Modelling Time-Varying Exchange Rate Dependence Using the Conditional Copula

Andrew J. Patton*
University of California, San Diego

First version: 29 December, 2000. This version: 26 May, 2002

Evidence that the distributions of many common economic variables are non-normal has been widely reported. When faced with such empirical evidence two problems arise: the first is proposing alternative, more palatable, density specifications. The second is finding measures of the dependence between two (or more) variables that are more informative than linear correlation, as when the joint distribution of the variables of interest is not elliptical the usual correlation coefficient is no longer sufficient to describe the dependence structure. In this paper we use the theory of copulas to address these problems. Sklar's theorem (1959) shows that an n-dimensional joint distribution function may be decomposed into its n marginal distributions, and a *copula*, which completely describes the dependence between the n variables. We verify that Sklar's theorem may be extended to conditional distributions, and illustrate how to construct, estimate and evaluate models of time-varying multivariate conditional densities using copulas through an application to Deutsche mark - U.S. dollar and Yen - U.S. dollar exchange rate returns. We find significant evidence that the conditional dependence between these exchange rates is asymmetric: dependence is higher during appreciations of the U.S. dollar against the mark and the yen than during depreciations of the dollar. We also find strong evidence of a structural break in the conditional copula following the introduction of the euro.

KEYWORDS: time series, copulas, exchange rates, dependence, density forecasting.
J.E.L. Codes: C32, C51, C52, F31.

---

*The author would like to thank Graham Elliott, Rob Engle, Marjorie Flavin, Raffaella Giacomini, Tony Hall, Joshua Rosenberg, Kevin Sheppard, Alan Sorensen, Ross Starr, Allan Timmermann, and seminar participants at the 2001 North American Econometric Society meetings in Maryland, Western Economic Association meetings in San Francisco, 'Advanced Computing for Financial Markets' conference in Bangor, Wales, and at UCSD, UC-Riverside, University of Technology Sydney, Monash University and the Commonwealth Scientific and Industrial Research Organisation Computing and Mathematical Sciences Division for their comments and suggestions. 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. This paper may be downloaded from http://www.econ.ucsd.edu/~apatton/research.html.

1 Introduction

Evidence that the distributions of many common economic variables are non-normal has been widely reported, as far back as Mills (1927). Common examples of deviations from normality are excess kurtosis (or fat tails) and skewness in univariate distributions. Recent studies have also reported deviations from *multivariate* normality, in the form of asymmetric dependence. One example of asymmetric dependence is where two asset returns exhibit greater correlation during market downturns than market upturns, as reported in Erb, *et al.* (1994), Longin and Solnik (2001) and Ang and Chen (2002). When faced with such empirical evidence two problems arise: the first is proposing alternative, more palatable, density specifications. The second is finding measures of the dependence between two (or more) variables that more informative than linear correlation, as when the joint distribution of the variables of interest is non-elliptical the usual correlation coefficient is no longer sufficient to describe the dependence structure.

In this paper we make use of a theorem due to Sklar (1959), which shows that any n-dimensional joint distribution function may be decomposed into its n marginal distributions, and a *copula*, which completely describes the dependence between the n variables[1]. By using an extension of Sklar's theorem we are able to exploit the success we have had in the modelling of univariate densities by first specifying models for the marginal distributions of a multivariate distribution of interest, and then specifying a copula. As an example, consider the modelling of the joint distribution of two exchange rates: the Student's t distribution has been found to provide a reasonable fit to the conditional univariate distribution of daily exchange rate returns, see Bollerslev (1987) amongst others. A natural starting point in the modelling of the joint distribution of two exchange rates might then be a bivariate t distribution. However, the standard bivariate Student's t distribution has the restrictive property that both marginal distributions have the same degrees of freedom parameter. Studies such as Bollerslev (1987) have shown that different exchange rates have different degrees of freedom parameters. In our empirical section we show that even the flexible BEKK model for the conditional variance-covariance matrix, due to Engle and Kroner (1995), estimated assuming a bivariate t density fails goodness-of-fit tests when estimated on the Deutsche mark - U.S. dollar and the Yen - U.S. dollar exchange rates. The condition that both exchange rate returns have the same degrees of freedom parameter is simply too restrictive. Note also that this is possibly the most ideal situation: where both assets turn out to have univariate distributions from the same family, the Student's t, and very similar degrees of freedom, 5.9 for the mark and 4.3 for the yen. We could imagine situations where the two variables of interest have quite different marginal distributions, a stock return and an exchange rate for example, where no obvious choice for the bivariate density

[1] The word *copula* comes from Latin for a 'link' or 'bond', and was coined by Sklar (1959), who first proved the theorem that a collection of marginal distributions can be 'coupled' together via a copula to form a multivariate distribution. It has been given various names, such as *dependence function* (Galambos, 1978 and Deheuvels, 1978), *uniform representation* (Kimeldorf and Sampson, 1975, and Hutchinson and Lai, 1990) or *standard form* (Cook and Johnson, 1981).

exists. Decomposing the multivariate distribution into the marginal distributions and the copula allows for the construction of better models of the individual variables than would be possible if we constrained ourselves to look only at existing multivariate distributions.

An alternative to a parametric specification of the multivariate distribution would of course be a non-parametric estimate, which can accommodate all possible distributional forms. The main drawback with the non-parametric approach is the lack of precision that occurs when the dimension of the distribution of interest is moderately large (say over four), or when we consider multivariate distributions conditioned on a state vector (as is the case in this paper). The trade-off for this lack of precision is the fact that a parametric specification may be mis-specified. It is for this reason that we devote a great deal of attention to tests of goodness-of-fit of the proposed specifications.

This paper makes two main contributions. Firstly, we show in Section 2 that the existing theory of (unconditional) copulas may be extended to the conditional case, thus allowing us to use copula theory in the modelling of time-varying conditional dependence. Time variation in the conditional first and second moments of economic time series has been widely reported, and so allowing for time variation in the conditional dependence between economic time series seems natural. The second contribution of this paper is to show how we may use the theory of conditional copulas for multivariate density modelling. We examine daily Deutsche mark - U.S. dollar (DM-USD) and Japanese yen - U.S. dollar (Yen-USD) exchange rates over the period January 1991 to December 2001. The modelling of the entire conditional joint distribution of these exchange rates, rather than just, say, the conditional means, variances and linear correlation, has a number of attractive features: given the conditional joint distribution we can, of course, obtain the conditional means, variances and correlation, so this type of modelling nests solely modelling conditional moments. Also, we can obtain the time-paths of *any* other dependence measure of interest, such as rank correlation, which can capture non-linear dependence, or measures of dependence in the extremes, such as tail dependence[2]. Further, there are economic situations where the entire conditional joint density is required, such as the pricing of financial options with multiple underlying assets, see Rosenberg (1999) or in the calculation of portfolio Value-at-Risk (VaR), see Hull and White (1998), or in a forecast situation where the loss function of the forecast's end-user is unknown.

In our empirical application we report two main findings. We find significant evidence that the conditional dependence between the DM-USD and Yen-USD exchange rates is asymmetric, in that they are more dependent during appreciations of the U.S. dollar (or alternatively, during depreciations of the mark and the yen) than during depreciations of the U.S. dollar. We also find very strong evidence of a structural break in the conditional copula following the introduction of the euro in January 1999: the level of dependence drops substantially, the dynamics of conditional dependence change, and the dependence structure goes from asymmetric to approximately symmetric.

Despite the fact that copulas were introduced as a means of isolating the dependence structure of

[2]This measure will be discussed in more detail in Section 4. Dependence during extreme events has been the subject of much analysis in the financial contagion literature, see Hartmann, *et al.* (2001) amongst others.

a multivariate distribution over forty years ago, it is only recently that they attracted the attention of economists. In the last few years numerous papers have appeared, using copulas to analyse such topics as multivariate option pricing, portfolio Value-at-Risk, models of default risk, selectivity bias, nonlinear autoregressive dependence and contagion[3]. To our knowledge, this paper is the first to consider copulas for time-varying conditional distributions, emphasize the importance of formal goodness-of-fit testing for copulas and marginal distributions, and to employ statistical tests comparing the goodness-of-fit of competing non-nested copulas.

The structure of the remainder of this paper is as follows. Section 2 introduces the theory of the conditional copula, and Section 3 discusses some of the issues regarding the evaluation and comparison of copula models. In Section 4 we apply the theory of conditional copulas to a model of the time-varying joint distribution of the Deutsche mark - U.S. dollar and Yen - U.S. dollar exchange rates. In that section we discuss the construction and evaluation of time-varying conditional copula models. Finally, in Section 5 we summarise our results, and discuss possible extensions. All proofs are contained in Appendix A.

2 The Theory of the Conditional Copula

In this section we introduce the theory of copulas, and discuss its usefulness for multivariate density modelling and forecasting. We firstly review the existing theory on copulas for unconditional random variables, and then extend the analysis to allow for conditioning variables. We derive the properties of conditional distributions and conditional copulas from the properties of the underlying unconditional distributions. Firstly we must define the notation and some technical details.

Let (X, Y, W) be random variables on a complete probability space $(- \,, \mathcal{F}, H^*)$, where $- \equiv \mathbb{R} \times \mathbb{R} \times \mathbb{R}^j$, j is some finite integer, $\mathcal{F} \equiv \mathcal{B}\left(\mathbb{R} \times \mathbb{R} \times \mathbb{R}^j\right)$ is the Borel σ-field generated by $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^j$, and H^* is a probability measure. The variables of interest are X and Y and the conditioning variable is W. Though in this paper we focus on bivariate distributions, it should be noted that the theory of copulas is applicable to the more general multivariate case. Let the conditional distribution of (X, Y) given W, (that is, $(X, Y)\,|W$) be denoted H, and let the conditional marginal distributions of $X|W$ and $Y|W$ be denoted F and G respectively. We will assume in this paper that the distribution function, H^* is sufficiently smooth for all required derivatives to exist, and that F, G and H are continuous. The latter assumptions are not necessary, but making them allows us to introduce copula theory in a more intuitive manner. 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 $\bar{\mathbb{R}} \equiv \mathbb{R} \cup \{\pm\infty\}$. We adopt

[3]In statistics see Clayton (1978), Cook and Johnson (1981), Oakes (1982, 1989), Genest and Rivest (1993), Genest, *et al.,* (1995), Capéraà, *et al.,* (1997), Denuit and Genest (1999) and Fine and Jiang (2000). In economics and finance see Frees, *et al.* (1996), Frees and Valdez (1998), Rosenberg (1999, 2000), Bouyé *et al.* (2000a), Cherubini and Luciano (2000, 2001), Costinot *et al.* (2000), Li (2000), Scaillet (2000), Embrechts, *et al.* (2001), Frey and McNeil (2001), Rockinger and Jondeau (2001), Sancetta and Satchell (2001), Smith (2001) and Patton (2002a).

the usual convention of denoting random variables in upper case, X_t, and realisations of random variables in lower case, x_t.

2.1 The unconditional copula

Let us assume in this section that $j = 0$ and so we have no conditioning variables, thus $X \sim F$, $Y \sim G$ and $(X, Y) \sim H$. We define the (two-dimensional) unconditional copula of (X, Y) below.

Definition 1 (Unconditional copula) *The unconditional copula of (X, Y), where $X \sim F$ and $Y \sim G$, and F and G are continuous, is the joint distribution function of $U \equiv F(X)$ and $V \equiv G(Y)$.*

The two variables U and V are known as the 'probability integral transforms' of X and Y. Fisher (1932) showed that the random variable $U = F(X)$ has the $Unif\,(0,1)$ distribution, regardless of the original distribution, F.[4] This is replicated in Theorem 1 below. Firstly we will define a distribution function, and the 'quasi-inverse' of a distribution function. The second condition below refers to the 'H-volume' of a box in $\bar{\mathbb{R}}^k$, denoted by V_H. This is simply the probability of observing a point in the box, and in the case that the probability density function exists this condition reduces to the function H always implying a non-negative density.

Definition 2 (Multivariate distribution function) *A k-dimensional distribution function is a function H with domain $\bar{\mathbb{R}}^k$ such that*

1. $H(\mathbf{x}) = 0$ *for all* $\mathbf{x} \in \bar{\mathbb{R}}^k$ *such that at least one element of* $\mathbf{x} = -\infty$, *and* $H(\infty, \infty, ..., \infty) = 1$,
2. $V_H(\mathbf{B}) \equiv \sum_{i=1}^{2^k} sgn(\mathbf{c})\, H(\mathbf{c}) \geq 0$ *for* $\mathbf{B} = [a_1, b_1] \times [a_2, b_2] \times ... \times [a_k, b_k]$, *the "k-box" with vertices* $\mathbf{c}$, *where* c_i *is either equal to* a_i *or* b_i *for* $i = 1, 2, ..., k$, *and*

$$sgn(\mathbf{c}) = \begin{array}{ll} 1, & \text{if } c_k = a_k \text{ for an even number of } k's \\ -1, & \text{if } c_k = a_k \text{ for an odd number of } k's \end{array}$$

Remark 1 (Univariate distribution function) *In the univariate case these conditions simplify to:*

1. $F(-\infty) = 0$ *and* $F(\infty) = 1$,
2. $F(x_2) - F(x_1) \geq 0$ *for all* $x_1, x_2 \in \bar{\mathbb{R}}$ *such that* $x_1 \leq x_2$.

.

[4] The probability integral transform has also been used in the context of goodness-of-fit tests as far back as the 1930s, see K. Pearson (1933) for example. More recently Diebold, *et al.* (1998) extended the probability integral transform theory to the time series case, and proposed using it in the evaluation of density forecasts. We will discuss this further in Section 3.

Remark 2 (Bivariate distribution function) *In the bivariate case these conditions simplify to:*

1. $H(x, -\infty) = H(-\infty, y) = 0$ *for all* $(x, y) \in \bar{\mathbb{R}} \times \bar{\mathbb{R}}$*, and* $H(\infty, \infty) = 1$,
2. $V_H([x_1, x_2] \times [y_1, y_2]) \equiv H(x_2, y_2) - H(x_1, y_2) - H(x_2, y_1) + H(x_1, y_1) \geq 0$ *for all* $x_1, x_2, y_1, y_2 \in \bar{\mathbb{R}}$ *such that* $x_1 \leq x_2$ *and* $y_1 \leq y_2$.

Remark 3 (Trivariate distribution function) *In the trivariate case these conditions simplify to:*

1. $H(x, y, -\infty) = H(x, -\infty, w) = H(-\infty, y, w) = 0$ *for all* $(x, y, w) \in \bar{\mathbb{R}} \times \bar{\mathbb{R}} \times \bar{\mathbb{R}}$ *and* $H(\infty, \infty, \infty) = 1$,
2. $V_H([x_1, x_2] \times [y_1, y_2] \times [w_1, w_2]) \equiv H(x_2, y_2, w_2) - H(x_2, y_2, w_1) - H(x_2, y_1, w_2) - H(x_1, y_2, w_2) + H(x_2, y_1, w_1) + H(x_1, y_2, w_1) + H(x_1, y_1, w_2) - H(x_1, y_1, w_1)$ *for all* $x_1, x_2, y_1, y_2, w_1, w_2 \in \bar{\mathbb{R}}$ *such that* $x_1 \leq x_2$, $y_1 \leq y_2$ *and* $w_1 \leq w_2$.

Note that the marginal distributions are extracted from the bivariate distribution (for example) as follows:

$$F(x) \equiv H(x, \infty) \text{ for all } x \in \bar{\mathbb{R}} \tag{1}$$

$$G(y) \equiv H(\infty, y) \text{ for all } y \in \bar{\mathbb{R}} \tag{2}$$

With the formal definition of a distribution function in hand, we now define the 'quasi-inverse' of a distribution function.

