

ESTIMATION OF MULTIVARIATE MODELS FOR TIME SERIES OF POSSIBLY DIFFERENT LENGTHS

ANDREW J. PATTON*

London School of Economics, Financial Markets Group, Houghton Street, London WC2A 2AE, UK

SUMMARY

We consider the problem of estimating parametric multivariate density models when unequal amounts of data are available on each variable. We focus in particular on the case that the unknown parameter vector may be partitioned into elements relating only to a marginal distribution and elements relating to the copula. In such a case we propose using a multi-stage maximum likelihood estimator (MSMLE) based on all available data rather than the usual one-stage maximum likelihood estimator (1SMLE) based only on the overlapping data. We provide conditions under which the MSMLE is *not less* asymptotically efficient than the 1SMLE, and we examine the small sample efficiency of the estimators via simulations. The analysis in this paper is motivated by a model of the joint distribution of daily Japanese yen–US dollar and euro–US dollar exchange rates. We find significant evidence of time variation in the conditional copula of these exchange rates, and evidence of greater dependence during extreme events than under the normal distribution. Copyright © 2006 John Wiley & Sons, Ltd.

1. INTRODUCTION

The economy cannot be relied upon to generate data in neatly overlapping samples. In financial economics, for instance, cases of unequal amounts of data arise in many interesting applications: the analysis of developed markets and emerging markets; collections of assets that include recently floated companies or companies that went bankrupt; any collection of assets with some dominated in euros and some not.

In this paper we consider the estimation of parametric multivariate density models involving variables with histories of differing lengths. Motivation in economics and finance for multivariate density models beyond the multivariate normal distribution has been provided by the plethora of papers presenting evidence against the assumption of normality for economic variables, starting with Mills (1927) and continuing through to today.¹ Parametric density models are used widely in financial risk management, see Jorion (1997), Duffie and Pan (1997) and Diebold *et al.* (1999), and in macroeconomic forecasting, see Tay and Wallis (2000), Clements (2002) and Wallis (2003).

We consider multivariate models with an unknown parameter vector that may be partitioned into elements relating only to the marginal distributions and elements only relating to the copula. This partition is possible in many common multivariate models, such as vector autoregressions and some multivariate GARCH models. If such a partition is not possible, the familiar one-stage maximum likelihood estimator (1SMLE) using only the overlapping data is the natural estimator to

* Correspondence to: Andrew J. Patton, Financial Markets Group, London School of Economics, Houghton Street, London WC2A 2AE, UK. E-mail: a.patton@lse.ac.uk

Contract/grant sponsor: UCSD Project in Econometric Analysis.

¹ See, *inter alia*, Mendelbrot (1963), Fama (1965), Bollerslev (1987), Richardson and Smith (1993), Erb *et al.* (1994), Longin and Solnik (2001), Ang and Chen (2002) and Bae *et al.* (2003).

employ. When this partitioning is possible, however, great computational savings may be achieved by employing a multi-stage maximum likelihood estimator (MSMLE) using all available data. Furthermore, under conditions provided in Section 2, the MSMLE may not be less efficient than the 1SMLE. Essentially this result relies on the information in the non-overlapping data on one or more of the variables offsetting the loss of information from estimating elements of the parameter vector separately rather than simultaneously. If the differences in the sample sizes available for each variable are asymptotically negligible then, as expected, our results simplify to the standard MSMLE case, see Pagan (1986), Newey and McFadden (1994) or White (1994).

Numerous other authors have considered the problem of non-overlapping data and missing observations. Harvey *et al.* (1998) suggest using the Kalman filter, while Kofman and Sharpe (2000) discuss using the EM algorithm and its Bayesian alternative, the imputation posterior method, for general estimation problems. Little and Rubin (1987) and Weeks (1999) present various methods of dealing with missing observations and numerous further references on the topic.² Anderson (1957) and Stambaugh (1997) suggest using the marginal/conditional distribution decomposition of a joint distribution, for the case of *iid* multivariate normal random variables. The approach of these latter two papers is the most closely related to ours. Our assumption that the multivariate model is constructed using copulas allows us to deal with non-normality, with more irregular data sets, and to further simplify the estimation of the model.

It should be pointed out that the theory presented in this paper is only applicable in the case that the starting dates and ending dates of the series do not contain any information for the parameters of interest that is not contained in the observed data. If the cause of the missing data, the 'missing-data mechanism' in the terminology of Little and Rubin (1987), is related to the data-generating process, then analysing only the observable data will lead to biased inference. A standard example of a non-ignorable missing-data mechanism is when an observation is censored if its value is greater or less than a certain value, see Tobin (1958) for example. Examples in finance where the missing-data mechanism is not ignorable may be found in Brown *et al.* (1995), Goetzmann and Jorion (1999), Weeks (1999) and Kofman and Sharpe (2000), *inter alia*.

The main contribution of this paper is a comparison of the MSMLE estimated using all available data with the usual 1SMLE estimated using only the overlapping data, using both asymptotic theory and small sample simulations. We focus on the bivariate case, but all of our methods extend naturally to higher dimensions. We show that the existing two-stage maximum likelihood framework, see Newey and McFadden (1994) and White (1994), requires only simple extensions to handle this type of irregular data set. We present a sufficient condition for the MSMLE to be *not less* asymptotically efficient than the 1SMLE. In finite samples we find that the MSMLE compares favourably with the 1SMLE for a collection of data-generating processes chosen to be reflective of daily asset returns. Specifically, we find that the use of non-overlapping data generally substantially improves the small sample efficiency of the estimator of the parameters relating to the data with the longer history, without a significant deterioration in the efficiency of the estimators of the remaining parameters. Unlike some alternative methods for dealing with irregular data sets, the MSMLE is easily implemented in standard software packages, and these efficiency results provide further motivation for its use on such data sets.

The second contribution of the paper is the application of the estimator to a model of the joint distribution of daily Japanese yen–US dollar and euro–US dollar exchange rates. These are the

² For studies of particular econometric estimators using non-overlapping and missing data, see Linton (2005), Lynch and Wachter (2004), Schmidt (1977) and Swamy and Mehta (1975), for example.

two most frequently traded exchange rates, making up almost half of total turnover in the FX market,³ and this paper is one of the first to model these important variables jointly. We find significant evidence of time variation in the conditional copula of these exchange rates, indicating that the dependence between the yen and the euro is time-varying. We allow for a non-zero probability of joint extreme movements via the use of the Student's t copula, and find that it is a significant improvement over the normal copula. Our results on the conditional dependence structure between these currencies may be used to improve international portfolio decisions, risk management decisions involving international holdings, or for stress testing risk management models.

The remainder of the paper is organized as follows. In Section 2 we present the multi-stage maximum likelihood estimator and provide a brief introduction to copulas. In Section 3 we present the results of a Monte Carlo study of the small sample properties of the estimator and in Section 4 we apply the estimator to a model of the joint distribution of yen/dollar and euro/dollar exchange rates. We conclude in Section 5. Assumptions required for the consistency and asymptotic normality of the estimator are presented in Appendix A, and proofs are collected in Appendix B.

2. MULTI-STAGE ESTIMATION OF COPULA-BASED MODELS

2.1. Irregular Data Sets

Let us denote the two variables of interest as X and Y . In this paper we allow for the situation that the amount of data available on X is possibly different to that available on Y , which is also possibly different to the amount of overlapping data on both X and Y . This scenario is depicted in Figure 1. Let X denote the variable with the most data available. We will denote the number of observations on X , Y and the common sample as n_x , n_y and n_c respectively. All data lengths are assumed to be (fixed) functions of n_x . We consider cases where $n_y/n_x \rightarrow \lambda_y$ and $n_c/n_x \rightarrow \lambda_c$ as $n_x \rightarrow \infty$, where $0 < \lambda_y \leq 1$ and $0 < \lambda_c \leq 1$. If $\lambda_y = \lambda_c = 1$ then the differences in sample sizes are asymptotically negligible. If $\lambda_y < 1$ and/or $\lambda_c < 1$, then the differences in sample sizes are asymptotically non-negligible. In some cases, such as in our study of the joint density of the yen/dollar and euro/dollar exchange rates, it may appear more accurate to label the

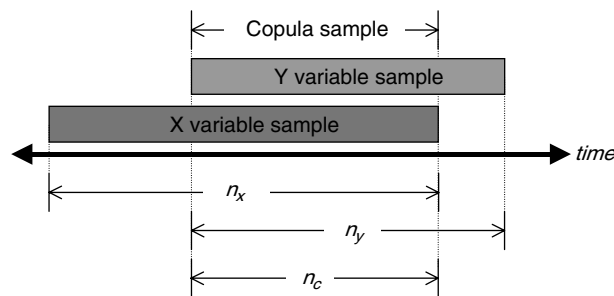


Figure 1. One possible scenario where the amounts of data available on each individual variable are different, as is the amount of data available for the estimation of the copula

³ Source: Bank for International Settlements (2002). The figures quoted here are for April 2001. Euro/USD and yen/USD turnover accounted for approximately 29.5% and 19.3% respectively of the total turnover.

differences in sample sizes as asymptotically negligible.⁴ However, we would argue that treating the differences in sample sizes as asymptotically *non-negligible*, i.e. allowing $\lambda_y < 1$ and $\lambda_c < 1$, generates asymptotic theory that better reflects the small sample reality, and hopefully yields distribution results that are a closer approximation to the true unknown small sample distribution of the parameter estimates. This is because the differences in sample sizes, which are present in our finite sample, are maintained as we move towards asymptopia.

2.2. Multivariate Models and Copulas

We focus on multivariate density models constructed using copulas. The theory of copulas dates back to Sklar (1959), who showed that one may decompose a joint distribution into its k univariate marginal distributions and a copula, which describes the dependence between the variables. Below we focus on the bivariate case, though both the theory of copulas and the estimation methods presented here extend quite naturally to the general multivariate case. We will assume in this paper that the distribution functions F , G and H are continuous, and sufficiently smooth for all required derivatives to exist.

In the bivariate case, if we let $X \sim F$, $Y \sim G$ and $(X, Y) \sim H$, then we may write:

$$H(x, y) = C(F(x), G(y)) \quad (1)$$

$$h(x, y) = f(x) \cdot g(y) \cdot c(F(x), G(y)) \quad (2)$$

where C is the copula of (X, Y) , and c is the copula density. As usual, we denote a distribution function (*cdf*) with an upper case letter and a density function (*pdf*) with a lower case letter. A copula links marginal distributions together to form a joint distribution, and completely describes the dependence between the variables. By construction, it is a multivariate distribution function with Uniform $(0, 1)$ marginals. See Joe (1997) or Nelsen (1999) for an introduction to copulas.

