suggestions made in the literature on how to deal with unequal amounts of data: Harvey, *et al.* (1997) suggest using the Kalman filter, and Kofman and Sharpe (2000) discuss using the EM algorithm and its Bayesian alternative, the Imputation Posterior method. Anderson (1957) and Stambaugh (1997) suggest using the marginal/conditional distribution decomposition of a joint distribution, for the case of *i.i.d.* multivariate normal random variables. The relationship between our method and theirs is discussed in the following section. Little and Rubin (1987) present many different methods of dealing with missing observations, and provide numerous further references on the topic.

3.1 The estimator

Our two-stage (or more accurately, 'three-stage') maximum likelihood estimator is denoted $\hat{\theta}_n$, and its components are given below.

$$\hat{\varphi}_{n_x} \equiv \arg\max_{\varphi\in\Phi} \; n_x^{-1}\sum_{t=1}^{n_x}\log f_t\left(Z^t;\varphi\right) \tag{9}$$

$$\hat{\gamma}_{n_y} \equiv \arg\max_{\gamma\in\Gamma} \; n_y^{-1}\sum_{t=1}^{n_y}\log g_t\left(Z^t;\gamma\right) \tag{10}$$

$$\hat{\kappa}_{n_c} \equiv \arg\max_{\kappa\in\mathcal{K}} \; n_c^{-1}\sum_{t=1}^{n_c}\log c_t\left(Z^t;\hat{\varphi}_{n_x},\hat{\gamma}_{n_y},\kappa\right) \tag{11}$$

$$\hat{\theta}_n \equiv \left[\hat{\varphi}'_{n_x},\hat{\gamma}'_{n_y},\hat{\kappa}'_{n_c}\right]' \tag{12}$$

We show that this estimator is consistent for θ_0, and that it is asymptotically normal. The asymptotic variance-covariance matrix of this estimator is slightly different that of the standard two-stage MLE, and below it is discussed in some detail.

The method of estimation presented in this paper relies on the copula decomposition of a joint distribution, repeated in equations (13) and (14) below. The method of Anderson (1957), Little and Rubin (1987), and Stambaugh (1997) uses the marginal/conditional decomposition in equations (15) and (16) below.

$$h_t\left(x_t,y_t;\theta_0\right) = f_t\left(x_t;\varphi_0\right)\cdot g_t\left(y_t;\gamma_0\right)\cdot c_t\left(F_t\left(x_t;\varphi_0\right),G_t\left(y_t;\gamma_0\right);\kappa_0\right), \quad \text{so} \tag{13}$$

$$\mathcal{L}_{XY}\left(\theta_0\right) = \mathcal{L}_X\left(\varphi_0\right)+\mathcal{L}_Y\left(\gamma_0\right)+\mathcal{L}_C\left(\varphi_0,\gamma_0,\kappa_0\right) \tag{14}$$

$$h_t\left(x_t,y_t;\theta_0\right) = f_t\left(x_t;\varphi_0\right)\cdot h_{t,y|x}\left(y_t|x_t;\gamma_0,\kappa_0\right), \quad \text{so} \tag{15}$$

$$\mathcal{L}_{XY}\left(\theta_0\right) = \mathcal{L}_X\left(\varphi_0\right)+\mathcal{L}_{Y|X}\left(\varphi_0,\gamma_0,\kappa_0\right) \tag{16}$$

For the multivariate normal distribution, Anderson (1957) showed that one could obtain an estimate of θ_0 by first maximising $\mathcal{L}_X$ using all n_x observations to obtain an estimate of φ_0, and then maximising $\mathcal{L}_{Y|X}$ using $n_y = n_c$ observations, conditioning on the estimate of φ_0. Our estimator applies to *any* joint distribution, subject to satisfying regularity conditions presented in Appendix 2, and simplifies estimation one step further, by decomposing the conditional likelihood function, $\mathcal{L}_{Y|X}$, into the marginal likelihood of Y, $\mathcal{L}_Y$, and the copula likelihood, $\mathcal{L}_C$. This additional decomposition reduces the computational difficulty of estimation, and allows for more irregular data sets[5].

It must be pointed out that all of the following results rely on the assumption that the data generating process of the variables is known up to a vector of undetermined parameters, θ_0. This is obviously quite a restrictive assumption. The econometrics literature contains some work on the estimation of models of the conditional mean that are robust to misspecification of the conditional

[5]The method of Anderson (1957) and Little Rubin (1987) cannot deal with the data situation presented in Figure 1; they require that $n_y = n_c$.

variance and/or the conditional density, see Gourieroux, *et al.*, (1984) for example, and on the estimation of models of the conditional mean and variance that are robust to misspecification of the conditional density, see Bollerslev and Wooldridge (1992). To the author's knowledge, corresponding results for the estimation of conditional densities are not available, though such results would be of great interest. We leave this for future work.

We now present the consistency and asymptotic normality results for the two-stage estimator presented above. These results are based on the work of Newey and McFadden (1994) and White (1994), both of which provide thorough reviews of two-stage maximum likelihood estimation theory. All assumptions for this section are collected in Appendix 2, and all proofs are in Appendix 3.

Theorem 3 (Consistency of $\hat{\varphi}_{n_x}$ and $\hat{\gamma}_{n_y}$) *Let assumptions 1 to 6 hold. Then* $\hat{\varphi}_{n_x} \xrightarrow{p} \varphi_0$ *and* $\hat{\gamma}_{n_y} \xrightarrow{p} \gamma_0$ *as* $n \to \infty$*.*

Theorem 4 (Consistency of $\hat{\kappa}_{n_c}$) *Let the assumptions of the previous theorem hold, and let assumption 7 hold in addition. Then* $\hat{\kappa}_{n_c} \xrightarrow{p} \kappa_0$ *as* $n \to \infty$*.*

Theorem 5 (Asymptotic normality of $\hat{\theta}_n$) *Let the assumptions of the previous theorems hold, and let assumption 8 hold in addition. Then*

$$B_n^{0-1/2} \cdot \mathcal{N}^{1/2} \cdot A_n^0 \cdot \left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{D}} N\left(0, I_s\right) \tag{17}$$

where

$$\mathcal{N} \equiv \begin{bmatrix} n_x \cdot I_p & 0 & 0 \\ 0 & n_y \cdot I_q & 0 \\ 0 & 0 & n_c \cdot I_r \end{bmatrix}, \tag{18}$$

$$A_n^0 \equiv \begin{bmatrix} n_x^{-1} \sum_{t=1}^{n_x} E\left[\nabla_{\varphi\varphi} \log f_t^0\right] & 0 & 0 \\ 0 & n_y^{-1} \sum_{t=1}^{n_y} E\left[\nabla_{\gamma\gamma} \log g_t^0\right] & 0 \\ n_c^{-1} \sum_{t=1}^{n_c} E\left[\nabla_{\varphi\kappa} \log c_t^0\right] & n_c^{-1} \sum_{t=1}^{n_c} E\left[\nabla_{\gamma\kappa} \log c_t^0\right] & n_c^{-1} \sum_{t=1}^{n_c} E\left[\nabla_{\kappa\kappa} \log c_t^0\right] \end{bmatrix} \tag{19}$$

$$B_n^0 \equiv var\left[\sum_{t=1}^{n} \left[n_x^{-1/2} s_{1t}^{0\prime}, n_y^{-1/2} s_{2t}^{0\prime}, n_c^{-1/2} s_{3t}^{0\prime}\right]^{\prime}\right]$$

$$= \begin{bmatrix} n_x^{-1} \sum_{t=1}^{n_x} E\left[s_{1t}^0 \cdot s_{1t}^{0\prime}\right] & (n_x n_y)^{-1/2} \sum_{t=1}^{n_y} E\left[s_{1t}^0 \cdot s_{2t}^{0\prime}\right] & (n_x n_c)^{-1/2} \sum_{t=1}^{n_c} E\left[s_{1t}^0 \cdot s_{3t}^{0\prime}\right] \\ (n_x n_y)^{-1/2} \sum_{t=1}^{n_y} E\left[s_{2t}^0 \cdot s_{1t}^{0\prime}\right] & n_y^{-1} \sum_{t=1}^{n_y} E\left[s_{2t}^0 \cdot s_{2t}^{0\prime}\right] & (n_y n_c)^{-1/2} \sum_{t=1}^{n_c} E\left[s_{2t}^0 \cdot s_{3t}^{0\prime}\right] \\ (n_x n_c)^{-1/2} \sum_{t=1}^{n_c} E\left[s_{3t}^0 \cdot s_{1t}^{0\prime}\right] & (n_y n_c)^{-1/2} \sum_{t=1}^{n_c} E\left[s_{3t}^0 \cdot s_{2t}^{0\prime}\right] & n_c^{-1} \sum_{t=1}^{n_c} E\left[s_{3t}^0 \cdot s_{3t}^{0\prime}\right] \end{bmatrix} \tag{20}$$

where $s_{1t}^0 \equiv \nabla_\varphi \log f_t\left(Z^t;\varphi_0\right)$, $s_{2t}^0 \equiv \nabla_\gamma \log g_t\left(Z^t;\gamma_0\right)$, $s_{3t}^0 \equiv \nabla_\kappa \log c_t\left(Z^t;\theta_0\right)$, $f_t^0 \equiv f_t\left(Z^t;\varphi_0\right)$, $g_t^0 \equiv g_t\left(Z^t;\gamma_0\right)$ *and* $c_t^0 \equiv c_t\left(Z^t;\theta_0\right)$.

Remark 1 *If $n_x = n_y = n_c \equiv n$, then the result of the above proposition simplifies to*

$$B_n^{0^{-1/2}} \cdot A_n^0 \cdot \sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{D}} N\left(0, I_s\right) \tag{21}$$

with B_n^0 and A_n^0 as defined above.

The matrix $A_n^{0^{-1}} B_n^0 A_n^{0^{-1\prime}}$ is also known in the theory of estimating equations (or inference functions) as the 'Godambe information matrix', named for Godambe (1960) who introduced the theory of estimating equations, if we take the right-hand sides of equations (9) to (11) as our 'estimating equations'. McLeish and Small (1988) provide an overview of the theory estimating equations, and Bera and Bilias (2001) present some of linkages between these estimation methods and method more common in econometrics. Joe and Xu (1996) and Xu (1996) use the estimating equations framework for their two-stage estimator of copula models, however we elected to use the maximum likelihood framework owing to its familiarity to econometricians, and the large body of econometric theory already existing on this estimator.

3.2 Estimation of the covariance matrix

We discuss in this section the consistent estimation of the covariance matrix of the two-stage estimator presented above. Following White (1994), we say that if $\ddot{V}_n^{-1/2}\sqrt{n}(\ddot{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N\left(0, I\right)$, then the asymptotic covariance matrix of the estimator $\ddot{\theta}_n$ is $\ddot{V}_n$, or that $avar(\ddot{\theta}_n) = \ddot{V}_n$. For the two-stage estimator we have that $B_n^{0^{-1/2}} \cdot \mathcal{N}^{1/2} \cdot A_n^0 \cdot (\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N\left(0, I\right)$. Notice that the root n scaling factor is wedged between the Hessian and the inverse square root of the outer product of the scores. We can modify it as follows

$$\begin{aligned} B_n^{0^{-1/2}} \cdot \mathcal{N}^{1/2} \cdot A_n^0 \cdot \left(\hat{\theta}_n - \theta_0\right) &= B_n^{0^{-1/2}} \cdot n_c^{-1/2} \cdot \mathcal{N}^{1/2} \cdot A_n^0 \cdot \sqrt{n_c}\left(\hat{\theta}_n - \theta_0\right) \\ &\equiv B_n^{0^{-1/2}} \cdot \mathcal{N}^{*1/2} \cdot A_n^0 \cdot \sqrt{n_c}\left(\hat{\theta}_n - \theta_0\right) \end{aligned}$$

Thus the asymptotic covariance matrix is $A_n^{0^{-1}} \cdot \mathcal{N}_\infty^{*-1/2} \cdot B_n^0 \cdot \mathcal{N}_\infty^{*-1/2} \cdot A_n^{0^{-1\prime}}$, where $\mathcal{N}_\infty^* \equiv \lim_{n\to\infty} \mathcal{N}^*$. The form of $\mathcal{N}_\infty^*$ will depend on the relationship between n_x, n_y and n_c as $n \to \infty$. If it is assumed that the difference between n_c and n_x and n_y is constant as $n \to \infty$, then $\mathcal{N}_\infty^* = I_s$, and so $avar(\hat{\theta}_n)$ takes the same form as the standard two-stage maximum likelihood estimator. If instead the ratio between n_c and n_x, and n_c and n_y is assumed constant, then $\mathcal{N}_\infty^* = \begin{bmatrix} d_x \cdot I_p & 0 & 0 \\ 0 & d_y \cdot I_q & 0 \\ 0 & 0 & I_r \end{bmatrix}$, where $\lim_{n\to\infty} \frac{n_x}{n_c} \equiv d_x \geq 1$ and $\lim_{n\to\infty} \frac{n_y}{n_c} \equiv d_y \geq 1$.

For a consistent estimate of the covariance matrix, we need a consistent estimate of both A_n^0 and B_n^0. In the following two lemmas, we provide conditions under which sample analogues of these matrices are consistent.

Lemma 1 (Consistency of $\hat{A}_n$) *Given assumptions 1, 2, 3 and 5, we have that* $\hat{A}_n - A_n^0 \xrightarrow{p} 0$, *where* A_n^0 *is given in Theorem 5 and*

$$\hat{A}_n \equiv \begin{bmatrix} n_x^{-1}\sum_{t=1}^{n_x} \nabla_{\varphi\varphi} \log \hat{f}_t & 0 & 0 \\ 0 & n_y^{-1}\sum_{t=1}^{n_y} \nabla_{\gamma\gamma} \log \hat{g}_t & 0 \\ n_c^{-1}\sum_{t=1}^{n_c} \nabla_{\varphi\kappa} \log \hat{c}_t & n_c^{-1}\sum_{t=1}^{n_c} \nabla_{\gamma\kappa} \log \hat{c}_t & n_c^{-1}\sum_{t=1}^{n_c} \nabla_{\kappa\kappa} \log \hat{c}_t \end{bmatrix} \tag{22}$$

where $\hat{f}_t \equiv f\left(Z^t; \hat{\varphi}_{n_x}\right)$, $\hat{g}_t \equiv g(Z^t; \hat{\gamma}_{n_y})$ *and* $\hat{c}_t \equiv c(Z^t; \hat{\theta}_n)$.

