Modelling Longitudinal Data using a Pair-Copula Decomposition of Serial Dependence

Michael Smith[1,⋆], Aleksey Min[2],

Claudia Czado[2] and Carlos Almeida[2]

1. Melbourne Business School, University of Melbourne

2. Zentrum Mathematik, Technische Universität München

September 18, 2009

⋆ Corresponding author; address for correspondence: Professor Michael Smith, Melbourne Business School, 200 Leicester Street, Carlton, VIC, 3053, Australia. Email: mike.smith@mbs.edu.

Modelling Longitudinal Data using a Pair-Copula Decomposition of Serial Dependence

Abstract

Copulas have proven to be very successful tools for the flexible modelling of cross-sectional dependence. In this paper we express the dependence structure of continuous time series data using a sequence of bivariate copulas. This corresponds to a type of decomposition recently called a 'vine' in the graphical models literature, where each copula is entitled a 'pair-copula'. We propose a Bayesian approach for the estimation of this dependence structure for longitudinal data. Bayesian selection ideas are used to identify any independence pair-copulas, with the end result being a parsimonious representation of a time-inhomogeneous Markov process of varying order. Estimates are Bayesian model averages over the distribution of the lag structure of the Markov process. Overall, the pair-copula construction is very general and the Bayesian approach generalises many previous methods for the analysis of longitudinal data. Both the reliability of the proposed Bayesian methodology, and the advantages of the pair-copula formulation, are demonstrated via simulation and two examples. The first is an agricultural science example, while the second is an econometric model for the forecasting of intraday electricity load. For both examples the Bayesian pair-copula model is substantially more flexible than longitudinal models employed previously.

Keywords: Longitudinal Copulas; Covariance Selection; Inhomogeneous Markov Process; D-vine; Bayesian Model Selection; Antedependent Model; Intraday Electricity Load

1 Introduction

Modelling multivariate distributions using copulas has proven to be increasingly popular. This is largely due to the flexibility that copula models provide, whereby the marginal distributions can be modelled arbitrarily, and any dependence captured by the copula. Joe (1997) and Nelsen (2006) provide introductions to copula models and their properties. While there are a large range of copulas from which to choose (Frees and Valdez, 1998), only a few are readily applicable to high dimensional problems. Copula built from elliptical distributions, such as the Gaussian (Song, 2000) or t copulas (Demarta and McNeil, 2005), are most popular in this case. However, these can prove restrictive and in the recent graphical models literature alternative multivariate copulas have been proposed that are constructed from series of bivariate copulas. There are a large number of permutations in which this can be undertaken, and Bedford and Cooke (2002) were the first to organize the different decompositions in a systematic way. They label the resulting multivariate copulas 'vines', while Aas, Czado, Frigessi and Bakken (2009) label the component bivariate copulas 'pair-copulas'; see Kurowicka and Cooke (2006) for a recent summary.

To date copula models have been employed largely to account for cross-sectional dependence. Applications to serial dependence in time series and longitudinal data are rare, although the potential is vast. For such data the marginal distribution of the process at each point in time can be modelled arbitrarily, while dependence over time is captured by a multivariate copula. This approach was suggested by Meester and MacKay (1994) in a low dimensional setting, while more recent examples include Lambert and Vandenhende (2002), Frees and Wang (2005) and Sun, Frees and Rosenberg (2008). However, all these authors employ multivariate copulas that do not fully exploit the time ordering of the margins. In this paper we aim to show that doing so results in a substantially more flexible representation that is both more insightful and allows for more efficient inference for continuous data.

We decompose the distribution of a continuous process at a point in time, conditional upon the past, into the product of a sequence of bivariate copula densities and the marginal density. We show that the resulting decomposition of the joint distribution is a D-vine (Bedford and Cooke, 2002), where each bivariate copula is recognisable as a pair-copula. Any mix of bivariate copula can be used to model the pair-copula densities, resulting in an extremely flexible modelling framework. When the underlying process exhibits Markovian properties, this can be accounted for by setting pair-copulas to the independence copula. For longitudinal data this results in a time-inhomogeneous Markov process with order that also varies over time. As we demonstrate here, not only does this produce greater insight into the underlying process, in high-dimensional longitudinal applications it can also lead to a substantial improvement in the quality of inference.

Bedford and Cooke (2002) give the theoretical construction of regular vines, however no estimation of pair copula parameters is attempted. Kurowicka and Cooke (2006) estimate Gaussian vine copula parameters by minimising the determinant of the correlation matrix. Aas et al. (2009) estimate pair-copula using maximum likelihood for Gaussian and non-Gaussian pair-copulas in both C-vine and D-vines. Min and Czado (2008) suggest a Bayesian method for the estimation of D-vines using Markov chain Monte Carlo (MCMC). In all cases cross-sectional dependence is examined, where the determination of an appropriate ordering of the dimensions for the decomposition remains an open problem. However, for time-ordered data this issue does not arise, and one of the insights of this paper is that a pair-copula decomposition is arguably more appropriate.

We suggest a Bayesian approach for the estimation of a pair-copula decomposition for longitudinal data. Indicator variables are introduced to identify which pair-copulas are independence copula. By doing so, we extend the existing Gaussian covariance selection methods to a flexible non-Gaussian framework, both in the longitudinal case (Pourahmadi, 1999; Smith and Kohn, 2002; Huang, Liu, Pourahmadi and Liu, 2006; Liu, Daniels and Marcus 2009) and more generally (Dempster, 1972; Yuan and Lin 2007). We suggest a Metropolis-Hastings scheme to generate both the indicator variable and dependence parameter(s) of a pair-copula jointly, where the proposal is based on a latent variable representation of the pair-copula parameter(s). Forming inference using Bayesian computational methods allows for a stochastic search over the space of the indicators. The full spectrum of posterior inference is available, including measures of dependency such as Kendall's tau. We also propose a diagnostic related to the distribution of the sum of the process over the time points. All estimates are model averages over the distribution of the order of the Markov process.

A simulation study using Gaussian, Gumbel and Clayton pair-copulas highlights the accuracy and reliability of the Bayesian procedure for both the selection and estimation of pair-copulas. The results show that selection can improve the estimated dependence structure, and that pair-copulas can provide substantial improvements over the common approach of using a multivariate Gaussian copula. We demonstrate the usefulness of the approach using two real data examples. The first is a dataset on the liveweight of cows which has been examined previously in the longitudinal literature, but where we consider t and Gaussian pair-copula models. A high level of parsimony is identified, accounting for which substantially perturbs the estimated dependence structure. The second example is a longitudinal model for intraday electricity load in the Australian state of New South Wales. Here, marginal Gaussian regressions are employed to account for day type, seasonal, trend and temperature effects, as well as intraday heteroscedasticity. However, intraday dependence is captured flexibly using Gaussian, Gumbel and Clayton pair-copulas. A time-varying Markovian structure with varying order is identified, confirming the structure of longitudinal models currently used in the energy forecasting literature (Cottet and Smith, 2003; Soares and Medeiros, 2008). We show that the choice of pair-copula type has a substantial impact on the accuracy of intraday forecasts.

The rest of the paper is organised as follows. In Section 2 we outline the pair-copula decomposition of the joint distribution of time-ordered data. As an illustration we show how a familiar Gaussian AR(2) process can be written in this form. We then outline the likelihood for longitudinal data with time dependence modelled using a pair-copula decomposition with selection. Section 3 discusses the priors employed and the computation of posterior inference. Also discussed are measures of dependence and a diagnostic for the quality of fit. Section 4 contains the simulation study, while Section 5 contains the real data examples. Section 6 concludes the paper by placing it in the context of existing models and methods.