One of the uses of this theorem in econometric modelling is in the construction of flexible multivariate distributions: we may combine a mix of k marginal distributions of any form with any copula to form a valid multivariate distribution. Most existing multivariate distributions are simple extensions of univariate distributions, and often have the restrictive property that all of the marginal distributions are of the same type (by its construction, all marginal distributions of a multivariate Student's t_6 are univariate Student's t_6 , for example). If the individual variables of interest were known to be best fitted by different univariate distributions, the choice of a suitable joint distribution was difficult.⁵ The application of copula theory to the analysis of economic problems is a fast-growing field: examples of work in this field include Chen and Fan (2002), Chen *et al.*, (2004), Cherubini and Luciano (2002), Embrechts *et al.* (2003), Fermanian and Scaillet (2003), Li (2000), Patton (2004, 2005), Rockinger and Jondeau (2001) and Rosenberg (2003).

There is a large body of work on the estimation theory underlying the numerous applications of copula theory that have appeared in the statistics literature, see Oakes (1982), Genest and Rivest

⁴ The euro was introduced in January 1999, and so we have less data available on that currency than we do on the yen. But from now on we will receive new data on both exchange rates each day and so as time progresses it is true that n_y/n_x and n_c/n_x will converge to one.

⁵ As recently as Farebrother (1992), for example, it was a considerable challenge in econometrics to construct an asymmetric bivariate density with common marginal densities. Employing copula theory renders the task almost trivial: select any asymmetric copula and use it to link any two marginal distributions of the same type. Suitable copulas include the Clayton and the Gumbel copulas, see Joe (1997) or Nelsen (1999).

(1993), Genest *et al.* (1995), Shih and Louis (1995), Joe and Xu (1996), Capéraà *et al.* (1997) and Glidden (2000). This theory, however, was developed for applications where the data could be assumed to be independent and identically distributed (i.i.d.), an assumption that is rejected for almost every economic time series. We employ the two-stage maximum likelihood framework of Newey and McFadden (1994) and White (1994) to overcome this restriction.

Let the conditional distribution $(X_t, Y_t)|\mathcal{F}_{t-1}$ be parameterized as $H_t(\theta_0) = C_t(F_t(\varphi_0), G_t(\gamma_0); \kappa_0)$. We assume that H_t is known, but that θ_0 must be estimated. In general $h_t = h_t(x_t, y_t; Z^{t-1}, \theta_0)$ where Z^{t-1} is a vector of elements of \mathcal{F}_{t-1} . For simplicity, we will write $h_t(x_t, y_t; Z^{t-1}, \theta_0)$ as $h_t(Z^t; \theta_0)$, although of course not all of the elements of Z^t will be required. Similarly, we will write $f_t(Z^t; \varphi_0)$, $g_t(Z^t; \gamma_0)$ and $c_t(F_t(Z^t; \varphi_0), G_t(Z^t; \gamma_0); Z^{t-1}, \kappa_0)$ or $c_t(Z^t; \theta_0)$. Assume for now that the data are fully overlapping and that $n_x = n_y = n_c \equiv n$. Then

$$h_t(Z^t; \theta_0) = f_t(Z^t; \varphi_0) \cdot g_t(Z^t; \gamma_0) \cdot c_t(F_t(Z^t; \varphi_0), G_t(Z^t; \gamma_0); Z^{t-1}, \kappa_0) \tag{3}$$

$$\begin{aligned} \mathcal{L}_{XY}(\theta_0) &\equiv n^{-1} \sum_{t=1}^n \log h_t(Z^t; \theta_0) \\ &= n^{-1} \sum_{t=1}^n \log f_t(Z^t; \varphi_0) + n^{-1} \sum_{t=1}^n \log g_t(Z^t; \gamma_0) \\ &\quad + n^{-1} \sum_{t=1}^n \log c_t(F_t(Z^t; \varphi_0), G_t(Z^t; \gamma_0); Z^{t-1}, \kappa_0) \end{aligned} \tag{4}$$

$$\equiv \mathcal{L}_X(\varphi_0) + \mathcal{L}_Y(\gamma_0) + \mathcal{L}_C(\varphi_0, \gamma_0, \kappa_0) \tag{5}$$

where $\varphi_0 \in \text{int}(\Phi) \subseteq \mathbb{R}^p$, $\gamma_0 \in \text{int}(\Gamma) \subseteq \mathbb{R}^q$, $\kappa_0 \in \text{int}(\mathcal{K}) \subseteq \mathbb{R}^r$ and so $\theta_0 \equiv [\varphi_0', \gamma_0', \kappa_0']' \in \text{int}(\Theta) \equiv \text{int}(\Phi) \times \text{int}(\Gamma) \times \text{int}(\mathcal{K}) \subseteq \mathbb{R}^{p+q+r} \equiv \mathbb{R}^s$, where $\text{int}(\mathcal{A})$ is the interior of the set \mathcal{A} .

It will not always be the case that the parameter vector θ_0 decomposes neatly into three components associated with the first margin, second margin and the copula. Examples of common models where such a decomposition is possible are vector autoregressions and certain multivariate GARCH models, including the CCC model of Bollerslev (1990), who was the first to propose multi-stage estimation of multivariate GARCH models, and the DCC model of Engle (2002). Common models where the decomposition is not possible are multivariate ARMA models and some other multivariate GARCH models such as the BEKK model, see Engle and Kroner (1995) and Bauwens *et al.* (2003). Parameter restrictions across marginal distributions can also cause this decomposition to fail to hold.

2.3. The Estimator

Our multi-stage maximum likelihood estimator, MSMLE, is denoted $\hat{\theta}_n$, and its components are given below.⁶ To simplify notation, we assume that all samples (on X , Y and the common sample) start at $t = 1$ and run through until $t = n_x$, n_y and n_c respectively.

$$\hat{\varphi}_{n_x} \equiv \arg \max_{\varphi \in \Phi} n_x^{-1} \sum_{t=1}^{n_x} \log f_t(Z^t; \varphi) \tag{6}$$

⁶ When a quantity, such as the MSMLE, depends on the different sample sizes, n_x , n_y and n_c , we will denote it simply with a subscript n .

$$\hat{\gamma}_{n_y} \equiv \arg \max_{\gamma \in \Gamma} n_y^{-1} \sum_{t=1}^{n_y} \log g_t(Z^t; \gamma) \quad (7)$$

$$\hat{\kappa}_{n_c} \equiv \arg \max_{\kappa \in \mathcal{K}} n_c^{-1} \sum_{t=1}^{n_c} \log c_t(Z^t; \hat{\varphi}_{n_x}, \hat{\gamma}_{n_y}, \kappa) \quad (8)$$

$$\hat{\theta}_n \equiv [\hat{\varphi}'_{n_x}, \hat{\gamma}'_{n_y}, \hat{\kappa}'_{n_c}]' \quad (9)$$

Allowing for differing sample sizes causes no complications beyond the standard case for proving the consistency of this estimator, and so we do not present these results. We present below an asymptotic normality result for the MSMLE presented above. This result is a simple extension of the two-stage MLE framework discussed in Newey and McFadden (1994) and White (1994), both of which provide thorough reviews of maximum likelihood estimation theory. The result relies on standard regularity conditions required for asymptotic normality of an MLE, presented in Appendix A. The asymptotic covariance matrix of this estimator is slightly different to that of the standard two-stage MLE; the key difference introduced when allowing for histories of different lengths is the presence of the matrix $\mathcal{N}^{1/2}$, rather than \sqrt{n} as in the standard case.

Theorem 1 *Denote the number of observations on X, Y and the common sample as n_x , n_y and n_c respectively. Let $n_x \geq n_y \geq n_c$ and let $n_y/n_x \rightarrow \lambda_y$ and $n_c/n_x \rightarrow \lambda_c$ as $n_x \rightarrow \infty$, where $0 < \lambda_y \leq 1$ and $0 < \lambda_c \leq 1$. Under regularity conditions presented in Appendix A, the estimator $\hat{\theta}_n$ defined in equation (9) satisfies:*

$$B_n^{0-1/2} \cdot \mathcal{N}^{1/2} \cdot A_n^0 \cdot (\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, I_s) \quad (10)$$

where I_k is a $k \times k$ identity matrix, and

$$\mathcal{N} \equiv \begin{bmatrix} n_x \cdot I_p & 0 & 0 \\ 0 & n_y \cdot I_q & 0 \\ 0 & 0 & n_c \cdot I_r \end{bmatrix} \quad (11)$$

$$Hess_n^0 \equiv \begin{bmatrix} n_x^{-1} \sum_{t=1}^{n_x} \nabla_{\varphi\varphi} \log f_t^0 & 0 & 0 \\ 0 & n_y^{-1} \sum_{t=1}^{n_y} \nabla_{\gamma\gamma} \log g_t^0 & 0 \\ n_c^{-1} \sum_{t=1}^{n_c} \nabla_{\varphi\kappa} \log c_t^0 & n_c^{-1} \sum_{t=1}^{n_c} \nabla_{\gamma\kappa} \log c_t^0 & n_c^{-1} \sum_{t=1}^{n_c} \nabla_{\kappa\kappa} \log c_t^0 \end{bmatrix} \quad (12)$$

$$OPG_n^0 \equiv \begin{bmatrix} n_x^{-1} \sum_{t=1}^{n_x} s_{1t}^0 \cdot s_{1t}^{0'} & (n_x n_y)^{-1/2} \sum_{t=1}^{n_y} s_{1t}^0 \cdot s_{2t}^{0'} & (n_x n_c)^{-1/2} \sum_{t=1}^{n_c} s_{1t}^0 \cdot s_{3t}^{0'} \\ (n_x n_y)^{-1/2} \sum_{t=1}^{n_y} s_{2t}^0 \cdot s_{1t}^{0'} & n_y^{-1} \sum_{t=1}^{n_y} s_{2t}^0 \cdot s_{2t}^{0'} & (n_y n_c)^{-1/2} \sum_{t=1}^{n_c} s_{2t}^0 \cdot s_{3t}^{0'} \\ (n_x n_c)^{-1/2} \sum_{t=1}^{n_c} s_{3t}^0 \cdot s_{1t}^{0'} & (n_y n_c)^{-1/2} \sum_{t=1}^{n_c} s_{3t}^0 \cdot s_{2t}^{0'} & n_c^{-1} \sum_{t=1}^{n_c} s_{3t}^0 \cdot s_{3t}^{0'} \end{bmatrix} \quad (13)$$

$$\text{and } A_n^0 = E [Hess_n^0], B_n^0 = E [OPG_n^0] \quad (14)$$

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

Remark 1 *If $n_x = n_y = n_c \equiv n$, then the result above simplifies to the standard case:*

$$B_n^{0-1/2} \cdot A_n^0 \cdot \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, I_s) \quad (15)$$

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

Following White (1994), we say that if $\ddot{V}_n^{-1/2}\sqrt{n}(\ddot{\theta}_n - \theta_0) \xrightarrow{D} N(0, I)$, then the asymptotic covariance matrix of the estimator $\ddot{\theta}_n$ is \ddot{V}_n , or that $\text{avar}(\ddot{\theta}_n) = \ddot{V}_n$. For the MSMLE we have $B_n^{0^{-1/2}} \cdot \mathcal{N}^{*1/2} \cdot A_n^0 \cdot \sqrt{n_c}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, I)$, where $\mathcal{N}^* = n_c^{-1}\mathcal{N}$. Thus the asymptotic covariance matrix is $A_n^{0^{-1}} \cdot \mathcal{N}_\infty^{*-1/2} \cdot B_n^0 \cdot \mathcal{N}_\infty^{*-1/2} \cdot A_n^{0^{-1'}}$, where

$$\mathcal{N}_\infty^* \equiv \lim_{n_x \rightarrow \infty} \mathcal{N}^* = \begin{bmatrix} \lambda_c^{-1} \cdot I_p & 0 & 0 \\ 0 & \lambda_y \lambda_c^{-1} \cdot I_q & 0 \\ 0 & 0 & I_r \end{bmatrix} \tag{16}$$

If the differences in sample sizes are asymptotically negligible then $\mathcal{N}_\infty^* = I_s$, and $\text{avar}(\hat{\theta}_n)$ takes the same form as the standard MSMLE.