Lemma 2 (Consistency of $\hat{B}_n$) *Given assumptions 1, 2, and 9, then* $\hat{B}_n - B_n^0 \xrightarrow{p} 0$, *where* B_n^0 *is given in Theorem 5 and*

$$\hat{B}_n \equiv \begin{bmatrix} n_x^{-1}\sum_{t=1}^{n_x} \hat{s}_{1t} \cdot \hat{s}_{1t}' & (n_x n_y)^{-1/2}\sum_{t=1}^{n_y} \hat{s}_{1t} \cdot \hat{s}_{2t}' & (n_x n_c)^{-1/2}\sum_{t=1}^{n_c} \hat{s}_{1t} \cdot \hat{s}_{3t}' \\ (n_x n_y)^{-1/2}\sum_{t=1}^{n_y} \hat{s}_{2t} \cdot \hat{s}_{1t}' & n_y^{-1}\sum_{t=1}^{n_y} \hat{s}_{2t} \cdot \hat{s}_{2t}' & (n_y n_c)^{-1/2}\sum_{t=1}^{n_c} \hat{s}_{2t} \cdot \hat{s}_{3t}' \\ (n_x n_c)^{-1/2}\sum_{t=1}^{n_c} \hat{s}_{3t} \cdot \hat{s}_{1t}' & (n_y n_c)^{-1/2}\sum_{t=1}^{n_c} \hat{s}_{3t} \cdot \hat{s}_{2t}' & n_c^{-1}\sum_{t=1}^{n_c} \hat{s}_{3t} \cdot \hat{s}_{3t}' \end{bmatrix}$$

where $\hat{s}_{1t} \equiv \nabla_\varphi \log f_t\left(Z^t; \hat{\varphi}_{n_x}\right)$, $\hat{s}_{2t} \equiv \nabla_\gamma \log g_t\left(Z^t; \hat{\gamma}_{n_y}\right)$ *and* $\hat{s}_{3t} \equiv \nabla_\kappa \log c_t\left(Z^t; \hat{\kappa}_{n_c}\right)$.

Theorem 6 (Consistency of the covariance matrix estimator) *Given assumptions 1 to 3, 5, 6 and 9 then* $\hat{A}_n^{-1} \cdot \mathcal{N}^{*-1/2} \cdot \hat{B}_n \cdot \mathcal{N}^{*-1/2} \cdot \hat{A}_n^{-1\prime}$ *is a consistent estimator of the asymptotic covariance matrix of* $\hat{\theta}_n$.

Remark 2 *The consideration of the case that* $n_x \neq n_y \neq n_c$ *lead to a slightly more complicated form for the covariance matrix of the two-stage maximum likelihood estimator. In the case that* $n_x = n_y = n_c \equiv n$, *the appropriate covariance matrix estimator is* $\hat{A}_n^{-1} \cdot \hat{B}_n \cdot \hat{A}_n^{-1\prime}$.

In finite samples we use the asymptotic covariance matrix of the estimator as an approximation to the true finite sample covariance matrix. For the estimator presented above the covariance matrix estimate is: $\hat{A}_n^{-1} \cdot \mathcal{N}^{-1/2} \cdot \hat{B}_n \cdot \mathcal{N}^{-1/2} \cdot \hat{A}_n^{-1\prime}$, which simplifies to $n^{-1}\hat{A}_n^{-1} \cdot \hat{B}_n \cdot \hat{A}_n^{-1\prime}$ in the case that $n_x = n_y = n_c \equiv n$.

3.3 Cross-marginal distribution restrictions

Consider now the case that we cannot separate the parameters of the two marginal distributions from each other. One possible cause of this is the presence of cross-marginal distribution restrictions, or the dependence of one marginal on some function of the residual of the other marginal (such as a multivariate moving average model, or a multivariate GARCH model, for example). In this section we show how to adapt the estimator presented above when we cannot separate the parameters of the marginal distributions from each other, but we can separate them from the parameters of the copula.

Assume now that $H(\theta) = C(F(\varphi,\gamma), G(\varphi,\gamma);\kappa)$, and so $\mathcal{L}_{XY}(\theta_0) = \mathcal{L}_X(\varphi_0,\gamma_0) + \mathcal{L}_Y(\varphi_0,\gamma_0) + \mathcal{L}_C(\varphi_0,\gamma_0,\kappa_0)$. We may still achieve some computational simplification by first estimating $[\varphi_0', \gamma_0']'$, and then estimating κ_0 conditional on the estimates of $[\varphi_0', \gamma_0']'$. The estimator is denoted $\hat{\theta}_{n_c}$:

$$\begin{aligned}
\left[\hat{\varphi}_{n_c}', \hat{\gamma}_{n_c}'\right]' &\equiv \underset{[\varphi',\gamma']'\in\Phi\times\Gamma}{\arg\max}\; n_c^{-1}\sum_{t=1}^{n_c}\log f_t\left(Z^t;\varphi,\gamma\right) + \; n_c^{-1}\sum_{t=1}^{n_c}\log g_t\left(Z^t;\varphi,\gamma\right) \\
\hat{\kappa}_{n_c} &\equiv \underset{\kappa\in\mathcal{K}}{\arg\max}\; n_c^{-1}\sum_{t=1}^{n_c}\log c_t\left(Z^t;\hat{\varphi}_{n_c},\hat{\gamma}_{n_c},\kappa\right) \\
\hat{\theta}_{n_c} &\equiv \left[\hat{\varphi}_{n_c}', \hat{\gamma}_{n_c}', \hat{\kappa}_{n_c}'\right]'
\end{aligned}$$

In the first stage of this method we assume that X_t and Y_t are independent, and so maximising the joint likelihood is equivalent to maximising the sum of the marginal likelihoods. In the second stage we model the dependence of X_t and Y_t directly, by maximising the copula likelihood, conditioning on the marginal parameter estimates. In the case that X_t and Y_t truly are independent, the two-stage method is fully efficient.

The consistency results of the previous section hold, and the asymptotic normality results need just a minor modification:

Theorem 7 (Asymptotic normality of $\hat{\theta}_{n_c}$) *Let the assumptions of Theorem 5 hold. Then* $B_n^{0-1/2} \cdot A_n^0 \cdot \sqrt{n_c}\left(\hat{\theta}_{n_c} - \theta_0\right) \xrightarrow{\mathcal{D}} N(0, I_s)$, *where*

$$\begin{aligned}
B_n^0 &\equiv var\left[\sum\nolimits_{t=1}^{n}\left[n_c^{-1/2}s_{12t}^{0\prime}, n_c^{-1/2}s_{3t}^{0\prime}\right]'\right] \\
&= n_c^{-1}\begin{bmatrix} \sum_{t=1}^{n_c} E\left[s_{12t}^0 \cdot s_{12t}^{0\prime}\right] & \sum_{t=1}^{n_c} E\left[s_{12t}^0 \cdot s_{3t}^{0\prime}\right] \\ \sum_{t=1}^{n_c} E\left[s_{3t}^0 \cdot s_{12t}^{0\prime}\right] & \sum_{t=1}^{n_c} E\left[s_{3t}^0 \cdot s_{3t}^{0\prime}\right] \end{bmatrix}, \\
A_n^0 &\equiv n_c^{-1}\begin{bmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & 0 \\ A_{31} & A_{32} & A_{33} \end{bmatrix}
\end{aligned}$$

where $s_{12t}^0 \equiv \begin{bmatrix} \nabla_\varphi \log f_t\left(Z^t;\varphi_0,\gamma_0\right) + \nabla_\varphi \log g_t\left(Z^t;\varphi_0,\gamma_0\right) \\ \nabla_\gamma \log f_t\left(Z^t;\varphi_0,\gamma_0\right) + \nabla_\gamma \log g_t\left(Z^t;\varphi_0,\gamma_0\right) \end{bmatrix}$, $s_{3t}^0 \equiv \nabla_\kappa \log c_t\left(Z^t;\theta_0\right)$, $A_{11} \equiv \sum_{t=1}^{n_c} E\left[\nabla_{\varphi\varphi}\log f_t^0\right] + E\left[\nabla_{\varphi\varphi}\log g_t^0\right]$, $A_{12} \equiv \sum_{t=1}^{n_c} E\left[\nabla_{\gamma\varphi}\log f_t^0\right] + E\left[\nabla_{\gamma\varphi}\log g_t^0\right]$, $A_{21} \equiv \sum_{t=1}^{n_c} E\left[\nabla_{\varphi\gamma}\log f_t^0\right]$-$E\left[\nabla_{\varphi\gamma}\log g_t^0\right]$, $A_{22} \equiv \sum_{t=1}^{n_c} E\left[\nabla_{\gamma\gamma}\log f_t^0\right] + E\left[\nabla_{\gamma\gamma}\log g_t^0\right]$, $A_{31} \equiv \sum_{t=1}^{n_c} E\left[\nabla_{\varphi\kappa}\log c_t^0\right]$, $A_{32} \equiv \sum_{t=1}^{n_c} E\left[\nabla_{\gamma\kappa}\log c_t^0\right]$ *and* $A_{33} \equiv \sum_{t=1}^{n_c} E\left[\nabla_{\kappa\kappa}\log c_t^0\right]$.

The estimator of the covariance matrix of $\hat{\theta}_{n_c}$ is an obvious adjustment of that for $\hat{\theta}_n$. We now turn to an analysis of the efficiency of the two-stage estimator.

3.4 Efficiency of the estimator

In the situation that we have the same amount of data available for all three estimators, $\hat{\varphi}$, $\hat{\gamma}$ and $\hat{\kappa}$, it is well known (see Le Cam, 1956, for example) that the one-stage maximum likelihood estimator is the most efficient estimator, in that it attains the minimum asymptotic variance bound. In this section we compare the asymptotic efficiency of the two-stage estimator discussed in Section 3.1 with the one-stage maximum likelihood estimator. The asymptotic efficiency of two asymptotically normal estimators are compared by examining the difference of their asymptotic covariance matrices. The small sample efficiency of these estimators is compared in the next section.

In this section we also discuss the one-step adjusted two-stage maximum likelihood estimator, see Newey and McFadden (1994) or White (1994), for our case. The adjusted estimator is a single step modification of the two-stage estimator requiring no maximisations, which achieves the minimum asymptotic variance bound.

The one-stage maximum likelihood estimator is denoted $\hat{\theta}^{eff}_{n_c}$, as it is known to be the most efficient estimator when $n_x = n_y = n_c$. It is based by necessity on the common sample period. Let $M^0_{n_c}$ denote the asymptotic variance of $\theta^{eff}_{n_c}$, thus representing the minimum asymptotic variance bound. Define $D^0_n \equiv A^{0^{-1}}_n \cdot B^0_n \cdot A^{0^{-1}\prime}_n$. As presented in the previous section, the asymptotic covariance matrix of the two-stage estimator is $A^{0^{-1}}_n \cdot \mathcal{N}^{*-1/2}_\infty \cdot B^0_n \cdot \mathcal{N}^{*-1/2}_\infty \cdot A^{0^{-1}\prime}_n$. If the difference between the sample sizes is constant, then $A^{0^{-1}}_n \cdot \mathcal{N}^{*-1/2} \cdot B^0_n \cdot \mathcal{N}^{*-1/2} \cdot A^{0^{-1}\prime}_n - D^0_n \xrightarrow{p} 0$, and the two-stage maximum likelihood estimator is asymptotically less efficient than the one-stage maximum likelihood estimator, regardless of the magnitude of the difference in the (finite) sample sizes. If the ratio between the sample sizes is constant, however, there exist situations in which the two-stage maximum likelihood estimator is asymptotically *not less* efficient than the one-stage estimator. This is made clear in the following proposition.

Proposition 1 *Let our two-stage estimator be denoted* $\hat{\theta}_n$ *and let the one-stage maximum likelihood estimator be denoted* $\hat{\theta}^{eff}_{n_c}$. *The asymptotic covariance matrices of these two estimators are* $A^{0^{-1}}_n \cdot \mathcal{N}^{*-1/2}_\infty \cdot B^0_n \cdot \mathcal{N}^{*-1/2}_\infty \cdot A^{0^{-1}\prime}_n$ *and* $M^0_{n_c}$ *respectively. Let* $D^0_n \equiv A^{0^{-1}}_n \cdot B^0_n \cdot A^{0^{-1}\prime}_n$. *If* $\lim_{n\to\infty} \frac{n_x}{n_c} \equiv d_x > 1$ *or* $\lim_{n\to\infty} \frac{n_y}{n_c} \equiv d_y > 1$ *and if* d_x *or* d_y *is 'sufficiently large', then the two-stage maximum likelihood estimator is* not less efficient *than the one-stage estimator. If we let* M_{ij} *denote the* $(i,j)^{th}$ *element of the matrix* $M^0_{n_c}$ *and similarly for* C^0_n, *then a sufficient condition is that* $d_x > \frac{D_{ii}}{M_{ii}}$ *for some* $i \in [1,p]$, *or that* $d_y > \frac{D_{jj}}{M_{jj}}$ *for some* $j \in [p+1,q]$.

The intuition behind this is that we must have enough extra observations on the marginal distributions to offset the loss of information incurred by estimating each marginal distribution separately. If this is the case then the two-stage estimator will be more efficient in the estimation of the marginal parameters. Regardless of the amount of extra information available on the marginal distributions, the two-stage estimates of the copula parameters will always be less asymptotically efficient than the one-stage estimates. It is this fact that leads us now to consider a modification of the two-stage estimator that improves its efficiency.

3.4.1 One-step adjustment of the two-stage estimator

Newey and McFadden (1994) and White (1994) discuss a method of adjusting any consistent, asymptotically normal estimator to make it fully efficient. This method involves taking a single iteration of the Newton-Raphson algorithm from the parameter estimate towards the true parameter.

Theorem 8 (Fully efficient two-stage estimator) *Let Assumptions 1 to 5, and 8 hold, and let the adjusted two-stage estimator be defined as below.*

$$\hat{\theta}_n^* \equiv \hat{\theta}_n - \hat{A}_n^{-1} \cdot \begin{bmatrix} n_x^{-1} \sum_{t=1}^{n_x} s_{1t}\left(Z^t; \hat{\varphi}_{n_x}\right) \\ n_y^{-1} \sum_{t=1}^{n_y} s_{2t}\left(Z^t; \hat{\gamma}_{n_y}\right) \\ n_c^{-1} \sum_{t=1}^{n_c} s_{3t}\left(Z^t; \hat{\theta}_n\right) \end{bmatrix} \tag{23}$$

where $\hat{\theta}_n$ is defined in equation (12) and $\hat{A}_n$, s_{1t}, s_{2t} and s_{3t} are as defined in Lemmas 1 and 2. Then $\hat{\theta}_n^*$ attains the minimum asymptotic variance bound.

The above result is very powerful, in that it shows that we may employ the computationally simpler two-stage estimator presented above, modify it without need of further optimisation, and achieve a fully efficient maximum likelihood estimator. An estimator for the asymptotic covariance matrix of $\hat{\theta}_n^*$ is given below.

Proposition 2 (Covariance matrix estimator for $\hat{\theta}_n^*$) *Given assumptions 1 to 3, 5, 6 and 9 then the asymptotic covariance matrix of $\hat{\theta}_n^*$ defined in Theorem 8 may be consistently estimated by $\hat{M}_{n_c}$, where*

$$\begin{aligned} \hat{M}_{n_c} &\equiv \hat{B}_n^{*-1}, \\ \hat{B}_n^{*-1} &\equiv n_c^{-1} \sum_{t=1}^{n_c} \begin{bmatrix} \hat{s}_{1t}^* \cdot \hat{s}_{1t}^{*\prime} & \hat{s}_{1t}^* \cdot \hat{s}_{2t}^{*\prime} & \hat{s}_{1t}^* \cdot \hat{s}_{3t}^{*\prime} \\ \hat{s}_{2t}^* \cdot \hat{s}_{1t}^{*\prime} & \hat{s}_{2t}^* \cdot \hat{s}_{2t}^{*\prime} & \hat{s}_{2t}^* \cdot \hat{s}_{3t}^{*\prime} \\ \hat{s}_{3t}^* \cdot \hat{s}_{1t}^{*\prime} & \hat{s}_{3t}^* \cdot \hat{s}_{2t}^{*\prime} & \hat{s}_{3t}^* \cdot \hat{s}_{3t}^{*\prime} \end{bmatrix} \end{aligned}$$

and $\hat{s}_{1t}^* \equiv \hat{s}_{1t} + \nabla_\varphi \log c_t\left(Z^t; \hat{\theta}_n^*\right)$, $\hat{s}_{2t}^* \equiv \hat{s}_{2t} + \nabla_\gamma \log c_t\left(Z^t; \hat{\theta}_n^*\right)$ and $\hat{s}_{3t}^* = \hat{s}_{3t}$.