2 The Model

2.1 Pair-copula Construction for Time Series

Consider a univariate time series $\boldsymbol{X} = \{X_1, \ldots, X_T\}$ of continuously distributed data observed at T possibly unequally-spaced points in time. If the underlying process is Markovian, then this can be exploited by selecting models for the conditionals in the decomposition of the joint distribution of $\boldsymbol{X}$:

$$f(\boldsymbol{x}) = \prod_{t=2}^{T} f(x_t|x_{t-1}, \ldots, x_1) f(x_1)\,. \tag{2.1}$$

For example, the assumption that $X_t|X_{t-1}, \ldots, X_1 \sim N(\sum_{j=1}^{p} \rho_j X_{t-j}, \sigma^2)$ for $t > p$ corresponds to a simple Gaussian autoregression. However, copula modelling ideas can be used to express each conditional more generally. For $s < t$ there always exists a bivariate density $c_{t,s}$ on $[0,1]^2$, such that

$$\begin{aligned} f(x_t, x_s|x_{t-1}, \ldots, x_{s+1}) &= c_{t,s}(F(x_t|x_{t-1}, \ldots, x_{s+1}), F(x_s|x_{t-1}, \ldots, x_{s+1})) \\ &\times f(x_t|x_{t-1}, \ldots, x_{s+1}) f(x_s|x_{t-1}, \ldots, x_{s+1})\,. \end{aligned} \tag{2.2}$$

Here, $F(x_t|x_{t-1}, \ldots, x_{s+1})$ and $F(x_s|x_{t-1}, \ldots, x_{s+1})$ are the conditional distribution functions of X_t and X_s, respectively. Note that this is the theorem of Sklar (1959) with conditioning set $\{X_{t-1}, \ldots, X_{s+1}\}$. By setting $s = 1$, application of equation (2.2) results in the expression

$$f(x_t|x_{t-1}, \ldots, x_1) = c_{t,1}(F(x_t|x_{t-1}, \ldots, x_2), F(x_1|x_{t-1}, \ldots, x_2))f(x_t|x_{t-1}, \ldots, x_2)\,.$$

Repeated application with $s = 2, 3, \ldots, t-1$ leads to the following decomposition of the conditional density

$$\begin{aligned} f(x_t|x_{t-1}, \ldots, x_1) &= \prod_{j=1}^{t-2} \{c_{t,j}(F(x_t|x_{t-1}, \ldots, x_{j+1}), F(x_j|x_{t-1}, \ldots, x_{j+1}))\} \\ &\times \; c_{t,t-1}(F(x_t), F(x_{t-1}))f(x_t)\,, \end{aligned} \tag{2.3}$$

where $F(x_t)$ and $f(x_t)$ are the marginal distribution function and density of X_t, respectively.

For simplicity we denote $u_{t|j} \equiv F(x_t|x_{t-1}, \ldots, x_j)$ and $u_{j|t} \equiv F(x_j|x_t, \ldots, x_{j+1})$, where $j < t$. They correspond to projections backwards and forwards $t - j$ steps, respectively. By also denoting $u_{t|t} \equiv F(x_t)$, the joint density at equation (2.1) can be written as

$$f(\boldsymbol{x}) = \prod_{t=2}^{T} \left\{ \prod_{j=1}^{t-1} \{c_{t,j}(u_{t|j+1}, u_{j|t-1}; \theta_{t,j})\}\, f(x_t) \right\} f(x_1)\,, \tag{2.4}$$

which is a product of $T(T-1)/2$ bivariate densities $c_{t,j}$, each with parameter vector $\theta_{t,j}$, and the T marginal densities.

Equation (2.4) can be recognised as a 'D-vine' and is one of a wider class of vine decompositions recently discussed in the context of graphical models by Bedford and Cooke (2002) and others. In this literature, the bivariate densities $c_{t,j}$, and corresponding distribution functions $C_{t,j}$, are called 'pair-copulas'. The notation used makes the conditioning set explicit; for example, $c_{t,j|t-1,t-2,\ldots,j+1}$ would denote the copula density in equation (2.4). This is essential for differentiating between vine decompositions of general vectors $\boldsymbol{X}$. However, it is not necessary to uniquely identify the pair-copulas of the D-vine decomposition when the elements of $\boldsymbol{X}$ are time-ordered. By adopting the shorter notation for the pair-copulas we are following the same notational convention used for partial correlations when measuring conditional linear correlation.

The most challenging aspect of the D-vine representation is the evaluation of $u_{t|j+1}$ and $u_{j|t-1}$ in equation (2.4). The following property (Joe, 1996; p.125) proves useful in this regard:

Lemma

Let $u_1 = F(x_1|y)$ and $u_2 = F(x_2|y)$ be conditional distribution functions, and $F(x_1, x_2|y) = C(u_1, u_2; \theta)$, where C is a bivariate copula function with parameters θ, then

$$F(x_1|x_2, y) = h(u_1|u_2; \theta)\,, \text{ where } h(u_1|u_2; \theta) \equiv \frac{\partial C(u_1, u_2; \theta)}{\partial u_2}\,.$$

Proof is given in Appendix A, while Aas et al. (2009) provide analytical expressions of h for a variety of popular bivariate copulas. For $j < t$, direct application of the lemma results in the following recursive relationships:

$$u_{t|j} = F(x_t|x_{t-1}, \ldots, x_j) = h_{t,j}(u_{t|j+1}|u_{j|t-1}; \theta_{t,j})\,, \tag{2.5}$$

$$u_{j|t} = F(x_j|x_t, \ldots, x_{j+1}) = h_{t,j}(u_{j|t-1}|u_{t|j+1}; \theta_{t,j}), \tag{2.6}$$

where $h_{t,j}(u_1|u_2; \theta_{t,j}) = \frac{\partial}{\partial u_2} C_{t,j}(u_1, u_2; \theta_{t,j})$ and $C_{t,j}$ is the distribution function corresponding to pair-copula density $c_{t,j}$. We label equation (2.6) a forwards recursion and equation (2.5) a backwards recursion, and from these it can be seen that $u_{t|j}$ and $u_{j|t}$ are functions not only of $\theta_{t,j}$, but also of the parameters of other pair-copulas. The recursions give the following algorithm for the evaluation of the values of $u_{t|j}$ and $u_{j|t}$ employed in equation (2.4):

Algorithm 1

Step (1): For $t = 1, \ldots, T$ set $u_{t|t} = F(x_t)$

Step (2): For $k = 1, \ldots, T-1$ and $i = k+1, \ldots, T$

Backwards Step: $u_{i|i-k} = h_{i,i-k}(u_{i|i-k+1}|u_{i-k|i-1}; \theta_{i,i-k})$

Forwards Step: $u_{i-k|i} = h_{i,i-k}(u_{i-k|i+1}|u_{i|i-k+1};\theta_{i,i-k})$

Note that step (2) involves the evaluation of the $T(T-1)/2$ functions $h_{t,j}$, for $j < t$, twice. Figure 1 depicts the dependencies between $u_{t|j}, u_{j|t}$ resulting from the recursions in Algorithm 1.

—–Figure 1 about here—–

2.1.1 Illustration: Stationary Gaussian AR(2)

To illustrate, consider the familiar stationary Gaussian AR(2) process where $X_t|X_{t-1}, X_{t-2} \sim N(\phi_1 X_{t-1} + \phi_2 X_{t-2}, \sigma^2)$. Hamilton (1994; pp.56-58) outlines the properties of this series, where the marginal distribution $X_t \sim N(0,\sigma_0^2)$, $\mathrm{Cov}(X_t, X_{t-j}) = \sigma_0^2\rho_j$, $\rho_1 = \phi_1/(1-\phi_2)$, $\rho_2 = \phi_1\rho_1 + \phi_2$ and $\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2}$ for $3 \le j \le T$. In the D-vine at equation (2.4) this lag 2 Markov process has pairwise copula densities $c_{t,j}(u_1,u_2;\theta_{t,j}) = 1$, for $t \ge 4$, $j = 1,\ldots,t-3$, with the remaining being the bivariate Gaussian copula

$$c_{Ga}(u_1,u_2;\theta) = (1-\theta^2)^{-1/2}\exp\left\{\frac{-\theta^2(w_1^2+w_2^2) - 2\theta w_1 w_2}{2(1-\theta^2)}\right\},$$

where $w_1 = \Phi^{-1}(u_1)$, $w_2 = \Phi^{-1}(u_2)$ and Φ is the standard normal distribution function (Song, 2000).