Under standard conditions, see White (1994), the asymptotic covariance matrix may be consistently estimated using the Hessian and the outer-product of the scores evaluated at the MSMLE. Thus one estimator of $V[\hat{\theta}_n]$ is $\text{Hess}_n(\hat{\theta}_n)^{-1} \mathcal{N}^{-1/2} \text{OPG}_n(\hat{\theta}_n) \mathcal{N}^{-1/2} \text{Hess}_n(\hat{\theta}_n)^{-1}$. If $n_x = n_y = n_c$, then this estimator simplifies to $n_x^{-1} \text{Hess}_n(\hat{\theta}_n)^{-1} \text{OPG}_n(\hat{\theta}_n) \text{Hess}_n(\hat{\theta}_n)^{-1}$, which is the usual so-called ‘sandwich estimator’ of the variance of a MLE.

2.4. A Comparison of the Efficiency of the 1SMLE and MSMLE

The asymptotic efficiency of two asymptotically normal estimators can be compared by examining the difference of their asymptotic covariance matrices: if the matrix difference is positive (negative) semi-definite, then the first estimator is asymptotically less (more) efficient than the second estimator. The small sample efficiency of these estimators is compared via a Monte Carlo experiment in the next section.

In the case that $n_x = n_y = n_c$ it is well known (see Le Cam, 1956, for example) that the 1SMLE is the most efficient estimator, in that it attains the minimum asymptotic variance bound, while the MSMLE does not attain this bound. Similarly, if the difference between the sample sizes is asymptotically negligible then the MSMLE is asymptotically less efficient than the 1SMLE, regardless of the magnitude of the difference in the (finite) sample sizes. However, if the difference between the sample sizes is asymptotically non-negligible then the following proposition shows that there exist situations in which the MSMLE is asymptotically *not less* efficient than the 1SMLE.

Proposition 1 *Let the MSMLE be denoted $\hat{\theta}_n$, with asymptotic covariance matrix $A_n^{0^{-1}} \cdot \mathcal{N}_\infty^{*-1/2} \cdot B_n^0 \cdot \mathcal{N}_\infty^{*-1/2} \cdot A_n^{0^{-1'}}$. Let the 1SMLE and its asymptotic covariance matrix be denoted $\hat{\theta}_{n_c}^{eff}$ and $M_{n_c}^0$ respectively. Define $D_n^0 \equiv A_n^{0^{-1}} \cdot B_n^0 \cdot A_n^{0^{-1'}}$.*

If $\lim_{n_x \rightarrow \infty} n_c/n_x \equiv \lambda_c < 1$ or $\lim_{n_y \rightarrow \infty} n_c/n_y \equiv \lambda_{cy} < 1$ and if λ_c or λ_{cy} is ‘sufficiently small’, then the MSMLE is not less efficient than the 1SMLE. If we let M_{ij} denote the (i, j) th element of the matrix $M_{n_c}^0$ and similarly for D_n^0 , then a sufficient condition is that $\lambda_c < M_{ii}/D_{ii}$ for some $i \in [1, p]$, or that $\lambda_{cy} < M_{jj}/D_{jj}$ for some $j \in [p + 1, q]$.

Thus there are situations where neither estimator is more efficient than the other: the MSMLE is a more efficient estimator of the parameters in the marginal distribution (in the proof of the above

proposition this is for the first parameter of the first marginal distribution) while the 1SMLE is a more efficient estimator of the copula parameters. The intuition behind this is that it is possible to have enough extra observations on the marginal distributions to offset the loss of information incurred by estimating each marginal distribution separately. If this is the case then the MSMLE will be more efficient in the estimation of the marginal parameters. Regardless of the amount of extra information available on the marginal distributions, the MSMLE for the copula parameters will always be weakly less asymptotically efficient than the 1SMLE.

If X_t and Y_t are conditionally independent then estimating the parameters of the two marginal distributions separately involves no loss of information. In fact, by estimating the two marginal distribution models separately we correctly impose the assumption that the variables are independent. Thus it is not surprising that the MSMLE is *weakly more efficient* than the 1SMLE in this case (the proof is straightforward and omitted in the interest of brevity). If $n_x = n_y = n_c$ then the two estimators are equally asymptotically efficient in this case.

3. SMALL SAMPLE PROPERTIES

In this section we present the results of a Monte Carlo study of the small sample properties of the estimators discussed above for a representative collection of DGPs. Parametric density models have been used for macroeconomic variables such as inflation and GDP growth, see Tay and Wallis (2000), Clements (2002) and Wallis (2003), and for asset returns, see Bollerslev (1987), Hansen (1994), Diebold *et al.* (1998, 1999) and Hong *et al.* (2004). The simulation DGPs in this paper are designed to reflect the stylized facts about daily asset returns: weak serial dependence in the conditional mean, and highly persistent conditional variance.

3.1. Simulation Design

We consider three different DGPs. All three DGPs are bivariate distributions, with both marginals being conditionally normal with the same AR(1)–GARCH(1,1) specifications:

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

The DGPs differ in the amount of dependence between the two variables. We examine the case that the variables have the Clayton copula, with the copula parameter chosen so as to imply rank correlations of 0.25, 0.50 and 0.75. See Nelsen (1999) for more on this copula.

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

We do not consider DGPs with time-varying conditional dependence, nor time-varying higher-order marginal moments, in order to keep the simulation tractable. In addition to the three DGPs, we consider six possible data situations: $n_x = 1500$ and 3000 , and $n_y/n_x = 0.25, 0.50$ and 0.75 . In

all cases we assume that $n_c = n_y$. The estimators considered are the MSMLE, $\hat{\theta}_n \equiv [\hat{\varphi}'_{n_x}, \hat{\gamma}'_{n_y}, \hat{\kappa}'_{n_c}]'$, and the standard 1SMLE, $\hat{\theta}_{n_c}^{eff}$. We will compare the estimators by looking at their mean squared error⁷ (MSE) over 1000 replications.

3.2. Results

We computed the ratio of MSEs of the MSMLE to the 1SMLE for each of the 11 parameters of the model.⁸ A ratio of less than one indicates that the MSMLE has a lower MSE than the 1SMLE. To simplify interpretation we present only a summary of the complete results in Table I; the complete results are presented in Patton (2002). For the summary results, we present the average of the first marginal distribution's five parameter MSE ratios, and similarly for the second marginal distribution.⁹ The copula contains only one parameter, and so we present the actual ratio of MSEs in this case.

Table I. Efficiency of the MSMLE relative to the 1SMLE

This table presents the ratio of the mean-squared error of the multi-stage maximum likelihood estimator (MSMLE) of a given parameter to the one-stage maximum likelihood estimator (1SMLE) of that parameter. We present the average ratios across the five parameters in each marginal distribution, and the actual ratio for the (single) copula parameter. n_x is the number of observations on the first margin, and n_y/n_x is the ratio of the number of observations on the second margin to those on the first. We set $n_c = n_y$. ρ is the rank correlation between the two variables. All simulations were done with 1000 replications.

	$n_x = 1500$			$n_x = 3000$		
	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.75$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.75$
	$n_y/n_x = 0.25$					
First margin	0.16	0.23	0.48	0.21	0.27	0.58
Second margin	1.07	1.94	3.07	1.18	1.74	4.07
Copula	0.24	1.03	1.42	0.97	0.87	0.99
	$n_y/n_x = 0.50$					
First margin	0.42	0.71	0.77	0.55	0.66	1.27
Second margin	1.51	1.96	3.81	1.19	1.50	2.83
Copula	0.93	0.90	1.08	1.01	1.03	1.01
	$n_y/n_x = 0.75$					
First margin	0.75	1.64	2.21	0.79	1.20	2.10
Second margin	1.07	1.52	3.49	1.08	1.56	2.75
Copula	0.92	0.95	0.94	0.98	0.97	1.01

⁷ Recall that the MSE of an estimator is defined as $MSE(\hat{\theta}) \equiv R^{-1} \sum (\hat{\theta}_i - \theta_0)^2$ where $\hat{\theta}$ is the estimator, θ_0 is the true parameter, $\hat{\theta}_i$ is the estimate based on the i th Monte Carlo replication, and R is the number of replications.

⁸ In a previous version of this paper, see Patton (2002), we also studied the one-step adjusted multi-stage maximum likelihood estimator, see Newey and McFadden (1994) or White (1994). This estimator is a single step modification of the MSMLE, which achieves the minimum asymptotic variance bound and may be applied to any consistent and asymptotically normal estimator. This estimator is asymptotically fully efficient, and so has the same asymptotic covariance matrix as the 1SMLE. The small sample efficiency of this estimator is often adversely affected by the lack of precision in the estimated Hessian that is required for its computation. This was true in our case: this estimator generally performed much worse than both the 1SMLE and the MSMLE. We do not present these results here.

⁹ The complete results in Patton (2002) confirm that examining only the mean ratio for each marginal distribution does not distort the general conclusions from this simulation.

For the low overlap case ($n_y/n_x = 0.25$) we can see that for no level of rank correlation was the 1SMLE as good as the MSMLE for the parameters of the first margin; all MSE ratios are less than one. For the parameters of the second margin the MSMLE ranged from slightly worse (MSE ratios of 1.07 and 1.18) to much worse (MSE ratios of 3.07 and 4.07) depending on the degree of dependence between the variables. For the copula parameter the MSMLE was comparable in finite sample accuracy to the 1SMLE: in four out of six cases it performed slightly better, and in the remaining cases it performed slightly worse. Thus we see that although asymptotically the MSMLE is known to be less efficient than the 1SMLE for the copula parameter, in small samples the improved estimates of the first margin parameters often outweigh the loss of information incurred through multi-stage estimation.