Estimation of even moderately-sized bivariate time-varying conditional density models is quite a computational burden, and the estimation of time-varying conditional density models of higher dimension may be near impossible. The methods presented in this section allow for the fully efficient estimation of the parameters of a time-varying conditional density model in a less computationally burdensome manner. If the unknown parameters may be separated into those associated with a particular marginal distribution or with the copula, then we may estimate the elements of the parameter vector separately. These estimates are still consistent and asymptotically normal, under some conditions, though are not efficient. By applying the one-step adjustment of the estimator presented above, we obtain a fully efficient estimator. It is hoped that the methods presented here enable future researchers to more easily estimate multivariate models of higher dimensions than the current standard of two.

4 Small sample properties

In this section we present the results of a Monte Carlo study of the small sample properties of the estimators discussed above. A general study of the small sample properties of these estimators is not possible; each particular data generating process (DGP) must be considered separately.

4.1 Simulation design

We consider three different DGPs, with specifications that were constructed to resemble those commonly found in models of daily financial data: weak dependence in the conditional mean, and highly persistent conditional variance. All three DGPs are bivariate distributions, with both marginals being conditionally normal with the same AR(1)-GARCH(1,1) specifications:

$$
\begin{aligned}
X_t &= 0.01 + 0.05 X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N\left(0, h_t^x\right) \\
h_t^x &= 0.05 + 0.1 \varepsilon_{t-1}^2 + 0.85 h_{t-1}^x \\
Y_t &= 0.01 + 0.05 Y_{t-1} + \eta_t, \quad \eta_t \sim N\left(0, h_t^y\right) \\
h_t^y &= 0.05 + 0.1 \eta_{t-1}^2 + 0.85 h_{t-1}^y
\end{aligned}
$$

The DGPs differ in the amount of dependence between the two variables. We examine the case that the variables have the Clayton copula, with the copula parameter chosen so as to imply rank correlations of 0.25, 0.50 and 0.75; low, medium and high dependence.

$$
\begin{aligned}
\left(X_t, Y_t\right) | \mathcal{F}_{t-1} &\sim H \equiv Clayton\left(Normal, Normal; \kappa\right) \\
\kappa &= 0.41,\ 1.10 \text{ or } 2.50.
\end{aligned}
$$

We do not consider DGPs with time-varying conditional dependence, nor time-varying higher order marginal moments, in order to keep the simulation tractable: we compare the two-stage estimator with the one-stage estimator, the latter estimator being quite computationally difficult even for relatively simple models such as the one above.

In addition to the three DGPs, we consider six possible data situations: $n_x = 1500$ and 3000, and $n_y/n_x = 0.25$, 0.50 and 0.75. In all cases we assume that $n_c = n_y$. The three estimators considered are the two-stage estimator, $\hat{\theta}_n \equiv \left[\hat{\varphi}'_{n_x}, \hat{\gamma}'_{n_y}, \hat{\kappa}'_{n_c}\right]'$, the one-step efficient two-stage estimator, $\hat{\theta}^*_n$, and the standard one-stage estimator, $\hat{\theta}^{eff}_{n_c}$. We will compare the estimators by looking at their mean squared error (MSE):

$$MSE\left(\hat{\theta}\right) \equiv \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_i - \theta_0\right)^2$$

where $\hat{\theta}$ is the estimator, θ_0 is the true parameter, $\hat{\theta}_i$ is the estimate based on the i^{th} Monte Carlo replication and $N = 1000$ is the number of Monte Carlo replications.

4.2 Results

Let us firstly contemplate what we expect to find. We expect that $\hat{\varphi}_{n_x}$ has a lower MSE than $\hat{\varphi}^{eff}_{n_c}$ for small n_y/n_x and for low dependence cases. This is because the cost of using only the common sample is higher, and the gains from using information on Y are lower. In no case would we expect that $\hat{\gamma}_{n_y}$ is better than $\hat{\gamma}^{eff}_{n_c}$, given the above set-up. For $\hat{\kappa}_{n_c}$ versus $\hat{\kappa}^{eff}_{n_c}$ it is not clear what to expect: theoretically we know that the one-stage estimator is more efficient, but in small samples there may be gains from using a better estimate of φ_0 than obtainable from the one-stage estimator. If the gains exist at all, we would expect them to be greatest for small n_y/n_x and low levels of dependence. For small samples it is also not clear how $\hat{\theta}^*_n$ will compare to $\hat{\theta}^{eff}_{n_c}$, as both are asymptotically fully efficient. The fact that $\hat{\theta}^*_n$ requires an estimate of the covariance matrix of $\hat{\theta}_n$ may lead to some differences, however.

We computed the ratio of MSEs of the two-stage estimator to the one-stage estimator, and the one-step efficient estimator to the one-stage estimator for each of the eleven parameters of the model. A ratio of less than one indicates that the estimator has a lower MSE than the one-stage estimator. To simplify interpretation, we discuss only a summary of the complete results, presented in Tables 1 to 3. The complete results are in Tables 4 to 6. For the summary results, we present the average of the first marginal distribution's five parameter MSE ratios, and similarly for the second marginal distribution[6]. The copula contains only one parameter, and so we present the actual ratio of MSEs in this case. Table 1 presents the results for the two-stage estimator and Table 2 presents the results for the one-step efficient two-stage estimator.

[INSERT TABLE 1 HERE]

For the low dependence case (rank correlation of 0.25) we can see that for none of the combinations of n_x and n_y considered was the one-stage estimator as good as the two-stage estimator for the parameters of the first margin; all MSE ratios are less than one. The two-stage estimator was slightly worse than the one-stage estimator for the parameters of the second margin, with the MSE

[6]The complete results in Tables 4 to 6 confirm that examining only at the mean ratio for each marginal distribution does not distort the general conclusions.

ratios ranging from 1.07 to 1.51. For the copula parameter the two-stage estimator was slightly better than the one-stage estimator in all but one case, and was much better than the one-stage parameter in the case that $n_x = 1500$ and $n_y/n_x = 0.25$. Thus we see that although asymptotically the two-stage estimator is known to be less efficient than the one-stage estimator for the copula parameter, in small samples the improved estimates of the first margin parameters outweigh the loss of information incurred through multi-stage estimation.

As expected, greater dependence generally leads to higher MSE ratios - the loss of information from using the two-stage estimator rather than the one-stage estimator is greater for rank correlations of 0.50 and 0.75. Notice, however, that the MSE ratio of the copula parameters do not change very much with the level of dependence. In most cases this ratio is close to one, indicating that in terms of this parameter the two estimators are equally good.

Overall the results presented in Table 1 are surprisingly positive: the MSEs of the two-stage estimator are, in many cases, smaller than those of the one-stage estimator, and in the cases where the two-stage estimator MSEs are greater, this increase is moderate. Of course, these results are specific to the DGPs selected for this Monte Carlo study; similar results cannot be assumed for all possible DGPs.

While the results for the two-stage estimator were surprisingly good, the results for the one-step efficient estimator are surprisingly bad. In only a couple of cases did the one-step efficient estimator outperform the one-stage estimator, and in numerous cases the ratio of MSEs was very large, in some cases greater than 100. Asymptotically we know that the one-step efficient estimator is fully efficient, and so is better, asymptotically, than the two-stage estimator. It appears, however, that the use of an estimated Hessian and vector of scores leads to poor small sample properties in the one-step efficient estimator[7,8]. The MSE ratios for the parameters of the first margin are generally good, presumably owing to the larger amount of data and the better Hessian and score vector estimates. The MSE ratios for the copula parameter are also moderate for the larger data sets and higher n_y/n_x ratios for the same reason.

[INSERT TABLE 2 HERE]

[7]There are two potential sources of error here: the first is the standard small sample variability of estimates, and the second is the possible numerical error introduced by using numerical derivatives rather than analytical derivatives. Analytical derivatives for the model considered would be very cumbersome to derive and program due to the probability integral transformations that are employed. Given this, it seems likely that most researchers estimating a copula model would use numerical derivatives rather than analytical derivatives, making our use of numerical derivatives appropriate.

[8]The small sample properties of the one-step efficient estimator may be improved by employing a modification of this estimator proposed by Newey (1987). We do not investigate this possibility here. That Newey's modification still requires the use of estimates of the hessian and scores suggests the possibility that the modified one-step efficient estimator would merely match the performance of the two-stage estimator, rather than outperform it.

The very high MSE ratios for the one-step efficient estimator are driven by the fact that the parameter estimates are in many cases outside of the feasible region. As $n \to \infty$ this problem does not arise, but clearly in finite samples one needs to be conscious of it. In Table 3 below we present the MSE ratios for a 'modified' one-step efficient estimator: we set the parameter estimate to be just inside the boundary of the feasible region in the case that the one-step efficient estimator is outside the boundary[9].

[INSERT TABLE 3 HERE]

The results presented in Table 3 indicate that the modified one-step efficient estimator has better small sample properties than the original one-step efficient estimator, however it is still worse than the two-stage estimator.

On the whole, these results suggest that the two-stage estimator is a reasonable alternative to the fully efficient one-stage estimator. There were numerous situations where the two-stage estimator actually outperformed the one-stage estimator, as the two-stage estimator is able to exploit all available information on both variables. In the cases where a loss of efficiency was incurred this loss was moderate. Our simulation results also suggest that the two-stage estimator is preferable to the one-step efficient two-stage estimator. The need for estimates of the Hessian and scores for the one-step efficient estimator induces greater mean squared error in the finite samples considered.

5 A model of the euro and yen exchange rates

In this section we apply the methods discussed above to a model of the conditional joint distribution of daily Japanese yen - U.S. dollar and euro - U.S. dollar exchange rates. The data set employed runs from January 1991 to June 2001 for the yen, and from January 1999 to June 2001 for the euro, so that $n_x = 2695$ and $n_y = n_c = 643$. The data are plotted in Figure 2. It is possible that the fact that the euro came out on January 1, 1999, rather than some other date, carries useful information on the conditional distribution of the euro/dollar exchange rate, i.e. that the missing-data mechanism cannot be ignored. We will assume, however, that we can ignore the missing-data mechanism[10].

[9] For the GARCH inequality constraint, $\alpha + \beta < 1$, we do the following. Let $\hat{\alpha}_n^*$ and $\hat{\beta}_n^*$ be the one-step efficient parameter estimates. The modified one-step efficient parameter estimates are set to: $\hat{\alpha}_n^\bullet \equiv 0.9999\hat{\alpha}_n^* / \left(\hat{\alpha}_n^* + \hat{\beta}_n^*\right)$ and $\hat{\beta}_n^\bullet \equiv 0.9999\hat{\beta}_n^* / \left(\hat{\alpha}_n^* + \hat{\beta}_n^*\right)$.

[10] For example, if enough countries now using the euro had failed to meet the requirements laid down for joining, it is conceivable that the emergence of the euro would have been delayed. Thus, the fact that such a delay did not occur may carry information on the economic performance of the countries now using the euro, and possibly also on the conditional distribution of the euro itself. This possibility is left for future research.

[INSERT FIGURE 2 HERE]

The yen/dollar and euro/dollar rates are the two most frequently traded exchange rates, representing over 50% of total foreign exchange turnover[11] in 1998. The significance of these two exchange rates in the global foreign exchange market and the fact that there exist quite different amounts of data on each of these variables make them a prime application for the estimator introduced above: market participants can neither wait for more euro data to arrive, nor are they willing to throw away the additional information they have on the yen. Using the estimator from Section 3, we can use all available data on each exchange rate.

The Student's t distribution has previously been found to provide a good fit to individual exchange rates, see Bollerslev (1987) and Patton (2001a) for example, and so we employ it for the marginal distributions of both the yen and the euro exchange rates. For the yen margin an AR(1,10) model was estimated for the conditional mean, and a GARCH(1,1) model was estimated for the variance. The euro data exhibited no significant time variation in either the conditional mean or the conditional variance, and so these conditional moments were set to constants[12]. The estimated parameters and standard errors for these marginal distributions are presented in Table 7 below.

[INSERT TABLE 7 HERE]

The figures under the heading 'efficient two-stage' are the parameter estimates found by applying the one-step adjustment to the standard two-stage estimates, as discussed in Section 3.4.1. These estimates require a specification for the copula, and above we present the results obtained using Clayton's copula[13], which is discussed below. The reader can see that the parameter estimates do not differ greatly between the standard two-stage and the efficient two-stage methods, though the standard errors are quite different: specifically, the standard errors on the yen parameters are much greater for the efficient two-stage estimator than the standard two-stage estimator. Clearly in our case the gains from using a fully efficient estimator are outweighed by the fact that we cannot use all available data on the yen. Given this fact, and the finding in Section 4 that the two-stage estimator has better small sample properties than the one-step efficient estimator, we will concentrate below on the results from the standard two-stage estimator.

[11] Source: Bank for International Settlements (1999). The BIS only produces this report every three years, and so no figures for the euro itself are yet available. The 1998 figure for the individual currencies now in the euro is approximately 30.8%. The corresponding figure for the yen - U.S. dollar exchange rate is 20.9%.

[12] The absence of significant time variation in the conditional mean and variance is almost certainly due to the small amount of data available. With time, and more data, we expect that this hypothesis will be rejected.

[13] The results obtained using the other two copulas considered in this paper were not substantially different, and so are not presented. They are available from the author upon request.

The evaluation of the goodness-of-fit of the models for the marginal distributions is of critical importance: the joint distribution of the transformed variables, $U_t \equiv F_t(Z^t; \hat{\varphi}_{n_x})$ and $V_t \equiv G_t(Z^t; \gamma_{n_y})$, will be modelled with a copula, which has by construction margins that are Uniform$(0,1)$. If the marginal distribution models are misspecified then the variables U_t and V_t will not be uniform and the copula will be misspecified. In light of this, we employ a number of tests of the marginal specifications. It should be pointed out that all of the tests we employ ignore the impact of sampling variability in the parameter estimates on the distributions of $\{U_t\}_{t=1}^n$ and $\{V_t\}_{t=1}^n$. The first two tests follow Diebold, *et al.* (1998), who suggested testing that $U_t \sim i.i.d.\ Unif(0,1)$ and $V_t \sim i.i.d.\ Unif(0,1)$ in two stages: firstly testing that U_t and V_t are *i.i.d.* via LM tests, and then testing that they are Uniform(0,1). We test the *i.i.d.* assumption by regressing $(U_t - \bar{U}_n)^k$ and $(V_t - \bar{V}_n)^k$ on twenty lags of both variables for $k = 1, 2, 3, 4$. We test the $Unif(0,1)$ hypothesis via the well-known Kolmogorov-Smirnov test. The results of these tests are presented in Table 8 below. As this table shows, both marginal distribution models pass both tests.