Substituting these pairwise copula into the D-vine gives

$$\begin{aligned} f(\boldsymbol{x}) &= \prod_{t=2}^{T}\left\{c_{Ga}(u_{t|t-1}, u_{t-2|t-1};\theta_{t,t-2}) f(x_t|x_{t-1})\right\} f(x_1) \\ &= \prod_{t=2}^{T}\left\{c_{Ga}(u_{t|t-1}, u_{t-2|t-1};\theta_{t,t-2}) c_{Ga}(u_{t|t}, u_{t-1|t-1};\theta_{t,t-1}) f(x_t)\right\} f(x_1)\,. \end{aligned}$$

Here $X_t \sim N(0,\sigma_0^2)$, so that $u_{t|t} = F(x_t) = \Phi^{-1}(x_t/\sigma_0)$. The pairwise copula parameters are (partial) correlation coefficients, with $\theta_{t,t-2} = \mathrm{Corr}(X_t, X_{t-2}|X_{t-1}) = (\rho_2 - \rho_1^2)/(1-\rho_1^2)$ and

$\theta_{t,t-1} = \text{Corr}(X_t, X_{t-1}) = \rho_1$.

To compute values of arguments of the copula densities, first note that $h_{t,j}(u_1|u_2;\theta_{t,j}) = u_1$ when $c_{t,j}(u_1, u_2; \theta_{t,j}) = 1$. Following Aas et al. (2009) for the Gaussian pair-copula

$$h_{Ga}(u_1|u_2;\theta) = \frac{\partial C_{Ga}(u_1, u_2; \theta)}{\partial u_2} = \Phi\left(\frac{\Phi^{-1}(u_1) - \theta\Phi^{-1}(u_2)}{\sqrt{1-\theta^2}}\right).$$

Starting with initial conditions $u_{t|t} = \Phi(x_t/\sigma_0)$ for $t = 1, \ldots, T$, the recursions in Algorithm 1 are run with $h_{t,j}$ as given above. For example, employing the recursions in this case gives for $t \geq 3$:

$$\begin{aligned}
u_{t|t-2} &= h_{Ga}(u_{t|t-1}|u_{t-2|t-1};\theta_{t,t-2})\,, \text{ where} \\
u_{t|t-1} &= h_{Ga}(u_{t|t}, u_{t-1|t-1};\theta_{t,t-1}) \text{ and} \\
u_{t-2|t-1} &= h_{Ga}(u_{t-1|t-1}|u_{t-2|t-2};\theta_{t-1,t-2})\,.
\end{aligned}$$

Employing the initial conditional conditions and expressions for $\theta_{t,t-1}$ and $\theta_{t-1,t-2}$ above, with some algebra $u_{t|t-2}$ can be expressed in terms of ϕ_1 and ϕ_2 as

$$u_{t|t-2} = \Phi\left((x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2})/\sigma^2\right)\,,$$

which agrees with the conditional distribution function $F(x_t|x_{t-1}, x_{t-2})$ from the definition of an AR(2).

2.2 Conditional Distributions and Simulation

From equation (2.5) $F(x_t|x_{t-1}, \ldots, x_1) = u_{t|1} = h_{t,1}(u_{t|2}|u_{1|t-1};\theta_{t,1})$, where $u_{t|2} = F(x_t|x_{t-1}, \ldots, x_2)$ is a function of x_t, but $u_{1|t-1}$ is not. Repeated use of equation (2.5) provides expressions for $u_{t|2}, \ldots, u_{t|t-1}$, and by noting that $u_{t|t} = F(x_t)$, the conditional distribution function can be expressed as

$$F(x_t|x_{t-1},\ldots,x_1) = h_{t,1} \circ h_{t,2} \circ \ldots h_{t,t-1} \circ F(x_t)\,. \tag{2.7}$$

To evaluate $h_{t,j}(\cdot|u_{j|t-1},\theta_{t,j})$, for $j = t-1,\ldots,1$, the values $u_{1|t-1},\ldots,u_{t-1|t-1}$ also need computing, which can be obtained by running Algorithm 1, but with $T = t$. The expression at equation (2.7) can be used to provide the efficient algorithm below for simulating from the multivariate copula via the method of composition. We simulate T independent uniforms $w_1,\ldots,w_T$, and compute $x_1 = F^{-1}(w_1)$ and $x_t = F^{-1}(w_t|x_{t-1},\ldots,x_1)$ for $t = 2,\ldots,T$, so that $\boldsymbol{x}$ has density at (2.4).

Algorithm 2

For $t = 1,\ldots,T$:

Step (1): Generate $w_t \sim \text{Uniform}(0,1)$

Step (2): If $t = 1$ set $x_1 = F^{-1}(w_1)$, otherwise set $x_t = F^{-1} \circ h^{-1}_{t,t-1} \circ \ldots \circ h^{-1}_{t,1}(w_t)$

Step (3): Set $u_{t|t} = F(x_t)$, and if $t > 1$ compute:

$$u_{t|j} = h_{t,j}(u_{t|j+1}|u_{j|t-1};\theta_{t,j}) \text{ for } j = t-1,\ldots,1$$
$$u_{j|t} = h_{t,j}(u_{j|t-1}|u_{t|j+1};\theta_{t,j}) \text{ for } j = 1,\ldots,t-1$$

We note that the functions $h^{-1}_{t,j}$ are easily computed analytically for all commonly used bivariate copula. Moreover, Algorithm 2 can be adjusted to produce an iterate from the conditional distribution $F(x_T, x_{T-1},\ldots,x_{t_0+1}|x_{t_0},\ldots,x_1)$ simply by skipping Steps (1) and (2) for $t = 1,\ldots,t_0$, but not Step (3). This can be useful in computing forecasts, particularly when the vector is longitudinal as we demonstrate in Section 5.2. Both Kurowicka and Cooke (2007) and Aas et al. (2009) give algorithms that are equivalent to Algorithm 2, although the former do not provide an expression for the conditional distribution function, while that of the latter is less succinct.

2.3 Longitudinal Data and Pair-Copula Selection

While the decomposition at equation (2.4) provides a flexible representation for time series data generally, we focus here on the longitudinal case. That is, where there are n independent observations $\boldsymbol{x} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ on a dependent time series vector $\boldsymbol{x}_i = (x_{i,1}, \ldots, x_{i,T})$. In the case where the number of pair-copulas, $T(T-1)/2$, is large compared to the number of scalar observations nT, it can prove hard to obtain reliable estimates without imposing strong restrictions, and thus a data-driven method that allows for parsimony is useful. We do so by following the Bayesian variable selection literature (see Clyde and George 2004 for a recent summary) and introduce indicator variables $\Lambda = \{\gamma_{t,s}; (t,s) \in \mathcal{I}\}$, where $\mathcal{I} = \{(t,s)|t = 2, \ldots, T\,; s < t\}$ and

$$
\begin{aligned}
c_{t,s}(u_1, u_2; \theta_{t,s}) &= 1 && \text{iff} \quad \gamma_{t,s} = 0 \\
c_{t,s}(u_1, u_2; \theta_{t,s}) &= c^{\star}_{t,s}(u_1, u_2; \theta_{t,s}) && \text{iff} \quad \gamma_{t,s} = 1.
\end{aligned}
$$

In the above $c^{\star}_{t,s}$ is a pre-specified bivariate copula density, such as a Gaussian, t, Gumbel or Clayton; see Frees and Valdez (1998; p.25) for a list of common choices. While there is no reason why the pair-copula cannot vary with (t,s), for simplicity we assume $c^{\star}_{t,s}$ are all of the same form in our empirical work and therefore drop the subscripts of the copula density $c^{\star}$ and corresponding distribution function $C^{\star}$.

When $\gamma_{t,s} = 0$ the copula is the independence copula $C_{t,s}(u_1, u_2; \theta_{t,s}) = u_1 u_2$, and implies that $h_{t,s}(u_1|u_2; \theta_{t,s}) = u_1$. Therefore, the parameter γ determines the form of the time series dependency. For example, if $\gamma_{t,s} = 0$ for all $s \leq t - p$, then $f(x_t|x_{t-1}, \ldots, x_1) = f(x_t|x_{t-1}, \ldots, x_{t-p})$ and the process is Markov of order p. In general, γ determines a parsimonious dependence structure that can vary with time t, extending antedependent models for longitudinal data (Gabriel, 1962; Pourahmadi, 1999; Smith and Kohn, 2002) and covariance selection for Gaussian copulas (Pitt, Chan and Kohn 2006).