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

$$F^{(-1)}(u) = \inf\{x : F(x) \geq u\}, \text{ for } u \in [0, 1]. \tag{3}$$

We are now ready to present Fisher's result on the probability integral transform of a random variable with a continuous distribution function. This result can be found in Casella and Berger (1990), for example.

Theorem 1 (Fisher, 1932) *Let* X *have a continuous cdf* F. *Then* $U = F(X)$ *has the* $Unif(0, 1)$ *distribution, regardless of the original distribution,* F.

All proofs are in Appendix 1. If $(U, V) \sim C$, then Theorem 1 and Definition 1 show that the marginal distributions of C must be $Unif(0, 1)$. Thus a copula is a joint distribution of two $Unif(0, 1)$ random variables. With this result we can show that a copula must have the following properties. We focus on the bivariate case; the corresponding k-dimension case is derived in a similar fashion.

Proposition 1 (Properties of an unconditional copula) *A two-dimensional copula has the following properties:*

1. *It is a function* $C : [0,1] \times [0,1] \rightarrow [0,1]$
2. $C(u,0) = C(0,v) = 0$ *for all* $u, v \in [0,1]$
3. $C(u,1) = u$ *and* $C(1,v) = v$ *for all* $u, v \in [0,1]$
4. $V_C([u_1,u_2] \times [v_1,v_2]) \equiv C(u_2,v_2) - C(u_1,v_2) - C(u_2,v_1) + C(u_1,v_1) \geq 0$ *for all* $u_1, u_2, v_1, v_2 \in [0,1]$ *such that* $u_1 \leq u_2$ *and* $v_1 \leq v_2$.

One motivation for analysing the copula of two random variables is that it allows us to focus on the dependence between random variables in a general way. The copula contains information from the joint distribution that is not contained in the marginal distributions. By transforming X and Y to U and V we filter out the information in the marginal distributions, because these variables will be $Unif(0,1)$ regardless of F and G. The information in H that is not in the marginal distributions is all of the *dependence* information. Thus C contains all of the information on the dependence between X and Y, but no information on the univariate characteristics of X or Y. It is for this reason that the copula of (X,Y) is alternatively known as the 'dependence function' of (X,Y), see Galambos (1978) and Deheuvels (1978) for example.

A further motivation, and the one driving our interest in copulas for multivariate density modelling and forecasting, is given by Sklar (1959). Sklar's theorem is the main result in copula theory, and we provide it below.

Theorem 2 (Sklar's theorem for continuous distributions) *Let F be the distribution of X, G be the distribution of Y, and H be the joint distribution of (X,Y). Assume that F and G are continuous. Then there exists a unique copula C such that*

$$H(x,y) = C(F(x), G(y)), \quad \forall\ (x,y) \in \bar{\mathbb{R}} \times \bar{\mathbb{R}} \tag{4}$$

Conversely, if we let F and G be distribution functions and C be a copula, then the function H defined by equation (4) is a bivariate distribution function with marginal distributions F and G.

It is the converse of the Sklar's theorem that is the most interesting for multivariate density modelling. It implies that we may link together *any* two univariate distributions, of any type (not necessarily from the same family), with *any* copula and we will have defined a valid bivariate distribution. The usefulness of this result stems from the fact that while in the economics and statistics literatures we have a vast selection of flexible parametric univariate distributions, the set of parametric multivariate distributions available is much smaller. This is possibly due to the increased complexity of these functions, or possibly due to lack of demand for flexible parametric multivariate distributions in the past. There are a great deal of parametric copulas for us to use in multivariate density modelling, see Hutchinson and Lai (1990), Chapter 5 of Joe (1997) and Chapter 4 of Nelsen (1999) for example, and so the set of possible parametric multivariate distributions is increased dramatically by Sklar's theorem, using only functions that have already been reported. For example, let M, N and P be the number of parametric multivariate distributions, univariate distributions and copulas previously reported in the literature, and note that $N >> M$. With Sklar's theorem the set of possible parametric multivariate distributions is increased from M to $N^2 \cdot P >> M$. Of course not all of these distributions will be useful empirically, but increasing the set of possible models clearly improves our chances of finding density models that provide an adequate fit to the data under analysis.

With the following corollary to Sklar's theorem, given in Nelsen (1999) for example, the set of possible parametric multivariate distributions increases even further, as we are able to extract the copula from any given multivariate distribution and use it independently of the marginal distributions of the original distribution.

Corollary 1 *Let H be any bivariate distribution with continuous marginal distributions F and G. Let $F^{(-1)}$ and $G^{(-1)}$ denote the (quasi-) inverses of the marginal distributions. Then there exists a unique copula $C : [0,1] \times [0,1] \to [0,1]$ such that*

$$C(u,v) = H\left(F^{(-1)}(u), G^{(-1)}(v)\right), \quad \forall (u,v) \in [0,1] \times [0,1] \tag{5}$$

From the converse of Sklar's theorem we know that given any two marginal distributions and any copula we have a joint distribution. From the above corollary we know that from any given joint distribution we can extract the implied copula and marginal distributions. This corollary allows us to extract, for example, the 'normal copula' from a standard bivariate normal distribution.

2.2 The conditional copula

In economics, and particularly in economic time series analysis, the case that the random variables of interest are conditioned on some pre-determined variables is an important one. In this section we show how the existing results in copula theory may be extended to allow for conditioning variables. We derive the properties of conditional joint distributions and the conditional copula from the properties of unconditional distributions and copulas. For the purposes of exposition we will assume below that the dimension of the conditioning variable, W, is 1.

Proposition 2 (Properties of a univariate conditional distribution function) *Let the joint distribution of (X, W) be F_{xw}, the marginal distribution of W be F_w, and the support of W be $\mathcal{W}$. The conditional univariate distribution of $X|W$, denoted F, is defined as*

$$F(x|w) \equiv f_w(w)^{-1} \cdot \frac{\partial F_{xw}(x,w)}{\partial w} \tag{6}$$

and satisfies the following properties:

1. $F(-\infty|w) = 0$ *and* $F(\infty|w) = 1$ *for each* $w \in \mathcal{W}$

2. $F(x_2|w) - F(x_1|w) \geq 0$ *for all* $x_1 \leq x_2 \in \bar{\mathbb{R}}$ *and each* $w \in \mathcal{W}$.

We can obtain in a similar fashion the properties of conditional bivariate distribution functions.

Proposition 3 (Properties of a bivariate conditional distribution function) *Let the joint distribution of* (X, Y, W) *be* H_{xyw}*, the marginal distribution of* W *be* F_w*, and the support of* W *be* $\mathcal{W}$*. The conditional bivariate distribution of* $(X, Y)|W$*, denoted* H*, is defined as*

$$H(x, y|w) \equiv f_w(w)^{-1} \cdot \frac{\partial H_{xyw}(x, y, w)}{\partial w} \tag{7}$$

and satisfies the following properties:

1. $H(x, -\infty|w) = H(-\infty, y|w) = 0$*, and* $H(\infty, \infty|w) = 1$ *for all* $(x, y) \in \bar{\mathbb{R}} \times \bar{\mathbb{R}}$ *and each* $w \in \mathcal{W}$,

2. $V_H([x_1, x_2] \times [y_1, y_2]|w) \equiv H(x_2, y_2|w) - H(x_1, y_2|w) - H(x_2, y_1|w) + H(x_1, y_1|w) \geq 0$ *for all* $x_1, x_2, y_1, y_2 \in \bar{\mathbb{R}}$*, such that* $x_1 \leq x_2$*,* $y_1 \leq y_2$ *and each* $w \in \mathcal{W}$.

The conditional marginal distributions of X and Y are defined as $F(x|w) \equiv H(x, \infty|w)$, and $G(y|w) \equiv H(\infty, y|w)$. Unlike the conditional bivariate distribution function, we cannot obtain the conditional copula of $(X, Y)|W$ from the copula of (X, Y, W). The 'conditional copula of $(X, Y)|W$' is defined below.

Definition 4 (Conditional copula) *The conditional copula of* $(X, Y)|W$*, where* $X|W \sim F$ *and* $Y|W \sim G$*, is the conditional joint distribution function of* $U \equiv F(X|W)$ *and* $V \equiv G(Y|W)$ *given* W.

A two-dimensional conditional copula is derived from any distribution function C^* (the joint distribution of U, V and W) such that the conditional joint distribution of the first two variables given the remaining variables is a copula for all values of the conditioning variables.

Proposition 4 (Properties of a conditional copula) *A two-dimensional conditional copula is a function* $C : [0, 1] \times [0, 1] \times \mathcal{W} \to [0, 1]$ *with the following properties:*

1. $C(u, 0|w) = C(0, v|w) = 0$*, and* $C(u, 1|w) = u$ *and* $C(1, v|w) = v$*, for every* u, v *in* $[0, 1]$ *and each* $w \in \mathcal{W}$,

2. $V_C([u_1, u_2] \times [v_1, v_2]|w) \equiv C(u_2, v_2|w) - C(u_1, v_2|w) - C(u_2, v_1|w) + C(u_1, v_1|w) \geq 0$ *for all* $u_1, u_2, v_1, v_2 \in [0, 1]$*, such that* $u_1 \leq u_2$ *and* $v_1 \leq v_2$ *and each* $w \in \mathcal{W}$.

The first property in Proposition 4 provides the lower bound on the distribution function, and ensures that the conditional marginal distributions, $C(u, 1|w)$ and $C(1, v|w)$, are uniform. The condition that V_C is non-negative has the same interpretation as for distribution functions: it simply ensures that the conditional probability of observing a point in the region $[u_1, u_2] \times [v_1, v_2]$ is non-negative. A two-dimensional conditional copula, then, is the conditional joint distribution of two conditionally Uniform$(0, 1)$ random variables.

We now move on to an extension of the key result in the theory of copulas: Sklar's (1959) theorem for conditional distributions:

Theorem 3 (Sklar's theorem for continuous conditional distributions) *Let F be the conditional distribution of $X|W$, G be the conditional distribution of $Y|W$, and H be the joint conditional distribution of $(X, Y)|W$. Assume that F and G are continuous in x and y. Then there exists a unique conditional copula C such that*

$$H(x, y|w) = C(F(x|w), G(y|w)|w), \quad \forall \ (x, y) \in \bar{\mathbb{R}} \times \bar{\mathbb{R}} \text{ and each } w \in \mathcal{W} \tag{8}$$

Conversely, if we let F be the conditional distribution of $X|W$, G be the conditional distribution of $Y|W$, and C be a conditional copula, then the function H defined by equation (8) is a conditional bivariate distribution function with conditional marginal distributions F and G.

The only complication introduced when extending Sklar's theorem to conditional distributions is that the conditioning variable(s), W, must be the same for both marginal distributions and the copula. This is important in the construction of conditional density models using copula theory. Failure to use the same conditioning variable for F, G and C will, in general, lead to a failure of the function H to satisfy the conditions for it to be a joint distribution function.

The density function equivalent of (8) is useful for maximum likelihood analysis, and is easily obtained, provided that F and G are differentiable, and H and C are twice differentiable.

$$\begin{aligned}
h(x, y|w) &\equiv \frac{\partial^2 H(x, y|w)}{\partial x \partial y} \\
&= \frac{\partial F(x|w)}{\partial x} \cdot \frac{\partial G(y|w)}{\partial y} \cdot \frac{\partial^2 C(F(x|w), G(y|w)|w)}{\partial u \partial v} \\
h(x, y|w) &\equiv f(x|w) \cdot g(y|w) \cdot c(u, v|w), \quad \forall (x, y, w) \in \bar{\mathbb{R}} \times \bar{\mathbb{R}} \times \mathcal{W}, \text{ so} && (9) \\
\mathcal{L}_{XY} &= \mathcal{L}_X + \mathcal{L}_Y + \mathcal{L}_C && (10)
\end{aligned}$$

where $u \equiv F(x|w)$, and $v \equiv G(y|w)$, $\mathcal{L}_{XY} \equiv \log h(x, y|w)$, $\mathcal{L}_X \equiv \log f(x|w)$, $\mathcal{L}_Y \equiv \log g(y|w)$, and $\mathcal{L}_C \equiv \log c(u, v|w)$.

We can also obtain an equivalent to Corollary 1 for the conditional case.

Definition 5 *If F is a conditional distribution function then the inverse with respect to its first argument is defined as:*

$$F^{(-1)}(u|w) = \inf\{x : F(x|w) \geq u\}, \text{ for } u \in [0, 1] \text{ and each } w \in \mathcal{W} \tag{11}$$

Corollary 2 *Let H be any conditional bivariate distribution with marginal distributions F and G that are continuous in their first arguments. Let $F^{(-1)}$ and $G^{(-1)}$ denote the (quasi-) inverses of the conditional marginal distributions with respect to their first arguments. Then there exists a unique conditional copula $C : [0,1] \times [0,1] \times \mathcal{W} \rightarrow [0,1]$ such that*

$$C(u, v|w) = H\left(F^{(-1)}(u|w), G^{(-1)}(v|w)|w\right), \quad \forall (u,v) \in [0,1] \times [0,1] \text{ and each } w \in \mathcal{W} \tag{12}$$

And so from any bivariate conditional distribution we may extract the implied conditional copula.

2.3 Examples of some copulas

To provide some idea as to the flexibility that copula theory gives us, we now consider various joint distributions, all with standard normal marginal distributions and all implying a linear correlation coefficient, ρ, of 0.5. The contour plots of these distributions are presented in Figure 1. In the upper left corner of this figure is the standard bivariate normal distribution with $\rho = 0.5$. The other elements of this figure show the dependence structures implied by other copulas, with each copula calibrated so as to also yield $\rho = 0.5$. It is quite clear that knowing the marginal distributions and linear correlation is not sufficient to describe a joint distribution: Clayton's copula, for example, has contours that are quite peaked in the negative quadrant, implying greater dependence for joint negative events than for joint positive events. Gumbel's copula implies just the opposite. The functional form of the symmetrised Joe-Clayton will be given in Section 4; the remaining copula functional forms may be found in Joe (1997).

[INSERT FIGURE 1 HERE]

3 Evaluation of time series density models

Before moving on to developing models for the conditional copula, we must first establish a means of evaluating and comparing their goodness-of-fit. Measures of goodness-of-fit are not only of importance for evaluating the fit of a proposed copula, but for testing the specification of the marginal distributions. Modelling of the conditional copula requires that the models for the marginal distributions to be indistinguishable from the true marginal distributions. Say we use $\tilde{F}$ and $\tilde{G}$ rather than F and G in the marginal distribution modelling stage. This will lead to $\tilde{U} \equiv \tilde{F}(X|W)$ and $\tilde{V} \equiv \tilde{G}(Y|W)$ being non-Uniform$(0,1)$, and so any copula model applied to the joint distribution of $\left(\tilde{U}, \tilde{V}\right)|W$ will automatically be mis-specified, by Proposition 4. Thus testing for marginal distribution mis-specification is a critical step in constructing multivariate distribution models using copulas.

3.1 Evaluation

As discussed above, a copula may be viewed as the joint distribution of two uniform random variables, and so the evaluation of copula models is a special case of the more general problem of evaluating multivariate density models. The density model (or forecast) evaluation literature is relatively young, and no single method has emerged as best. Studies by Diebold, *et al.* (1998) and Diebold, *et al.* (1999) focus on the probability integral transforms of the data in the evaluation of the density model, and so are clearly relevant in evaluating copula models. We use the tests of Diebold, *et al.* (1998, 1999) and propose and employ a new test, described below.