As one would expect *a priori*, greater dependence and ratios of n_y/n_x closer to unity generally lead to higher MSE ratios: the loss of information from using the MSMLE rather than the 1SMLE is greater for rank correlations of 0.50 and 0.75, while the gains are smaller if there is larger overlap between the two variables' histories. Notice, however, that the MSE ratios of the copula parameters do not change very much with the level of dependence or the degree of overlap. In most cases this ratio is close to one, indicating that in terms of this parameter the two estimators are approximately equally good.

Overall the simulation results suggest that the MSMLE is a good alternative to 1SMLE. In addition to being much more attractive computationally,¹⁰ there were numerous situations where the MSMLE actually outperformed the 1SMLE, as the MSMLE exploits all available information on both variables. In the cases where a loss of efficiency was incurred, this loss was generally moderate.

4. A MODEL OF THE EURO AND YEN EXCHANGE RATES

In this section we apply the methods discussed above to a flexible model of the conditional joint distribution of daily Japanese yen–US dollar and euro–US dollar exchange rates. Flexible models of joint distributions for these exchange rates have numerous potential applications: to improve portfolio decisions, see Ang and Bekaert (2002) or Patton (2004), to improve risk measurement and management, see Rosenberg and Schuermann (2004), or to price derivative securities with multiple underlying assets, see Cherubini and Luciano (2002) and Rosenberg (2003). For a review of copula-based methods in finance more generally, see Cherubini *et al.* (2004).

The data set employed runs from 1 January 1991 to 30 June 2003 for the yen, and from 1 January 1999 to 30 June 2003 for the euro, so $n_x = 3210$ and $n_y = n_c = 1159$. The data are plotted in Figures 2 and 3. It is possible that the fact that the euro was introduced on 1 January 1999, rather than some other date, carries useful information on the conditional distribution of the euro/dollar exchange rate. We will assume, however, that we can ignore the missing-data mechanism.¹¹ As usual, we will analyse the log-difference of the exchange rates.

¹⁰ For example, using Matlab 7 on a Pentium 4 2.4 GHz machine, the estimation of the 1SMLE and its standard errors for the time-varying Student's t copula presented in the next section took 7 minutes 22 seconds, while the estimation of the MSMLE and its standard errors for the same model took 1 minute 16 seconds; almost six times faster.

¹¹ For example, if enough countries now using the euro had failed to meet the requirements laid down for joining, it is conceivable that the emergence of the euro would have been delayed. Thus, the fact that such a delay did not occur may carry information on the economic performance of the countries now using the euro, and possibly also on the conditional distribution of the euro itself.

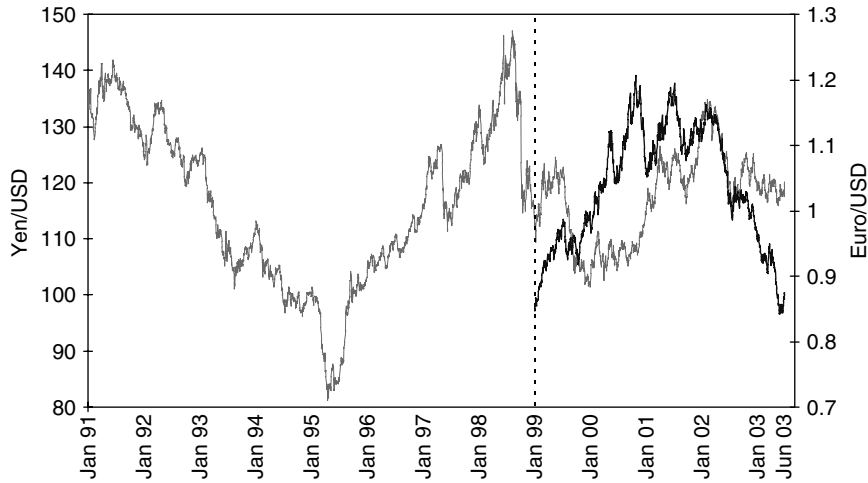


Figure 2. The daily yen/dollar and euro/dollar exchange rates over the sample period, 1 January 1991 to 30 June 2003

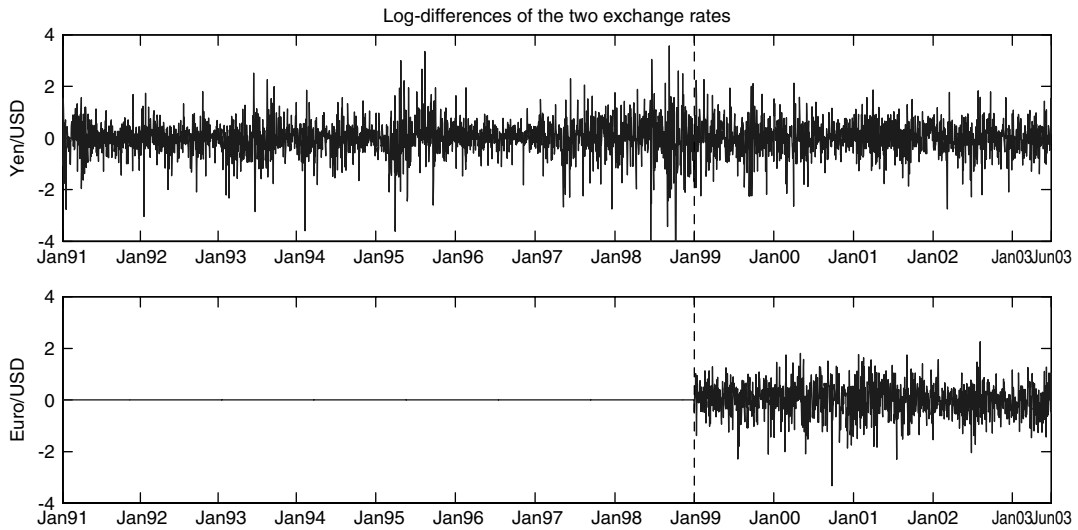


Figure 3. Log-differences (multiplied by 100) of the yen/dollar and euro/dollar exchange rates over the sample period, 1 January 1991 to 30 June 2003

The significance of these two exchange rates in the global foreign exchange market, and the fact that there exist quite different amounts of data on each of these variables, motivated our analysis of the estimator above: market participants can neither wait for more euro data to arrive, nor are they willing to throw away the additional information they have on the yen. The sample rank and linear correlation coefficients between the exchange rate returns for the overlapping period are 0.26 and 0.25. Thus this application most resembles the low dependence ($\rho = 0.25$, our case: $\hat{\rho} = 0.26$), long first sample ($n_x = 3000$, our case: $n_x = 3210$) with small to medium overlapping period

($n_y/n_x = n_c/n_x = 0.25$ or 0.50 , our case: $n_y/n_x = n_c/n_x = 0.36$) in our simulation. If the true DGP for the yen/USD and euro/USD exchange rates resembles our simulated DGP, then Table I suggests that we should expect the MSMLE to be a more efficient estimator of the parameters of the yen margin than the 1SMLE, with the MSE ratio being between 0.21 and 0.55. The MSMLE is expected to be a slightly less efficient estimator of the euro margin parameters, with an MSE ratio of around 1.18, and for the copula parameters we expect the MSMLE and 1SMLE to be approximately equally accurate. For the purposes of comparison we will present results using both the MSMLE and the 1SMLE.¹²

The Student's t distribution has previously been found to provide a good fit to individual exchange rates, see Bollerslev (1987) and Patton (2005) for example, and so we employ it for the marginal distributions of both the yen and the euro exchange rates. For the yen margin an AR(1,10) model¹³ was estimated for the conditional mean, and a GARCH(1,1) model, see Engle (1982) and Bollerslev (1986), was estimated for the variance. We tested for the presence of a structural break in the parameters of the yen/dollar marginal model upon the introduction of the euro, and found some evidence of a structural break in the degrees of freedom parameter, which changed from 3.95 to 4.52. Breaks in the remaining parameters were not significant at the 5% level and so we imposed that they are constant. Estimation of the yen model via 1SMLE uses only post-euro data and so effectively imposes that all parameters of the model changed upon the introduction of the euro. The fact that the remaining parameters did not significantly change indicates that we lose information by only using post-euro data on the yen, and this is reflected in the standard errors on the 1SMLE versus the MSMLE. The euro data exhibited no statistically significant time variation in either the conditional mean or the conditional variance, and so these conditional moments were set to constants.¹⁴ The estimated parameters and standard errors for these marginal distributions are presented in Table II.

It is interesting to note that although the Student's t distribution provides a good fit to both exchanges rates, MSMLE of the degrees of freedom parameters for each margin are different; 4.52 for the yen in the post-euro period and 7.67 for the euro. A test for the significance of the difference in these estimates yields a p -value of 0.06.¹⁵ Tests for the significance of this difference using the 1SMLE are not significant at the 10% level, reflecting the greater estimation error in these estimates.

The evaluation of the goodness-of-fit of the models for the marginal distributions is of critical importance in this application: the joint distribution of the transformed variables, $U_t \equiv F_t(Z_t'; \hat{\varphi}_{n_x})$ and $V_t \equiv G_t(Z_t'; \gamma_{n_y})$, will be modelled with a copula. If the marginal distribution models are misspecified then the variables U_t and V_t will not be uniform and the copula will be misspecified.

¹² The 1SMLE results are those obtained when the marginal distribution models described here are combined with the time-varying Student's t copula described below.

¹³ An AR(1) model was originally specified for the conditional mean, but specification tests indicated that the 10th lag was important for this exchange rate, and so an AR(1,10) was used.

¹⁴ Whilst zero serial correlation in exchange rate returns is not unusual, the lack of volatility clustering is surprising. ARCH LM tests of serial correlation in squared returns up to five lags yielded a p -value of 0.18. More generally, for lags from 1 to 20 this test indicated no volatility clustering. There was, however, significant serial correlation in absolute returns, perhaps because absolute returns are a less noisy proxy for conditional variance. The constant volatility model passes all of our model diagnostic tests and so we continue to use it. As more data on the euro becomes available it is likely that volatility clustering in daily data will become more pronounced and a conditional volatility model will need to be employed.

¹⁵ The construction of this test statistic requires the specification of the copula model. The p -value reported here is that obtained when the time-varying Student's t copula is used. The p -values obtained using the other copula models considered in this paper differed by less than 0.001.

Table II. Results for the marginal distributions

This table presents the estimated parameters and asymptotic standard errors of the marginal distribution models for the yen–US dollar and euro–US dollar exchange rates. An asterisk denotes that the parameter is significantly different from zero at the 5% level, except for the degrees of freedom parameters, for which the null is that the inverse of the parameter is equal to zero. We allow for a structural break in the degrees of freedom parameter in the yen model on 1 January 1999, and denote the pre-euro estimate by ν_x^1 and the post-euro estimate by ν_x^2 .