The likelihood $f(\boldsymbol{x}|\Theta, \Lambda) = \prod_{i=1}^{n} f(\boldsymbol{x}_i|\Theta, \Lambda)$, where $\Theta = \{\theta_{t,s}; (t,s) \in \mathcal{I}\}$ and

$$f(\boldsymbol{x}_i|\Theta, \Lambda) = \prod_{t=2}^{T} \left\{ \prod_{j=1}^{t-1} \left\{ \left(c^\star(u_{i,t|j+1}, u_{i,j|t-1}; \theta_{t,j}) \right)^{\gamma_{t,j}} \right\} f(x_{i,t}) \right\} f(x_{i,1}) \, .$$

Here, the conditional copula data

$$\begin{aligned} u_{i,t|j+1} &= F(x_{i,t}|x_{i,t-1}, \ldots, x_{i,j+1}) \text{ and} \\ u_{i,j|t-1} &= F(x_{i,j}|x_{i,t}, \ldots, x_{i,j+1}) \, , \end{aligned} \tag{2.8}$$

are computed using Algorithm 1 applied separately to each observation $\boldsymbol{x}_i$. To speed the computation the following adjustment to Step (2) of Algorithm 1 can be employed:

$$h_{i,i-k}(u_1|u_2; \theta_{i,i-k}) = u_1 \text{ if } \gamma_{i,i-k} = 0$$

$$h_{i,i-k}(u_1|u_2; \theta_{i,i-k}) = h^\star(u_1|u_2; \theta_{i,i-k}) \equiv \frac{\partial}{\partial u_2} C^\star(u_1, u_2; \theta_{i,i-k}) \text{ if } \gamma_{i,i-k} = 1 \, .$$

Exploiting this observation in Algorithm 1 substantially increases execution speed when the proportion of zeros in γ is high; something that is likely to be the case in many longitudinal studies.

When the marginal distributions are Gaussian, the framework nests a wide range of longitudinal models. When $c^\star$ is a bivariate Gaussian copula, then the longitudinal vector follows a Gaussian AR(p) when $\gamma_{t,j} = 0$ for $t > p$ and $j > t - p$, while $\gamma_{t,j} = 1$ otherwise. When the Gaussian pair-copula parameters $\{\theta_{t,j}|t > p, j > t-p\}$ vary with t, a time-varying parameter autoregressive model is obtained. If the elements of γ vary, then the model is further extended to an antedependent model. However, by choosing non-Gaussian pair-copula densities $c^\star$, the approach allows for more complex models of dependence; which is something we show can have a considerable impact in our empirical work.

3 Bayesian Inference

3.1 Priors

The prior on Λ can be chosen to represent a preference for shorter Markov orders by setting the marginal priors $\pi(\gamma_{t,s} = 1) \propto \delta^{(t-s)}$, for $0 < \delta < 1$. Similarly, an informative prior can be used to ensure that $\gamma_{t,s} = 0$ if $\gamma_{t,s-1} = 0$. However, in our empirical work we do neither and simply place equal marginal prior weight upon each indicator. As observed by Kohn et al. (2001) such a prior can still prove highly informative when $N = T(T-1)/2$ is large. For example, if $K_\Lambda = \sum_{(t,s)\in\mathcal{I}} \gamma_{t,s}$ is the number of non-zero elements of Λ, then assuming the flat prior $\pi(\Lambda) = 2^{-N}$ puts very high prior weight on values of Λ which have $K_\Lambda \approx N/2$. To avoid this, beta priors can be employed (see, for example, Liu et al. 2009), although we adopt the prior

$$\pi(\Lambda) = \frac{1}{N+1} \begin{pmatrix} N \\ K_\Lambda \end{pmatrix}^{-1}.$$

which has been used successfully in the component selection literature (Cripps et al., 2005). It results in equal marginal priors, uniform prior weight on $\pi(K_\Lambda) = 1/(1+N)$, and the conditional prior

$$\pi(\gamma_{t,s} | \{\Lambda \backslash \gamma_{t,s}\}) \propto \frac{\Gamma(N - K_\Lambda + 1)\Gamma(K_\Lambda + 1)}{\Gamma(N+1)}. \tag{3.1}$$

The priors of the dependence parameters $\theta_{t,s}$ vary according to choice of copula function $C^\star$. When Gaussian pair-copulas are employed, the $\theta_{t,s}$ are partial correlations, and independent beta priors adopted as suggested by Daniels and Pourahmadi (2009) or volume based priors as in Pitt, Chan and Kohn (2006). When non-Gaussian pair-copulas are used, following equation (2.3), the parameters $\theta_{t,s}$ capture conditional dependence more generally. Unless mentioned otherwise, we employ independent flat priors on the domain of these dependence parameters. This extends the approaches suggested by Joe (2006) and Daniels and Pourahmadi (2009) for modeling covariance matrices, and we show that this is an effective strategy for a range of copula functions in our empirical work.

3.2 Sampling Scheme

Given the margins, we generate iterates from the joint posterior $f(\Lambda, \Theta|\boldsymbol{x})$ by introducing latent variables $\tilde{\theta}_{t,s}$, for $(t,s) \in \mathcal{I}$, such that

$$\theta_{t,s} = \begin{cases} \tilde{\theta}_{t,s} & \text{if } \gamma_{t,s} = 1 \\ \theta^\dagger & \text{if } \gamma_{t,s} = 0\,. \end{cases}$$

Here, $\theta^\dagger$ is the bivariate copula parameter value that corresponds to the independence copula $c^\star(u_1, u_2; \theta^\dagger) = 1$. For example, for the Gaussian copula this is the correlation coefficient and $\theta^\dagger = 0$, for the Gumbel $\theta^\dagger = 1$ and for the t copula with correlation ρ and degrees of freedom ν, $\theta^\dagger = (\rho = 0, \nu)$, with $\nu \to \infty$.

Note that $(\theta_{t,s}, \gamma_{t,s})$ is known exactly from $(\tilde{\theta}_{t,s}, \gamma_{t,s})$, and that $\pi(\tilde{\theta}_{t,s}|\gamma_{t,s} = 1) \propto \pi(\theta_{t,s}|\gamma_{t,s} = 1)$. We also assume prior independence between the latent and indicator variables, so that $\pi(\tilde{\Theta}, \Lambda) = \pi(\Lambda) \prod_{(t,s)\in\mathcal{I}} \pi(\tilde{\theta}_{t,s})$, with $\tilde{\Theta} = \{\tilde{\theta}_{t,s}; (t,s) \in \mathcal{I}\}$. We evaluate the posterior distribution using MCMC. The sampling scheme consists of Metropolis-Hastings (MH) steps that traverse the latent and indicator variable space by generating sequentially each pair $(\tilde{\theta}_{t,s}, \gamma_{t,s})$ for $t = 2, \ldots, T$ and $s = 1, \ldots, t-1$.

In the case where the pair-copula has a single dependency parameter, we adopt a MH proposal that is independent in $\gamma_{t,s}$ and $\tilde{\theta}_{t,s}$, so that $q(\tilde{\theta}, \gamma) = q_1(\gamma) q_2(\tilde{\theta})$. When there are multiple dependency parameters for a pair-copula, we simply generate each parameter independently in the same manner. Kohn, Smith and Chan (2001) compare the relative efficiency of a number of choices for q_1 in the regression variable selection problem, while Nott and Kohn (2005) look at the efficiency of adaptive sampling methods. In this paper we consider two choices for q_1. The first corresponds to the simple proposal $q_1(\gamma = 1) = q_1(\gamma = 0) = 1/2$, while the second is Sampling Scheme 2 proposed by Kohn et al. (2001).

This was the most computationally efficient scheme suggested by the authors, and employs the conditional prior as the proposal. For clarity, we label these two sampling schemes SS1 and SS2, respectively, and we examine their relative computational efficiency empirically in Section 5.1. In both cases we use a random walk proposal for q_2, with $\tilde{\theta}_{t,s}$ generated using a t-distribution with d degrees of freedom and scale τ^2. Appendix B outlines the sampling steps in more detail.