Let us denote the two transformed series as $\{u_t\}_{t=1}^{T}$ and $\{v_t\}_{t=1}^{T}$, where $u_t \equiv F_t(x_t|\mathcal{F}_{t-1})$ and $v_t \equiv G_t(y_t|\mathcal{F}_{t-1})$, for $t = 1, 2, ..., T$. Diebold, *et al.* (1998) showed that for a time series of probability integral transforms will be *i.i.d.* $Unif(0,1)$ if the sequence of densities is correct, and proposed testing the specification of a density model by testing whether or not the transformed series was *i.i.d.*, and $Unif(0,1)$ in two separate stages[5]. We follow this suggestion, and test the independence of the first four moments of U_t and V_t, by regressing $(u_t - \bar{u})^k$ and $(v_t - \bar{v})^k$ on 20 lags of both $(u_t - \bar{u})^k$ and $(v_t - \bar{v})^k$, for $k = 1, 2, 3, 4$. We test the hypothesis that the transformed series are $Unif(0,1)$ via the Kolmogorov-Smirnov (K-S) test[6].

There are two drawbacks of the above approach to evaluating a density model: the main drawback is that we must test the correctness of the density model separately from testing for serial dependence in the transformed variables[7]. The second drawback is that the fact that the Kolmogorov-Smirnov test has lower power in the tails of the distribution than in the centre, see Stephens (1986). We propose here an alternative test, which draws on the interval forecasting literature and quantifies the intuition that Diebold, *et al.* (1998) suggest can be gained by looking at the empirical histograms of the transformed data. Diebold, *et al.* suggest that by comparing the number of observations in each bin, otherwise known as a 'hit' in that bin, with what would be expected under the null hypothesis we may gain some insight as to where the model fails, if at all. This form of evaluation has its roots in K. Pearson's (1900) χ^2 test, see D'Agostino and Stephens

[5] It should be noted that these tests were developed for the case when the parameters of the proposed model are *known*, and not estimated from the sample. Constructing the variables u_t and v_t using parameter estimates is not innocuous. Indeed, it was known as far back as David and Johnson (1948) that when the probability integral transform is taken with respect to the true distribution but using estimated parameters the resulting random variable does *not* have the $Unif(0,1)$ distribution; instead it has a distribution that depends on the distribution of the original random variable. The implications for these specification tests are that we need, as some authors in the past have done, see Engle and Manganelli (1999) and Diebold *et al.* (1998), to interpret the tests as being *conditional* on the estimated parameters. These tests, then, ignore any estimation error in the parameters. The best we can hope for is that for large sample sizes the impact of the estimation uncertainty is small.

[6] See Shao (1999) for the theory underlying this test.

[7] Berkowitz (2001) proposed one solution to this problem. He suggested that instead of testing the $\{u_t\}_{t=1}^{T}$ series, say, we may define a new series: $\left\{z_t \equiv \Phi^{-1}(u_t)\right\}_{t=1}^{T}$, where Φ^{-1} is the inverse cdf of a standard normal distribution. The null hypothesis that $\{u_t\}_{t=1}^{T}$ is *i.i.d.* $Unif(0,1)$ may be tested by testing that $\{z_t\}_{t=1}^{T}$ is *i.i.d.* $N(0,1)$, which is possibly easier due to the large number of tests of normality available.

(1986) for more details.

In the following test we decompose a density model into a set of 'region' models ('interval' models in the univariate case), each of which should be correctly specified under the null hypothesis that the entire density is correctly specified. The specification introduced below is an extension of the 'hit' regressions of Christoffersen (1998) and Engle and Manganelli (1999), proposed to evaluate interval forecasts, such as Value-at-Risk forecasts. We will describe the test below in a general setting, and discuss the details of implementation in Section 4.5.

Let W_t be the (possibly multivariate) random variable under analysis, and denote the support of W_t by $\mathcal{S}$. Let $\{R_j\}_{j=0}^{K}$ be regions in $\mathcal{S}$ such that $R_i \cap R_j = \emptyset$ if $i \neq j$, and $\cup_{j=0}^{K} R_j = \mathcal{S}$. Let π_{jt} be the true probability that $W_t \in R_j$ and let p_{jt} be the probability suggested by the model[8]. Finally, let $\Pi_t \equiv [\pi_{0t}, \pi_{1t}, ..., \pi_{Kt}]'$ and $P_t \equiv [p_{0t}, p_{1t}, ..., p_{Kt}]'$. Under the null hypothesis that the model is correctly specified we have that $P_t = \Pi_t$ for $t = 1, 2, ..., T$. Let us define the variables to be analysed in the tests as $Hit_t^j \equiv \mathbf{1}\{X_t \in R_j\}$, where $\mathbf{1}\{A\}$ takes the value 1 if the argument, A, is true and zero elsewhere, and $M_t \equiv \sum_{j=0}^{K} j \cdot \mathbf{1}\{X_t \in R_j\}$.

We may test that the model is adequately specified in each of the $K+1$ regions individually via tests of the hypothesis $H_0 : Hit_t^j \sim i.n.i.d.$[9] $Bernoulli\,(p_{jt})$ versus $H_1 : Hit_t^j \sim Bernoulli\,(\pi_{jt})$, where π_{jt} is a function of both p_{jt}, and other elements of the time $t-1$ information set thought to possibly have explanatory power for the probability of a hit. This is where our test differs from those presented in Christoffersen (1998) and Engle and Manganelli (1999). Christoffersen (1998) proposed modelling π_{jt} as a first-order Markov chain to check for first-order serial dependence of the hits, while Engle and Manganelli (1999) proposed using a linear probability model to determine if other variables, such as lagged hits and also lagged levels of the Value-at-Risk, had significant predictive value. The Markov chain approach suffers from the drawback that it is difficult to check for the influence of other variables or longer lags, while Engle and Manganelli's (1999) model may be improved relatively easily by using a better model for the hits than a linear probability model. We propose using a logit model for the hits, which yields more efficient parameter estimates, and thus hopefully a more powerful test[10]. The model we propose for π_{jt} is:

[8] Given the similarity between this test and Pearson's χ^2 test it would not be surprising to find that the power of the test is maximised when the probability mass in each region is equal. For a univariate density model this is a simple task, however it may be a more difficult task in the more general multivariate case. Also, it may be that the researcher has a particular interest in certain regions of the support (the lower tails, for example, which are important for VaR estimation) being correctly specified. For these reasons we consider the case where the probability mass in each region is possibly unequal.

[9] "$i.n.i.d.$" stands for "independent but not identically distributed".

[10] If we wished instead to retain the simplicity of the test of Engle and Manganelli (1999) we could employ an alternative extension: If we define $Hit_t^* \equiv (p_t\,(1-p_t))^{-1/2} \cdot (Hit_t - p_t)$, then we may use OLS to regress Hit_t^* on a constant and variables in the time-t information set in the same manner as Engle and Manganelli (1999). The test that all of the parameters in the Hit^* regression are zero would also be conducted in the same fashion. Standardising the variance of the dependent variable in the hit regression, in addition to standardising the mean as in Engle and Manganelli (1999), is necessary as the conditional variance of Hit_t under the null is $p_t\,(1-p_t)$, and thus if p_t is time-varying this causes Hit_t to be heteroscedastic. In the case that p_t is constant this concern obviously does not

$$\pi_{jt} = \pi_j\left(Z_{jt}, \beta_j, p_{jt}\right) = \Lambda\left(\lambda_j\left(Z_{jt}, \beta_j\right) - \ln\left[\frac{1-p_{jt}}{p_{jt}}\right]\right) \tag{13}$$

where $\Lambda(x) \equiv \frac{1}{1+e^{-x}}$ is the logistic transformation, Z_{jt} is a matrix containing elements from the information set at time $t-1$, β_j is a $(k_j \times 1)$ vector of parameters to be estimated, and λ_j is any function of regressors and parameters such that $\lambda_j(Z, 0) = 0$ for all Z. The condition on λ_j is imposed so that when $\beta_j = 0$ we have that $\pi_{jt} = \pi_j(Z_{jt}, 0, p_{jt}) = p_{jt}$, and thus the competing hypotheses may be expressed as $\beta_j = 0$ versus $\beta_j \neq 0$. The parameter β_j may be found via maximum likelihood, where the likelihood function to be maximised is: $\mathcal{L}\left(\pi_j\left(Z_j, \beta_j, p_j\right)|Hit^j\right) = \Sigma_{t=1}^{T} Hit_t^j \cdot \ln \pi_j\left(Z_{jt}, \beta_j, p_{jt}\right) + \left(1 - Hit_t^j\right) \cdot \ln\left(1 - \pi_j\left(Z_{jt}, \beta_j, p_{jt}\right)\right)$. The test is then conducted as a likelihood ratio test, where $LR_j \equiv -2 \cdot \left(\mathcal{L}\left(p_j|Hit^j\right) - \mathcal{L}\left(\pi_j\left(Z_j, \hat{\beta}_j, p_j\right)|Hit^j\right)\right) \sim \chi^2_{k_j}$ under the null hypothesis that the model is correctly specified in region R_j.

We may test whether the proposed density model is correctly specified in all $K+1$ regions simultaneously by testing the hypothesis $H_0 : M_t \sim i.n.i.d.\ Multinomial(P_t)$ versus $H_1 : M_t \sim Multinomial(\Pi_t)$, where again we specify Π_t to be a function of both P_t and variables in the time $t-1$ information set. We propose the following specification for the elements of Π_t:

$$\begin{aligned}
\pi_t^1 &= \pi^1\left(Z_t, \boldsymbol{\beta}, P_t\right) = \Lambda\left(\lambda_1\left(Z_{1t}, \beta_1\right) - \ln\left[\frac{1-p_{1t}}{p_{1t}}\right]\right) && (14)\\
\pi_t^j &= \pi^j\left(Z_t, \boldsymbol{\beta}, P_t\right) \\
&= \left(1 - \sum\nolimits_{i=1}^{j-1} \pi_{it}\right) \cdot \Lambda\left(\lambda_j\left(Z_{jt}, \beta_j\right) - \ln\left[\frac{1 - \sum_{i=1}^{j} p_{it}}{p_{jt}}\right]\right), \quad for\ j = 2, ..., K && (15)\\
\pi_t^0 &= 1 - \sum\nolimits_{j=1}^{K} \pi_{jt} && (16)
\end{aligned}$$

where $\Lambda(x) \equiv \frac{1}{1+e^{-x}}$ is the logistic transformation, $Z_t \equiv [Z_1, ..., Z_K]'$ and $\boldsymbol{\beta} \equiv [\beta_1, ..., \beta_K]'$.

Let the length of $\boldsymbol{\beta}$ be denoted K_β. This expression for Π_t is specified so that $\Pi_t(Z_t, \mathbf{0}, P_t) = P_t$ for all Z_t. Further, it allows each of the elements of Π_t to be a function of a set of regressors, Z_{jt}, while ensuring that each $\pi_{jt} \geq 0$ and that $\Sigma_{j=0}^{K} \pi_{jt} = 1$. Again the competing hypotheses may be expressed as $\boldsymbol{\beta} = 0$ versus $\boldsymbol{\beta} \neq 0$. The likelihood function to be maximised to obtain the parameter $\boldsymbol{\beta}$ is $\mathcal{L}\left(\Pi\left(Z, \boldsymbol{\beta}, P\right)|Hit\right) = \Sigma_{t=1}^{T} \Sigma_{j=0}^{K} \ln \pi_{jt} \cdot \mathbf{1}\left\{M_t = j\right\}$. The joint test may also be conducted as a likelihood ratio test: $LR_{ALL} \equiv -2 \cdot \left(\mathcal{L}\left(P|Hit\right) - \mathcal{L}\left(\Pi\left(Z, \hat{\boldsymbol{\beta}}, P\right)|Hit\right)\right) \backsim \chi^2_{K_\beta}$ under the null hypothesis that the model is correctly specified in all K regions.

3.2 Comparison

The comparison of alternative copula models is complicated by the fact that they are generally non-nested. This means that standard methods of comparison, such as a likelihood ratio test, are not available to us. We employ a method recently proposed by Rivers and Vuong (2002) to

arise.

overcome this problem. These authors present results for the comparison of non-nested models for time series data, with few restrictions on the performance measure used to compare the models.

Rivers and Vuong (2002) show that under some conditions the mean of the difference in performance measures for two models is asymptotically normal. When the performance measure is the log-likelihood function, the asymptotic variance of the log-likelihood ratio is simple to compute; we do so using a Newey-West (1987) heteroskedasticity and autocorrelation consistent variance estimator. We use a variety of 'truncation lengths' (the number of lags used in correcting for autocorrelation and heteroskedasticity) to determine the appropriate length.

4 An Application of the Conditional Copula

In this section we apply the theory of conditional copulas to the modelling of the conditional bivariate distribution of the daily Deutsche mark - U.S. dollar and Japanese yen - U.S. dollar exchange rate returns over the period January 2, 1991 to December 31, 2001. This represents the post-unification era in Germany (East and West Germany were united in late 1989, and some financial integration was still being carried out during 1990) and includes the first three years of the euro's reign as the official currency of Germany[11]. These two exchange rates are the two most heavily traded, representing close to 50% of total foreign exchange trading volume (see the Bank for International Settlements, 1996). Given their status, the DM-USD and Yen-USD exchange rates have been relatively widely studied, see Andersen and Bollerslev (1998), Diebold *et al.* (1999), Andersen *et al.* (2001), amongst others.

In addition to the economic interest in these series, they also represent a statistically interesting pair of series. The multivariate GARCH literature, for example, has numerous examples of studies reporting evidence of time-varying variances and covariances for exchange rates, see the papers mentioned above, and also Bollerslev (1987, 1990), amongst many others. This evidence may suggest that the entire conditional density and dependence structure is time-varying. In modelling the conditional marginal distributions we will assume that the conditional means evolve according to an autoregressive process and that the conditional variances evolve according to a GARCH(1,1) process, see Engle (1982) and Bollerslev (1986). The standardised innovations in both margins are assumed to be independent and identically distributed (*i.i.d.*) as Student's t random variables for the entire sample period.

There are many ways of capturing possible time variation in the conditional copula. We will assume that the functional form of the copula remains fixed over the sample while the parameters vary according to some evolution equation. An alternative to this may allow also for time variation in the functional form, using a regime switching model for example.

[11] The mark was still used for transactions in Germany until the end of 2001, but the mark/euro exchange rate was fixed on January 1, 1999, and all international transactions are denominated in euros. See the European Central Bank web site (http://www.ecb.int) for more information.

4.1 Description of the Data

As mentioned above, the data set used for this analysis comprises daily Deutsche mark - U.S. dollar and Japanese yen - U.S. dollar exchange rates over the eleven-year period from 2 January 1991 to 31 December, 2000, giving us 2819 observations; 2046 from the eight-year period prior to the introduction of the euro and 773 from the three-year post-euro period. The data were taken from the database of Datastream International. As usual, we take the log-difference of each exchange rate. Table 1 below presents some summary statistics of the data.

[INSERT TABLE 1 HERE]

The above table shows that neither exchange rate had a significant trend over either period, both means being very small relative to the standard deviation of each series. Both series also exhibit slight negative skewness, and excess kurtosis. The Jarque-Bera test of the normality of the unconditional distribution of each exchange rate strongly rejects unconditional normality in both periods. We also test for the presence of serial correlation up to the 20^{th} lag in the squared returns, an indication of ARCH-type heteroskedasticity, via the ARCH LM test of Engle (1982). Both exchange rates exhibit strong evidence of serial correlation in the squared returns in the pre-euro period, but only the yen exchange rate appears to exhibit such evidence in the post-euro period. The unconditional correlation coefficient between these two exchange rate returns indicates relatively high linear dependence prior to the introduction of the euro, and weaker dependence afterwards.

4.2 The Model

In specifying a model of the bivariate density of DM-USD and Yen-USD exchange rates we must specify three models: the models for the marginal distributions of each exchange rate, and the model for the conditional copula.