[INSERT TABLE 8 HERE]

We employ two further tests, suggested in Patton (2001a). These tests jointly test the hypotheses of *i.i.d.* and uniformity via 'hit' tests. The support of the distribution is divided into five regions, R_j, according to quantiles, with boundaries at 0, 0.1, 0.25, 0.75, 0.9 and 1, representing the extreme upper and lower tails, the intermediate upper and lower tails, and the centre of the distribution. The hit random variable is defined as taking the value 1 if the transformed variable (U_t or V_t) lies in the region and zero else. That is: $Hit_{jt}^X \equiv \mathbf{1}\{U_t \in R_j\}$ and $Hit_{jt}^Y \equiv \mathbf{1}\{V_t \in R_j\}$, for $j = 1, 2, .., 5$. Under the null hypothesis that $U_t \sim i.i.d.\ Unif(0,1)$ we have that $Hit_{jt}^X \sim i.i.d.$ $Bernoulli\left(\Pr_{\hat{F}}\left[Hit_{jt}^X\right]\right)$. We may then test this hypothesis for each of the five regions. Testing each region separately enables us to see if deficiencies exist in the model's fit in particular regions (such as in the tails if regions 1 and 5 are misspecified, or in uncaptured skewness, if both upper or both lower regions are misspecified). We may also test the joint hypothesis that all five regions are well specified via a multinomial test, also described in Patton (2001a). We perform these tests, including as regressors the number of hits in the last day, week and month to test for serial dependence, and a constant to test for misspecification of the conditional density. The results are presented in Table 9 below.

[INSERT TABLE 9 HERE]

The above table shows that both margins pass the individual hit tests for all five regions and pass the joint test also. We thus conclude that the marginal distributions are adequately modelled, and proceed to the modelling of the copula.

There is a vast literature in statistics on the generation of families of copulas, though only a few have been used in modelling. For the purposes of comparison we will estimate three of the more common copulas: the Gaussian, or normal, copula, the Plackett copula and the Clayton[14] copula. The functional forms of these three copulas are presented in Appendix 4, and further details on them may be found in Nelsen (1999) or Joe (1997). The normal copula is the copula associated with the bivariate normal distribution, and thus is the dependence function implicitly assumed whenever the bivariate normal distribution is used. Plackett's (1965) copula is symmetric, like the normal copula, but exhibits less dependence in the (bivariate) tails of the distribution. Clayton's (1978) copula is an asymmetric copula, exhibiting greater dependence in the negative tail than in the positive tail. In Figure 3 we present diagrams to help illustrate the three copulas. Copulas themselves are difficult to develop intuition for graphically, and so we present instead the joint distributions formed by using particular copulas to link together two standard normal random variables. In all cases the parameter of the copula is calibrated to yield a linear correlation coefficient of 0.5. The three left panels present contour plots of the three joint densities. These panels clearly show the asymmetry implied by Clayton's copula, and show that both the normal and Plackett copulas are symmetric. The three right panels examine the different dependence structures by looking at the conditional density of one variable, X, given the other, Y, for three possible values of Y: -1.96, 0 and 1.96. These panels show that Clayton's copula has greater dependence for $Y < 0$, and less dependence for $Y > 0$. They also show that the conditional density using Plackett's copula is skewed for $Y \neq 0$, and that it has more probability mass at zero than the normal when $Y = 0$.

[INSERT FIGURE 3 HERE]

Little evidence of time variation in the conditional copula was found[15], and so only constant versions of the copulas were fitted. See Patton (2001a and 2001b) for examples of time-varying conditional copula models of exchange rates and stock returns, respectively. Specification tests, discussed below, indicate that the assumption of a constant conditional copula could not be rejected for this data set. The estimated copula parameters and copula likelihoods are presented in Table 10 below.

[INSERT TABLE 10 HERE]

This table shows that the Clayton copula had the best fit, in terms of the likelihood function. As these copulas are non-nested, however, we cannot conduct standard likelihood ratio tests to determine whether the improvement in the copula likelihood was significant. In the place of a standard likelihood ratio test, we employed a recent test proposed by White (2000), and implemented by Sullivan, *et al.*, (1999), Sullivan, *et al.*, (2001) and Hansen and Lunde (2001), called the

[14] Joe (1997) refers to this copula as the Kimeldorf and Sampson copula.

[15] As for the euro margin, we expect that this is due to the small amount of data available. It should be noted that although the copula and the euro margin are both assumed constant, the presence of time variation in the yen margin makes our model a time-varying conditional density model.

'reality check'. This test may be used to determine whether a particular model is as good as the best alternative model considered, according to some performance measure. We used the likelihood ratio as our performance measure. The reality check was originally developed to control for data snooping, where researchers search over thousands of models to find a good fit, but it can be equally well applied in our situation with just three models. We employed the bootstrap version of this test, and used the stationary bootstrap of Politis and Romano (1994) to deal with our non-*i.i.d.* data. The reader is referred to White (2000) for a detailed description of the implementation of this test. We performed the test three times, once with each model as the null, and found that the normal copula was the only one that could be rejected: the p-value for this test was 0.08 indicating that at the 10% alpha level we had evidence that the normal copula was not as good as the best alternative. Neither the Plackett nor the Clayton copulas could be rejected as being equal to the best, with p-values of 0.23 and 0.57 respectively.

Finally, we conducted specification tests on the three copulas, using the bivariate extension of the hit tests used for the marginal models. We defined seven regions of interest in the support of the copula, depicted in Figure 4, and tested the goodness-of-fit in the individual regions and in all regions simultaneously for all three copula models.

[INSERT FIGURE 4 HERE]

In Table 11 below we present the results of the specification tests for the three copula models. These results show that all three copulas passed the joint test of correct specification in all regions, but only the Clayton copula passed all seven individual region tests; the normal and the Plackett copulas failed at the 10% alpha level in region 1, the extreme negative bivariate tail. This is precisely the region where the Clayton copula differs most from the other two copulas, in that the Clayton copula suggests strong dependence in the lower tail, while the normal and Plackett copulas do not. It appears that the dependence between the yen-dollar and euro-dollar exchange rates is better modelled by a copula that captures the increased dependence in this negative tail, such as the Clayton copula.

[INSERT TABLE 11 HERE]

The increased dependence in the lower (joint) tail of the bivariate distribution of the yen/dollar and euro/dollar exchange rates suggests that these variables are more dependent during depreciations of the dollar than during appreciations of the dollar. This is in contrast with the results of Patton (2001a), who found that the yen/dollar and German mark/dollar exchange rates were more dependent during appreciations of the dollar than depreciations. The implication of these two results is that the type of asymmetry (greater dependence in the lower joint tail than the upper joint tail, or vice versa) in the dependence function of exchange rates may be time-varying. In our short sample on the euro we did not detect any time-variation in dependence, however for longer time series switches in the type of asymmetry may be important, and need to be captured in the copula model.

6 Conclusion

This paper proposed a two-stage maximum likelihood estimator for parametric copula models for time series, in the framework of Newey and McFadden (1994) and White (1994). The use of this estimator greatly eases the computational burden associated with the estimation of time-varying multivariate density models. We showed in this paper that the estimator is consistent and asymptotically normal, under standard conditions, and provided a consistent estimator of its covariance matrix.

The estimator is also flexible enough that the case that unequal amounts of data are available on each variable is easily handled. Numerous situations exists where we have differing amounts of data on the variables of interest: models of developed and emerging markets, models of recently floated stocks and the market portfolio, and models involving the euro. Our estimator may be interpreted as an extension of that of Anderson (1957) and Stambaugh (1997) to more irregular data sets, and to non-normal, serially dependent random variables.

A benefit of this estimator is that, when adjusted as in equation (23), it attains the minimum asymptotic variance bound, and so is fully efficient. A simulation study also showed that the unadjusted two-stage estimator has good finite sample properties. Possibly the main drawback of this method is that a complete specification of the conditional density is required, and that this specification must be assumed to be correct. Many of the copula estimators in the statistics literature, see Genest and Rivest (1993), Genest, *et al.,* (1995) and Capéraà, *et al.,* (1997), were developed so that the researcher was able to obtain an estimate of the copula without specifying the marginal distributions. The extension of these methods to the time series case is not straightforward: even if we abstract from the shape of the marginal distributions we must still model the dynamics, and this is not easily done in a nonparametric fashion. The impact of misspecification of one or more marginal distributions is an important problem, and will be addressed in future research.

We applied our estimator to a model of the joint distribution of daily Japanese yen - U.S. dollar and euro - U.S. dollar exchange rates. These rates are the two most frequently traded exchange rates, and have the characteristic that we have much more data available on the yen than we do on the euro. We estimated three different copula models and found some evidence that the Clayton copula, which allows for asymmetry in the dependence structure, provided a better fit than the other two copulas, which impose symmetric dependence. Applications of the theory of copulas appears fertile ground for research. We hope that the results presented in this paper provide some assistance in its application to the modelling of economic time series.

7 Appendix 1: Proofs for Section 2

Proof of Theorem 1. See the proof of Theorem 2.1.4 of Casella and Berger (1990) ■

Proof of Theorem 2. See the proof of Theorem 1 of Patton (2001a). ■

Proof of Corollary 1. See the proof of Corollary 1 of Patton (2001a). ■

8 Appendix 2: Assumptions for Section 3

Presented below are the assumptions required at some stage in the proofs of the theorems. They are collected here for convenience and ease of reference. Most of these assumptions are based on those presented in White (1994). In addition to the assumptions below we make the usual assumptions that observed data are a realisation of a stochastic process on a complete probability space and that all functions are measurable.

Assumption 1 (Conditions on the log-likelihoods) *(a)(i) For each $\varphi \in \Phi$, $E\left[\log f\left(Z^t, \varphi\right)\right]$ exists and is finite, $t = 1, 2, ...$;*

(ii) For each $\gamma \in \Gamma$, $E\left[\log g\left(Z^t, \gamma\right)\right]$ exists and is finite, $t = 1, 2, ...$;

(iii) For each $\theta \in \Theta$, $E\left[\log c\left(Z^t, \theta\right)\right]$ exists and is finite, $t = 1, 2, ...$;

(b)(i) $E\left[\log f\left(Z^t, \cdot\right)\right]$ is continuous on Φ, $t = 1, 2, ...$;

(ii) $E\left[\log g\left(Z^t, \cdot\right)\right]$ is continuous on Γ, $t = 1, 2, ...$;

(iii) $E\left[\log c\left(Z^t, \cdot\right)\right]$ is continuous on Θ, $t = 1, 2, ...$;

(c) $\left\{\log f\left(Z^t, \theta\right)\right\}$, $\left\{\log g\left(Z^t; \gamma\right)\right\}$ and $\left\{\log c\left(Z^t; \theta\right)\right\}$ obey the weak uniform law of large numbers.

Assumption 2 *$\left\{n_x^{-1} \sum_{t=1}^{n_x} E\left[\log f\left(Z^t; \varphi\right)\right]\right\}$ and $\left\{n_y^{-1} \sum_{t=1}^{n_y} E\left[\log g\left(Z^t; \gamma\right)\right]\right\}$ are $\mathcal{O}(1)$ uniformly on Φ and Γ respectively, and $\left\{n_x^{-1} \sum_{t=1}^{n_x} E\left[\log f\left(Z^t; \cdot\right)\right]\right\}$ and $\left\{n_y^{-1} \sum_{t=1}^{n_y} E\left[\log g\left(Z^t; \cdot\right)\right]\right\}$ have unique maximisers φ_0 and γ_0 interior to Φ and Γ.*

Assumption 3 *$f\left(Z^t; \cdot\right)$, $g\left(Z^t; \cdot\right)$ and $c\left(Z^t; \cdot\right)$ are continuously differentiable of order 2 on Φ, Γ and Θ respectively almost surely, $t = 1, 2, ...$*

Assumption 4 (Conditions on the scores) *(a)(i) For all $\varphi \in \Phi$, $E\left[n_x^{-1} \sum_{t=1}^{n_x} s_1\left(Z^t; \varphi\right)\right] < \infty$*

(ii) For all $\gamma \in \Gamma$, $E\left[n_y^{-1} \sum_{t=1}^{n_y} s_2\left(Z^t; \gamma\right)\right] < \infty$

(iii) For all $\theta \in \Theta$, $E\left[n_c^{-1} \sum_{t=1}^{n_c} s_3\left(Z^t; \theta\right)\right] < \infty$

(b)(i) $E\left[n_x^{-1} \sum_{t=1}^{n_x} s_1\left(Z^t; \varphi\right)\right]$ is continuous on Φ uniformly in $n_x = 1, 2, ...$

(ii) $E\left[n_y^{-1}\sum_{t=1}^{n_y} s_2\left(Z^t;\gamma\right)\right]$ *is continuous on* Γ *uniformly in* $n_y = 1, 2, ...$

(iii) $E\left[n_c^{-1}\sum_{t=1}^{n_c} s_3\left(Z^t;\theta\right)\right]$ *is continuous on* Θ *uniformly in* $n_c = 1, 2, ...$

(c)(i) $\left\{s_1\left(Z^t;\varphi\right)\right\}$ *obeys the weak uniform law of large numbers*

(ii) $\left\{s_2\left(Z^t;\gamma\right)\right\}$ *obeys the weak uniform law of large numbers*

(iii) $\left\{s_3\left(Z^t;\theta\right)\right\}$ *obeys the weak uniform law of large numbers*

where $s_1\left(Z^t;\varphi\right) \equiv \nabla_\varphi \log f\left(Z^t;\varphi\right)$, $s_2\left(Z^t;\gamma\right) \equiv \nabla_\gamma \log g\left(Z^t;\gamma\right)$ *and* $s_3\left(Z^t;\theta\right) \equiv \nabla_\kappa \log c\left(Z^t;\theta\right)$ *are the vectors of scores.*

Assumption 5 (Conditions on the hessians) *(a)(i) For all* $\varphi \in \Phi$, $E\left[n_x^{-1}\sum_{t=1}^{n_x} \nabla_{\varphi\varphi} \log f\left(Z^t;\varphi\right)\right] < \infty$, $n_x = 1, 2, ...$

(ii) For all $\gamma \in \Gamma$, $E\left[n_y^{-1}\sum_{t=1}^{n_y} \nabla_{\gamma\gamma} \log g\left(Z^t;\gamma\right)\right] < \infty$, $n_y = 1, 2, ...$

(iii) For all $\theta \in \Theta$, $E\left[n_c^{-1}\sum_{t=1}^{n_c} \nabla_{\kappa\kappa} \log c\left(Z^t;\theta\right)\right]$, $E\left[n_c^{-1}\sum_{t=1}^{n_c} \nabla_{\varphi\kappa} \log c\left(Z^t;\theta\right)\right]$ *and* $E\left[n_c^{-1}\sum_{t=1}^{n_c} \nabla_{\gamma\kappa} \log c\left(Z^t;\theta\right)\right]$ *are* $< \infty$, $n_c = 1, 2, ...$

(b)(i) $E\left[n_x^{-1}\sum_{t=1}^{n_x} \nabla_{\varphi\varphi} \log f\left(Z^t;\varphi\right)\right]$ *is continuous on* Φ *uniformly in* $n_x = 1, 2, ...$

(ii) $E\left[n_y^{-1}\sum_{t=1}^{n_y} \nabla_{\gamma\gamma} \log g\left(Z^t;\gamma\right)\right]$ *is continuous on* Γ *uniformly in* $n_y = 1, 2, ...$

(iii) $E\left[n_c^{-1}\sum_{t=1}^{n_c} \nabla_{\kappa\kappa} \log c\left(Z^t;\theta\right)\right]$, $E\left[n_c^{-1}\sum_{t=1}^{n_c} \nabla_{\varphi\kappa} \log c\left(Z^t;\theta\right)\right]$ *and* $E\left[n_c^{-1}\sum_{t=1}^{n_c} \nabla_{\gamma\kappa} \log c\left(Z^t;\theta\right)\right]$ *are continuous on* Θ *uniformly in* $n_c = 1, 2, ...$

(c)(i) $\left\{\nabla_{\varphi\varphi} \log f\left(Z^t;\varphi\right)\right\}$ *obeys the weak uniform law of large numbers*

(ii) $\left\{\nabla_{\gamma\gamma} \log g\left(Z^t;\gamma\right)\right\}$ *obeys the weak uniform law of large numbers*

(iii) $\left\{\nabla_{\kappa\kappa} \log c\left(Z^t;\theta\right)\right\}$, $\left\{\nabla_{\varphi\kappa} \log c\left(Z^t;\theta\right)\right\}$ *and* $\left\{\nabla_{\gamma\kappa} \log c\left(Z^t;\theta\right)\right\}$ *obey the weak uniform law of large numbers.*

Assumption 6 $E\left[n_c^{-1}\sum_{t=1}^{n_c} \nabla_{\kappa\kappa} \log c\left(Z^t;\theta\right)\right]$, $E\left[n_c^{-1}\sum_{t=1}^{n_c} \nabla_{\varphi\kappa} \log c\left(Z^t;\theta\right)\right]$ *and* $E\left[n_c^{-1}\sum_{t=1}^{n_c} \nabla_{\gamma\kappa} \log c\left(Z^t;\theta\right)\right]$ *are* $\mathcal{O}(1)$ *and negative definite uniformly in* n.