Because the likelihood can be computed in closed form, it is straightforward to estimate any marginal parameters, joint with the copula parameters, by appending additional MH steps as outlined by Pitt, Chan and Kohn (2006) for Gaussian copula. However, in practise it is well known that joint estimation does not meaningfully affect the estimated dependence structure; see Silva and Lopes (2008) for an empirical demonstration with bivariate copulas. In our empirical work we condition on any marginal estimates and focus on studying inference for the serial dependence structure on $[0,1]^T$.

3.3 Posterior Inference and Diagnostics

The sampling schemes are run for a burnin period, and then J iterates $\{\Theta^{[j]}, \Lambda^{[j]}\} \sim f(\Theta, \Lambda | \boldsymbol{x})$ collected. From these posterior inference is computed, including posterior means which we use in our empirical work as point estimates. Of particular interest here is $\mathrm{pr}(\gamma_{t,s} = 0 | \boldsymbol{x}) \approx \frac{1}{J} \sum_j (1 - \gamma_{t,s}^{[j]})$, which is the estimate of the marginal probability that the (t,s)th pair-copula is the independence copula.

Because in our analysis we consider different pair-copula families it is important to measure dependence on a common metric. There is an extensive literature on measures of concordence, with an comprehensive summary given by Nelsen (2006; Chapter 5). We focus on Kendall's tau, which is defined for each pair $(t,s) \in \mathcal{I}$ as

$$\tau_{t,s} = \mathrm{pr}\left((X_t^1 - X_t^2)(X_s^1 - X_s^2) > 0\right) - \mathrm{pr}\left((X_t^1 - X_t^2)(X_s^1 - X_s^2) < 0\right)$$

where (X_t^i, X_s^i) are independent copies of (X_t, X_s) for $i = 1, 2$. For a bivariate copula

$$\tau_{t,s} = 4 \int_0^1 \int_0^1 C_{t,s}(u_1, u_2; \theta_{t,s}) \mathrm{d}C_{t,s}(u_1, u_2; \theta_{t,s}) - 1\,.$$

For the independence copula $\tau_{t,s} = 0$, while Kendall's tau can be expressed as a function of $\theta_{t,s}$ for numerous common bivariate copula (Lindskog, McNeil and Schmock, 2003). For Gaussian and t copulas $\tau_{t,s} = \arcsin\left(\frac{2}{\pi}\theta_{t,s}\right)$, for the Clayton $\tau_{t,s} = \theta_{t,s}/(\theta_{t,s} + 2)$ and for the Gumbel $\tau_{t,s} = 1 - \theta_{t,s}^{-1}$. In these cases we can write Kendall's tau as $\tau_{t,s}(\theta_{t,s})$ and compute it's posterior mean as

$$\begin{aligned} E(\tau_{t,s}|\boldsymbol{x}) &= \int \tau_{t,s}(\theta_{t,s}) f(\theta_{t,s}|\boldsymbol{x}) \mathrm{d}\theta_{t,s} \\ &= \int \tau_{t,s}(\tilde{\theta}_{t,s}) f(\tilde{\theta}_{t,s}, \gamma_{t,s} = 1|\boldsymbol{x}) \mathrm{d}\tilde{\theta}_{t,s} \approx \frac{1}{J} \sum_{j=1}^{J} \tau_{t,s}(\tilde{\theta}_{t,s}^{[j]}) \gamma_{t,s}^{[j]}\,. \end{aligned}$$

This expression makes it clear that the posterior mean is a model average over the model indicator $\gamma_{t,s}$.

To judge the adequacy of the parametric copula fit we employ a diagnostic tool based on the sum of the transformed marginal variables and constructed in a Monte Carlo manner as follows. First, simulate iterates of $\boldsymbol{U} = (U_1, \ldots, U_T)$, where $U_t = F_t(X_t)$, using Algorithm 2 appended to the end of each sweep of the sampling scheme. After convergence we select every twentieth iterate to obtain an approximately independent sample $\{\boldsymbol{U}^{[1]}, \ldots, \boldsymbol{U}^{[K]}\}$ from the fitted pair-copula. For each of these iterates we compute the sum $S = \sum_{j=1}^{T} \Phi^{-1}(U_j)$. This sum is both is highly sensitive to the dependence structure of the fitted pair-copula model and, like Kendall's tau, comparable across different pair-copula families.

The iterates $\{S^{[1]}, \ldots, S^{[K]}\}$ form a sample from the density $f(S|\boldsymbol{x})$ that is approximately independent and from which a kernel density estimate (KDE) can be computed. This can be compared to two different benchmarks. The first is where the elements of $\boldsymbol{X}$ are assumed independent, so that $U_1, \ldots, U_T$ are independent uniforms and $S \sim N(0, T)$. The second is the empirically observed distribution of S. This is given as the KDE based on the sample $S_i = \sum_{t=1}^{T} \Phi^{-1}(u_{i,t}^{\text{obs}})$, for $i = 1, \ldots, n$, where $u_{i,t}^{\text{obs}} = \hat{F}_t(x_{it})$ and $\hat{F}_t(x_{it})$ is the empirical distribution function of the data $\{x_{1,t}, \ldots, x_{n,t}\}$. A parametric model that more adequately fits the observed dependence in the data will have $f(S|\boldsymbol{x})$ closer to this second benchmark distribution.

4 Simulation Study

We study the performance of the selection approach using a small simulation study. We consider the Gaussian, Clayton and Gumbel bivariate copulas, where the latter two have copula densities:

$$\begin{aligned} c_{Clay}(u_1, u_2; \theta) &= (1+\theta)(u_1 u_2)^{(-1-\theta)} \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta - 2}, \text{ for } \theta > 0, \text{ and} \\ c_{Gum}(u_1, u_2; \theta) &= \frac{\kappa^{(-2+2/\theta)} (\log u_1 \log u_2)^{(\theta-1)} \left[1 + (\theta - 1)\kappa^{-1/\theta}\right]}{u_1 u_2 \exp\left(\kappa^{1/\theta}\right)}, \text{ for } \theta \geq 1, \end{aligned}$$

and $\kappa = (-\log u_1)^\theta + (-\log u_2)^\theta$; see Nelsen (2006) for an introduction to the properties of these two copula. We assume the marginal distributions are known, and focus on the effectiveness of the approach to estimate the dependency structure on $[0,1]^T$. We simulate from the following three models, each with $T = 7$ margins and $n = 100$ observations on the longitudinal vector:

Model A: Zero mean Gaussian AR(1) with autoregressive coefficient 0.85 and unit variance disturbances. The margins are therefore (known) $N(0, 1/(1-0.85^2))$ distributions, $\gamma_{t,t-1} = 1$ and $\gamma_{t,s} = 0$ for $t - s > 1$ and $\theta_{t,t-1} = \rho$. Figure 2(a) depicts the resulting values of Kendall's tau for each pair-copula.

Model B: Clayton pair-copula model with $\gamma_{t,s} = 1$ for all $t - s \leq 2$, and zero otherwise. This corresponds to a second order time-inhomogenous Markov process, with pair-copula dependence parameters set so that the values of Kendall's tau are as depicted in Figure 2(b).

Model C: Gumbel pair-copula model with dependency structure as depicted in Figure 2(c), which specifies a time-inhomogenous Markov process with varying order.

——Figure 2 about here——

We simulate 50 datasets from the three models and fit each with the following estimators:

Estimator E1: Estimation with selection and correctly specified pair-copula type.

Estimator E2: Estimation without selection (so that $\gamma_{t,s} = 1$ for $(t, s) \in \mathcal{I}$) and correctly specified pair-copula type.

Estimator E3: Estimation with selection and incorrectly specified pair-copula type (Clayton for Model A; Gumbel for Model B; and Gaussian for Model C).

Estimator E4: Estimation with a Gaussian copula without selection.

Estimator E1 is our proposed method, while estimators E2, E3 & E4 are for comparison. We note that estimators E2 and E4 are the same for model A only. Figure 2 provides a summary of the reliability of the pair-copula selection procedure of estimator E1. To quantify this, for each pair-copula we compute the mean posterior probability of being dependent over the simulation $\bar{P}_{t,s} = \frac{1}{50}\sum_{i=1}^{50} P_{t,s}(i)$, where $P_{t,s}(i)$ is the posterior probability that $\gamma_{t,s} = 1$ in the ith dataset. Panels (d)-(f) plot these values for all pair-copulas and the three models, indicating that the Bayesian selection approach appears highly accurate. To confirm this, we also examine the performance of the approach for classification using a simple threshold. For each replicated dataset we classify each pair-copula as being dependent when $\text{pr}(\gamma_{t,s} = 1|\boldsymbol{x}) > 0.5$, or the independence copula otherwise. Over the three models, all pair-copula and all simulations, 99.5% of dependent pair-copula and 98.8% of independence pair-copula were correctly classified by estimator E1.