	MSMLE		ISMLE	
	Coeff.	Std error	Coeff.	Std error
Yen margin				
μ_x	0.018	0.010	0.036	0.019
ϕ_{1x}	-0.011	0.017	-0.000	0.026
ϕ_{10x}	0.056*	0.017	0.027	0.026
ω_x	0.008*	0.003	0.009	0.006
β_x	0.944*	0.012	0.959*	0.021
α_x	0.045*	0.009	0.021*	0.009
ν_x^1	3.952*	0.315	—	—
ν_x^2	4.521*	0.379	6.167*	0.871
Euro margin				
μ_y	0.016	0.018	0.025	0.019
σ_y^2	0.404*	0.022	0.347*	0.020
ν_y	7.672*	1.634	8.308*	1.084

We employ two simple tests proposed by Diebold *et al.* (1998) for evaluating the marginal distribution models. Diebold *et al.* suggested testing that $U_t \sim$ i.i.d. $Unif(0, 1)$ and $V_t \sim$ i.i.d. $Unif(0, 1)$ in two stages: firstly testing that U_t and V_t are i.i.d. via LM tests, and then testing that they are Uniform(0, 1). We test the i.i.d. assumption by regressing $(U_t - \bar{U}_n)^k$ and $(V_t - \bar{V}_n)^k$ on 20 lags of both variables for $k = 1, 2, 3, 4$. We test the $Unif(0, 1)$ hypothesis via the well-known Kolmogorov–Smirnov test. The results of these tests are presented in Table III. As this table shows, both marginal distribution models pass both tests. We thus conclude that the marginal distributions are adequately modelled, and proceed to the modelling of the copula.

There is a vast literature in statistics on the generation of families of copulas, though only a few have been used in econometric models. For the purposes of comparison we will estimate two copulas: the Gaussian, or normal, copula and the Student’s t copula. The normal copula is the copula associated with the bivariate normal distribution, and is the dependence function implicitly assumed whenever the bivariate normal distribution is used. Similarly, the Student’s t copula is the dependence structure associated with the bivariate Student’s t distribution. The functional forms of these two copula densities are given below:

$$c_N(u, v; \rho) = \frac{1}{\sqrt{1 - \rho^2}} \exp \left\{ -\frac{\rho^2 (\Phi^{-1}(u))^2 + \Phi^{-1}(v)^2 - 2\rho\Phi^{-1}(u)\Phi^{-1}(v)}{2(1 - \rho^2)} \right\} \quad (17)$$

Table III. LM tests of serial independence and Kolmogorov–Smirnov tests of the marginal densities

The first four rows of this table present the p -values from LM tests of the independence of the first four moments of the variables U_t and V_t , described in the text, which correspond to the yen and euro distribution models respectively. We regress $(U_t - \bar{U})^k$ and $(V_t - \bar{V})^k$ on 20 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_{40}^2 . p -values of less than 0.05 indicate significant serial dependence in these variables and suggest that the distribution dynamics are misspecified. The final row presents the p -values from a Kolmogorov–Smirnov test on U_t and V_t . p -values of less than 0.05 indicate that the shape of the distribution model is misspecified.

	Yen margin	Euro margin
First moment	0.85	0.51
Second moment	0.87	0.70
Third moment	0.78	0.67
Fourth moment	0.67	0.59
K–S test	0.59	0.39

$$c_T(u, v; \rho, \nu) = \frac{\Gamma((\nu + 2)/2)/\Gamma(\nu/2)}{\nu\pi t(\tilde{x}; \nu)t(\tilde{y}; \nu)\sqrt{1 - \rho^2}} \left(1 + \left(\frac{\tilde{x}^2 + \tilde{y}^2 - 2\rho\tilde{x}\tilde{y}}{\nu(1 - \rho^2)} \right)^{-(\nu+2)/2} \right) \quad (18)$$

for $\rho \in (-1, 1)$ and $\nu > 0$, where Φ^{-1} is the inverse *cdf* of a $N(0, 1)$ random variable, $t(\cdot; \nu)$ is the *pdf* of a Student's t random variable with ν degrees of freedom, $\tilde{x} \equiv T^{-1}(u; \nu)$, $\tilde{y} \equiv T^{-1}(v; \nu)$ where $T^{-1}(\cdot; \nu)$ is the inverse *cdf* of a Student's t random variable with ν degrees of freedom.

The Student's t copula generalizes the normal copula by allowing for non-zero dependence in the extreme tails. This type of dependence is measured by 'upper tail dependence', τ^U , and 'lower tail dependence', τ^L , see Joe (1997):

$$\tau^L \equiv \lim_{\varepsilon \rightarrow 0} \Pr [U \leq \varepsilon | V \leq \varepsilon] = \lim_{\varepsilon \rightarrow 0} \Pr [V \leq \varepsilon | U \leq \varepsilon] = \lim_{\varepsilon \rightarrow 0} \frac{C(\varepsilon, \varepsilon)}{\varepsilon} \quad (19)$$

$$\tau^U \equiv \lim_{\delta \rightarrow 1} \Pr [U > \delta | V > \delta] = \lim_{\delta \rightarrow 1} \Pr [V > \delta | U > \delta] = \lim_{\delta \rightarrow 1} \frac{1 - 2\delta + C(\delta, \delta)}{1 - \delta} \quad (20)$$

Two random variables exhibit lower tail dependence, for instance, if $\tau^L > 0$. This would imply a non-zero probability of observing an extremely large appreciation of the yen against the dollar together with an extremely large appreciation of the euro against the dollar. The normal copula imposes that this probability is zero. The two parameters of the Student's t copula, ρ and ν , jointly determine the amount of dependence between the variables in the extremes. Being a symmetric

copula, the dependence between the variables during extreme appreciations is restricted to be the same as during extreme depreciations,¹⁶ and is given by:

$$\tau_T^U = \tau_T^L = 2T \left(-\sqrt{\nu+1} \sqrt{\frac{1-\rho}{1+\rho}}; \nu+1 \right) \quad (21)$$

While there was no evidence of time variation in the conditional density of the euro/dollar exchange rate, there does appear to be evidence of time variation in the conditional copula between the euro/dollar and the yen/dollar exchange rates. In Figure 4 we plot simple rolling window correlations between $\Phi^{-1}(U_t)$ and $\Phi^{-1}(V_t)$ over the period July 1999 to June 2003, where Φ^{-1} is the inverse standard normal *cdf*. For the first two years the correlation was approximately zero, while for the last two years it averaged close to 0.5.

We elected to model time variation in the conditional copula parameter in a manner similar to that used by a GARCH model to capture time-varying volatility, or by the DCC model to capture time-varying conditional correlations. For both copulas we set the correlation parameter at time t as a function of a constant, the correlation parameter at time $t-1$, and some forcing variable. For correlation parameters the natural forcing variables to use are $\Phi^{-1}(u_{t-1}) \cdot \Phi^{-1}(v_{t-1})$ and $T^{-1}(u_{t-1}; \nu) \cdot T^{-1}(v_{t-1}; \nu)$. We elected to use the average over 10 lags due to the presence of the 10th lag in the yen margin model. Time variation in the parameters of the conditional copulas

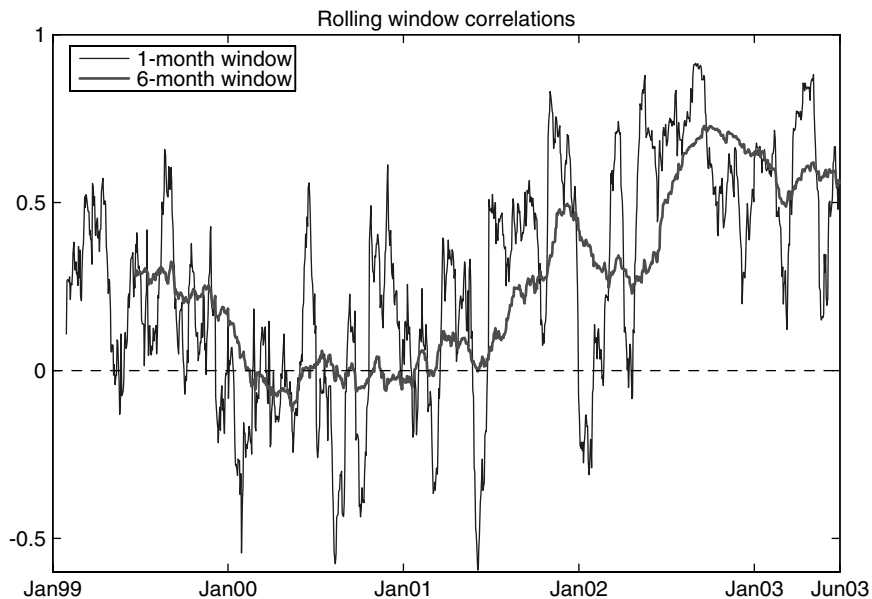


Figure 4. Rolling window correlations through the overlapping sample period, January 1999 to June 2003

¹⁶ In unreported results, we also employed the symmetrized Joe–Clayton (SJC) copula proposed in Patton (2005). The SJC copula has two parameters, one determining τ^L and the other determining τ^U , thus allowing for asymmetric dependence. However, the asymmetry was not significant at conventional levels. Further, a test comparing the constant and time-varying SJC copulas to the constant and time-varying Student's t copulas showed no significant difference in goodness-of-fit. We do not report the results for the SJC copula in the interests of brevity.

are modelled as:

$$\text{Normal: } \rho_t = \tilde{\Lambda} \left(\omega_N + \beta_N \rho_{t-1} + \alpha_N \frac{1}{10} \sum_{j=1}^{10} \Phi^{-1}(u_{t-j}) \Phi^{-1}(v_{t-j}) \right) \quad (22)$$

$$\text{Student's } t: \rho_t = \tilde{\Lambda} \left(\omega_T + \beta_T \rho_{t-1} + \alpha_T \frac{1}{10} \sum_{j=1}^{10} T^{-1}(u_{t-j}; \nu) T^{-1}(v_{t-j}; \nu) \right) \quad (23)$$

where Φ^{-1} is the inverse *cdf* of a standard normal, $T^{-1}(\cdot, \nu)$ is the inverse *cdf* of a Student's *t* random variable with ν degrees of freedom, and $\tilde{\Lambda}(x) = (1 - e^{-x}) / (1 + e^{-x})$ is the modified logistic function. This function is used to ensure that the correlation coefficients remain in $(-1, 1)$ at all times. The above specification is in the spirit of Hansen's (1994) autoregressive conditional density model. The degrees of freedom parameter in the Student's *t* copula was assumed to be constant for simplicity.