4.2.1 The models for the marginal distributions

The models employed for the marginal distributions are presented below. We will denote the log-difference of the DM-USD exchange rate as the variable X_t, and the log-difference of the Yen-USD exchange rate as the variable Y_t.

$$X_t = \mu_x + \phi_{1x} X_{t-1} + \varepsilon_t \tag{17}$$

$$\sigma_{x,t}^2 = \omega_x + \beta_x \sigma_{x,t-1}^2 + \alpha_x \varepsilon_{t-1}^2 \tag{18}$$

$$\sqrt{\frac{\upsilon_x}{\sigma_{x,t}^2(\upsilon_x - 2)}} \cdot \varepsilon_t \sim i.i.d.\ t_{\upsilon_x} \tag{19}$$

$$
\begin{aligned}
Y_t &= \mu_y + \phi_{1y} Y_{t-1} + \phi_{10y} Y_{t-10} + \eta_t && (20)\\
\sigma^2_{y,t} &= \omega_y + \beta_y \sigma^2_{y,t-1} + \alpha_y \eta^2_{t-1} && (21)\\
& \sqrt{\frac{\upsilon_y}{\sigma^2_{y,t}(\upsilon_y - 2)}} \cdot \eta_t \sim i.i.d.\ t_{\upsilon_y} && (22)
\end{aligned}
$$

That is, the marginal distribution for the DM-USD exchange rate is assumed to be completely characterised by an AR(1), t-GARCH(1,1) specification, while the marginal distribution for the Yen-USD exchange rate is assumed to be characterised by an AR(1,10)-t-GARCH(1,1) specification[12]. In our case it happened that we only needed univariate models for these two marginal distributions (no lags of the 'other' variable appear in either variable's model). This will not always be so, and it should be checked for each individual case. We will call the above specifications the 'copula models' for the marginal distributions, as they are to be used with the copula models introduced below.

4.2.2 For Comparison: Normal and Student's t BEKK models

For the purposes of comparison we also estimate two existing alternative models (the estimation results are not presented in the interests of parsimony, but are available from the author upon request). For both of the additional models we first model the conditional means of the two exchange rate returns series, using the models in equations (17) and (20). We then estimate a flexible multivariate GARCH models on the residuals: the BEKK model introduced by Engle and Kroner (1995). This model is written as:

$$
\Sigma_t = C'C + B'_{t-1}\Sigma_{t-1}B + A'e'_{t-1}e_{t-1}A \tag{23}
$$

where $\Sigma_t \equiv \begin{bmatrix} \sigma^2_{x,t} & \sigma_{xy,t} \\ \sigma_{xy,t} & \sigma^2_{y,t} \end{bmatrix}$, $C \equiv \begin{bmatrix} c_{11} & 0 \\ c_{12} & c_{22} \end{bmatrix}$, $B \equiv \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$, $A \equiv \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, $e_t \equiv \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}$, $\sigma^2_{x,t}$ is the conditional variance of X at time t, and $\sigma_{xy,t}$ is the conditional covariance between X and Y at time t. The two models differ in their assumption regarding the joint distribution of the residuals: the first model assumes bivariate normality, while the second assumes a bivariate Student's t distribution.

We include these models as benchmark density models obtained using techniques previously presented in the literature. When coupled with bivariate Student's t innovations the BEKK model is one of the most flexible conditional multivariate distribution models currently available. The main cost of the BEKK models is that they quickly become unwieldy in higher dimension problems[13]

[12]The marginal distribution specification tests, described in Section 4.3, suggested that the model for the conditional mean of the Yen-dollar exchange rate return needed the tenth autoregressive lag. This lag was not required for the DM-dollar exchange rates.

[13]Kearney and Patton (2000) estimated a five-dimension BEKK model on european exchange rates. We have not seen any applications of the BEKK model to problems of higher dimensions than this.

and are quite difficult to estimate even for bivariate problems when the Student's t distribution is assumed.

4.2.3 The models for the copula

In selecting a copula to use, we must have a clear idea of the properties of the data under analysis. Many of the copulas presented in the statistics literature are best suited to variables that take on extreme values in only one direction: survival times (Clayton, 1978), concentrations of particular chemicals (Cook and Johnson, 1981 and Genest and Rivest, 1993), flood data (Oakes, 1989). However, exchange rates have extremes in both directions: large positive *and* negative movements.

For the purposes of comparison, we will specify and estimate two alternative copulas, the 'symmetrised Joe-Clayton' copula and the normal (or Gaussian) copula, both with and without time variation. The normal copula may be considered the benchmark copula in economics. The most commonly employed distributional assumption in economics is that of normality, and so when looking for a copula to take as a benchmark the normal copula seems the most reasonable.

The reason for our interest in the symmetrised Joe-Clayton specification is that while it nests symmetry as a special case, it does not impose symmetric dependence on the variables as the normal copula does. The distribution of individual exchange rates has been found to be symmetric in many cases, but no investigation of bivariate symmetry has been undertaken to our knowledge. Asymmetric dependence in stock returns has been reported by a number of authors, see Erb, *et al.* (1994), Longin and Solnik (2001) and Ang and Chen (2002), but no such evidence has been reported for exchange rates. One possible factor driving asymmetric dependence in exchange rates may be an asymmetry in the behaviour of the central banks. For example, if the DM depreciates against the U.S. dollar, the Bank of Japan may intervene to ensure a corresponding depreciation of the yen against the dollar to maintain the competitiveness of Japanese exports to the U.S. with German exports to the U.S. When the DM appreciates against the dollar, however, there may be no incentive for the Bank of Japan to seek an appreciation of the yen against the dollar. This type of behaviour would induce an asymmetry in the dependence structure between these exchange rates, and a similar scenario for east Asian currencies is described in Takagi (1999). Of course, there are numerous other possible scenarios, including the one that there is *no* asymmetry in the dependence structure.

The symmetrised Joe-Clayton copula The first copula that will be used is a modification of the 'BB7' copula of Joe (1997). We refer to the BB7 copula as the Joe-Clayton copula, as it is constructed by taking a particular Laplace transformation of Clayton's copula[14]. The Joe-Clayton copula is:

[14] For more details on the construction of this copula or on Laplace transformations in copula theory, the reader is referred to Joe (1997).

$$
\begin{aligned}
C_{JC}(u,v|\tau^U,\tau^L) &= 1-\left(\left\{[1-(1-u)^{\kappa}]^{-\gamma}+[1-(1-v)^{\kappa}]^{-\gamma}-1\right\}^{-1/\gamma}\right)^{1/\kappa} && (24)\\
\text{where } \kappa &= 1/\log_2\left(2-\tau^U\right)\\
\gamma &= -1/\log_2\left(\tau^L\right)\\
\text{and } \tau^U &\in (0,1)\,,\ \tau^L\in(0,1) && (25)
\end{aligned}
$$

Joe (1997) asserts that this copula has a number of nice properties: it collapses to Clayton's copula when $\kappa = 1$, and the Fréchet-Hoeffding upper bound[15] is approached when either parameter approaches infinity. The Joe-Clayton copula has two parameters, τ^U and τ^L, which are measures of dependence known as *tail dependence*. These measures of dependence are defined below.

Definition 6 *If the limit*

$$\lim_{\varepsilon\to 0}\Pr[U\le\varepsilon|V\le\varepsilon]=\lim_{\varepsilon\to 0}\Pr[V\le\varepsilon|U\le\varepsilon] = \lim_{\varepsilon\to 0}C(\varepsilon,\varepsilon)/\varepsilon=\tau^L$$

exists, then the copula C exhibits lower tail dependence if $\tau^L \in (0,1]$ and no lower tail dependence if $\tau^L = 0$. Similarly, if the limit

$$\lim_{\delta\to 1}\Pr[U>\delta|V>\delta]=\lim_{\delta\to 1}\Pr[V>\delta|U>\delta]=\lim_{\delta\to 1}(1-2\delta+C(\delta,\delta))/(1-\delta)=\tau^U$$

exists, then the copula C exhibits upper tail dependence if $\tau^U \in (0,1]$ and no upper tail dependence if $\tau^U = 0$.

Tail dependence captures the behaviour of the random variables during extreme events. Informally, it measures the probability that we will observe an extremely large positive (negative) realisation of one variable, given that the other variable also took on an extremely large positive (negative) value. As an example, the bivariate normal distribution (and thus the normal copula) has $\tau^U = \tau^L = 0$ for correlation not equal to one, meaning that in the extreme tails of the distribution the variables are independent. The Joe-Clayton copula allows upper and lower tail dependence to range anywhere from zero to one.

One major drawback of the Joe-Clayton copula is that even when the two tail dependence measures are equal there is still some (slight) asymmetry in the Joe-Clayton copula, due simply to the functional form of this copula. A more desirable alternative would have the tail dependence measures completely determining the presence (or not) of asymmetry. To this end, we propose the 'symmetrised Joe-Clayton' copula:

$$C_{SJC}\left(u,v|\tau^U,\tau^L\right)=0.5\cdot\left(C_{JC}\left(u,v|\tau^U,\tau^L\right)+C_{JC}\left(1-u,1-v|\tau^L,\tau^U\right)+u+v-1\right) \tag{26}$$

Clearly, the symmetrised Joe-Clayton (SJC) is just a slight modification of the original Joe-Clayton copula, but by construction it is symmetric when $\tau^U = \tau^L$. From an empirical perspective we believe that the fact the SJC copula nests symmetry as a special case makes it a more interesting specification than the Joe-Clayton copula.

[15] The Fréchet-Hoeffding upper bound is a theoretical upper bound on the value that a joint distribution can take at any given point. This upper bound corresponds to perfect positive dependence between the two random variables.

Parameterising time-variation in the conditional copula We allow for time-variation in the conditional copula by allowing the parameters of the copula to evolve through time according to a particular equation. The difficulty in specifying how the parameters evolve over time lies in defining the forcing variable for the evolution equation . Unless the parameter has some interpretation, as the parameters of the Gaussian and SJC copulas do, it is very difficult to know what might (or should) influence it to change. We propose the evolution equations for the symmetrised Joe-Clayton copula:

$$\tau_t^U = \Lambda\left(\omega_U + \beta_U \tau_{t-1}^U + \alpha_U \cdot \frac{1}{10}\sum_{j=1}^{10}|u_{t-j} - v_{t-j}|\right) \tag{27}$$

$$\tau_t^L = \Lambda\left(\omega_L + \beta_L \tau_{t-1}^L + \alpha_L \cdot \frac{1}{10}\sum_{j=1}^{10}|u_{t-j} - v_{t-j}|\right) \tag{28}$$

where $\Lambda(x) \equiv \frac{1}{1+e^{-x}}$ is the logistic transformation, used to keep τ^U and τ^L in $(0,1)$ at all times.

In the above equations we propose that the upper and lower tail dependence parameters each follow something akin to a restricted ARMA(1,10) process. The right hand side of the model for the tail dependence evolution equation contains an autoregressive term, $\beta_U \tau_{t-1}^U$ and $\beta_L \tau_{t-1}^L$, and a forcing variable. Identifying a forcing variable for a time-varying limit probability is somewhat difficult. We propose using the mean absolute difference between u_t and v_t over the previous ten observations as a forcing variable[16]. The intuition behind this can be explained with the aid of Figure 2. If X and Y are perfectly positively dependent (otherwise known as 'comonotonic') then the transformed variables U and V will all lie on the main diagonal of the unit square. The absolute value of the difference between u_t and v_t is proportional to the minimum distance from the point (u_t, v_t) to the main diagonal, and we use the mean absolute difference between u_t and v_t over the previous ten observations as an indication of how far from comonotonicity the data were.

The second copula considered, the normal copula, is the dependence function associated with bivariate normality, and is extracted via Corollary 2.

$$C(u,v|\rho) = \int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left\{\frac{-(r^2 - 2\rho rs + s^2)}{2(1-\rho^2)}\right\} dr\ ds, \ -1 < \rho < 1 \tag{29}$$

where Φ^{-1} is the inverse of the standard normal $c.d.f.$

The transformations $\Phi^{-1}(u) = \Phi^{-1} \circ F(x)$ and $\Phi^{-1}(v) = \Phi^{-1} \circ G(y)$ transform the variables X and Y, which are distributed according to F and G, into standard normal random variables. The normal copula takes as arguments the standard normal transforms of X and Y, and assumes

[16] A few variations on this forcing variable were tried, such as weighting the observations by how close they are to the extremes, or by using an indicator variable for whether the observation was in the first, second, third or fourth quadrant. No significant improvement was found, and so we have elected to use the simplest model.

that they are *jointly* normally distributed. This is how we are able to back out the dependence implied by bivariate normality. We propose the following evolution equation for ρ_t:

$$\rho_t = \tilde{\Lambda}\left(\omega_\rho + \beta_\rho \cdot \rho_{t-1} + \alpha \cdot \frac{1}{10}\sum_{j=1}^{10} \Phi^{-1}(u_{t-j}) \cdot \Phi^{-1}(v_{t-j})\right) \tag{30}$$

where $\tilde{\Lambda}(x) \equiv \frac{1-e^{-x}}{1+e^{-x}}$ is the modified logistic transformation, designed to keep ρ_t in $(-1, 1)$ at all times.

Equation (30) reveals that we again assume that the copula parameter follows an ARMA(1,10)-type process: we include ρ_{t-1} as a regressor to capture any persistence in the dependence parameter, and the mean of the product of the last ten observations of the transformed variables $\Phi^{-1}(u_{t-j})$ and $\Phi^{-1}(v_{t-j})$, to capture any variation in dependence[17].

4.2.4 Estimating the model

Maximum likelihood is the natural estimation procedure to use in this context: in specifying models for the two marginal distributions and the copula, we have defined a joint distribution function for the two exchange rates, and thus a joint likelihood, as shown in equation (10). The procedure employed to develop the joint distribution lends itself naturally to multi-stage estimation of the model. Although estimating all of the coefficients simultaneously yields the most efficient estimates, the large number of parameters can make numerical maximisation of the likelihood function difficult. In this paper we make use of the multi-stage maximum likelihood estimator presented in Patton (2001). Under the usual conditions the estimates obtained are consistent and asymptotically normal.

4.2.5 Allowing for a structural break: the euro

On the 1^{st} of January, 1999 the euro was introduced and eleven European countries agreed to an irrevocable conversion rate between their currencies and the new euro; the conversion rate for the Deutsche mark is 1 euro = 1.95583 marks[18]. The data used in this study includes 773 observations in the period following the introduction of the euro.

We examine the impact on the introduction of the euro on the joint distribution of the DM-USD and Yen-USD exchange rates by allowing the parameters of the joint distribution to change between the pre- and post-euro subsamples. Note that allowing the parameters to change pre- and post-euro is equivalent to using an information set of $\mathcal{F}_t = \sigma(x_t, y_t, w_{t+1}, x_{t-1}, y_{t-1}, ..., x_{t-9}, y_{t-9})$,

[17] Averaging $\Phi^{-1}(u_{t-j}) \cdot \Phi^{-1}(v_{t-j})$ over the previous ten lags was done to keep the copula specification here comparable with that of the time-varying symmetrised Joe-Clayton copula.

[18] The eleven participating nations are: Austria, Belgium, Finland, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Spain. The complete list of conversion rates for all eleven currencies now linked to the euro may be found at http://www.ecb.int/press/pr981231-2.htm.

where W_t takes the value 0 in the pre-euro sample and 1 in the post-euro sample. Recall that the same information set must be used for both margins and the copula, meaning that we must test for a structural break in the DM margin, the Yen margin and the copula. To minimise the number of additional parameters in the new models, we conducted tests for the significance of the change in each parameter, and imposed constancy on those parameters that were not significantly different in the two periods.