To show that the method also produces reliable estimates of the dependence structure, for each pair-copula we estimate the bias $\hat{b}(\tau_{t,s}) = \frac{1}{50}\sum_{i=1}^{50}\left(\tau_{t,s}(i) - \tau^*_{t,s}\right)$, where $\tau^*_{t,s}$ is the true value and $\tau_{t,s}(i)$ the posterior mean for the ith dataset of Kendall's tau for pair-copula $c_{t,s}$. Figure 2(g)-(i) reports these estimated biases, with most being zero to two decimal places.

——Table 1 about here——

For comparison, Table 1 reports the mean bias $\hat{b}(\tau_{t,s})$ for all combinations of the 3 models and 4 estimators, both for dependent and independence pair-copula. Estimator E1 exhibits a substantially lower bias than the other estimators in every case. Table 1 also reports the width $\hat{w}(\tau_{t,s})$ of the posterior probability interval for $\tau_{t,s}$, defined for each dataset as follows. Order the iterates $\{\tau_{t,s}(\theta_{t,s}^{[1]}),\ldots,\tau_{t,s}(\theta_{t,s}^{[J]})\}$ from smallest to largest, and then compute the Monte Carlo estimate of the 90% posterior probability interval by counting off the lower and upper 5% of the iterates. The mean width, computed across all pair-copula and simulation replicates, is reported for all cases and for dependent and independence pair-copulas. Again, estimator E1 has substantially lower widths than the other estimators, and by at least an order of magnitude for the independence pair-copulas. Overall, the simulation suggests that the selection approach works well, and can improve the estimated dependence structure. In every case, pair-copulas with selection out-performed by an order of magnitude the simple alternative of fitting a Gaussian copula.

5 Empirical Applications

5.1 Cow Liveweight

We use our method to identify the time series dependence structure of the liveweight of $n = 25$ cows measured at $m = 23$ unequally-spaced points in time. The data are discussed by Diggle, Liang and Zeger (1994, p.100) and Smith and Kohn (2002), and are the result of a longitudinal study with a 2×2 factorial design. In both studies the logarithm of the liveweight is modelled as having a quadratic time trend for each of the four treatment groups, which we also assume here. Diagnostics indicate that a Gaussian distribution fits the marginal data particularly well, and with this choice of parametric distribution for the margins we estimate a pair-copula decomposition assuming independence in the copula data across cows, but not across time.

Bivariate t copulas are used, so that

$$c^{\star}(u_1, u_2; \theta) = \frac{\Gamma\left(\frac{\nu+2}{2}\right)/\Gamma\left(\frac{\nu}{2}\right)}{t_7(x_1)t_7(x_2)\nu\pi(1-\rho^2)^{1/2}}\left(1 + \frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{\nu(1-\rho^2)}\right)^{-(\nu+1)/2},$$

where $\theta = (\rho, \nu)$, $x_1 = T_\nu^{-1}(u_1)$, $x_2 = T_\nu^{-1}(u_2)$ and t_ν and T_ν are the density and distribution functions of a univariate student t with ν degrees of freedom. Given the sample size, it is difficult to estimate the degrees of freedom for each of the pair-copula, and we first fix $\nu = 7$. Table 2 reports the posterior means $E(\rho_{t,s}|\boldsymbol{x})$, but only for pair-copulas where the posterior probabilities $\mathrm{pr}(\gamma_{t,s} = 1|\boldsymbol{x}) > 0.5$. The resulting pattern is sparse, indicating a Markovian structure with varying order; for example, at time $t = 22$ the order is 1, while at $t = 21$ the order is 6. However, even when the order is longer, serial dependence is still specified in a parsimonious fashion. For example, even though at time $t = 15$ the order is 14, there are only 4 previous periods that affect the value of X_{15}, conditional upon the previous values. At all times of the day $E(\rho_{t,t-1}|\boldsymbol{x}) > 0$, capturing strong positive dependence between X_t and X_{t-1}; something that has been noted in previous analyses.

——Tables 2 and 3 about here——

Table 3 shows the proportion of times different transitioning occured for all indicators during estimation via both SS1 and SS2. Ignoring minor computations, SS2 involves 1.45 times as many evalutions of the likelihood as SS1. This transfers directly into increased execution speed, as confirmed in Table 3, with SS2 taking 1.47 times as long to complete 1000 sweeps the sampling scheme as SS1.

The sample size of this data are moderate relative to the length of the longitudinal vector. It this situation pair-copula selection has the potential to impact substantially on the estimated dependence structure. To study this, we estimate the same longitudinal model, but without selection. Figure 3 plots the difference between the posterior means $E(\tau_{t,s}|\boldsymbol{x})$ obtained with no selection, and those obtained with selection. The differences are substantial, and the impact on the estimated dependence structure of selection is considerable in this case.

——Figure 3 about here——

We also relax the t copula specification, and allow the degrees of freedom $\nu_{t,s}$ for each pair-copula to vary. Uninformative priors lead to wide posterior distributions for $\nu_{t,s}$, with the sample size here being insufficient to yield tight intervals. Therefore, we assume an informative prior $\nu_{t,s} \sim N(9.5, 1.5^2)$, constrained so that $\nu_{t,s} > 5$. As before $\text{pr}(\gamma_{t,t-1} = 1|\boldsymbol{x}) \approx 1$ throughout, and the other pair-copulas that are identified as dependent is largely unchanged. Figure 4 plots the posterior means $E(\nu_{t,s}|\gamma_{t,s} = 1, \boldsymbol{x})$ for those pair-copula that are likely to be dependent with $\text{pr}(\gamma_{t,s} = 1|\boldsymbol{x}) > 0.5$.

——Figure 4 above here——

For comparison, we also estimate the dependence structure using Gaussian pair-copulas. A similarly sparse structure is determined, with 57 indicators with posterior probabilities

$\text{pr}(\gamma_{t,s} = 1|\boldsymbol{x}) > 0.5$, compared to 52 for the t_7 pair-copula model. Again, $\text{pr}(\gamma_{t,t-1} = 1|\boldsymbol{x}) \approx 1$ throughout.

The pair-copula analysis is more flexible than that by Diggle et al. (1994), who assume a parametric time series model, and also that by Smith and Kohn (2002). The latter authors assume a normalilty, where selection is considered for the non-fixed elements of the Cholesky factor of the disturbance precision matrix. When Gaussian pair-copulas are employed, a zero at the (t, s)th element of the Cholesky factor corresponds to an independence copula $c_{t,s} = 1$ in the decomposition at (2.4). However, for a Gaussian analysis of the cow liveweight data, the pattern found in the Cholesky factor does not match exactly that found using Gaussian pair-copulas because of the different prior structures employed in the two approaches.

5.2 New South Wales Intraday Electricity Load

Modelling and forecasting electricity load at an intraday resolution is an important problem faced by all electricity utilities; see Soares and Medeiros (2008) for a recent discussion. When observed intraday, load has strong daily, weekly, and yearly periodic behaviour, along with meteorologically induced variation (Harvey and Koopman, 1993). Numerous models have been proposed for intraday load, but some of the most successful are longitudinal (Ramananthan et al. 1997; Cottet and Smith, 2003). We model electricity load in the Australian state of New South Wales (NSW) observed every two hours between 2 January 2002 and 27 June 2005. The data were used previously by Panagiotelis and Smith (2008), who employ a longitudinal model with multivariate Gaussian disturbances over the day.[1] We also use a longitudinal model, but where the intraday dependence is captured by a more flexible pair-copula formulation.