From the above specifications for the conditional copula model the important question of stationarity arises, which is required for the consistency and asymptotic normality of the MSMLE. Whilst the conditions required for stationarity of univariate AR and GARCH processes have been widely studied, see Hamilton (1994) for discussion and references, comparable conditions for multivariate nonlinear processes such as those in equations (22) and (23) are not available. Carrasco and Chen (2002) and Meitz and Saikkonen (2004) present conditions for general classes of univariate nonlinear processes, which include as special cases GARCH, stochastic volatility and autoregressive conditional duration (ACD) processes. Similar results for the multivariate case would be very useful, but are beyond the scope of this paper. We are thus left to simply assume that the conditions for stationarity hold.¹⁷

We estimated constant and time-varying versions of these copulas, and report the parameters, standard errors and log-likelihood values in Table IV.¹⁸ A number of insights are possible from these results. Firstly, the best fitting constant copula is the Student's *t* copula with an estimated degrees of freedom parameter of 6.27, implying substantial joint fat tails. The implied tail dependence coefficient for the Student's *t* copula is $\hat{\tau}^U = \hat{\tau}^L = 0.08$, with a 95% confidence interval of [0.02, 0.15]. Thus when one variable takes an extreme value there is about an 8% chance of the other variable taking an extreme value. A simple likelihood ratio (LR) test for the significance of the improvement in the log-likelihood of the Student's *t* copula over the normal copula yields a *p*-value of less than 0.001, indicating that the constant Student's *t* copula yields a significantly better fit than the normal copula.

In Figure 4 we saw evidence of time variation in the conditional copula, and so we now turn to the time-varying conditional copula specifications. Comparing the time-varying normal with the time-varying Student's *t* we again see a substantial improvement in the likelihood, and a LR test again yields a *p*-value of less than 0.001 in favour of the Student's *t* copula. The estimated degrees of freedom parameter for the time-varying Student's *t* copula is larger than for the

¹⁷ A heuristic method of determining whether stationarity is a reasonable assumption is to simulate the conditional copula using the estimated parameters. One can then plot the resulting time series for ρ_t to see whether it diverges to ± 1 , or exhibits other clear violations of stationarity. We did this for the parameter estimates presented below and found no clear violations of stationarity. This is of course no substitute for having theoretical sufficient conditions for stationarity.

¹⁸ We do not present the copula log-likelihood values for the ISMLE as these are not comparable across copula models; only the *joint* log-likelihood values for each model are comparable for this estimator.

Table IV. Copula model results

This table presents the estimated copula parameter and the value of the copula likelihood, \mathcal{L}_C , at the optimum for the copula models considered in this paper. An asterisk denotes that the parameter is significantly different from zero at the 5% level, except for the degrees of freedom parameters, for which the null is that the inverse of the parameter is equal to zero. Copula likelihoods for the 1SMLE are not comparable across copula models and so are not reported.

Model	Parameter	MSMLE			ISMLE	
		Estimate	Std error	\mathcal{L}_C	Estimate	Std error
Constant normal	ρ_N	0.268*	0.031	41.96	0.194*	0.023
Constant Student's t	ρ_T	0.279*	0.030	53.77	0.202*	0.025
	ν	6.273*	1.452		11.026*	2.280
Time-varying normal	ω_N	0.219	0.116	61.49	0.153	0.099
	α_N	0.627*	0.274		0.303*	0.135
	β_N	0.695	0.610		0.736	0.649
Time-varying Student's t	ω_T	0.315*	0.124	71.08	0.207	0.111
	α_T	0.587*	0.194		0.288*	0.138
	β_T	0.300	0.550		0.409	0.707
	ν_T	7.388*	1.901		11.861*	1.731

constant Student's t copula, indicating that some joint extreme events are generated by time-varying correlations rather than by joint fat tails, in the same way that time-varying heteroskedasticity can explain some of the excess kurtosis in the unconditional distribution of individual asset returns. (Recall that kurtosis is inversely related to the degrees of freedom parameter for a Student's t random variable.) A test that the degrees of freedom parameter is the same for both exchange rates and for the time-varying Student's t copula in the post-euro period yields a p -value of 0.095. So while components of different multivariate Student's t densities fit these exchange rate data well, a single multivariate Student's t density appears too restrictive a specification.

Testing for the significance of time variation in the conditional copula is complicated by the presence of a nuisance parameter that is unidentified under the null hypothesis of no time variation.¹⁹ In our case we may test this null by first noting that the autoregressive parameter in the Student's t copula model, for example, denoted β_T , is not significantly different from zero. Thus we could re-estimate this model imposing that $\beta_T = 0$, making this model for the conditional correlation parameter resemble an ARCH model. Testing for the significance of time variation in such a model is done via a simple test that α_T equals zero. Following such a procedure yields a p -value of less than 0.001, indicating that time variation in the conditional copula is indeed significant.

In Figure 5 we plot the time path of the conditional correlation implied by the time-varying Student's t copula model.²⁰ For comparison we also plot the conditional correlation obtained from the dynamic conditional correlation (DCC) model of Engle (2002). This model for the conditional

¹⁹ Methods for overcoming this problem have been proposed for simpler time series models, see Andrews and Ploberger (1994) for example. Notice that we do not need to constrain $\alpha_T \geq 0$, as is needed in GARCH models, and so $\alpha_T = 0$ does not represent a boundary of the parameter space, meaning we avoid this potential problem.

²⁰ The conditional correlation between the exchange rates is not the same as the correlation parameter in the Student's t copula model, as the marginal distributions and the copula have different degrees of freedom parameters. We obtain the conditional correlation implied by the model via simulation.

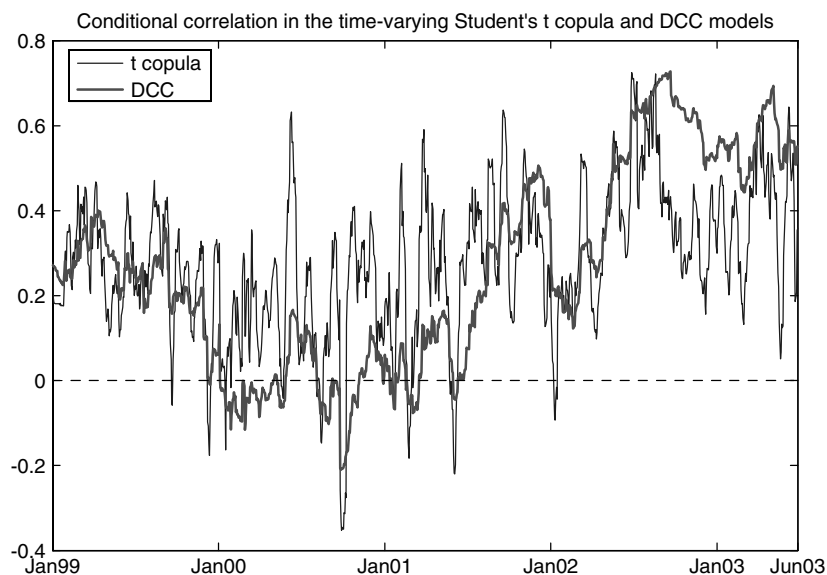


Figure 5. Conditional correlation estimates from the time-varying Student's t copula model and the DCC model over the period January 1999 to June 2003

covariance matrix is also specified in stages; with models for the conditional variance estimated first, and the model for the conditional correlation estimated in a final stage, thus satisfying the restriction that the parameter vector may be partitioned into parts relating to the marginal distribution and the copula. We used the same mean and variance models as presented above, and employed a simple DCC(1,1) specification. Figure 5 shows that the two models yield similar estimates of the conditional correlation over the sample period: both yield correlation estimates around or below zero during 2000, for example, and both indicate that correlation rose substantially since then.

While the estimates of conditional correlation are qualitatively similar from the Student's t copula and the DCC models, the estimates of conditional tail dependence are very different. Taken as a model for the entire density, with margins and copula assumed to be normal, the DCC model implies zero tail dependence for every day, as does any model using a normal copula.²¹ In Figure 6 we plot the estimated conditional tail dependence from the constant and time-varying Student's t copula models. As in the conditional correlations, we see substantial time variation in tail dependence, ranging from 0.003 in September 2000, to 0.28 in July 2002. Figure 6 thus indicates that the probability of these joint extreme movements can range from near zero to over one-quarter.

To test the goodness-of-fit of these copulas we follow the method proposed in Diebold *et al.* (1999). This method involves using the conditional *cdf* of U_t given V_t , and V_t given U_t . Let $W_{1t} \equiv C_{1|2}(U_t|V_t)$ and $W_{2t} \equiv C_{2|1}(V_t|U_t)$. If the copula is correctly specified then the time series

²¹ The DCC model was originally proposed as a model for the conditional covariance matrix, rather than for the entire conditional distribution. The most obvious copula to use in combination with a DCC model is the normal copula, but this is by no means the only possibility. Indeed, it is possible to use the DCC specification of the conditional correlation matrix in any other copula model that has a correlation-type matrix as a parameter; the Student's t copula for example.

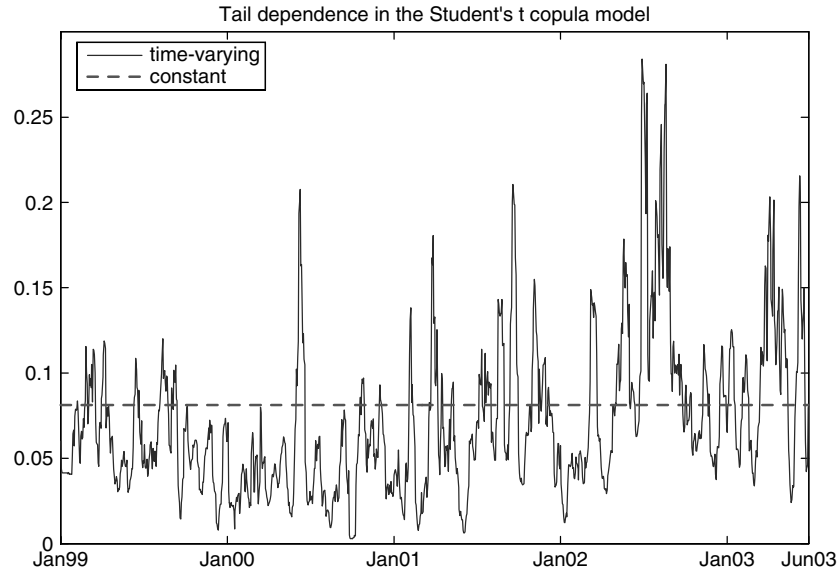


Figure 6. Conditional tail dependence estimates from the Student's t copula models over the period January 1999 to June 2003

Table V. LM tests of serial independence and Kolmogorov–Smirnov tests of the copula

This table presents the p -values from LM tests of serial independence of the first four moments of the variables (U_t, W_{2t}) and (V_t, W_{1t}) , described in the text, for the four copula models considered. We use 20 lags in all tests. 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.