Assumption 7 *Let* $\hat{\varphi}$ *and* $\hat{\gamma}$ *be consistent estimators of* φ_0 *and* γ_0. *Then* $\left\{n_c^{-1}\sum_{t=1}^{n_c} \log c\left(Z^t;\hat{\varphi},\hat{\gamma},\kappa\right)\right\}$ *has a unique maximiser* κ_0 *interior to* $\mathcal{K}$.

Let us simplify notation for the following assumption: let $s_{1t}^0 \equiv s_1\left(Z^t;\varphi_0\right)$, and $\hat{s}_{1t} \equiv s_1\left(Z^t;\hat{\varphi}_{n_x}\right)$. Similarly for s_{2t} and s_{3t}. Let us define $\log g\left(Z^t;\gamma\right) = \nabla_\gamma \log g\left(Z^t;\gamma\right) = 0$ for $t > n_y$ and $\log c\left(Z^t;\theta\right) = \nabla_\kappa \log c\left(Z^t;\theta\right) = 0$ for $t > n_c$ to deal with time indices beyond the sample sizes available.

Assumption 8 *The double array* $\left\{\left[n_x^{-1/2}s_{1t}^{0\prime}, n_y^{-1/2}s_{2t}^{0\prime}, n_c^{-1/2}s_{3t}^{0\prime}\right]'\right\}$ *obeys the central limit theorem with covariance matrix* B_n^0, *given below, where* B_n^0 *is* $\mathcal{O}(1)$ *and positive definite.*

$$B_n^0 \equiv \begin{bmatrix} n_x^{-1}\sum_{t=1}^{n_x} E\left[s_{1t}^0 \cdot s_{1t}^{0\prime}\right] & (n_x n_y)^{-1/2}\sum_{t=1}^{n_y} E\left[s_{1t}^0 \cdot s_{2t}^{0\prime}\right] & (n_x n_c)^{-1/2}\sum_{t=1}^{n_c} E\left[s_{1t}^0 \cdot s_{3t}^{0\prime}\right] \\ (n_x n_y)^{-1/2}\sum_{t=1}^{n_y} E\left[s_{2t}^0 \cdot s_{1t}^{0\prime}\right] & n_y^{-1}\sum_{t=1}^{n_y} E\left[s_{2t}^0 \cdot s_{2t}^{0\prime}\right] & (n_y n_c)^{-1/2}\sum_{t=1}^{n_c} E\left[s_{2t}^0 \cdot s_{3t}^{0\prime}\right] \\ (n_x n_c)^{-1/2}\sum_{t=1}^{n_c} E\left[s_{3t}^0 \cdot s_{1t}^{0\prime}\right] & (n_y n_c)^{-1/2}\sum_{t=1}^{n_c} E\left[s_{3t}^0 \cdot s_{2t}^{0\prime}\right] & n_c^{-1}\sum_{t=1}^{n_c} E\left[s_{3t}^0 \cdot s_{3t}^{0\prime}\right] \end{bmatrix} \quad (24)$$

The above definition of the covariance matrix B_n^0 is the natural extension of the standard definition to the case of unequal amounts of data, and reduces to the standard case when $n_x = n_y = n_c$. To see where the unusual scaling figures come from, recall that the covariance matrix is defined as

$$B_n^0 \equiv var\left[\sum\nolimits_{t=1}^{n}\left[n_x^{-1/2}s_{1t}^{0\prime}, n_y^{-1/2}s_{2t}^{0\prime}, n_c^{-1/2}s_{3t}^{0\prime}\right]'\right]$$

Noting that the expectation of the scores are zero at the true parameter, and expanding the above expression for the variance yields equation (24).

Let $B_n(\theta)$ be the matrix B_n evaluated at the point θ, and so B_n^0 defined above equals $B_n(\theta_0)$. We use this definition in the following assumption.

Assumption 9 ***(a)*** *The elements of* B_n *are finite and continuous on* Θ *uniformly in* $n = 1, 2, ...$

(b) *The elements of* $\left\{\left[s_{1t}^{0\prime}, s_{2t}^{0\prime}, s_{3t}^{0\prime}\right]' \cdot \left[s_{1t}^{0\prime}, s_{2t}^{0\prime}, s_{3t}^{0\prime}\right]\right\}$ *obey the weak uniform law of large numbers.*

Andrews (1988), Gallant and White (1988), White (1994) and White (2001) provide some results on laws of large numbers for dependent, heterogeneously distributed random variables that may be used to satisfy assumption 9 (b).

9 Appendix 3: Proofs for Section 3 (draft)

Proof of Theorem 3. See proof of Theorem 3.13 of White (1994). ∎

Proof of Theorem 4. See proof of Theorem 3.10 of White (1994). ∎

Proof of Theorem 5. For the complete proof of the standard two-stage maximum likelihood case see the proof of Theorem 6.11 of White (1994). Below we provide a sketch of the modifications that need to be made to accommodate the differing sample sizes.

Firstly, some more notation: $A(\theta) =$

$$\begin{bmatrix} n_x^{-1}\sum_{t=1}^{n_x} E\left[\nabla_{\varphi\varphi}\log f_t\left(Z^t;\varphi\right)\right] & 0 & 0 \\ 0 & n_y^{-1}\sum_{t=1}^{n_y} E\left[\nabla_{\gamma\gamma}\log g_t\left(Z^t;\gamma\right)\right] & 0 \\ n_c^{-1}\sum_{t=1}^{n_c} E\left[\nabla_{\varphi\kappa}\log c_t\left(Z^t;\theta\right)\right] & n_c^{-1}\sum_{t=1}^{n_c} E\left[\nabla_{\gamma\kappa}\log c_t\left(Z^t;\theta\right)\right] & n_c^{-1}\sum_{t=1}^{n_c} E\left[\nabla_{\kappa\kappa}\log c_t\left(Z^t;\theta\right)\right] \end{bmatrix}$$

so $A(\theta_0) = A_n^0$.

As usual, the proof starts by taking a Taylor series expansion of the scores evaluated at the estimated parameters about the scores evaluated at the true parameters, which equal zero due to the assumption that the true parameters lie in the interior of Θ.

$$0=\begin{bmatrix} n_x^{-1}\sum_{t=1}^{n_x} s_{1t}\left(Z^t;\hat{\varphi}_{n_x}\right) \\ n_y^{-1}\sum_{t=1}^{n_y} s_{2t}\left(Z^t;\hat{\gamma}_{n_y}\right) \\ n_c^{-1}\sum_{t=1}^{n_y} s_{3t}\left(Z^t;\hat{\varphi}_{n_x},\hat{\gamma}_{n_y},\hat{\kappa}_{n_c}\right) \end{bmatrix}=\begin{bmatrix} n_x^{-1}\sum_{t=1}^{n_x} s_{1t}\left(Z^t;\varphi_0\right) \\ n_y^{-1}\sum_{t=1}^{n_y} s_{2t}\left(Z^t;\gamma_0\right) \\ n_c^{-1}\sum_{t=1}^{n_y} s_{3t}\left(Z^t;\theta_0\right) \end{bmatrix}+A\left(\bar{\theta}_n\right)\cdot\begin{bmatrix} \hat{\varphi}_{n_x}-\varphi_0 \\ \hat{\gamma}_{n_y}-\gamma_0 \\ \hat{\kappa}_{n_c}-\kappa_0 \end{bmatrix}$$

where $\bar{\theta}_n \equiv \lambda\hat{\theta}_n+(1-\lambda)\,\theta_0$, and $\lambda\in[0,1]$. So

$$\begin{aligned} A\left(\bar{\theta}_n\right)\cdot\begin{bmatrix} \hat{\varphi}_{n_x}-\varphi_0 \\ \hat{\gamma}_{n_y}-\gamma_0 \\ \hat{\kappa}_{n_c}-\kappa_0 \end{bmatrix} &= -\begin{bmatrix} n_x^{-1}\sum_{t=1}^{n_x} s_{1t}\left(Z^t;\varphi_0\right) \\ n_y^{-1}\sum_{t=1}^{n_y} s_{2t}\left(Z^t;\gamma_0\right) \\ n_c^{-1}\sum_{t=1}^{n_y} s_{3t}\left(Z^t;\theta_0\right) \end{bmatrix} \\ \mathcal{N}^{1/2}\cdot A\left(\bar{\theta}_n\right)\cdot\begin{bmatrix} \hat{\varphi}_{n_x}-\varphi_0 \\ \hat{\gamma}_{n_y}-\gamma_0 \\ \hat{\kappa}_{n_c}-\kappa_0 \end{bmatrix} &= -\mathcal{N}^{1/2}\cdot\begin{bmatrix} n_x^{-1}\sum_{t=1}^{n_x} s_{1t}\left(Z^t;\varphi_0\right) \\ n_y^{-1}\sum_{t=1}^{n_y} s_{2t}\left(Z^t;\gamma_0\right) \\ n_c^{-1}\sum_{t=1}^{n_y} s_{3t}\left(Z^t;\theta_0\right) \end{bmatrix} \\ B_n^{0-1/2}\cdot\mathcal{N}^{1/2}\cdot A\left(\bar{\theta}_n\right)\cdot\begin{bmatrix} \hat{\varphi}_{n_x}-\varphi_0 \\ \hat{\gamma}_{n_y}-\gamma_0 \\ \hat{\kappa}_{n_c}-\kappa_0 \end{bmatrix} &= -B_n^{0-1/2}\cdot\begin{bmatrix} n_x^{-1/2}\sum_{t=1}^{n_x} s_{1t}\left(Z^t;\varphi_0\right) \\ n_y^{-1/2}\sum_{t=1}^{n_y} s_{2t}\left(Z^t;\gamma_0\right) \\ n_c^{-1/2}\sum_{t=1}^{n_y} s_{3t}\left(Z^t;\theta_0\right) \end{bmatrix} \\ B_n^{0-1/2}\cdot\mathcal{N}^{1/2}\cdot A\left(\theta_0\right)\cdot\begin{bmatrix} \hat{\varphi}_{n_x}-\varphi_0 \\ \hat{\gamma}_{n_y}-\gamma_0 \\ \hat{\kappa}_{n_c}-\kappa_0 \end{bmatrix} &= -B_n^{0-1/2}\cdot\begin{bmatrix} n_x^{-1/2}\sum_{t=1}^{n_x} s_{1t}\left(Z^t;\varphi_0\right) \\ n_y^{-1/2}\sum_{t=1}^{n_y} s_{2t}\left(Z^t;\gamma_0\right) \\ n_c^{-1/2}\sum_{t=1}^{n_y} s_{3t}\left(Z^t;\theta_0\right) \end{bmatrix}+o_p(1) \\ &\xrightarrow{\mathcal{D}} N(0,I) \end{aligned}$$

by assumption 8 and Theorem 8.10 of Lehmann and Casella (1998, p58). ■

Proof of Lemma 1. See the proof of Corollary 3.8 of White (1994). ■

Proof of Lemma 2. See the proof of Theorem 8.26 (i) of White (1994). ■

Proof of Theorem 6. We must show that $\hat{A}_n^{-1}\cdot\mathcal{N}^{*-1/2}\cdot\hat{B}_n\cdot\mathcal{N}^{*-1/2}\cdot\hat{A}_n^{-1\prime}-A_n^{0^{-1}}\cdot\mathcal{N}_\infty^{*-1/2}\cdot B_n^0\cdot\mathcal{N}_\infty^{*-1/2}\cdot A_n^{0^{-1\prime}}\xrightarrow{p}0$. The above two lemmas showed that $\hat{A}_n-A_n^0\xrightarrow{p}0$ and $\hat{B}_n-B_n^0\xrightarrow{p}0$, and by definition we have that $\mathcal{N}^*\to\mathcal{N}_\infty^*$. The result then follows from Assumption 6 and Proposition 2.30 of White (2001). ■

Proof of Theorem 7. Follows that of Theorem 5 suitably modified. ■

Proof of Proposition 1. The asymptotic efficiency of asymptotically normal estimators are compared by analysing their asymptotic covariance matrices. Specifically, the one-stage estimator is more efficient than the two-stage estimator if $\left(avar(\hat{\theta}_n) - avar(\hat{\theta}_n^{eff})\right)$ is a positive semi-definite matrix.