For every two hour period $(t = 1, \ldots, 12)$ electricity load $L_{i,t}$ on day i is modelled with

[1] The data are currently available at URL `works.bepress.com/michael_smith/`

the marginal Gaussian regression

$$L_{i,t} = \alpha_t^1 + \alpha_t^2 i + \beta_t' Z_{i,t} + \alpha_t^3 |T_{i,t} - 18.3| + \epsilon_{i,t}\,,\ \epsilon_{i,t} \sim N(0, \sigma_t^2)\,. \tag{5.1}$$

Here, α_t^1 and α_t^2 measure level and linear time trend and $Z_{i,t}$ is a design matrix for the 12 seasonal polynomials and 14 day type dummy variables listed in Panagiotelis and Smith (2008). The effect of air temperature[2] $T_{i,t}$ is nonlinear with a minimum at 18.3C (65F), which is a commonly employed functional form in the demand modelling literature (Pardo, Meneu and Valor, 2002). Each marginal model is estimated using using maximum likelihood. Residual plots show that the marginal models remove the strong signal in the load data and quantile plots indicate that the marginal Gaussian assumption in equation (5.1) is appropriate.

To account for the strong intraday correlation a pair-copula decomposition with $\boldsymbol{x}_i = (L_{i,1}, \ldots, L_{i,12})'$ is used, where the first element corresponds to load at 03:30, which is the approximate time of the overnight low in demand. Figure 5 contains the results when Gaussian pair-copulas are employed with Bayesian selection. Panels (a) and (b) contain the Monte Carlo estimates of $\mathrm{pr}(\gamma_{t,s} = 1|\boldsymbol{x})$ and corresponding posterior means $E(\tau_{t,s}|\boldsymbol{x})$, for $(t, s) \in \mathcal{I}$. The results show that there is strong positive dependency between load at time t and the previous time $t-1$, with $\mathrm{pr}(\gamma_{t,t-1} = 1|\boldsymbol{x}) \approx 1$ throughout. Nevertheless, the dependence structure is sparse, with $\mathrm{pr}(\gamma_{t,s} = 1|\boldsymbol{x}) < 0.75$ for 33 of the 66 pair-copulas. Panel (c) depicts the differences in posterior means $E(\tau_{t,s}|\boldsymbol{x})$ with and without selection. While selection perturbs the dependence structure, it does not do so as substantially as for the liveweight data. This is because the sample size $n = 1096$ is large relative to the length of the longitudinal vector.

——Figures 5 and 6 about here——

Only four pair-copula in Figure 5(b) capture negative dependence, which is only minor

[2] The temperature $T_{i,t}$ is ambient air temperature in degree centigrade at Bankstown airport in western Sydney, which is the considered the centroid of demand in NSW by regulators.

with $E(\tau_{3,1}|\boldsymbol{x}) = -0.03$, $E(\tau_{6,4}|\boldsymbol{x}) = -0.13$, $E(\tau_{9,1}|\boldsymbol{x}) = -0.03$ and $E(\tau_{12,8}|\boldsymbol{x}) = -0.04$. Therefore, we also employ the Gumbel and Clayton pair-copulas, which restrict dependence to be positive. We compute posterior inference in both cases, and Figure 6 contains the equivalent plots as those produced for the Gaussian pair-copula. As before, strong dependencies between loads at time t and $t-1$ are captured with these two pair-copulas, although the dependence structure is more sparse in both cases than with the Gaussian pair-copula.

To judge the adequacy of the different parametric copula fit we employ the diagnostic based on the distribution of the sum discussed in Section 3.3. However, as our empirical benchmark we employ the copula data $u_{i,t} = \Phi((L_{i,t} - \hat{L}_{i,t})/\hat{\sigma}_t)$, computed over a 210 day long forecast period 3 January to 31 June 2005, so that $i = n, \ldots, (n + 210)$. These are the marginal predictive distributions with parameter values estimated from the in-sample data, but evaluted at the out-of-sample data points.

Figure 7 plots KDEs constructed from the thinned Monte Carlo samples of S for each of the three different pair-copulas considered. For comparison, also plotted is the distribution of S based on an assumption of independence and also that observed empirically over the forecast period. Ignoring the intraday dependence in the data leads to substantial under-statement of future variation in the sum. This translates directly into an under-statement in the variation of future daily total load, a quantity that is also important to electricity utilities. All three pair-copula models improve substantially on this benchmark. However, the Gaussian and Gumbel pair-copula models provide better forecasts than the Clayton pair-copula, which results in forecasts of the sum that are biased upwards markedly.

——Figures 7 and 8 about here——

Intraday forecasts of electricity load are essential for effective system management by electricity utilities. Forecasts for peak periods are made at mid-morning, and are typically much more accurate than those made prior to 06:00. Such forecasts can be constructed using a longitudinal model by evaluating the distribution $F(L_{i,12}, \ldots, L_{i,h+1}|L_{i,h}, \ldots, L_{i,1})$

for a horizon h in the mid-morning. Here, we select $h = 4$, which corresponds to 09:30, and evaluate the conditional distribution with the model parameters integrated out with respect to their posterior by appending Algorithm 2 to the end of the sampling scheme, but skipping Steps (1) and (2) for $t \leq h$. Figure 8 provides plots of the estimated marginal predictive densities $f(L_{i,t}|L_{i,4}, \ldots, L_{i,1})$ for $t = 6$ (13:30) and $t = 8$ (17:30) on 3 January 2008, which correspond to peak demand times. Unlike the earlier examples, selection has only a small impact, although pair-copula choice makes a great deal of difference, highlighting the importance of the type of pair-copula construction in longitudinal models.

6 Discussion

We argue in this paper that pair-copula constructions, and in particular the D-vine, are particularly suitable for the modeling of longitudinal data. This is unlike the more general graphical models case, where an ordering of the margins is required for a pair-copula decomposition. Our approach extends the current literature on covariance modeling for longitudinal data from the Gaussian case (Smith and Kohn, 2002; Huang, Liu, Pourahmadi and Liu 2006; Levina, Rothman and Zhu, 2008) to a wide range of non-Gaussian situations. Moreover, it extends the approach where general multivariate copula are employed to model longitudinal data (Meester and MacKay, 1994; Lambert and Vandenhende, 2002; Sun, Frees and Rosenberg, 2008) to fully exploit the time-ordering of the data. A paper close to ours in objective is Ibragimov and Lentzas (2008), who also develop multivariate copulas for time series data from a sequence of bivariate copulas for a Markov process. However, the copula constructed is not recognisable as a D-vine, and it is not clear how to compute inference for the resulting multivariate copula.

A Bayesian formulation is particularly appropriate for computing inference when the data have the potential to exhibit Markovian properties, by allowing rapid exploration of the high dimensional model space. Nevertheless, penalised maximum likelihood (Huang et al., 2006) or LASSO style estimators (Levina et al., 2008) are alternative shrinkage and covariance selection methods that also have potential for the efficient computation of inference for the pair-copula decomposition. Last, Pitt, Chan and Kohn (2006) suggest a Bayesian approach for covariance selection when using a Gaussian copula. Our approach extends Bayesian analysis to the more flexible pair-copula family when the data are longitudinal.

Acknowledgments

The work of Michael Smith was partially supported by an Australian Research Council Grant. Claudia Czado and Carlos Almeida gratefully acknowledge the financial support from the Deutsche Forschungsgemeinschaft (Cz 86/1-3: Statistical inference for high dimensional dependence models using pair-copulas).