	Constant copulas		Time-varying copulas	
	Normal	Student's t	Normal	Student's t
First moment	0.36	0.29	0.31	0.25
	0.16	0.12	0.16	0.12
Second moment	0.51	0.38	0.45	0.33
	0.17	0.11	0.17	0.12
Third moment	0.53	0.39	0.50	0.38
	0.14	0.09	0.14	0.10
Fourth moment	0.48	0.36	0.48	0.39
	0.11	0.08	0.09	0.08
K–S test	0.97	0.86	0.37	0.50
	0.93	0.94	0.33	0.31

$(\dots, U_t, W_{2t}, \dots) \sim \text{iid Unif}(0, 1)$ and $(\dots, V_t, W_{1t}, \dots) \sim \text{iid Unif}(0, 1)$. We test this null by again testing for serial correlation in the first four moments of these series, and by the K–S test for uniformity. The results of these tests are presented in Table V. The specification tests

provide no evidence against any of the four copula specifications. This is perhaps reflective of the difficulty in detecting deviations from the correct specification of a bivariate density with only 1159 observations.

5. CONCLUSION

This paper compared two maximum likelihood estimators (MLEs) of the parameters of a multivariate model for time series with histories of different lengths. Numerous situations exist where differing amounts of data are available on the variables of interest: models of developed and emerging markets, models of recently floated stocks and the market portfolio, and models involving the euro.

The benchmark estimator in this paper is the standard one-stage MLE estimated using only the overlapping data. The alternative estimator is a simple modification of the two-stage MLE of Newey and McFadden (1994) and White (1994). In addition to greatly simplifying the estimation of the model by breaking the estimation problem into smaller problems, the multi-stage estimator is designed to use all available data on each series. This estimator may be interpreted as an extension of Anderson (1957) and Stambaugh (1997) to more irregular data sets, and to non-normal, serially dependent random variables. We compared the theoretical asymptotic efficiency of the estimators, and provided a sufficient condition for the multi-stage estimator to be *not less* efficient than the one-stage estimator. Simply put, the information gained by using non-overlapping data must outweigh the information foregone by estimating the parameter vector in stages rather than simultaneously. We compared the small sample efficiency of the estimators via a Monte Carlo study designed to replicate the key features of daily financial time series, and found that the multi-stage estimator was often more efficient than the one-stage estimator, and in most cases not less efficient.

We applied our estimator to a model of the joint distribution of daily Japanese yen–US dollar and euro–US dollar exchange rates, over the period 1 January 1999 to 30 June 2003. These rates are the two most frequently traded exchange rates, and much more data is available on the yen than on the euro. We found that the Student's t distribution fits both exchange rates well, though the degrees of freedom parameter is different for the different exchange rates. We estimated two different copula models, with and without time variation in the parameters. Time variation in the conditional copula was found to be statistically significant, indicating that the assumption of constant correlations or a constant copula could lead to inferior portfolio or risk management decisions. Further, allowing for joint extreme events via the Student's t copula significantly improved the fit. The probability of joint extreme events varied over the sample period, from near zero to over one-quarter.

APPENDIX A: ASSUMPTIONS

Presented below are the assumptions required for Theorem 1. They are collected here for convenience and ease of reference. Most of these assumptions are based on those presented in White (1994). In addition to the assumptions below we make the usual assumptions that observed data are a realization of a stochastic process on a complete probability space and that all functions are measurable.

Assumption 1 (Conditions on the log-likelihoods)

(a)

- (i) For each $\varphi \in \Phi$, $E[\log f(Z^t, \varphi)]$ exists and is finite, $t = 1, 2, \dots$
- (ii) For each $\gamma \in \Gamma$, $E[\log g(Z^t, \gamma)]$ exists and is finite, $t = 1, 2, \dots$
- (iii) For each $\theta \in \Theta$, $E[\log c(Z^t, \theta)]$ exists and is finite, $t = 1, 2, \dots$

(b)

- (i) $E[\log f(Z^t, \cdot)]$ is continuous on Φ , $t = 1, 2, \dots$
- (ii) $E[\log g(Z^t, \cdot)]$ is continuous on Γ , $t = 1, 2, \dots$
- (iii) $E[\log c(Z^t, \cdot)]$ is continuous on Θ , $t = 1, 2, \dots$

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

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

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

Assumption 4 (Conditions on the scores)

(a)

- (i) For all $\varphi \in \Phi$, $E \left[n_x^{-1} \sum_{t=1}^{n_x} s_1(Z^t; \varphi) \right] < \infty$
- (ii) For all $\gamma \in \Gamma$, $E \left[n_y^{-1} \sum_{t=1}^{n_y} s_2(Z^t; \gamma) \right] < \infty$
- (iii) For all $\theta \in \Theta$, $E \left[n_c^{-1} \sum_{t=1}^{n_c} s_3(Z^t; \theta) \right] < \infty$

(b)

- (i) $E \left[n_x^{-1} \sum_{t=1}^{n_x} s_1(Z^t; \varphi) \right]$ is continuous on Φ uniformly in $n_x = 1, 2, \dots$
- (ii) $E \left[n_y^{-1} \sum_{t=1}^{n_y} s_2(Z^t; \gamma) \right]$ is continuous on Γ uniformly in $n_y = 1, 2, \dots$
- (iii) $E \left[n_c^{-1} \sum_{t=1}^{n_c} s_3(Z^t; \theta) \right]$ is continuous on Θ uniformly in $n_c = 1, 2, \dots$

(c)

- (i) $\{s_1(Z^t; \varphi)\}$ obeys the weak uniform law of large numbers
- (ii) $\{s_2(Z^t; \gamma)\}$ obeys the weak uniform law of large numbers
- (iii) $\{s_3(Z^t; \theta)\}$ obeys the weak uniform law of large numbers

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

Assumption 5 (Conditions on the Hessians)

(a)

- (i) For all $\varphi \in \Phi$, $E \left[n_x^{-1} \sum_{t=1}^{n_x} \nabla_{\varphi\varphi} \log f(Z^t; \varphi) \right] < \infty$, $n_x = 1, 2, \dots$
- (ii) For all $\gamma \in \Gamma$, $E \left[n_y^{-1} \sum_{t=1}^{n_y} \nabla_{\gamma\gamma} \log g(Z^t; \gamma) \right] < \infty$, $n_y = 1, 2, \dots$
- (iii) For all $\theta \in \Theta$, $E \left[n_c^{-1} \sum_{t=1}^{n_c} \nabla_{\kappa\kappa} \log c(Z^t; \theta) \right]$, $E \left[n_c^{-1} \sum_{t=1}^{n_c} \nabla_{\varphi\kappa} \log c(Z^t; \theta) \right]$ and $E \left[n_c^{-1} \sum_{t=1}^{n_c} \nabla_{\gamma\kappa} \log c(Z^t; \theta) \right]$ are $< \infty$, $n_c = 1, 2, \dots$

(b)

- (i) $E \left[n_x^{-1} \sum_{t=1}^{n_x} \nabla_{\varphi\varphi} \log f(Z^t; \varphi) \right]$ is continuous on Φ uniformly in $n_x = 1, 2, \dots$
- (ii) $E \left[n_y^{-1} \sum_{t=1}^{n_y} \nabla_{\gamma\gamma} \log g(Z^t; \gamma) \right]$ is continuous on Γ uniformly in $n_y = 1, 2, \dots$
- (iii) $E \left[n_c^{-1} \sum_{t=1}^{n_c} \nabla_{\kappa\kappa} \log c(Z^t; \theta) \right]$, $E \left[n_c^{-1} \sum_{t=1}^{n_c} \nabla_{\varphi\kappa} \log c(Z^t; \theta) \right]$ and $E \left[n_c^{-1} \sum_{t=1}^{n_c} \nabla_{\gamma\kappa} \log c(Z^t; \theta) \right]$ are continuous on Θ uniformly in $n_c = 1, 2, \dots$

(c)

- (i) $\{\nabla_{\varphi\varphi} \log f(Z^t; \varphi)\}$ obeys the weak uniform law of large numbers
- (ii) $\{\nabla_{\gamma\gamma} \log g(Z^t; \gamma)\}$ obeys the weak uniform law of large numbers
- (iii) $\{\nabla_{\kappa\kappa} \log c(Z^t; \theta)\}$, $\{\nabla_{\varphi\kappa} \log c(Z^t; \theta)\}$ and $\{\nabla_{\gamma\kappa} \log c(Z^t; \theta)\}$ obey the weak uniform law of large numbers.

Assumption 6 $\{A_n^0\}$, as defined in Theorem 1, is $\mathcal{O}(1)$ and negative definite uniformly in n_x .

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

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

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

$$B_n^0 \equiv \begin{bmatrix} n_x^{-1} \sum_{t=1}^{n_x} E[s_{1t}^0 \cdot s_{1t}^{0'}] & (n_x n_y)^{-1/2} \sum_{t=1}^{n_y} E[s_{1t}^0 \cdot s_{2t}^{0'}] & (n_x n_c)^{-1/2} \sum_{t=1}^{n_c} E[s_{1t}^0 \cdot s_{3t}^{0'}] \\ (n_x n_y)^{-1/2} \sum_{t=1}^{n_y} E[s_{2t}^0 \cdot s_{1t}^{0'}] & n_y^{-1} \sum_{t=1}^{n_y} E[s_{2t}^0 \cdot s_{2t}^{0'}] & (n_y n_c)^{-1/2} \sum_{t=1}^{n_c} E[s_{2t}^0 \cdot s_{3t}^{0'}] \\ (n_x n_c)^{-1/2} \sum_{t=1}^{n_c} E[s_{3t}^0 \cdot s_{1t}^{0'}] & (n_y n_c)^{-1/2} \sum_{t=1}^{n_c} E[s_{3t}^0 \cdot s_{2t}^{0'}] & n_c^{-1} \sum_{t=1}^{n_c} E[s_{3t}^0 \cdot s_{3t}^{0'}] \end{bmatrix} \quad (24)$$

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

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

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

Let $B_n(\theta)$ be the matrix B_n evaluated at the point θ , and so B_n^0 defined above equals $B_n(\theta_0)$.

Assumption 9

- (a) The elements of $B_n(\theta)$ are finite and continuous on Θ uniformly in $n_x = 1, 2, \dots$
- (b) The elements of $\{[s_{1t}^{0'}, s_{2t}^{0'}, s_{3t}^{0'}]' \cdot [s_{1t}^{0'}, s_{2t}^{0'}, s_{3t}^{0'}]\}$ obey the weak uniform law of large numbers.

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

APPENDIX B: PROOFS

Proof of Theorem 1. Here we provide a sketch of the modifications that need to be made to the standard MSMLE proof of asymptotic normality (see theorem 6.11 of White, 1994, for example) to accommodate the differing sample sizes.

Let

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

and $A(\theta) = E[Hess(\theta)]$, so $A(\theta_0) = A_n^0$.

As usual, the proof starts by taking a mean value expansion of the scores evaluated at the estimated parameters about the scores evaluated at the true parameters, which equal zero due to the assumption that the true parameters lie in $int(\Theta)$.