4.3 Results for the marginal distributions

The parameter estimates and standard errors for marginal distribution models are presented in Table 2. In the DM margin all parameters except for the degrees of freedom parameter changed significantly following the introduction of the euro. The drift term in the mean increased from 0.01 to 0.07, reflecting the sharp depreciation in the euro in its first three years. The GARCH parameters changed in such a way that the implied unconditional annualised standard deviation increased slightly from 12.25% to 12.77%. In the yen margin only the degrees of freedom changed, from 4.26 to 6.65, indicating a 'thinning' of the tails of the Yen - U.S. dollar exchange rate. In the pre-euro period the yen had fatter tails than the DM, but in the post-euro period this ordering was reversed. The t-statistic (p-value) for the significance of difference in the degrees of freedom parameters was 2.0877 (0.0368) for the pre-euro period and -0.4618 (0.6442) in the post-euro period, indicating a significant[19] difference prior to the break, but no significant difference afterwards[20]. The significant difference in degrees of freedom parameters in the pre-euro period is important as the bivariate Student's t distribution is unable to capture this: as mentioned in the introduction, a bivariate Student's t imposes the same degrees of freedom parameter on *both* marginal distributions (and also on the copula, as we know from Corollary 2). A bivariate t_6, for example, imposes that both marginals are univariate Student's t_6 distributions and that the copula is a 'Student's t_6' copula. This specification is rejected by the data for the pre-euro period.

[INSERT TABLE 2 HERE]

Although we will not present the parameter estimates from the two BEKK models estimated, we will include them in the specification tests. In Table 3 we present the LM tests for serial independence of the probability integral transforms, U and V, and the Kolmogorov-Smirnov (K-S) tests for the correctness of the density specification. We do so for the two sub-periods and for the entire sample. From the LM tests we can see that all three models for both exchange rates appear to adequately capture the time series dynamics. No rejections are found in any sub-period for any moment of the transformed residuals.

[19] All tests in this paper will be conducted at the 5% significance level.

[20] The variance matrix used here assumed the time-varying symmetrised Joe-Clayton copula was used to complete the joint distribution. Almost identical results were obtained when the time-varying normal copula was used.

The K-S tests indicate that the normal BEKK model is mis-specified, with five out of the six tests rejecting the null hypothesis of correct specification. The t-BEKK model passes all K-S tests at the 5% significance level, though on two occasions the p-value for the tests are close to 0.05. The copula model margins pass the K-S tests easily.

[INSERT TABLE 3 HERE]

In light of the possible low power of the K-S test, we employ the hit tests discussed in Section 3 to check for the correctness of the specification in particular regions of the support. We use five regions: the lower 10% tail, the interval from the 10^{th} to the 25^{th} quantile, the interval from the 25^{th} to the 75^{th} quantile, the interval from the 75^{th} to the 90^{th} quantile, and the upper 10% tail. These regions represent economically interesting subsets of the support - the upper and lower tails are notoriously difficult to fit, and so checking for correct specification there is important, while the middle 50% of the support contains the 'average' observations. We use as regressors ('Z_{jt}' in the notation of Section 3) a constant, to check that the model implies the correct proportion of hits, and three variables that count the number of hits in that region in the last day, week and month, to check that the model dynamics are correctly specified[21]. The λ_j functions are set to simple linear functions of the parameters and the regressors: $\lambda_j\left(Z_{jt}, \beta_j\right) = Z_{jt} \cdot \beta_j$. The results of the tests in the individual regions and the joint test for all five regions are presented in Table 4 below.

[INSERT TABLE 4 HERE]

Table 4 shows that the normal BEKK model is rejected in numerous individual regions, and fails the joint test for both exchange rates in the pre-euro period and over the entire sample. The t-BEKK model fails in four individual region tests, and fails the joint test for the yen in the pre-euro period and for the entire sample. The copula model margins fail two individual region tests, but pass all joint tests (though two of these have quite low p-values).

Overall, these specification tests show that conditional bivariate normality is easily rejected, and that the assumption of equal fatness-of-tails for these two exchange rates is rejected. Thus both the bivariate normal and bivariate Student's t distributions are mis-specified for these exchange rates.

4.4 Results for the copulas

We now present the results of the estimation of the normal and symmetrised Joe-Clayton (SJC) models. For the purposes of comparison we also present the results for these two copulas when no time variation in the copula parameters is assumed. It should be pointed out, though, that neither of these copulas are closed under temporal aggregation, so if the conditional copula of (X_t, Y_t) is

[21] We also conducted this test including as additional regressors three variables that counted the number of hits in the corresponding region of the other variable's support over the last day, week and month. The results did not change significantly.

normal or SJC, the unconditional copula will not in general be normal or SJC. The results are presented in Table 5.

[INSERT TABLE 5 HERE]

All of the parameters in the time-varying normal copula were found to significantly change following the introduction of the euro, and a test of the significance of a break for this copula yielded a p-value of 0.0000. Using quadrature[22] we computed the implied time path of conditional correlation between the two exchange rates, and present the results in Figure 4. This figure shows quite clearly the structural break in dependence that occurred upon the introduction of the euro. The level and the dynamics of (linear) dependence both clearly change. The p-value from the test for a change in level only was 0.0000, and the p-value from a test for a change in dynamics given a change in level was also 0.0000, confirming this conclusion.

[INSERT FIGURE 4 HERE]

For the time-varying SJC copula only the level of dependence was found to significantly change; the dynamics of the parameters of the conditional copula were not significantly different. The significance of the change in level was 0.0000. For the purposes of comparing the results for the SJC copula with the normal copula, we present in Figure 5 the conditional correlation between the two exchange rates implied by the SJC copula. The plot is similar to that in Figure 4, and the change in the level of dependence is again very clear.

[INSERT FIGURE 5 HERE]

In Figures 6 and 7 we present plots of the conditional tail dependence implied by the time-varying SJC copula model. Figure 6 confirms that the change in linear dependence also takes place in tail dependence, with average tail dependence dropping from 0.33 to 0.04 after the break. Figure 7 shows the degree of asymmetry in the conditional copula via the difference between the upper and lower conditional tail dependence measures. In this application, upper (lower) tail dependence measures the dependence between the exchange rates on days when the U.S. dollar is appreciating (depreciating). Our unconditional results suggested that the limiting probability of the dollar appreciating heavily against the mark, given that it has appreciated heavily against the yen, is about 20% greater than the corresponding depreciation probability[23], meaning that the exchange rates are less dependent in bad markets for the US dollar than in good markets.

[INSERT FIGURES 6 AND 7 HERE]

[22]We use Gauss-Legendre quadrature, with ten nodes for each margin, leading to a total of 100 nodes. See Judd (1998) for more on this technique.

[23]As tail dependence is a symmetric concept, it does not matter which of the two currencies one conditions on the dollar having appreciated against.

In Figure 7, we see that before the break conditional upper tail dependence was almost always greater than conditional lower tail dependence (their difference is almost always above zero) while after the break the reverse is true. In fact, on 94% of days before the break the implied conditional upper tail dependence was greater than conditional lower tail dependence, and conditional lower tail dependence was greater than upper tail dependence on every single day after the break. Because we used the same forcing variable in the evolution equations for both upper and lower dependence, we can formally test for the significance of asymmetry in the conditional copula. The p-value on this test is 0.0103 in the pre-euro sample and 0.1413 in the post-euro sample (and 0.0261 overall). Thus we have strong evidence that the conditional dependence structure between the DM-dollar and yen-dollar exchange rates was asymmetric over the period January 1991 to December 1998, a finding that has not been previously reported in the empirical exchange rate literature, and one that we would not have been able to capture with standard multivariate distributions like the normal or Student's t.

The presence of such an asymmetry has many possible economic implications. As we suggested above, such an asymmetry may be a result of an asymmetric reaction by the three relevant central banks to appreciations versus depreciations of their currency. This asymmetry may also have implications for the distribution of portfolios of international stocks and bonds: Patton (2002b) shows that asymmetric dependence can lead to skewed portfolios even when the individual assets have zero skewness, a case that seems quite relevant here, and recent work in the empirical finance literature has found that stocks that exhibit conditional skewness and asymmetric dependence carry a market premium, see Harvey and Siddique (2000) and Ang, *et al.* (2002).

4.5 Goodness-of-fit tests and comparisons

In Table 6 we present the results of the bivariate 'hit' tests. We divided the support of the copula into seven regions[24], each with an economic interpretation, and one 'remnant' region. The regions are presented graphically in Figure 3. Regions 1 and 2 correspond to the lower and upper 10% Value-at-Risk for each variable. The ability to correctly capture the probability of both exchange rates taking on extreme values simultaneously is of critical importance to portfolio managers and macroeconomists, amongst others. Regions 3 and 4 represent moderately large up and down days: days in which both exchange rates were somewhere between their 10^{th} and 25^{th}, or 75^{th} and 90^{th}, quantiles. Region 5 is the 'median' region: days when both exchange rates were in the middle 50% of their distributions. Regions 6 and 7 are the extremely asymmetric days, those days when one exchange rate was in the upper 25% of its distribution while the other was in the lower 25% of its distribution.

[INSERT FIGURE 3]

[24]Using rectangular regions makes computing the probability of a hit in that region implied by the copula model, $\hat{C}$, particularly simple: it is just the $\hat{C}$-volume of the region, defined in Section 2.

We again specify a simple linear function for λ_j, that is: $\lambda_j\left(Z_{jt}, \beta_j\right) = Z_{jt} \cdot \beta_j$, and we include in Z_{jt} a constant term, to capture any over- or under-estimation of the unconditional probability of a hit in region j, and three variables that count the number of hits that occurred in the past day, one week and one month, to capture any violations of the assumption that the hits are serially independent. The results for each of the seven regions, for the four models considered are presented below. For the joint test we define the zeroth region as that part of the support not covered by regions one to seven.

[INSERT TABLE 6 HERE]

Table 6 again shows that the normal BEKK model is rejected by the data. This model fails all three joint tests and numerous individual region tests. Note that, since the marginal distribution models were also rejected, it is possible that marginal mis-specification is causing the copula model mis-specification. The t-BEKK model passes all joint tests and all individual region tests, though some p-values are close to 0.05. Both the normal and the SJC copula models pass all individual region and joint tests easily.

Finally, we conducted likelihood ratio tests to compare the competing models. For the comparison of the normal BEKK and the t-BEKK we may use a standard likelihood ratio test as the normal BEKK model is nested within the t-BEKK model. This comparison yielded a p-value of 0.0000, clear evidence that the t-BEKK model is preferred to the normal BEKK model. The remaining comparisons all involve non-nested models, and so we use the Rivers and Vuong (2002) test described in Section 3.2. Using this test we also found that the normal and SJC copula specifications significantly out-performed the normal BEKK model, however no other comparisons were significant. The fact that both the normal and the SJC copula models pass the goodness-of-fit tests, and are not distinguishable using the Rivers and Vuong test indicates the difficulty these tests have in distinguishing between similar models, even with substantial amounts of data. The significance of the conditional asymmetry in the SJC copula, though, suggests that a density model with a symmetric copula, such as the normal copula, is mis-specified.

5 Conclusion

In this paper we verified that the existing theory of copulas may be extended to the conditional case, and applied it to a model of the time-varying conditional joint distribution of the daily Deutsche mark - U.S. dollar and Yen - U.S. dollar exchange rates, over the period from January 1991 to December 2001. Standard AR- tGARCH models were employed for the marginal distributions of each exchange rate, and two different copulas were estimated: the copula associated with the bivariate normal distribution, and the 'symmetrised Joe-Clayton' copula, which allows for asymmetric dependence in the joint distribution. Time-variation in the dependence structure between the two exchange rates was captured by allowing the parameters of the two copulas to vary over the sample period, employing an evolution equation similar to the GARCH model for conditional variances. For comparison, we also estimated models using the BEKK specification for the conditional covariance matrix coupled with a bivariate normal and a bivariate Student's t assumption for the standardised residuals.

Some attention was paid to tests of the relative goodness-of-fit of the copulas analysed. Goodness-of-fit testing in our study was complicated by the fact that we wished to test the adequacy of the entire density, rather than just a set of moments from this density, and by the fact that many of our models were non-nested. We introduced and employed an extension of the 'hit' tests of Christoffersen (1998) and Engle and Manganelli (1999) to test for the goodness-of-fit of the four models considered, and proposed a new test for evaluating the performance of multiple interval forecasts simultaneously. We also used the model selection test for non-nested models recently proposed by Rivers and Vuong (2002).

We found evidence against both the bivariate normal and bivariate Student's t distributional assumptions. However, a bivariate model with univariate Student's t marginal distributions was not rejected once we allowed each exchange rate to exhibit different thickness of tails. We proposed a new copula, the 'symmetrised Joe-Clayton' copula, which allows for the possibility of asymmetric dependence and nests symmetric dependence as a special case. In the pre-euro period we were able to reject the null hypothesis that the two exchange rates under analysis had a symmetric conditional dependence structure: specifically, dependence was greater during appreciations of the U.S. dollar (or alternatively, during depreciations of the mark and the yen) than during depreciations of the U.S. dollar. Finally, we reported strong evidence of a structural break in the conditional copula following the introduction of the euro in January 1999. The dependence between these exchange rates fell dramatically following the break, and the conditional dependence structure went from significantly asymmetric to approximately symmetric.

Many extensions of our analysis are possible. The use of conditional copulas in constructing higher dimension density forecasts is possible, though some care may be required to keep the model tractable. Also, other forms of time variation in the dependence between the variables may be explored: in this paper we considered allowing the parameter of the copula to vary through time, holding the form of the copula fixed. An alternative to this may be to consider conditional copulas that vary in functional form, perhaps in a regime switching model.

6 Appendix A: Proofs

Some of the results below are simplified with the following lemmas. These are adapted from Nelsen (1999).

Lemma 1 *Let $x_1 \leq x_2$ and $y_1 \leq y_2$, then*

$$\begin{aligned} \gamma^x(x) &\equiv H(x, y_2) - H(x, y_1) \text{ is a non-decreasing function of } x, \text{ and} \\ \gamma^y(y) &\equiv H(x_2, y) - H(x_1, y) \text{ is a non-decreasing function of } y. \end{aligned}$$

Proof of Lemma 1. From the definition of a bivariate distribution function we know that $H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1) \geq 0$, so $H(x_2, y_2) - H(x_2, y_1) \geq H(x_1, y_2) - H(x_1, y_1)$ for all $x_1, x_2, y_1, y_2 \in \mathbb{R}$ such that $x_1 \leq x_2, y_1 \leq y_2$. Similarly for $\gamma^y(y)$. ■

Lemma 2 *Let $x_1 \leq x_2$, $y_1 \leq y_2$ and $w_1 \leq w_2$ then*

$$\begin{aligned} \delta^x(z) &\equiv H_{xyw}(z, y_2, w_2) - H_{xyw}(z, y_2, w_1) - H_{xyw}(z, y_1, w_2) + H_{xyw}(z, y_1, w_1) \\ \delta^y(z) &\equiv H_{xyw}(x_2, z, w_2) - H_{xyw}(x_2, z, w_1) - H_{xyw}(x_1, z, w_2) + H_{xyw}(x_1, z, w_1) \\ \delta^w(z) &\equiv H_{xyw}(x_2, y_2, z) - H_{xyw}(x_1, y_2, z) - H_{xyw}(x_2, y_1, z) + H_{xyw}(x_1, y_1, z) \end{aligned}$$

are all non-decreasing functions of z.

Proof of Lemma 2. From the definition of a trivariate distribution function we know that $H_{xyw}(x_2, y_2, w_2) - H_{xyw}(x_2, y_2, w_1) - H_{xyw}(x_2, y_1, w_2) + H_{xyw}(x_2, y_1, w_1) \geq H_{xyw}(x_1, y_2, w_2) - H_{xyw}(x_1, y_2, w_1) - H_{xyw}(x_1, y_1, w_2) + H_{xyw}(x_1, y_1, w_1)$ for all $x_1 \leq x_2$, $y_1 \leq y_2$ and $w_1 \leq w_2$. Thus $\delta^x(x)$ is non-decreasing in x. Similarly for $\delta^y(y)$ and $\delta^w(w)$. ■

Lemma 3 $|H(x_2, y_2) - H(x_1, y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)|$.