Consider the case that $d_x > \frac{D_{11}}{M_{11}}$, so that $d_x^{-1} < \frac{M_{11}}{D_{11}}$. Notice that we may write the first $(p \times p)$ elements of the matrix $A_n^{0^{-1}} \cdot \mathcal{N}_\infty^{*-1/2} \cdot B_n^0 \cdot \mathcal{N}_\infty^{*-1/2} \cdot A_n^{0^{-1\prime}}$ as d_x^{-1} times the first $(p \times p)$ elements of the matrix D_n^0. Let $\boldsymbol{\lambda} = [\lambda, \mathbf{0}]$, where $\mathbf{0}$ is a column vector of $p+q+r-1$ zeros and $\lambda \in \mathbb{R}\backslash\{0\}$. Then the quadratic form

$$\begin{aligned}
\boldsymbol{\lambda}' \cdot \left(avar(\hat{\theta}_n) - avar(\hat{\theta}_n^{eff})\right) \cdot \boldsymbol{\lambda} &= \boldsymbol{\lambda}' \cdot \left(A_n^{0^{-1}} \cdot \mathcal{N}_\infty^{*1/2} \cdot B_n^0 \cdot \mathcal{N}_\infty^{*1/2} \cdot A_n^{0^{-1\prime}} - M_{n_c}^0\right) \cdot \boldsymbol{\lambda} \\
&= \lambda' \cdot \left(d_x^{-1} D_{11} - M_{11}\right) \cdot \lambda \\
&< \lambda' \cdot \left(\frac{M_{11}}{D_{11}} D_{11} - M_{11}\right) \cdot \lambda \\
&= 0
\end{aligned}$$

Whereas if we let $\dot{\boldsymbol{\lambda}} = [\mathbf{0}, \lambda]$ then we find:

$$\begin{aligned}
\dot{\boldsymbol{\lambda}}' \cdot \left(avar(\hat{\theta}_n) - avar(\hat{\theta}_n^{eff})\right) \cdot \dot{\boldsymbol{\lambda}} &= \dot{\boldsymbol{\lambda}}' \cdot \left(A_n^{0^{-1}} \cdot \mathcal{N}_\infty^{*-1/2} \cdot B_n^0 \cdot \mathcal{N}_\infty^{*-1/2} \cdot A_n^{0^{-1\prime}} - M_n^0\right) \cdot \dot{\boldsymbol{\lambda}} \\
&= \lambda' \cdot (D_{s,s} - M_{s,s}) \cdot \lambda \\
&\geq 0 \text{ by the efficiency of the one-stage estimator}
\end{aligned}$$

Thus we find that the difference between the asymptotic covariance matrices under the above assumption is indefinite: neither estimator is more efficient than the other. The two-stage estimator is a more efficient estimator of the parameters in the marginal distribution (in the above case, for the first parameter of the first marginal distribution) while the one-stage estimator is a more efficient estimator of the copula parameters. ■

Proof of Theorem 8. The proof involves showing that $\sqrt{n_c}\left(\hat{\theta}_n^* - \hat{\theta}_n^{eff}\right) \xrightarrow{p} 0$ as $n_c \to \infty$. Allowing for differing sample sizes causes no difficulties here and the proof of Theorem 7.9 of White (1994) obtains. ■

Proof of Proposition 2. Theorem 8 gives us that $\hat{\theta}_n^*$ has the same asymptotic distribution as the one-stage maximum likelihood estimator $\hat{\theta}_n^{eff}$, and thus the same asymptotic covariance matrix. As we have assumed correct specification, we know that the asymptotic covariance matrix of $\hat{\theta}_n^{eff}$ is $\tilde{B}_n^{0^{-1}}$, the inverse of the Fisher information matrix. For this estimator $\tilde{B}_n^0$ takes the form

$$\tilde{B}_n^0 = n_c^{-1} \sum_{t=1}^{n_c} \begin{bmatrix} \tilde{s}_{1t}^0 \cdot \tilde{s}_{1t}^{0\prime} & \tilde{s}_{1t}^0 \cdot \tilde{s}_{2t}^{0\prime} & \tilde{s}_{1t}^0 \cdot \tilde{s}_{3t}^{0\prime} \\ \tilde{s}_{2t}^0 \cdot \tilde{s}_{1t}^{0\prime} & \tilde{s}_{2t}^0 \cdot \tilde{s}_{2t}^{0\prime} & \tilde{s}_{2t}^0 \cdot \tilde{s}_{3t}^{0\prime} \\ \tilde{s}_{3t}^0 \cdot \tilde{s}_{1t}^{0\prime} & \tilde{s}_{3t}^0 \cdot \tilde{s}_{2t}^{0\prime} & \tilde{s}_{3t}^0 \cdot \tilde{s}_{3t}^{0\prime} \end{bmatrix}$$

where $\tilde{s}_{1t}^0 \equiv \nabla_\varphi \log h_t\left(Z^t; \theta_0\right)$, $\tilde{s}_{2t}^0 \equiv \nabla_\gamma \log h_t\left(Z^t; \theta_0\right)$ and $\tilde{s}_{3t}^0 \equiv \nabla_\kappa \log h_t\left(Z^t; \theta_0\right)$, since this estimator maximises the complete joint likelihood, rather than the individual components of the likelihood. Notice that $\nabla_\varphi \log h_t\left(Z^t; \theta_0\right) = \nabla_\varphi \log f_t\left(Z^t; \varphi_0\right) + \nabla_\varphi \log c_t\left(Z^t; \theta_0\right)$, $\nabla_\gamma \log h_t\left(Z^t; \theta_0\right) =$

$\nabla_\gamma \log g_t\left(Z^t;\gamma_0\right) + \nabla_\gamma \log c_t\left(Z^t;\theta_0\right)$ and $\nabla_\kappa \log h_t\left(Z^t;\theta_0\right) = \nabla_\kappa \log c_t\left(Z^t;\theta_0\right) \equiv s_{3t}^0$ for the case that we may write $h_t\left(Z^t;\theta_0\right) = f_t\left(Z^t;\varphi_0\right) \cdot g_t\left(Z^t;\gamma_0\right) \cdot c_t\left(F_t\left(Z^t;\varphi_0\right), G_t\left(Z^t;\gamma_0\right);\kappa_0\right)$, which is what we have assumed.

Under the conditions given we have that the above scores evaluated at the true parameter θ_0 may be consistently estimated by the scores evaluated at the estimated parameter $\hat{\theta}_n^*$. It follows then that $\hat{B}_n^*$ is consistent for $\tilde{B}_n^0$, and thus that $\hat{M}_{n_c}$ is consistent for $M_{n_c}^0$. ■

10 Appendix 4: Normal, Plackett and Clayton copulas

Below we present the functional forms of the three copulas considered in this paper, namely the normal (or Gaussian) copula, Plackett's copula and Clayton's copula. For an extensive collection of bivariate copulas and some details on their characteristics the reader is referred to Chapter 5 of Joe (1997). We will denote the *c.d.f.* form of the copula with an upper case C, and the density form with a lower case c. The functions defined below are only for $(u,v) \in [0,1]^2$, that is, for observations in the support of the copula. The *c.d.f.* of each copula may be extended to the entire real line as described in equation (3), while the copula densities are as given below for $(u,v) \in [0,1]^2$ and zero for $(u,v) \notin [0,1]^2$.

Normal copula

$$\begin{aligned} C(u,v;\rho) &= \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left\{\frac{-\left(r^2+s^2-2\rho rs\right)}{2\left(1-\rho^2\right)}\right\} dr\ ds \\ c(u,v;\rho) &= \frac{1}{\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2-2\rho\cdot\Phi^{-1}(u)\cdot\Phi^{-1}(v)}{2\left(1-\rho^2\right)}\right\} \\ &\quad \cdot \exp\left\{-\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}\right\} \end{aligned}$$

where Φ^{-1} is the inverse of a univariate standard normal *c.d.f.*, for $\rho \in (-1,1)$.

Plackett's copula

$$\begin{aligned} C(u,v;\psi) &= \begin{cases} \frac{1+(\psi-1)(u+v)-\sqrt{(1+(\psi-1)(u+v))^2-4\psi(\psi-1)uv}}{2(\psi-1)} & \text{if } \psi \geq 0,\ \psi \neq 1 \\ u \cdot v & \text{if } \psi = 1 \end{cases} \\ c(u,v;\psi) &= \begin{cases} \frac{\psi(1+(\psi-1)(u+v-2uv))}{\sqrt{\left([1+(\psi-1)(u+v)]^2-4\psi(\psi-1)uv\right)^3}} & \text{if } \psi \geq 0,\ \psi \neq 1 \\ 1 & \text{if } \psi = 1 \end{cases} \end{aligned}$$

Clayton's copula

$$\begin{aligned} C(u,v;\delta) &= \begin{cases} \left(u^{-\delta}+v^{-\delta}-1\right)^{-1/\delta} & \text{if } \delta > 0 \\ u \cdot v & \text{if } \delta = 0 \end{cases} \\ c(u,v;\delta) &= \begin{cases} (1+\delta)(uv)^{-1-\delta}\left(u^{-\delta}+v^{-\delta}-1\right)^{-2-1/\delta} & \text{if } \delta > 0 \\ 1 & \text{if } \delta = 0 \end{cases} \end{aligned}$$

11 Tables

Table 1: *Ratio of two-stage estimator MSE to one-stage estimator MSE*

		$n_y/n_x = 0.25$	$n_y/n_x = 0.50$	$n_y/n_x = 0.75$
		Rank correlation = 0.25		
$n_x = 1500$	First margin	0.1639	0.4243	0.7524
	Second margin	1.0673	1.5109	1.069
	Copula	0.2422	0.9345	0.9244
$n_x = 3000$	First margin	0.2108	0.548	0.7893
	Second margin	1.1819	1.1905	1.0753
	Copula	0.9662	1.0063	0.9784
		Rank correlation = 0.50		
$n_x = 1500$	First margin	0.2312	0.7066	1.6421
	Second margin	1.9149	1.956	1.5235
	Copula	1.0268	0.8991	0.9467
$n_x = 3000$	First margin	0.2666	0.6586	1.2042
	Second margin	1.7432	1.5031	1.5561
	Copula	0.8707	1.0339	0.9714
		Rank correlation = 0.75		
$n_x = 1500$	First margin	0.4815	0.7746	2.2068
	Second margin	3.0706	3.8138	3.4899
	Copula	1.4216	1.0805	0.9381
$n_x = 3000$	First margin	0.5775	1.2669	2.0971
	Second margin	4.0657	2.8347	2.7536
	Copula	0.9913	1.0139	1.0123

Note: This table presents the ratio of the two-stage estimator small sample MSE to the one-stage estimator small sample MSE for each parameter of the joint distribution, for 1000 Monte Carlo replications. We present the average ratios for each marginal distribution, and the actual ratio for the (single) copula parameter. 'n_x' refers to the amount of data on the first margin, while 'n_y/n_x' is the ratio of the amount of data on the second margin to that on the first. We set $n_c = n_y$.

Table 2: *Ratio of one-step efficient estimator MSE to one-stage estimator MSE*

		$n_y/n_x = 0.25$	$n_y/n_x = 0.50$	$n_y/n_x = 0.75$
		Rank correlation = 0.25		
$n_x = 1500$	First margin	0.2196	0.6297	1.1946
	Second margin	112.00	46.514	3.0494
	Copula	4.1904	6.0718	1.0207
$n_x = 3000$	First margin	0.3978	1.1177	1.5469
	Second margin	23.389	6.8785	3.0242
	Copula	2.2087	1.5156	1.1176
		Rank correlation = 0.50		
$n_x = 1500$	First margin	0.3441	1.1493	2.7740
	Second margin	3863.6	149.91	4.1160
	Copula	13.713	2.9769	1.2843
$n_x = 3000$	First margin	0.5039	1.3346	2.5238
	Second margin	460.86	7.7345	4.4457
	Copula	3.9978	2.2217	1.3357
		Rank correlation = 0.75		
$n_x = 1500$	First margin	0.7710	1.1919	4.0205
	Second margin	455.46	28.485	10.026
	Copula	42.352	4.8292	1.6233
$n_x = 3000$	First margin	1.2666	2.7622	4.5771
	Second margin	77.767	15.107	8.0658
	Copula	9.4008	3.2043	1.6235

Note: This table presents the ratio of the one-step efficient two-stage estimator small sample MSE to the one-stage estimator small sample MSE for each parameter of the joint distribution, for 1000 Monte Carlo replications. We present the average ratios for each marginal distribution, and the actual ratio for the (single) copula parameter. 'n_x' refers to the amount of data on the first margin, while 'n_y/n_x' is the ratio of the amount of data on the second margin to that on the first. We set $n_c = n_y$.

Table 3: *Ratio of modified one-step efficient estimator MSE to one-stage estimator MSE*

		$n_y/n_x = 0.25$	$n_y/n_x = 0.50$	$n_y/n_x = 0.75$
		Rank correlation = 0.25		
$n_x = 1500$	First margin	0.2196	0.6297	1.1946
	Second margin	12.082	31.692	3.6890
	Copula	2.2677	1.7333	1.0207
$n_x = 3000$	First margin	0.3978	1.1177	1.5469
	Second margin	10.875	9.6398	3.0242
	Copula	1.8369	1.5156	1.1176
		Rank correlation = 0.50		
$n_x = 1500$	First margin	0.3441	1.1493	2.6678
	Second margin	61.708	20.5412	4.9941
	Copula	11.6722	2.9757	1.2843
$n_x = 3000$	First margin	0.5039	1.3346	2.5238
	Second margin	273.75	12.755	4.4457
	Copula	3.5549	2.2217	1.3357
		Rank correlation = 0.75		
$n_x = 1500$	First margin	0.7710	1.1917	4.0205
	Second margin	43.424	38.905	11.127
	Copula	34.791	4.3804	1.6233
$n_x = 3000$	First margin	1.2666	2.7622	4.5771
	Second margin	68.864	21.263	8.0658
	Copula	4.5003	3.2043	1.6235

Note: This table presents the ratio of the modified one-step efficient two-stage estimator small sample MSE to the one-stage estimator small sample MSE for each parameter of the joint distribution, for 1000 Monte Carlo replications. We present the average ratios for each marginal distribution, and the actual ratio for the (single) copula parameter. 'n_x' refers to the amount of data on the first margin, while 'n_y/n_x' is the ratio of the amount of data on the second margin to that on the first. We set $n_c = n_y$.

Table 4: *Ratio of estimator MSE to one-stage estimator MSE, rank correlation = 0.25*

	Two-stage	*One-step efficient*	*Two-stage*	*One-step efficient*
	$\mathbf{n}_x$**= 1500**		$\mathbf{n}_x$**= 3000**	
		$\mathbf{n}_y/\mathbf{n}_x$**= 0.25**		
μ_x	0.2323	0.2298	0.2585	0.2576
ϕ_x	0.2911	0.2860	0.2662	0.2616
ω_x	0.0349	0.0717	0.1096	0.2588
α_x	0.2015	0.3760	0.2684	0.7540
β_x	0.0597	0.1347	0.1512	0.4569
μ_y	0.9798	6.2254	0.9883	5.4261
ϕ_y	1.0446	7.7988	1.0623	5.7681
ω_y	1.0389	97.336	1.2272	55.915
α_y	1.2395	375.19	1.3074	16.065
β_y	1.0338	73.445	1.3243	33.769
κ_c	0.2422	4.1904	0.9662	2.2087
		$\mathbf{n}_y/\mathbf{n}_x$**= 0.50**		
μ_x	0.5039	0.5009	0.5128	0.5063
ϕ_x	0.5451	0.5360	0.6033	0.5928
ω_x	0.2026	0.4042	0.5273	1.2707
α_x	0.5719	1.0552	0.5414	1.5318
β_x	0.2979	0.6524	0.5552	1.6871
μ_y	0.9577	5.1416	0.9232	4.8861
ϕ_y	1.2059	6.4554	1.1373	6.0349
ω_y	2.1917	128.04	1.3163	11.047
α_y	1.2634	8.4137	1.2498	6.4429
β_y	1.9359	84.523	1.3261	5.9812
κ_c	0.9345	6.0718	1.0063	1.5156
		$\mathbf{n}_y/\mathbf{n}_x$**= 0.75**		
μ_x	0.7272	0.7241	0.7807	0.7748
ϕ_x	0.8142	0.8054	0.8513	0.8371
ω_x	0.6917	1.3817	0.7481	1.7849
α_x	0.8088	1.4706	0.7973	2.0635
β_x	0.7203	1.5914	0.7692	2.2745
μ_y	0.9637	2.1639	0.9857	2.2372
ϕ_y	1.1122	2.6064	1.0789	2.4776
ω_y	1.0220	3.8734	1.0792	3.2528
α_y	1.1820	3.6073	1.1425	3.9945
β_y	1.0647	2.9962	1.0902	3.1587
κ_c	0.9244	1.0207	0.9784	1.1176

Table 5: *Ratio of estimator MSE to one-stage estimator MSE, rank correlation = 0.50*