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

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

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

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

by theorem 8.10 of Lehmann and Casella (1998, p. 58) and assuming that the scores satisfy the conditions of a central limit theorem. ■

Proof of Proposition 1. The 1SMLE is more efficient than the MSMLE if $(\text{avar}(\hat{\theta}_n) - \text{avar}(\hat{\theta}_{n_c}^{eff}))$ is a positive semi-definite matrix. Consider the case that $\lambda_c < M_{11}/D_{11}$. Notice that we may write the first $(p \times p)$ elements of the matrix $A_n^{0^{-1}} \cdot \mathcal{N}_{\infty}^{*-1/2} \cdot \mathbf{B}_n^0 \cdot \mathcal{N}_{\infty}^{*-1/2} \cdot A_n^{0^{-1}}$ as λ_c times the first $(p \times p)$ elements of the matrix D_n^0 . Let $\boldsymbol{\beta} = [\beta, \mathbf{0}]$, where $\mathbf{0}$ is a column vector of $p + q + r - 1$ zeros and $\beta \in \mathbb{R} \setminus \{0\}$. Then the quadratic form

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

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

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

Thus the difference between the asymptotic covariance matrices is indefinite. ■

ACKNOWLEDGEMENTS

This article is based on chapter II of my PhD dissertation (Patton, 2002) and was previously circulated under the title ‘Estimation of copula models for time series of possibly different lengths’. I would like to thank the Editor (Tim Bollerslev), three anonymous referees and Graham Elliott, Rob Engle, Raffaella Giacomini, Giuseppe Ragusa, Kevin Sheppard and Allan Timmermann for useful comments and suggestions. The simulations presented in this paper were made possible by Vince Crawford and the UCSD Economics Experimental and Computational Laboratory. Financial support from the UCSD Project in Econometric Analysis fellowship is gratefully acknowledged.

REFERENCES

- Anderson TW. 1957. Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association* **52**(278): 200–203.
- Andrews DWK. 1988. Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory* **4**: 458–467.
- Andrews DWK, Ploberger W. 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62**(6): 1383–1414.
- Ang A, Bekaert G. 2002. International asset allocation with regime shifts. *Review of Financial Studies* **15**(4): 1137–1187.
- Ang A, Chen J. 2002. Asymmetric correlations of equity portfolios. *Journal of Financial Economics* **63**(3): 443–494.
- Bae K-H, Karolyi A, Stulz RM. 2003. A new approach to measuring financial contagion. *Review of Financial Studies* **16**(3): 717–764.
- Bank for International Settlements. 2002. *Triennial Central Bank Survey: Foreign Exchange and Derivatives Market Activity in 2001*, Basle.
- Bauwens L, Laurent S, Rombouts J. 2003. Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, forthcoming.
- Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*. **31**: 307–327.
- Bollerslev T. 1987. A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* **69**: 542–547.
- Bollerslev T. 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach. *Review of Economics and Statistics* **72**: 498–505.
- Brown SJ, Goetzmann WN, Ross SA. 1995. Survival. *Journal of Finance* **50**(3): 853–873.
- Capéraà P, Fougères A-L, Genest C. 1997. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84**(3): 567–577.
- Carrasco M, Chen X. 2002. Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* **18**: 17–39.
- Casella G, Berger RL. 1990. *Statistical Inference*. Duxbury Press: USA.
- Chen X, Fan Y. 2002. Semiparametric estimation of copula-based time series models. Working paper, Department of Economics, New York University.
- Chen X, Fan Y, Patton AJ. 2004. Simple tests for models of dependence between multiple financial time series, with applications to U.S. equity returns and exchange rates. Financial Markets Group, London School of Economics, Discussion Paper 483.
- Cherubini U, Luciano E. 2002. Bivariate option pricing with copulas. *Applied Mathematical Finance* **9**(2): 69–85.
- Cherubini U, Luciano E, Vecchiato W. 2004. *Copula Methods in Finance*. John Wiley & Sons: Chichester, UK.
- Clements MP. 2002. An evaluation of the survey of professional forecasters probability distributions of expected inflation and output growth. Manuscript, Department of Economics, University of Warwick.
- Diebold FX, Gunther T, Tay AS. 1998. Evaluating density forecasts with applications to financial risk management. *International Economic Review* **39**: 863–883.
- Diebold FX, Hahn J, Tay AS. 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.
- Duffie D, Pan J. 1997. An overview of value at risk. *Journal of Derivatives* **4**: 7–49.
- Embrechts P, Höing A, Juri A. 2003. Using copulae to bound the value-at-risk for functions of dependent risks. *Finance & Stochastics* **7**(2): 145–167.
- Engle RF. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**: 987–1008.
- Engle RF. 2002. Dynamic conditional correlation—a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* **20**(3): 339–350.
- Engle RF, Kroner K. 1995. Multivariate simultaneous GARCH. *Econometric Theory* **11**: 122–150.
- Erb CB, Harvey CR, Viskanta TE. 1994. Forecasting international equity correlations. *Financial Analysts Journal* **50**: 32–45.

- Fama EF. 1965. The behavior of stock-market prices. *Journal of Business* **38**: 34–105.
- Farebrother RW. 1992. A class of bivariate density functions with common marginals. *Econometric Theory* **8**(1): 145–146.
- Fermanian J-D, Scaillet O. 2003. Nonparametric estimation of copulas for time series. *Journal of Risk* **5**: 25–54.
- Gallant AR, White H. 1988. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell: Oxford.
- Genest C, Rivest L-P. 1993. Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association* **88**(423): 1034–1043.
- Genest C, Ghoudi K, Rivest L-P. 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**(3): 543–552.
- Glidden DV. 2000. A multi-stage estimator of the dependence parameter for the Clayton–Oakes model. *Lifetime Data Analysis* **6**: 141–156.
- Goetzmann WN, Jorion P. 1999. Re-emerging markets. *Journal of Financial and Quantitative Analysis* **34**(1): 1–32.
- Hamilton JD. 1994. *Time Series Analysis*. Princeton University Press: New Jersey.
- Hansen BE. 1994. Autoregressive conditional density estimation. *International Economic Review* **35**(3): 705–730.
- Harvey A, Koopman SJ, Penzer J. 1998. Messy time series. *Advances in Econometrics* **13**: 103–143.
- Hong Y, Li H, Zhao F. 2004. Out-of-sample performance of discrete-time spot interest rate models. *Journal of Business and Economic Statistics* **22**(4): 457–473.
- Joe H. 1997. *Multivariate Models and Dependence Concepts. Monographs in Statistics and Probability* 73. Chapman and Hall: London.
- Joe H, Xu JJ. 1996. The estimation method of inference functions for margins for multivariate models. Working paper, Department of Statistics, University of British Columbia.
- Jorion P. 1997. *Value-at-Risk*. Academic Press: San Diego.
- Kofman P, Sharpe I. 2000. Imputation methods for incomplete dependent variables in finance. Working paper, School of Finance and Economics, University of Technology, Sydney.
- Le Cam L. 1956. On the asymptotic theory of estimation and testing hypotheses. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1, 129–156.
- Lehmann EL, Casella G. 1998. *Theory of Point Estimation*, 2nd edn. Springer-Verlag: New York.
- Li DX. 2000. On default correlation: a copula function approach. *Journal of Fixed Income* **9**(4): 43–54.
- Linton O. 2005. Nonparametric inference for unbalanced time series data. *Econometric Theory* **21**(1): 143–157.
- Little RJA, Rubin DB. 1987. *Statistical Analysis with Missing Data*. John Wiley & Sons: New York.
- Longin F, Solnik B. 2001. Extreme correlation of international equity markets. *Journal of Finance* **56**(2): 649–676.
- Lynch AW, Wachter JA. 2004. Using samples of unequal length in generalized method of moments estimation. Working paper, NYU and University of Pennsylvania.
- Mandelbrot B. 1963. The variation of certain speculative prices. *Journal of Business* **26**: 394–419.
- Meitz M, Saikkonen P. 2004. Ergodicity, mixing, and the existence of moments of a class of Markov models with applications to GARCH and ACD models. SEE/EFI Working Paper No. 573, Stockholm School of Economics.
- Mills FC. 1927. *The Behaviour of Prices*. National Bureau of Economic Research: New York.
- Nelsen RB. 1999. *An Introduction to Copulas*. Springer-Verlag: New York.
- Newey WK, McFadden D. 1994. Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, Vol. 4, Engle RF, McFadden D (eds). North-Holland: Amsterdam.
- Oakes D. 1982. A model for association in bivariate survival data. *Journal of the Royal Statistical Association, Series B* **44**(3): 414–422.
- Pagan AR. 1986. Two stage and related estimators and their applications. *Review of Economic Studies* **8**: 517–538.
- Patton AJ. 2002. Applications of copula theory in financial econometrics. Unpublished PhD dissertation, Department of Economics, University of California, San Diego.
- Patton AJ. 2004. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics* **2**(1): 130–168.

- Patton AJ. 2005. Modelling time-varying exchange rate dependence. *International Economic Review*, forthcoming.
- Richardson M, Smith T. 1993. A test for multivariate normality in stock returns. *Journal of Business* **66**(2): 295–321.
- Rockinger M, Jondeau E. 2001. Conditional dependency of financial series: an application of copulas. Working paper, Department of Finance, HEC School of Management, France.
- Rosenberg JV. 2003. Nonparametric pricing of multivariate contingent claims. *Journal of Derivatives* **10**(3): 9–26.
- Rosenberg JV, Schuermann T. 2004. A general approach to integrated risk management with skewed, fat-tailed risk. *Journal of Financial Economics*.
- Schmidt P. 1977. Estimation of seemingly unrelated regressions with unequal numbers of observations. *Journal of Econometrics* **5**: 365–377.
- Shih JH, Louis TA. 1995. Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* **51**: 1384–1399.
- Sklar A. 1959. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris* **8**: 229–231.
- Stambaugh RF. 1997. Analyzing investments whose histories differ in length. *Journal of Financial Economics* **45**: 285–331.
- Swamy PAVB, Mehta JS. 1975. On Bayesian estimation of seemingly unrelated regressions when some observations are missing. *Journal of Econometrics* **3**: 157–169.
- Tay AS, Wallis KF. 2000. Density forecasting: a survey. *Journal of Forecasting* **19**: 235–254.
- Tobin J. 1958. Estimation of relationships for limited dependent variables. *Econometrica* **26**: 34–36.
- Wallis KF. 2003. Chi-squared tests of interval and density forecasts, and the Bank of England's fan charts. *International Journal of Forecasting* **19**: 165–175.
- Weeks M. 1999. Methods of imputation for missing data. Working paper, Department of Applied Economics, University of Cambridge.
- White H. 1994. *Estimation, Inference and Specification Analysis*. *Econometric Society Monographs* No. 22. Cambridge University Press: Cambridge.
- White H. 2001. *Asymptotic Theory for Econometricians*, Revised Edition. Academic Press: USA.