Proof of Lemma 3. By the triangle inequality we have:
$|H(x_2, y_2) - H(x_1, y_1)| \leq |H(x_2, y_2) - H(x_1, y_2)| + |H(x_1, y_2) - H(x_1, y_1)|$
Assume $x_1 \leq x_2$ and $y_1 \leq y_2$, then by the above lemma we have

$$\begin{aligned} H(x_2, y_1) - H(x_1, y_1) &\leq H(x_2, y_2) - H(x_1, y_2) \\ &\leq H(x_2, \infty) - H(x_1, \infty) \\ &= F(x_2) - F(x_1) \end{aligned}$$

Considering the case when $x_1 \geq x_2$ and applying the same logic leads us to

$$H(x_1, y_1) - H(x_2, y_1) \leq F(x_1) - F(x_2)$$

So we have

$$|H(x_2, y_1) - H(x_1, y_1)| \leq |F(x_2) - F(x_1)|$$

Similarly for y we find that

$$|H(x_1, y_2) - H(x_1, y_1)| \leq |G(y_2) - G(y_1)|$$

and so

$$\begin{aligned} |H(x_2, y_2) - H(x_1, y_1)| &\leq |H(x_2, y_2) - H(x_1, y_2)| + |H(x_1, y_2) - H(x_1, y_1)| \\ &\leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)| \end{aligned}$$

■

Lemma 4 $|H_{xyw}(x_2, y_2, w_2) - H_{xyw}(x_1, y_1, w_1)| \leq |F_x(x_2) - F_x(x_1)| + |G_y(y_2) - G_y(y_1)| + |F_w(w_2) - F_w(w_1)|$, *where* $(X, Y, W) \sim H_{xyw}$ *and* $X \sim F_x$, $Y \sim G_y$ *and* $W \sim F_w$.

Proof of Lemma 4. Follows using the same steps as the proof of Lemma 3. ■

Proof of Theorem 1. For $0 \leq u \leq 1$ we have:

$$\begin{aligned} \Pr[U \leq u] &= \Pr[F(X) \leq u] \\ &= \Pr\left[F^{(-1)}(F(X)) \leq F^{(-1)}(u)\right] \\ &= \Pr\left[X \leq F^{(-1)}(u)\right] \\ &\equiv F\left(F^{(-1)}(u)\right) \\ &= u \end{aligned}$$

For $u < 0$ we know that $\Pr[U \leq u] = 0$ as the range of F is $[0, 1]$. Similarly for $u > 1$. Thus U has the $Unif(0, 1)$ distribution. ■

Proof of Proposition 1.

1. That the domain of C is $[0, 1] \times [0, 1]$ is given by the fact that the ranges of F and G are both $[0, 1]$. Similarly the range of C is $[0, 1]$ as it is a bivariate distribution function.

2. In Theorem 1 we showed that U and V are $Unif(0, 1)$, thus $\Pr[U \leq 0] = \Pr[V \leq 0] = \Pr[U \leq 0 \cap V \leq v] = \Pr[U \leq u \cap V \leq 0] = 0$ for all $(u, v) \in [0, 1] \times [0, 1]$.

3. $C(u, 1) \equiv \Pr[U \leq u \cap V \leq 1] = \Pr[U \leq u] = u$, as U and V are $Unif(0, 1)$. Similarly for $C(1, v)$.

4. This follows from the fact that C is a bivariate distribution function on $[0, 1] \times [0, 1]$.

■

Proof of Theorem 2. Note that from Lemma 3 and the triangle inequality we have:

$$\begin{aligned} |H(x_2,y_2) - H(x_1,y_1)| &\leq |H(x_2,y_2) - H(x_2,y_1)| \\ &\quad + |H(x_1,y_2) - H(x_1,y_1)| \\ &\leq |F(x_2|w) - F(x_1|w)| + |G(y_2|w) - G(y_1|w)| \\ &= 0 \text{ when } x_1 = x_2 \text{ and } y_1 = y_2 \end{aligned}$$

Thus if $x_1 = x_2$ and $y_1 = y_2$ then $H(x_2,y_2) = H(x_1,y_1)$. The function C is defined by the set of ordered pairs: $\left\{((F(x),G(y)),\ H(x,y)) : (x,y) \in \bar{\mathbb{R}} \times \bar{\mathbb{R}}\right\}$.

That C is a copula must be verified: the domain of C is clearly $[0,1] \times [0,1]$, as this is the range of F and G. The range of C is similarly determined to be $[0,1]$ as this is the range of H. We now check the two conditions for C to be a copula, as given in Definition 1.

1.

$$\begin{aligned} C(u,0) &= H\left(F^{(-1)}(u), G^{(-1)}(0)\right) \\ &= H\left(F^{(-1)}(u), -\infty\right) \\ &= 0 \end{aligned}$$

Similarly for $C(0,v)$. Further,

$$\begin{aligned} C(u,1) &= H\left(F^{(-1)}(u), G^{(-1)}(1)\right) \\ &= H\left(F^{(-1)}(u), \infty\right) \\ &= F\left(F^{(-1)}(u)\right) \\ &= u \end{aligned}$$

Similarly for $C(1,v)$.

2. For this part, let $u_i = F(x_i)$, $v_i = G(y_i)$, and consider the points $x_1, x_2, y_1, y_2 \in R$ such that $x_1 \leq x_2$ and $y_1 \leq y_2$. Then,

$$\begin{aligned} V_C([u_1,u_2] \times [v_1,v_2]) &\equiv C(u_2,v_2) - C(u_1,v_2) - C(u_2,v_1) + C(u_1,v_1) \\ &= H(x_2,y_2) - H(x_1,y_2) - H(x_2,y_1) + H(x_1,y_1) \\ &\equiv V_H([x_1,x_2] \times [y_1,y_2]) \\ &\geq 0 \text{ since } H \text{ is a distribution function.} \end{aligned}$$

Thus the function C defined above is a copula.

The proof of the converse requires us to verify the conditions that make H a distribution function with marginal distributions F and G, given F and G are distribution functions, and C is a copula.

$$\begin{aligned}H(x,-\infty) &= C(F(x),G(-\infty))\\ &= C(F(x),0)\\ &= 0\end{aligned}$$

Similarly for $H(-\infty,y)$. Further,

$$\begin{aligned}H(x,\infty) &= C(F(x),G(\infty))\\ &= C(F(x),1)\\ &= F(x)\end{aligned}$$

Similarly for $H(\infty,y)$. Thus the marginal distributions H and F and G. We now need to show that $V_H \geq 0$.

$$\begin{aligned}V_H([x_1,x_2]\times[y_1,y_2]) &\equiv H(x_2,y_2)-H(x_1,y_2)-H(x_2,y_1)+H(x_1,y_1)\\ &= C(u_2,v_2)-C(u_1,v_2)-C(u_2,v_1)+C(u_1,v_1)\\ &\equiv V_C([u_1,u_2]\times[v_1,v_2])\\ &\geq 0 \text{ since } C \text{ is a copula}\end{aligned}$$

This completes the proof of the converse. Thus the function H is a bivariate distribution function with marginal distributions F and G. ◼

Proof of Corollary 1. This proof follows directly from that of Theorem 2, letting $x \equiv F^{(-1)}(u)$ and $y \equiv G^{(-1)}(v)$, and noting that $u = F\left(F^{(-1)}(u)\right)$ and $v = G\left(G^{(-1)}(v)\right)$ $\forall u,v \in [0,1]$. ◼

Proof of Proposition 2.

1.

$$\begin{aligned}F(-\infty|w) &= f_w(w)^{-1}\cdot\frac{\partial F_{xw}(x,w)}{\partial w}|_{x=-\infty}\\ &= f_w(w)^{-1}\cdot\frac{\partial}{\partial w}(F_{xw}(-\infty,w))\\ &= 0\end{aligned}$$

And

$$\begin{aligned}F(\infty|w) &= f_w(w)^{-1}\cdot\frac{\partial F_{xw}(x,w)}{\partial w}|_{x=\infty}\\ &= f_w(w)^{-1}\cdot\frac{\partial}{\partial w}(F_{xw}(\infty,w))\\ &\equiv f_w(w)^{-1}\cdot\frac{\partial}{\partial w}(F_w(w))\\ &= 1\end{aligned}$$

So the first conditions are satisfied.

2.

$$\begin{aligned} F(x_2|w) - F(x_1|w) &= f_w(w)^{-1} \cdot \left(\frac{\partial}{\partial w} (F_{xw}(x_2, w)) - \frac{\partial}{\partial w} (F_{xw}(x_1, w)) \right) \\ &= f_w(w)^{-1} \cdot \left(\frac{\partial}{\partial w} (F_{xw}(x_2, w) - F_{xw}(x_1, w)) \right) \\ &\geq 0 \end{aligned}$$

by Lemma 1 and the fact that f_w is a density function. Thus the properties of the conditional univariate distribution function are obtained directly from the joint distribution function.

■

Proof of Proposition 3. Let the joint distribution of (X, Y, W) be denoted H_{xyw}. and let the marginal distribution of W be F_w. Then

$$H(x, y|w) \equiv f_w(w)^{-1} \cdot \frac{\partial H_{xyw}(x, y, w)}{\partial w}$$

1.

$$\begin{aligned} H(x, -\infty|w) &= f_w(w)^{-1} \cdot \frac{\partial}{\partial w} (H_{xyw}(x, -\infty, w)) \\ &= 0 \end{aligned}$$

Similarly for $H(-\infty, y|w)$. And

$$\begin{aligned} H(\infty, \infty|w) &= f_w(w)^{-1} \cdot \frac{\partial}{\partial w} (H_{xyw}(\infty, \infty, w)) \\ &\equiv f_w(w)^{-1} \cdot \frac{\partial}{\partial w} (F_w(w)) \\ &= 1 \end{aligned}$$

2.

$$\begin{aligned} V_H([x_1, x_2] \times [y_1, y_2]|w) &\equiv H(x_2, y_2|w) - H(x_1, y_2|w) - H(x_2, y_1|w) + H(x_1, y_1|w) \\ &= f_w(w)^{-1} \cdot \Big(\frac{\partial}{\partial w} (H_{xyw}(x_2, y_2, w) - H_{xyw}(x_1, y_2, w) - \\ &\quad - H_{xyw}(x_2, y_1, w) + H_{xyw}(x_1, y_1, w)) \Big) \\ &\geq 0 \end{aligned}$$

by Lemma 2 and the fact that f_w is a density function.

■

Proof of Proposition 4. We obtain the properties of the conditional copula by deriving results for the joint distribution of $(F(X|W), G(Y|W), W)$. Let $U \equiv F(X|W)$ and $V \equiv G(Y|W)$.

We now consider a conditional version the probability integral transform described in the body of the paper. Consider the distribution of $U|W = w$:

$$\begin{aligned} \Pr[U \leq u|w] &= \Pr[F(X|w) \leq u|w] \\ &= \Pr\left[F^{(-1)}(F(X|w)|w) \leq F^{(-1)}(u|w)|w\right] \\ &= \Pr\left[X \leq F^{(-1)}(u|w)|w\right] \\ &\equiv F\left(F^{(-1)}(u|w)|w\right) \\ &= u \end{aligned}$$

for $u \in [0,1]$. For $u < 0$ $\Pr[U \leq u|w] = 0$ as the range of F is $[0,1]$. Similarly for $u > 1$. Thus the conditional distribution of $U \equiv F(X|W)|W = w$ is $Uniform(0,1)$. Note that this implies that the unconditional distribution of U is also $Uniform(0,1)$. This result hold similarly for V.

Let the joint distribution of (U,V,W) be denoted C^*, and let the joint distribution of (U,W) be denoted $\tilde{C}$. Notice that by the above result that $\tilde{C}(u,w) = u \cdot F_w(w)$. The conditional distribution of $(U,V)|W$ is given by:

$$C(u,v|w) \equiv f_w(w)^{-1} \cdot \partial_3 C^*(u,v,w)$$

where $\partial_3 C^*$ denotes the partial derivative of C^* with respect to its third argument. We do not write $\frac{\partial C^*(u,v,w)}{\partial w}$ as u and v are also functions of w.

We now verify the properties of C given in the proposition.

1.

$$\begin{aligned} C(u,0|w) &= f_w(w)^{-1} \cdot \partial_3(C^*(u,0,w)) \\ &= 0 \end{aligned}$$

Similarly for $C(0,v|w)$. Also

$$\begin{aligned} C(u,1|w) &= f_w(w)^{-1} \cdot \partial_3(C^*(u,1,w)) \\ &= f_w(w)^{-1} \cdot \partial_2\left(\tilde{C}(u,w)\right) \\ &= f_w(w)^{-1} \cdot u \cdot \frac{\partial}{\partial w}(F_w(w)) \\ &= u \end{aligned}$$

Similarly for $C(1,v|w)$. Thus the marginal distributions of $(U,V)|W \sim C$ are $Uniform(0,1)$.

2. The proof of part 2 follows the same reasoning as given to show part 2 in the proof of Proposition 3.

■

Proof of Theorem 3. We will only show the parts that differ from the proof of Sklar's theorem for the unconditional case.

Note that from Lemma 3 and the triangle inequality we have:

$$\begin{aligned}
|H(x_2, y_2|w_2) - H(x_1, y_1|w_1)| &\leq |H(x_2, y_2|w_2) - H(x_2, y_1|w_2)| \\
&\quad + |H(x_1, y_2|w_2) - H(x_1, y_1|w_2)| \\
&\quad + |H(x_1, y_1|w_2) - H(x_1, y_1|w_1)| \\
&\leq |F(x_2|w) - F(x_1|w)| + |G(y_2|w) - G(y_1|w)| \\
&\quad + |H(x_1, y_1|w_2) - H(x_1, y_1|w_1)| \\
&= |H(x_1, y_1|w_2) - H(x_1, y_1|w_1)| \text{ when } x_1 = x_2 \text{ and } y_1 = y_2 \\
&= f_w(w)^{-1} \cdot \left| \frac{\partial}{\partial w} \left(H_{xyw}(x_1, y_1, w_2) - H_{xyw}(x_1, y_1, w_1) \right) \right| \\
&= 0 \text{ when } w_1 = w_2
\end{aligned}$$

Thus if $x_1 = x_2$, $y_1 = y_2$ and $w_1 = w_2$ then $H(x_2, y_2|w_2) = H(x_1, y_1|w_1)$. The function C is defined by the set of ordered pairs: $\{((F(x|w), G(y|w), w),\ H(x, y|w)) : (x, y, w) \in \bar{\mathbb{R}} \times \bar{\mathbb{R}} \times \mathcal{W}\}$.

The remainder of the proof can be shown following the steps of the proof of Theorem 2. ■

Proof of Corollary 2. This proof follows directly from that of Theorem 3, letting $x \equiv F^{(-1)}(u|w)$ and $y \equiv G^{(-1)}(v|w)$, and noting that $u = F\left(F^{(-1)}(u|w)\,|w\right)$ and $v = G\left(G^{(-1)}(v|w)\,|w\right)$ $\forall u, v \in [0, 1]$ and each $w \in \mathcal{W}$. ■

7 Tables and Figures

Table 1: Summary statistics		
	DM-USD	Yen-USD
	Pre-euro	
Mean	0.0053	-0.0090
Std Dev	0.6757	0.7344
Skewness	-0.0149	-0.7486
Kurtosis	4.9642	9.2961
Jarque-Bera stat	327.35	3560.5
Jarque-Bera p-val	0.0000	0.0000
ARCH LM stat	104.05	131.99
ARCH LM p-val	0.0000	0.0000
Linear correlation	0.5085	
	Post-euro	
Mean	0.0358	0.0194
Std Dev	0.6617	0.6735
Skewness	-0.5031	-0.2294
Kurtosis	4.2830	4.2471
Jarque-Bera stat	84.593	55.967
Jarque-Bera p-val	0.0000	0.0000
ARCH LM stat	7.0850	25.253
ARCH LM p-val	0.2144	0.0001
Linear correlation	0.1240	

Note: This table presents some summary statistics of the data used in this paper. The data are 100 times the log-differences of the daily Deutsche mark - U.S. dollar and Japanese yen - U.S. dollar exchange rates. The sample period runs eleven years from January 1991 to December 2001, yielding 2819 observations in total; 2046 prior to the introduction of the euro on January 1, 1999 and 773 after the introduction of the euro.