	Two-stage	*One-step efficient*	*Two-stage*	*One-step efficient*
	n_x= 1500		**n_x= 3000**	
		n_y/n_x= 0.25		
μ_x	0.2512	0.2491	0.2715	0.2698
ϕ_x	0.3756	0.3714	0.3758	0.3709
ω_x	0.0902	0.1879	0.1241	0.2984
α_x	0.3129	0.6131	0.3703	0.9971
β_x	0.1258	0.2990	0.1916	0.5832
μ_y	1.0657	32.818	1.2447	6.6575
ϕ_y	1.5353	37.174	1.4986	8.3499
ω_x	2.8973	436.21	2.4352	1320.3
α_y	1.9235	18360	1.5584	250.74
β_y	2.1528	452.09	1.9792	718.27
κ_c	1.0268	13.713	0.8707	3.9978
		n_y/n_x= 0.50		
μ_x	0.5464	0.5425	0.5581	0.5544
ϕ_x	0.8418	0.8265	0.7321	0.7214
ω_x	0.6582	1.3151	0.6206	1.5039
α_x	0.7865	1.4819	0.7677	2.0066
β_x	0.7003	1.5804	0.6144	1.8866
μ_y	0.9710	5.2243	1.1331	5.6302
ϕ_y	1.6103	8.9915	1.4310	7.4175
ω_y	2.8625	477.96	1.7947	11.119
α_y	1.7234	37.273	1.4951	7.9972
β_y	2.6130	220.12	1.6615	6.5082
κ_c	0.8991	2.9769	1.0339	2.2217
		n_y/n_x= 0.75		
μ_x	0.8058	0.7968	0.8094	0.8033
ϕ_x	1.1858	1.1676	1.2523	1.2300
ω_x	3.1082	5.6013	1.4269	3.3519
α_x	1.1944	2.2870	1.2159	3.2775
β_x	1.9165	4.0172	1.3165	3.9560
μ_y	1.0865	2.5000	1.1060	2.5808
ϕ_y	1.4481	3.4861	1.6464	4.0011
ω_y	1.7967	5.8550	1.7373	5.2610
α_y	1.5795	4.3542	1.6066	5.5399
β_y	1.7066	4.3850	1.6842	4.8457
κ_c	0.9467	1.2843	0.9714	1.3357

Table 6: *Ratio of estimator MSE to one-stage estimator MSE, rank correlation = 0.75*

	Two-stage	*One-step efficient*	*Two-stage*	*One-step efficient*
	n_x= 1500		**n_x= 3000**	
	n_y/n_x= 0.25			
μ_x	0.2921	0.2896	0.2811	0.2790
ϕ_x	0.6762	0.6645	0.5910	0.5826
ω_x	0.4246	0.8552	0.6470	1.5659
α_x	0.5765	1.0606	0.7094	1.9049
β_x	0.4383	0.9849	0.6592	2.0008
μ_y	1.1884	9.1711	1.1063	5.7635
ϕ_y	2.8080	33.257	2.7113	14.901
ω_y	4.1007	949.22	7.9919	238.66
α_y	3.1254	782.74	2.9529	25.426
β_y	4.1306	502.88	5.5663	104.08
κ_c	1.4216	42.352	0.9913	9.4008
	n_y/n_x= 0.50			
μ_x	0.5581	0.5525	0.5841	0.5809
ϕ_x	1.1393	1.1236	1.2162	1.1996
ω_x	0.2172	0.4453	1.4974	3.5455
α_x	1.3921	2.5570	1.4794	3.8855
β_x	0.5663	1.2811	1.5576	4.5994
μ_y	1.1415	5.8964	1.1449	5.7048
ϕ_y	2.5993	14.392	2.4983	13.127
ω_y	7.0964	66.693	4.0292	25.208
α_y	3.0955	25.174	2.9409	17.395
β_y	5.1363	30.271	3.5603	14.101
κ_c	1.0805	4.8292	1.0139	3.2043
	n_y/n_x= 0.75			
μ_x	0.8779	0.8747	0.8327	0.8249
ϕ_x	1.9364	1.9115	1.9923	1.9494
ω_x	3.2606	6.6357	2.6665	6.3185
α_x	2.2108	4.3404	2.3814	6.1699
β_x	2.7484	6.3401	2.6128	7.6227
μ_y	1.1610	2.7589	1.1882	2.6908
ϕ_y	2.6674	6.3029	2.4473	5.7116
ω_y	5.4271	18.380	3.4318	11.270
α_y	3.4128	9.9011	3.2746	10.479
β_y	4.7810	12.786	3.4261	10.177
κ_c	0.9381	1.6233	1.0123	1.6235

Notes to Tables 4, 5 and 6: These tables present the ratio of the mean-squared error of the two-stage estimator and the one-step efficient estimator of a given parameter to the one-stage estimator of that parameter. A value less than (greater than) one indicates that the estimator has lower (higher) MSE than the one-stage estimator. $\mu_x, \phi_x, \omega_x, \alpha_x$ and β_x correspond to the mean, AR parameter, GARCH constant, GARCH innovation and GARCH smoothing parameters for the first margin. These are similarly defined for the second margin. κ_c indicates the parameter of the copula. n_x is the number of observations on the first margin, and n_y/n_x is the ratio of the number of observations on the second margin to those on the first. We set $n_c = n_y$. Tables 4, 5 and 6 present the results for rank correlations of 0.25, 0.50 and 0.75 respectively. All simulations were done with 1000 replications.

Table 7: *Results for the Marginal Distributions*

	Standard two-stage		*Efficient two-stage*	
	Coeff	Std Error	Coeff	Std Error
Yen Margin				
μ_x	0.0190	0.0109	0.0188	0.0277
ϕ_{1x}	−0.0059	0.0188	−0.0064	0.0372
ϕ_{10x}	0.0674	0.0179	0.0680	0.0356
ω_x	0.0067	0.0032	0.0068	0.0059
β_x	0.9419	0.0151	0.9418	0.0225
α_x	0.0465	0.0114	0.0466	0.0214
ν_x	4.7285	0.4190	4.7170	1.2182
Euro Margin				
μ_y	0.0803	0.0254	0.0782	0.0253
σ_y^2	0.4507	0.0345	0.4437	0.0414
ν_y	6.2016	1.6898	6.3586	1.6438

Note: This table presents the estimated parameters and asymptotic standard errors of the marginal distribution models for the yen - U.S. dollar and euro - U.S. dollar exchange rates. The efficient two-stage estimates are computed using Clayton's copula.

Table 8: *LM Tests of independence and Kolmogorov-Smirnov Tests of the density*		
	$(U_t - \bar{U})$	$(V_t - \bar{V})$
First moment	40.2020	32.9100
p-value	*0.5192*	*0.8173*
Second moment	36.7712	30.0963
p-value	*0.6690*	*0.8981*
Third moment	38.2055	30.3702
p-value	*0.6070*	*0.8914*
Fourth moment	40.5122	32.2420
p-value	*0.5056*	*0.8389*
K-S Stat	0.0315	0.0303
K-S p-value	*0.5410*	*0.5887*

Note: The first panel of this table presents the results of LM tests of the independence of the first four moments of the variables U_t and V_t, described in the text. We regress $(U_t - \bar{U})^k$ and $(V_t - \bar{V})^k$ on twenty lags of both variables, for $k = 1, 2, 3, 4$. The test statistic is $(T - 40) \cdot R^2$ for each regression, and is distributed under the null as χ^2_{40}. The second panel present the results of a Kolmogorov-Smirnov test on U_t and V_t.

Table 9: *Hit test results for the marginal distributions*		
	Yen / USD	Euro/USD
Test stat 1	1.4470	4.2744
p-value 1	*0.8360*	*0.3701*
Test stat 2	0.9549	4.0508
p-value 2	*0.9166*	*0.3992*
Test stat 3	6.5721	6.6988
p-value 3	*0.1603*	*0.1527*
Test stat 4	5.3716	0.9661
p-value 4	*0.2513*	*0.9149*
Test stat 5	4.9258	5.9009
p-value 5	*0.2950*	*0.2067*
Test stat ALL	15.8095	15.5528
p-value ALL	*0.4663*	*0.4846*

Note: 'Test stat' refers to the likelihood ratio statistic testing the null hypothesis that the model is correctly specified. 'P-value' refers to the area in the right tail of the distribution of the test statistic, a χ^2_4 random variable for the individual region tests and a χ^2_{16} random variable for the joint test. The numbers

1 through 5 refer to the regions of the marginal distribution support described in the text. 'ALL' refers to the joint test of all regions simultaneously.

Table 10: *Copula model results*

	Normal	*Plackett*	*Clayton*
Parameter estimate	0.0718	1.2754	0.0963
Log-likelihood	*1.7228*	*2.1136*	*2.9376*

Note: This table presents the estimated copula parameter and the value of the copula likelihood at the optimum for the three copula models considered in this paper: the normal copula, Plackett's copula and Clayton's copula.

Table 11: *Hit test results for the copula models*

	Normal	*Plackett*	*Clayton*
Test stat 1	7.8820	8.7042	5.9965
p-value 1	*0.0960*	*0.0889*	*0.1994*
Test stat 2	3.6064	3.5719	3.5453
p-value 2	*0.4619*	*0.4670*	*0.4710*
Test stat 3	4.6596	4.9653	4.6301
p-value 3	*0.3240*	*0.2909*	*0.3274*
Test stat 4	3.2530	3.3407	3.2459
p-value 4	*0.5164*	*0.5025*	*0.5176*
Test stat 5	1.4820	1.5375	1.6279
p-value 5	*0.8298*	*0.8200*	*0.8038*
Test stat 6	4.1717	4.1618	4.1754
p-value 6	*0.3833*	*0.3845*	*0.3828*
Test stat 7	0.8631	0.7688	0.8980
p-value 7	*0.9298*	*0.9426*	*0.9248*
Test stat ALL	25.5641	26.0446	23.9226
p-value ALL	*0.5970*	*0.5706*	*0.6856*

Note: 'Test stat' refers to the likelihood ratio statistic testing the null hypothesis that the model is correctly specified. 'P-value' refers to the area in the right tail of the distribution of the test statistic, a χ^2_4 random variable for the individual region tests and a χ^2_{28} random variable for the joint test. The numbers 1 through 7 refer to the regions of the copula support depicted in Figure 3. 'ALL' refers to the joint test of all regions simultaneously.

References

[1] Anderson, T. W., 1957, Maximum Likelihood Estimates for a Multivariate Normal Distribution when some Observations are Missing, *Journal of the American Statistical Association*, 52(278), 200-203.

[2] Andrews, Donald W. K., 1988, Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables, *Econometric Theory*, 4, 458-467

[3] Ang, Andrew, and Chen, Joe, 2001, Asymmetric Correlations of Equity Portfolios, forthcoming, *Journal of Financial Economics.*

[4] Bae, Kee-Hong, Karolyi, Andrew and Stulz, René M., 2000, A New Approach to Measuring Financial Contagion, Working Paper 2000-13, Fisher College of Business, Ohio State University.

[5] Bank for International Settlements, *Central Bank Survey of Foreign Exchange and Derivatives Market Activity*, Basle, May 1999.

[6] Bera, Anil K., and Bilias, Yannis, MM, ME, ML, EL, EF and GMM Approaches to Estimation: A Synthesis, working paper, Department of Economics, University of Illinois, Urbana-Champaign.

[7] Bollerslev, Tim, 1987, A Conditional Heteroscedastic Time Series Model for Speculative Prices and Rates of Return, *Review of Economics and Statistics*, 69, 542-547.

[8] Bollerslev, Tim, and Wooldridge, Jeffrey M., Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances, *Econometric Reviews*, 11, 143-172.

[9] Bouyé, Eric, Durrleman, Valdo, Nikeghbali, Ashkan, Riboulet, Gaël, and Roncalli, Thierry, 2000, Copulas for Finance: A Reading Guide and Some Applications, working paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais, France.

[10] Brown, Stephen J., Goetzmann, William N., and Ross, Stephen A., 1995, Survival, *Journal of Finance*, 50(3), 853-873.

[11] Campbell, Rachel, Koedijk, Kees, and Kofman, Paul, 2000, Increased Correlation in Bear Markets: A Downside Risk Perspective, Working Paper, Faculty of Business Administration, Erasmus University Rotterdam.

[12] Capéraà, P., Fougères, A.-L., and Genest, C., 1997, A Nonparametric Estimation Procedure for Bivariate Extreme Value Copulas, *Biometrika*, 84(3), 567-577.

[13] Casella, George, and Berger, Roger L., 1990, *Statistical Inference*, Duxbury Press, U.S.A.

[14] Clayton, D. G., 1978, A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence, *Biometrika*, 65(1), 141-151.

[15] Cook, R. Dennis, and Johnson, Mark E., 1981, A Family of Distributions for Modelling Non-Elliptically Symmetric Multivariate Data, *Journal of the Royal Statistical Society, Series B,* 43(2), 210-218.

[16] Denuit, Michel, and Genest, Christian, 1999, An Extension of Osuna's Model for Stress Caused by Waiting, *Journal of Mathematical Psychology*, 45, 115-130.

[17] Diebold, Francis X., Gunther T., and Tay, Anthony S., 1998, Evaluating Density Forecasts with Applications to Financial Risk Management, *International Economic Review*, 39, 863-883.

[18] Embrechts, Paul, Höing, Andrea and Juri, Alessandro, 2001, Using Copulae to Bound the Value-at-Risk for Functions of Dependent Risks, working paper, Department of Mathematics, ETHZ, Zurich, Switzerland.

[19] Erb, Claude B., Harvey, Campbell R., and Viskanta, Tadas E., 1994, Forecasting International Equity Correlations, *Financial Analysts Journal*, 50, 32-45.

[20] Fama, Eugene F., 1965, The Behavior of Stock-Market Prices, *Journal of Business*, 38, 34-105.

[21] Farebrother, R. W., 1992, A Class of Bivariate Density Functions with Common Marginals, *Econometric Theory*, 8(1), 145-146.

[22] Farebrother, R. W., 1993, A Class of Bivariate Density Functions with Common Marginals, *Econometric Theory*, 9(1), 148-149.

[23] Fine, J. P., and Jiang, H., 2000, On Association in a Copula with Time Transformations, *Biometrika*, 87(3), 559-571.

[24] Fischer, R. A., 1932, *Statistical Methods for Research Workers.*

[25] Gallant, A. Ronald, and White, Halbert, 1988, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Oxford: Basil Blackwell.

[26] Genest, C., and Rivest, L.-P., 1993, Statistical Inference Procedures for Bivariate Archimedean Copulas, *Journal of the American Statistical Association*, 88(423), 1034-1043.

[27] Genest, C., Ghoudi, K., and Rivest, L.-P., 1995, A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions, *Biometrika*, 82(3), 543-552.

[28] Glidden, David V., 2000, A Two-Stage Estimator of the Dependence Parameter for the Clayton-Oakes Model, *Lifetime Data Analysis*, 6, 141-156.

[29] Godambe, V. P., 1960, An Optimum Property of Regular Maximum Likelihood Estimation, *Annals of Mathematical Statistics*, 31(4), 1208-1211.

[30] Goetzmann, W. N., and Jorion, P., 1999, Re-emerging Markets, *Journal of Financial and Quantitative Analysis*, 34(1), 1-32.

[31] Gourieroux, C., Monfort, A., and Trognon, A., 1984, Pseudo-Maximum Likelihood Methods: Theory, *Econometrica*, 52, 681-700.

[32] Hansen, Peter Reinhard, and Lunde, Asger, 2001, A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?, Working Paper 01-04, Department of Economics, Brown University.