Table 2: Results for the marginal distributions				
	DM margin		Yen margin	
	Coeff	Std Error	Coeff	Std Error
Const_1	0.0125	0.0120	0.0225	0.0120
Const_2	0.0656	0.0229		
$\text{AR}(1)_1$	0.0032	0.0228	-0.0091	0.0220
$\text{AR}(1)_2$	0.0276	0.0322		
AR(10)	–	–	0.0679	0.0201
GARCH const_1	0.0046	0.0033	0.0077	0.0051
GARCH const_2	0.0011	0.0078		
Lagged variance_1	0.9344	0.0181	0.9394	0.0212
Lagged variance_2	0.9876	0.0304		
Lagged e_1^2	0.0579	0.0150	0.0475	0.0140
Lagged e_2^2	0.0107	0.0160		
nu_1	5.9256	0.7217	4.2554	0.3963
nu_2			6.6540	1.3959

Notes to Table 2 are on the following page.

Note to Table 2: Here we report the maximum likelihood estimates and asymptotic standard errors for the parameters of the two univariate exchange rate models. The subscripts refer to the period before or after the introduction of the euro on January 1, 1999. An empty cell indicates that the parameter did not change following the introduction of the euro. Note that only the Yen margin model contains an AR(10) parameter, and that this parameter did not change throughout the sample.

Table 3: LM tests of serial independence and Kolmogorov-Smirnov tests of the density

	Normal BEKK		Student's t BEKK		Copula margins	
	DM	Yen	DM	Yen	DM	Yen
			Pre-euro			
First moment	0.3861	0.7473	0.4011	0.6405	0.4233	0.8360
Second moment	0.8375	0.4498	0.8098	0.3788	0.8199	0.7107
Third moment	0.9488	0.3854	0.9454	0.3273	0.9421	0.6052
Fourth moment	0.9631	0.4384	0.9727	0.3643	0.9675	0.5751
K-S test	0.0024	0.0000	0.7940	0.0641	0.9889	0.7557
			Post-euro			
First moment	0.5294	0.5253	0.5554	0.5752	0.5416	0.7093
Second moment	0.5928	0.5082	0.5973	0.5914	0.5785	0.6318
Third moment	0.4978	0.3536	0.5068	0.4354	0.4870	0.4563
Fourth moment	0.3571	0.2306	0.3881	0.2923	0.3552	0.3125
K-S test	0.0225	0.0815	0.0924	0.6731	0.8111	0.8077
			Entire sample			
First moment	0.3067	0.5507	0.3292	0.4805	0.3852	0.4833
Second moment	0.7608	0.4640	0.7207	0.4251	0.7623	0.4152
Third moment	0.8698	0.4022	0.8320	0.3973	0.8482	0.3941
Fourth moment	0.9006	0.4002	0.8750	0.4175	0.8798	0.4260
K-S test	0.0002	0.0000	0.4252	0.0562	0.8469	0.4763
LL	-2765.50	-2861.60	-2701.50	-2743.83	-2699.45	-2733.11

Note: This table presents the p-values from LM tests of serial independence of the first four moments of the variables U_t and V_t, described in the text, from the three types of models: BEKK models for variance with normal and Student's t innovations, and marginal models to use with copulas. 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}. Any p-value less than 0.05 indicates a rejection of the null hypothesis that the particular model is well-specified. We also report the p-value from the Kolmogorov-Smirnov test for the adequacy of the distribution model, and the value of the log-likelihood at the optimum for these models.

Table 4: Hit test results for the marginal distributions						
	Normal BEKK		Student's t BEKK		Copula margins	
	DM	Yen	DM	Yen	DM	Yen
			Pre-euro			
Region 1	0.2027	0.0293	0.1749	0.0078	0.1829	0.0060
Region 2	0.0010	0.0007	0.0929	0.1922	0.3375	0.4154
Region 3	0.0000	0.0000	0.0368	0.3025	0.2520	0.7454
Region 4	0.8930	0.3934	0.7245	0.1997	0.8805	0.7300
Region 5	0.0176	0.0133	0.7552	0.4200	0.6975	0.0904
All regions	0.0000	0.0000	0.3306	0.0285	0.6581	0.0532
			Post-euro			
Region 1	0.6351	0.2379	0.7214	0.9004	0.6980	0.9508
Region 2	0.0449	0.7831	0.2492	0.8721	0.9275	0.9565
Region 3	0.0549	0.0672	0.6280	0.6793	0.7376	0.4142
Region 4	0.7666	0.7510	0.9039	0.6906	0.9827	0.5457
Region 5	0.4033	0.4677	0.0997	0.5835	0.0623	0.6613
All regions	0.1872	0.5156	0.5153	0.9664	0.8599	0.9766
			Entire sample			
Region 1	0.7069	0.0087	0.4460	0.0144	0.3381	0.0260
Region 2	0.0000	0.0003	0.0203	0.1358	0.2853	0.3655
Region 3	0.0000	0.0000	0.0769	0.1599	0.2982	0.3104
Region 4	0.8245	0.8258	0.8430	0.2781	0.9185	0.6888
Region 5	0.0268	0.0069	0.8515	0.2802	0.7087	0.0643
All regions	0.0000	0.0000	0.4007	0.0265	0.7991	0.0589

Note: We report the p-values from tests that the models are correctly specified. The test statistic is a χ^2_4 random variable for the individual region tests and a χ^2_{16} random variable for the joint tests. Any p-value less than 0.05 indicates a rejection of the null hypothesis that the particular model is well-specified. 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 5: Results for the copula models				
		Coeff	Std Error	L_C
Constant	ρ_1	0.5391	0.0148	360.628
Normal	ρ_2	0.1444	0.0386	
Constant	τ_1^U	0.3591	0.0247	
symmetrised	τ_1^L	0.2951	0.0271	354.413
Joe-Clayton	τ_2^U	0.0000	0.0000	
	τ_2^L	0.0719	0.0381	
	Const_1	-0.1611	0.0212	
Time-varying	Const_2	0.1642	1.3967	372.911
Normal	α_1	0.0600	0.0140	
	α_2	0.2858	0.0148	
	β_1	2.4872	0.3959	
	β_2	0.6101	0.0386	
	Const_1^U	-1.9105	0.0201	
	Const_2^U	-8.4501	0.0148	
Time-varying	α^U	-1.2534	0.0051	375.322
symmetrised	β^U	4.2345	0.0212	
Joe-Clayton	Const_1^L	2.1675	0.0140	
	Const_2^L	0.1026	0.0386	
	α^L	-8.3017	0.3959	
	β^L	-5.0944	1.3967	

Note: The subscripts refer to the period before or after the introduction of the euro on January 1, 1999. For the first three copulas all parameters changed, while for the fourth copula only the constant terms for the two tails dependence measures changed. $\mathcal{L}_C$ stands for the copula likelihood at the optimum.

Table 6: Hit test results for the copula models				
	Normal BEKK	Student's t BEKK	Normal copula	SJC copula
		Pre-euro		
Region 1	0.1143	0.1365	0.1807	0.2196
Region 2	0.1445	0.7821	0.9130	0.1005
Region 3	0.0161	0.0671	0.7424	0.7614
Region 4	0.5651	0.7424	0.7609	0.9097
Region 5	0.0000	0.5134	0.2737	0.6042
Region 6	0.3562	0.4798	0.7332	0.2665
Region 7	0.0208	0.0679	0.4967	0.2979
All regions	0.0000	0.0915	0.7780	0.3546
		Post-euro		
Region 1	0.1429	0.3857	0.1903	0.4970
Region 2	0.6616	0.9564	0.8134	0.8457
Region 3	0.3349	0.7801	0.7766	0.7782
Region 4	0.3416	0.9071	0.9401	0.9394
Region 5	0.0098	0.3976	0.7068	0.6404
Region 6	0.6038	0.6883	0.8478	0.9072
Region 7	0.0410	0.2149	0.3777	0.3788
All regions	0.0347	0.8375	0.9026	0.9624
		Entire sample		
Region 1	0.3807	0.1830	0.1485	0.2624
Region 2	0.1491	0.7952	0.8882	0.1348
Region 3	0.0080	0.0552	0.5842	0.6094
Region 4	0.8989	0.6642	0.7593	0.8909
Region 5	0.0000	0.8617	0.5131	0.8067
Region 6	0.1368	0.2629	0.9962	0.5975
Region 7	0.0019	0.0831	0.8139	0.2717
All regions	0.0000	0.1166	0.9236	0.5179

Note: We report the p-values from tests that the models are correctly specified. The test statistic is a χ_4^2 random variable for the individual region tests and a χ_{28}^2 random variable for the joint tests. Any p-value less than 0.05 indicates a rejection of the null hypothesis that the particular model is well-specified. 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] Andersen, Torben G., and Bollerslev, Tim, 1998, Answering the Skeptics:Yes, Standard Volatility Models Do Provide Accurate Forecasts, *International Economic Review*, 39(4), 885-905.

[2] Andersen, Torben G., Bollerslev, Tim, Diebold, Francis X., and Labys, Paul, 2001, The Distribution of Realized Exchange Rate Volatility, *Journal of the American Statistical Association*, 96, 42-55.

[3] Ang, Andrew, and Chen, Joe, 2002, Asymmetric Correlations of Equity Portfolios, *Journal of Financial Economics*, 63(3), 443-494.

[4] Ang, Andrew, Chen, Joseph, and Xing, Yuhang, 2001, Downside Risk and Expected Returns, working paper, Columbia Business School.

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

[6] Berkowitz, Jeremy, 2001, Evaluating the Forecasts of Risk Models, *Journal of Business and Economic Statistics*, 19, 465-474.

[7] Bollerslev, Tim, 1986, Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307-327.

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

[9] Bollerslev, Tim, 1990, Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach, *Review of Economics and Statistics*, 72, 498-505.

[10] Bouyé, Eric, Gaussel, Nicolas, and Salmon, Mark, 2000a, Investigating Dynamic Dependence using Copulae, working paper, City University Business School, London.

[11] Bouyè, Eric, Durrleman, Valdo, Nikeghbali, Ashkan, Riboulet, Gael and Roncalli, Thierry, 2000b, Copulas for Finance: A Reading Guide and Some Applications, Working Paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais. France.

[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] Cherubini, U., and Luciano, E., 2000, Bivariate option pricing with copulas, forthcoming, *Applied Mathematical Finance*.

[15] Cherubini, U., and Luciano, E., 2001, Value at Risk Trade-off and Capital Allocation with Copulas, *Economic Notes*, 30 (2).

[16] Christoffersen, Peter F., 1998, Evaluating Interval Forecasts, *International Economic Review*, 39(4), 841-862.

[17] 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.

[18] 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*, 43(2), 210-218.

[19] Costinot, Arnaud, Roncalli, Thierry and Teïletche, Jérôme, 2000, Revisiting the Dependence between Financial Markets with Copulas, Work in progress, Groupe de Recherche Opérationnelle, Crédit Lyonnais, France.

[20] D'Agostino, Ralph B., and Stephens, Michael A., eds., 1986, *Goodness-of-Fit Techniques*, Marcel Dekker Inc., New York.

[21] David, F. N., and Johnson, N. L., 1948, The Probability Integral Transformation when Parameters are Estimated from the Sample, *Biometrika*, 35(1), 182-190.

[22] Dawid, A. P., 1984, Statistical Theory: The Prequential Approach, *Journal of the Royal Statistical Society*, 92, 147-278.

[23] Deheuvels, P., 1978, La fonction de dépendence empirique et ses propriétés. Un test non paramétrique d'indépendence, *Acad. Roy. Belg. Bull. Cl. Sci.*, 65(5), 274-292.

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

[25] 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.

[26] Diebold, Francis X., Hahn, Jinyong, and Tay, Anthony S., 1999, Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange, *The Review of Economics and Statistics*, 81(4), 661-673.

[27] Embrechts, Paul, McNeil, Alexander and Straumann, Daniel, 1999, Correlation: Pitfalls and Alternatives, *RISK*, 12(5), 11-21.

[28] Embrechts, Paul, McNeil, Alexander and Straumann, Daniel, 2001, Correlation and Dependence Properties in Risk Management: Properties and Pitfalls, in M. Dempster, ed., *Risk Management: Value at Risk and Beyond*, Cambridge University Press.

[29] Engle, Robert F., 1982, Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation, *Econometrica,* 50, 987-1007.

[30] Engle, Robert F., and Kroner, Kenneth F., 1995, Multivariate Simultaneous Generalized ARCH, *Econometric Theory*, 11, 122-150.

[31] Engle, Robert F., and Manganelli, Simone, 1999, CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, Working Paper, Department of Economics, University of California, San Diego.

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

[33] Fisher, R. A., 1932, *Statistical Methods for Research Workers*.

[34] Frees, Edward W., Carriere, Jacques, and Valdez, Emiliano A., 1996, Annuity Valuation with Dependent Mortality, *Journal of Risk and Insurance*, 63, 229-261.

[35] Frees, Edward W., and Valdez, Emiliano A., 1998, Understanding Relationships Using Copulas, *North American Actuarial Journal*, 2(1), 1-25.

[36] Frey, Rüdiger, and McNeil, Alexander J., 2001, Modelling Dependent Defaults, working paper, Department of Mathematics, Federal Institute of Technology, ETHZ, Zurich, Switzerland.

[37] Galambos, J., 1978, *The Asymptotic Theory of Extreme Order Statistics*, John Wiley & Sons, New York.

[38] Genest, Christian, and Rivest, Louis-Paul, 1993, Statistical Inference Procedures for Bivariate Archimedean Copulas, *Journal of the American Statistical Association*, 88(423), 1034-1043.

[39] 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.

[40] Hartmann, P., Straetmans, S., and de Vries, C. G., 2001, Asset Market Linkages in Crisis Periods, Working Paper No. 71, European Central Bank.

[41] Harvey, Campbell R., and Siddique, Akhtar, 2000, Conditional Skewness in Asset Pricing Tests, *Journal of Finance*, 55(3), 1263-1295.

[42] Hull, John, and White, Alan, 1998, Value at Risk When Daily Changes in Market Variables are not Normally Distributed, *Journal of Derivatives*, 5(3), 9-19.

[43] Hutchinson, T. P. and Lai, C. D., 1990, *Continuous Bivariate Distributions, Emphasising Applications*, Rumsby, Sydney.

[44] Joe, Harry, 1997, *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

[45] Judd, Kenneth L., 1998, *Numerical Methods in Economics*, MIT Press, Cambridge, Massachusetts.

[46] Kearney, Colm, and Patton, Andrew J., 2000, Multivariate GARCH Modeling of Exchange Rate Volatility Transmission in the European Monetary System, *The Financial Review*, 41, 29-48.

[47] Kimeldorf, G. and Sampson, A. R., 1975, Uniform representations of bivariate distributions, *Communications in Statistics*, 4, 617-627.

[48] Kling, J. L., and Bessler, D. A., 1989, Calibration-based Predictive Distributions: An Application of Prequential Analysis to Interest Rates, Money, Prices and Output, *Journal of Business*, 62, 477-499.

[49] Li, David X., 2000, On Default Correlation: A Copula Function Approach, *Journal of Fixed Income*, 9(4), 43-54.

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

[51] Mills, Frederick C., 1927, *The Behaviour of Prices*, National Bureau of Economic Research, New York.

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

[53] Neyman, J., 1937, Smooth Tests for Goodness of Fit, *Skand. Aktuariedtidskrift*, 20, 149-199.

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

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

[56] Patton, Andrew J., 2001, Estimation of Copula Models for Time Series of Possibly Different Lengths, working paper, Department of Economics, University of California, San Diego.

[57] Patton, Andrew J., 2002a, On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation, working paper, Department of Economics, University of California, San Diego.

[58] Patton, Andrew J., 2002b, Skewness, Asymmetric Dependence, and Portfolios, working paper, Department of Economics, University of California, San Diego.

[59] Pearson, K., 1900, On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be Reasonably Supposed to have Arisen from Random Sampling, *Philosphical Magazine Series*, 50, 157-175.

[60] Pearson, K., 1933, On a Method of Determining Whether a Sample of Size n Supposed to Have Been Drawn from a Parent Population Having a Known Probability Integral has Probably Been Drawn at Random, *Biometrika*, 25, 3/4, 379-410.

[61] Pearson, E. S., 1938, The Probability Integral Transformation for Testing Goodness of Fit and Combining Independent Tests of Significance, *Biometrika*, 30, 1/2, 134-148.

[62] Press, William H., Flannery, Brian P., Teukolsky, Saul A., and Vetterling, William T., 1989, *Numerical Recipes*, Cambridge University Press, New York.

[63] Rivers, Douglas, Vuong, Quang, 2002, Model Selection Tests for Nonlinear Dynamic Models, *The Econometrics Journal*, 5(1), 1-39.

[64] 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.

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

[66] Rosenberg, Joshua V., 2000, Nonparametric Pricing of Multivariate Contingent Claims, Working Paper, Stern School of Business, New York University.

[67] Sancetta, Alessio, and Satchell, Stephen, 2001, Bernstein Approximations to the Copula Function and Portfolio Optimization, Working Paper 2001-5, Department of Applied Economics, University of Cambridge.

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

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

[70] Shao, Jun, 1999, *Mathematical Statistics*, Springer-Verlag, New York.

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

[72] Stephens, Michael A., 1986, Tests Based on EDF Statistics, in D'Agostino, Ralph B., and Stephens, Michael A., eds. *Goodness-of-Fit Techniques*, Marcel Dekker Inc., New York.

[73] Takagi, Shinji, 1999, The Yen and its East Asian Neighbors, 1980-95: Cooperation or Competition?, in Takatoshi Ito and Anne O. Krueger, eds., *Changes in Exchange Rates in Rapidly Developing Countries*, The University of Chicago Press, Chicago, 185-207.

[74] White, Halbert, 1994, *Estimation, Inference and Specification Analysis*, Econometric Society Monographs No. 22, Cambridge University Press, New York.

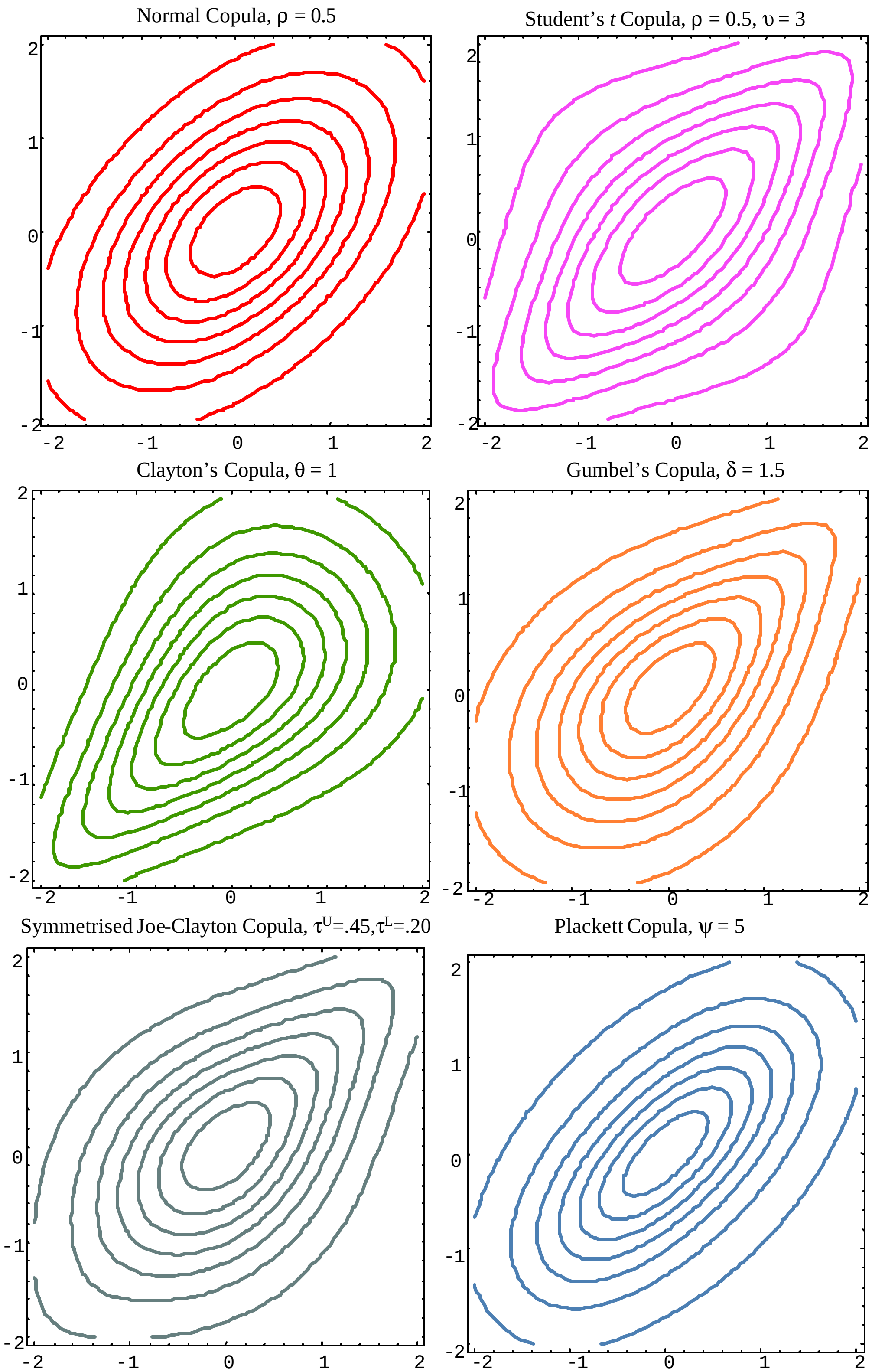

Figure 1: *Contour plots of various distributions all with standard normal marginal distributions and linear correlation coefficients of 0.5.*

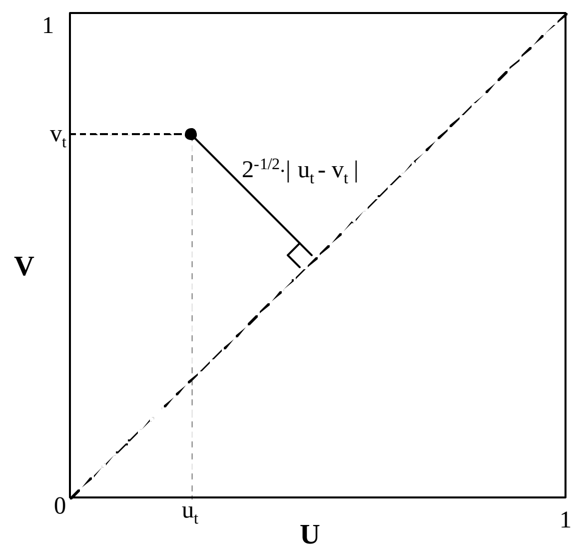

Figure 2: *Motivation for the choice of forcing variable in the specification of the time-varying symmetrised Joe-Clayton copula.*

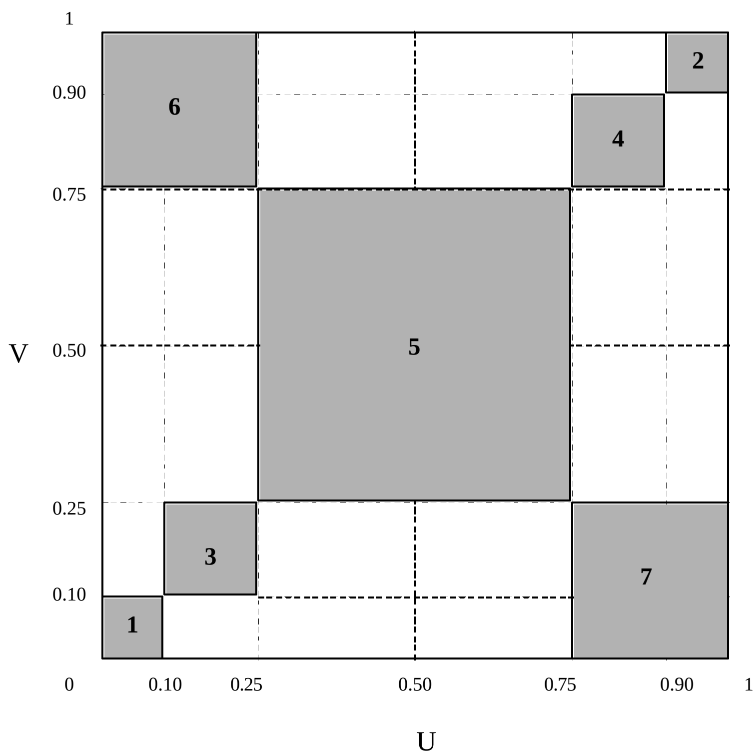

Figure 3: *Regions used in the hit tests*

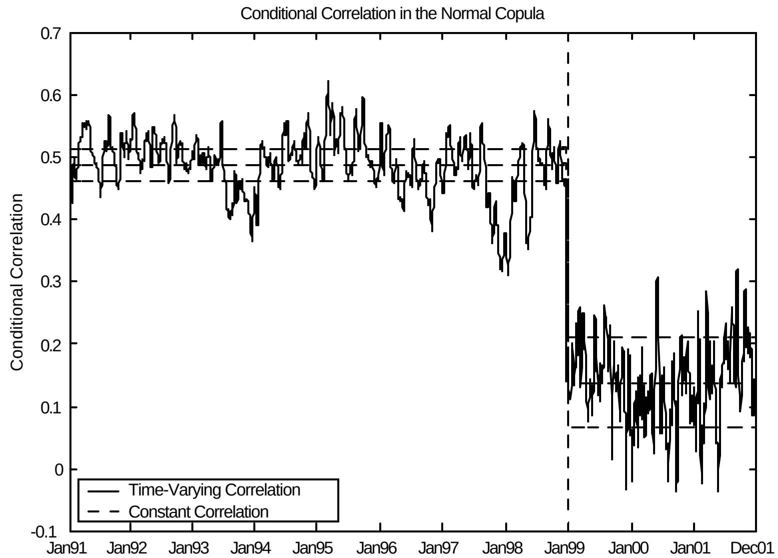

Figure 4: *Conditional correlation estimates from the Normal copulas allowing for a structural break at the introduction of the euro on January 1, 1999, with 95% confidence interval for the constant correlation case.*

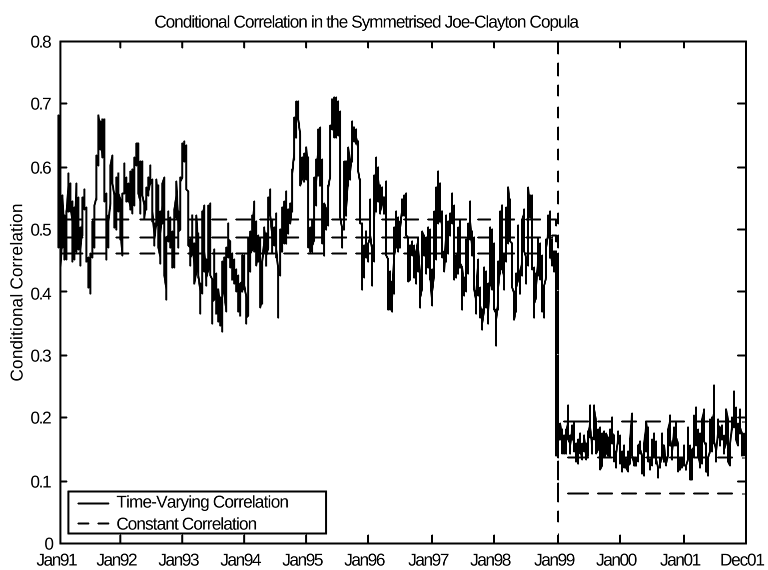

Figure 5: *Conditional correlation estimates from the symmetrised Joe-Clayton copulas allowing for a structural break at the introduction of the euro on January 1, 1999, with 95% confidence interval for the constant correlation case.*

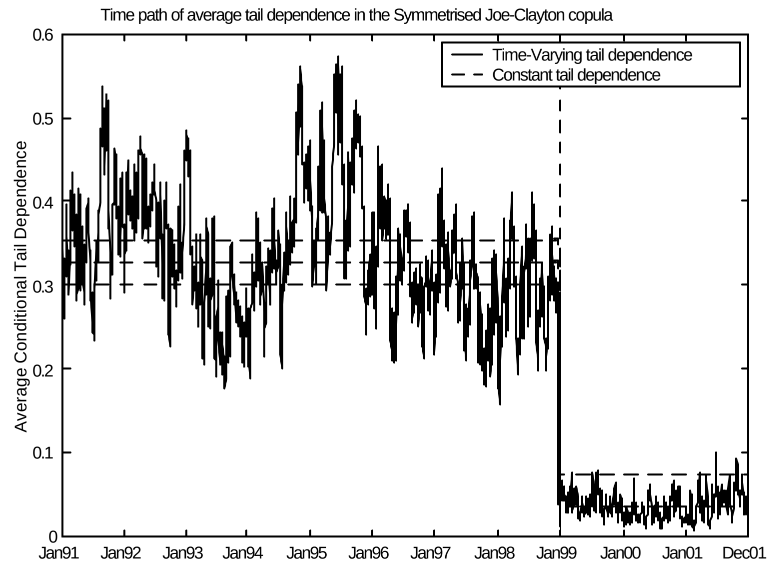

Figure 6: *Average tail dependence from the symmetrised Joe-Clayton copulas allowing for a structural break at the introduction of the euro on January 1, 1999.*

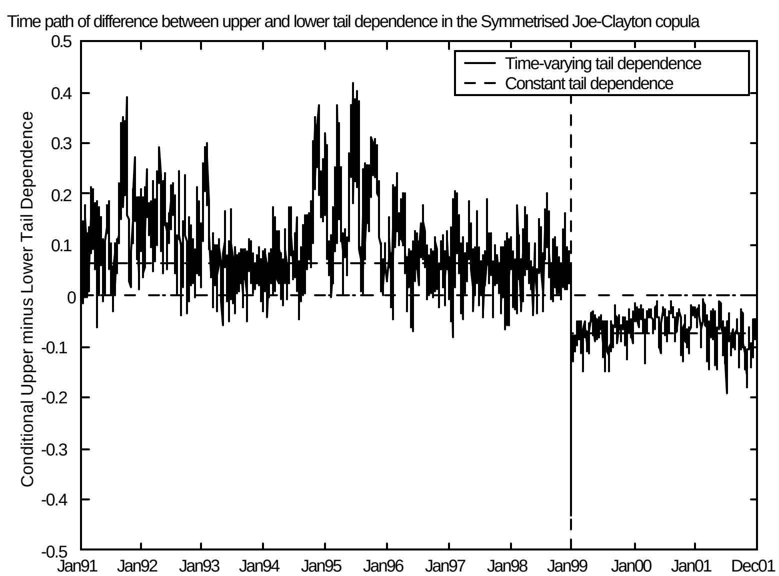

Figure 7: *Difference between upper and lower tail dependence from the symmetrised Joe-Clayton copulas allowing for a structural break at the introduction of the euro on January 1, 1999.*