[33] Harvey, Andrew, Koopman, Siem Jan, and Penzer, Jeremy, 1997, Messy Time Series: A Unified Approach, Discussion Paper EM/97/327, London School of Economics.

[34] Joe, Harry, 1997, *Multivariate Models and Dependence Concepts*, Monographs in Statistics and Probability 73, Chapman and Hall, London.

[35] Joe, Harry, and Xu, James J., 1996, The Estimation Method of Inference Functions for Margins for Multivariate Models, working paper, Department of Statistics, University of British Columbia.

[36] Kofman, Paul, and Sharpe, Ian, 2000, Imputation Methods for Incomplete Dependent Variables in Finance, working paper, School of Finance and Economics, University of Technology, Sydney.

[37] Le Cam, L., 1956, On the Asymptotic Theory of Estimation and Testing Hypotheses, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1, 129-156.

[38] Lehmann, E. L., and Casella, George, 1998, *Theory of Point Estimation*, Second Edition, Springer-Verlag, New York.

[39] Li, David X., 2001, On Default Correlation: A Copula Function Approach, Working Paper 99-07, The RiskMetrics Group.

[40] Little, R. J. A., and Rubin, D. B., 1987, *Statistical Analysis with Missing Data*, John Wiley and Sons, New York.

[41] Longin, François, and Solnik, Bruno, 2001, Extreme Correlation of International Equity Markets, *Journal of Finance*, 56(2), 649-676.

[42] Mandelbrot, Benoit, 1963, The variation of certain speculative prices, *Journal of Business*, 26, 394 - 419.

[43] McLeish, D. L., and Small, Christopher G., 1988, *The Theory and Applications of Statistical Inference Functions*, Springer-Verlag, New York.

[44] Oakes, David, 1982, A Model for Association in Bivariate Survival Data, *Journal of the Royal Statistical Association, Series B,* 44(3), 414-422.

[45] Oakes, David, 1989, Bivariate Survival Models Induced by Frailties, *Journal of the American Statistical Association*, 84(406), 487-493.

[46] Nelsen, Roger B., 1999, *An Introduction to Copulas*, Springer-Verlag, New York.

[47] Newey, Whitney, K., 1987, Asymptotic Properties of a One-Step Estimator Obtained from an Optimal Step Size, *Econometric Theory*, 3(2), 305.

[48] Newey, Whitney, K., 1988, Asymptotic Properties of a One-Step Estimator Obtained from an Optimal Step Size, *Econometric Theory*, 4(2), 359-361.

[49] Newey, Whitney K., and McFadden, Daniel, 1994, Large Sample Estimation and Hypothesis Testing, in Robert F. Engle and Daniel McFadden eds., *Handbook of Econometrics, Vol 4*, North-Holland, Amsterdam.

[50] Patton, Andrew J., 2001a, Modelling Time-Varying Exchange Rate Dependence Using the Conditional Copula, Working Paper 2001-09, Department of Economics, University of California, San Diego.

[51] Patton, Andrew J., 2001b, On the Importance of Skewness and Asymmetric Dependence in Stock Returns for Asset Allocation, working paper, Department of Economics, University of California, San Diego.

[52] Plackett, R. L., 1965, A Class of Bivariate Distributions, *Journal of the American Statistical Association*, 60, 516-522.

[53] Politis, Dimitris N., and Romano, Joseph P., 1994, The Stationary Bootstrap, *Journal of the American Statistical Association*, 89, 1303-1313.

[54] Richardson, Matthew, and Smith, Tom, A Test for Multivariate Normality in Stock Returns, *Journal of Business*, 66(2), 295-321.

[55] Rockinger, Michael, and Jondeau, Eric, 2001, Conditional Dependency of Financial Series: An Application of Copulas, working paper, Department of Finance, HEC School of Management, France.

[56] Rosenberg, Joshua V., 1999, Semiparametric Pricing of Multivariate Contingent Claims, working paper, Department of Finance, Stern School of Business, New York University.

[57] Rosenberg, Joshua V., 2000, Nonparametric Pricing of Multivariate Contingent Claims, working paper, Department of Finance, Stern School of Business, New York University.

[58] Scaillet, O., 2000, Nonparametric Estimation of Copulas for Time Series, working paper, Département des Sciences Economiques, Université Catholique de Louvain, Belgium.

[59] Schweizer, B., and Sklar, A., 1983, *Probabilistic Metric Spaces*, Elsevier Science, New York.

[60] Shih, Joanna H., and Louis, Thomas A., 1995, Inferences on the Association Parameter in Copula Models for Bivariate Survival Data, *Biometrics*, 51, 1384-1399.

[61] Sklar, A., 1959, Fonctions de répartition à n dimensionset leurs marges, *Publ. Inst. Statis. Univ. Paris*, 8, 229-231.

[62] Stambaugh, Robert F., 1997, Analyzing Investments whose Histories Differ in Length, *Journal of Financial Economics*, 45, 285-331.

[63] Sullivan, Ryan, Timmermann, Allan, and White, Halbert, 1999, Data-Snooping, Technical Trading Rules and the Bootstrap, *Journal of Finance*, 54, 1647-1692.

[64] Sullivan, Ryan, Timmermann, Allan, and White, Halbert, 2001, The Dangers of Data-Driven Inference: The Case of Calendar Effects in Stock Returns, *Journal of Econometrics*, forthcoming.

[65] White, Halbert, 1994, *Estimation, Inference and Specification Analysis*, Econometric Society Monographs No. 22, Cambridge University Press, Cambridge, U.K.

[66] White, Halbert, 2001, *Asymptotic Theory for Econometricians*, Second Edition, Academic Press, San Diego.

[67] Xu, James J., 1996, *Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data*, unpublished Ph.D. thesis, Department of Statistics, University of British Columbia.

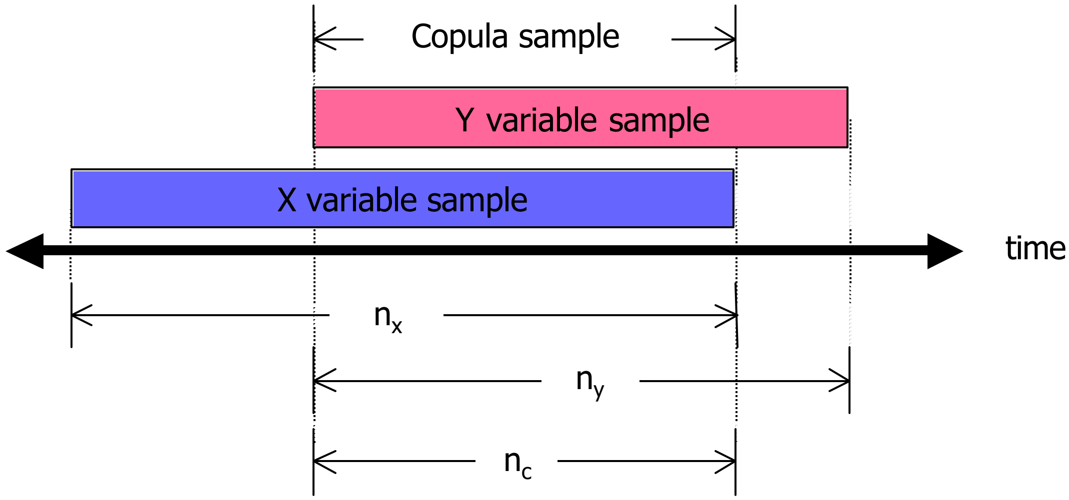

Figure 1: *One possible scenario where the amounts of data available on each individual variable are different, as is the amount of data available for the estimation of the copula.*

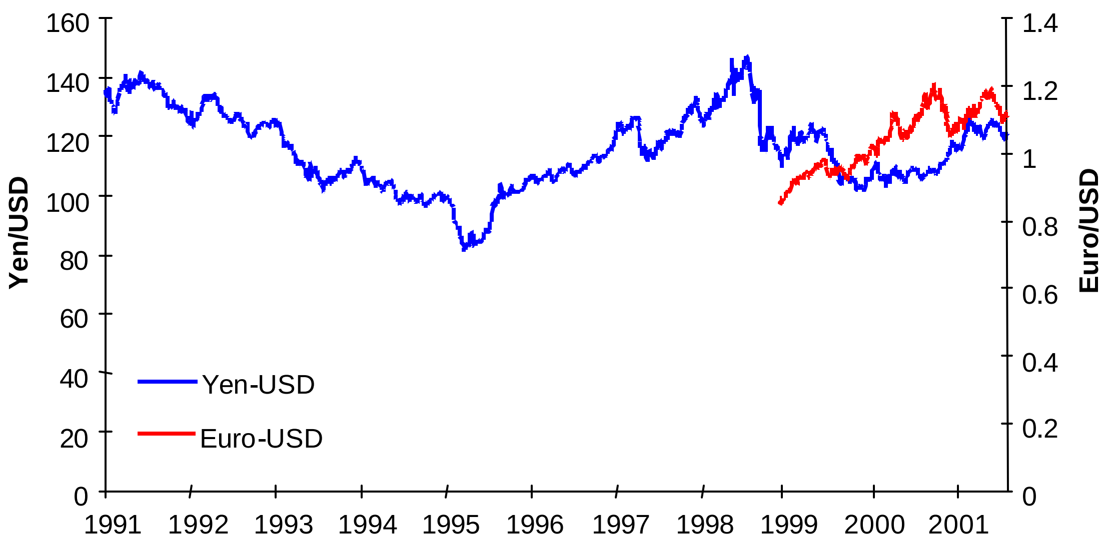

Figure 2: *Daily yen - U.S. dollar and euro - U.S. dollar exchange rates, from Jan 1991 to June 2001.*

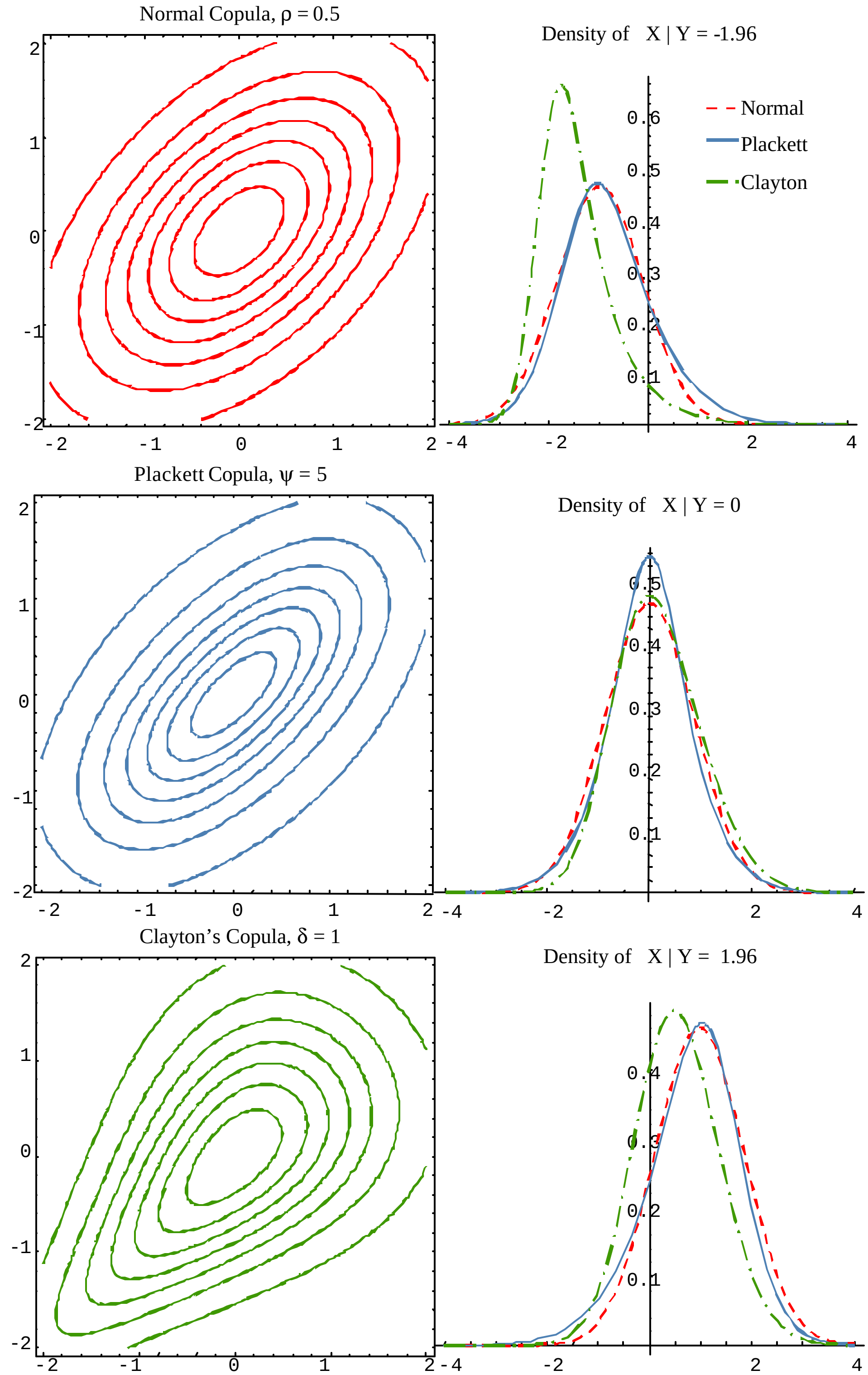

Figure 3: **Left Panels:** *Contour plots of three distributions all with standard normal marginal distributions and linear correlation coefficients of 0.5.* **Right Panels:** *The conditional density of X given Y for three values of Y when both X and Y are standard normal random variables, with copulas such that their correlation is 0.5.*

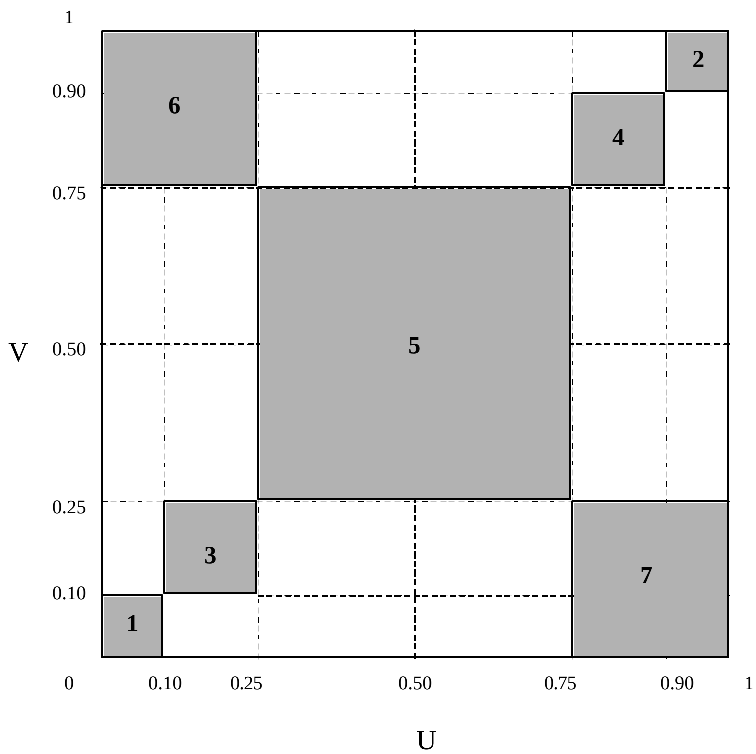

Figure 4: *Regions used in the hit tests*