Appendix A

To prove the Lemma in Section 2, note that

$$
\begin{aligned}
F(x_1|x_2, y) &= \int_0^{x_1} f(z_1|x_2, y)\mathrm{d}z_1 \\
&= \int_0^{x_1} \frac{\partial^2}{\partial z_1 \partial x_2} F(z_1, x_2|y)\mathrm{d}z_1 \frac{1}{f(x_2|y)} \,.
\end{aligned}
$$

Now because $F(z_1, x_2|y) = C(F(z_1|y), F(x_2|y); \theta)$,

$$
\int_0^{x_1} \frac{\partial^2}{\partial z_1 \partial x_2} F(z_1, x_2|y)\mathrm{d}z_1 = \frac{\partial}{\partial x_2} C(F(x_1|y), F(x_2|y); \theta) \,,
$$

so that:

$$
\begin{aligned}
F(x_1|x_2, y) &= \frac{\partial}{\partial u_2} C(F(x_1|y), F(x_2|y); \theta) \frac{\partial}{\partial x_2} F(x_2|y) \frac{1}{f(x_2|y)} \\
&= \frac{\partial}{\partial u_2} C(u_1, u_2; \theta) \,.
\end{aligned}
$$

Appendix B

Dropping the subscripts for notational convenience, each pair $(\tilde{\theta}, \gamma)$ is generated using a MH step in Section 3.2. The new iterate $(\tilde{\theta}^{\text{new}}, \gamma^{\text{new}})$ is accepted over the old $(\tilde{\theta}^{\text{old}}, \gamma^{\text{old}})$ with probability $\min(1, \alpha R)$, where R is an adjustment due to any bounds on the domain of θ. Denoting the conditional prior at equation (3.1) for the case when $\gamma = 1$ as π_1, and $\pi_0 = 1 - \pi_1$. If the likelihood in Section 2.3 is denoted as a function of the element (θ, γ) as

$L(\theta,\gamma)$ then α can be computed in four different cases as:

$$
\begin{aligned}
\alpha_{00} \equiv \alpha\left((\gamma^{\text{old}}=0,\tilde{\theta}^{\text{old}}) \to (\gamma^{\text{new}}=0,\tilde{\theta}^{\text{new}})\right) &= \frac{\pi(\tilde{\theta}^{\text{new}})}{\pi(\tilde{\theta}^{\text{old}})} \\
\alpha_{01} \equiv \alpha\left((\gamma^{\text{old}}=0,\tilde{\theta}^{\text{old}}) \to (\gamma^{\text{new}}=1,\tilde{\theta}^{\text{new}})\right) &= \frac{L(\tilde{\theta}^{\text{new}},\gamma^{\text{new}}=1)\pi(\tilde{\theta}^{\text{new}})\pi_1}{L(\gamma^{\text{old}}=0)\pi(\tilde{\theta}^{\text{old}})\pi_0} \times \frac{q_1(0)}{q_1(1)} \\
\alpha_{10} \equiv \alpha\left((\gamma^{\text{old}}=1,\tilde{\theta}^{\text{old}}) \to (\gamma^{\text{new}}=0,\tilde{\theta}^{\text{new}})\right) &= \frac{L(\gamma^{\text{new}}=0)\pi(\tilde{\theta}^{\text{new}})\pi_0}{L(\tilde{\theta}^{\text{old}},\gamma^{\text{old}}=1)\pi(\tilde{\theta}^{\text{old}})\pi_1} \times \frac{q_1(1)}{q_1(0)} \\
\alpha_{11} \equiv \alpha\left((\gamma^{\text{old}}=1,\tilde{\theta}^{\text{old}}) \to (\gamma^{\text{new}}=1,\tilde{\theta}^{\text{new}})\right) &= \frac{L(\tilde{\theta}^{\text{new}},\gamma^{\text{new}}=1)\pi(\tilde{\theta}^{\text{new}})}{L(\tilde{\theta}^{\text{old}},\gamma^{\text{old}}=1)\pi(\tilde{\theta}^{\text{old}})}\,.
\end{aligned}
$$

The likelihood L is not a function of θ when $\gamma=0$, while $L(\tilde{\theta},\gamma=1)=L(\theta,\gamma=1)$. If the prior for $\tilde{\theta}$ is uniform, as is the case in much of our empirical work, $\pi(\tilde{\theta}^{\text{new}})/\pi(\tilde{\theta}^{\text{old}})=1$. If θ is constrained to the domain (a,b), the factor

$$
R=\frac{T_d((b-\tilde{\theta}^{\text{old}})/\tau)-T_d((a-\tilde{\theta}^{\text{old}})/\tau)}{T_d((b-\tilde{\theta}^{\text{new}})/\tau)-T_d((a-\tilde{\theta}^{\text{new}})/\tau)}\,,
$$

where T_d is the distribution function of a standard t_d distribution. Note that the likelihood is not computed in the evaluation of α_{00}, and that with this proposal $\alpha_{00}=1$. Therefore, the more frequently this case arises, the faster the estimation. For SS1, the choice of q_1 further simplifies $\alpha_{01}=L(\tilde{\theta}^{\text{new}},\gamma^{\text{new}}=1)\pi_1/L(\gamma^{\text{old}}=0)\pi_0$, while $\alpha_{10}=L(\gamma^{\text{new}}=0)\pi_0/L(\tilde{\theta}^{\text{old}},\gamma^{\text{old}}=1)\pi_1$. For the choice of q_1 in SS2, $\alpha_{01}=L(\tilde{\theta}^{\text{new}},\gamma^{\text{new}}=1)/L(\gamma^{\text{old}}=0)$ and $\alpha_{10}=L(\gamma^{\text{new}}=0)/L(\tilde{\theta}^{\text{old}},\gamma^{\text{old}}=1)$.

References

Aas, K., Czado, C., Frigessi, A. and Bakken, H., (2009), ‘Pair-copula constructions of multiple dependence’, *Insurance: Mathematics and Economics*, 44, 2, 182-198.

Bedford, T. and Cooke, R. (2002), ‘Vines - a new graphical model for dependent random variables’, *Annals of Statistics*, 30, 1031-1068.

Clyde, M. and George, E., (2004), ’Model Uncertainty’, *Statistical Science*, 19, 81-94.

Cottet, R., and Smith, M., (2003), ‘Bayesian modeling and forecasting of intraday electricity load’, *Journal of the American Statistical Association*, 98, 839-849.

Cripps, E., Carter, C. and Kohn. R., (2005), ‘Variable Selection and Covariance Selection in Multivariate Regression Models’, in Dey, D. K. and Rao, C. R. (eds.), *Handbook of Statistics 25 Bayesian Thinking: Modeling and Computation*, Elsevier: North Holland, 519-552.

Daniels, M. and Pourahmadi, M., (2009), ‘Modeling covariance matrices via partial autocorrelations’, *Journal of Multivariate Analysis*, in press.

Demarta, S. and McNeil, A.J., (2005), ‘The t-copula and related copulas’, *International Statistical Review*, 73, 111-129.

Dempster, A., (1972), ‘Covariance selection’, *Biometrics*, 28, 157-75.

Diggle, P., Liang, K., and Zeger, S., (1994), *Analysis of Longitudinal Data*, Oxford, UK: Clarendon Press.

Embrechts, P., Lindskog, F., McNeil, A., (2003), ‘Modelling dependence with copulas and applications to risk management’, In: Rachev, S.T. (Ed.), *Handbook of Heavy Tailed Distributions in Finance*. North-Holland: Elsevier.

Frees, E.W. and E.A. Valdez (1998), “Understanding Relationships Using Copulas”, *North American Actuarial Journal*, 2, 1, 1-25.

Gabriel, K., (1962), ‘Ante-dependent analysis of an ordered set of variables’, *Ann. Math. Statist.*, 33, 201-212.

Hamilton, J., (1994), *Time Series Analysis*, New Jersey, USA: Princeton University Press.

Harvey, A. and Koopman, S., (1993), ‘Forecasting Hourly Electricity Demand Using Time-Varying Splines’, *Journal of the American Statistical Association*, vol.88, no.424, 1228-125.

Huang, J., Liu, N., Pourahmadi, M., and Liu, L. (2006), 'Covariance matrix selection and estimation via penalised normal likelihood' *Biometrika*, 93, 85-98.

Ibragimov, R. and Lentzas, G., (2008), 'Copulas and Long Memory', Harvard Institute of Economics Research Discussion Paper Number 2160.

Joe, H., (1996), 'Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters', In: Rüschendorf, L., Schweizer, B., Taylor, M.D. (Eds.), *Distributions with Fixed Marginals and Related Topics*.

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

Joe, H., (2006), Generating random correlation matrices based on partial correlations', *Journal of Multivariate Analysis*, 71, 2177-2189.

Kohn, R., Smith, M. and Chan, C., (2001), 'Nonparametric regression using linear combinations of basis functions', *Statistics and Computing*, 11, 313-322.

Kurowicka, D. and Cooke, R.M., (2006), *Uncertainty Analysis with High Dimensional Dependence Modelling*, Wiley: New York.

Kurowicka, D. and Cooke, R.M., (2007), 'Sampling algorithms for generating joint uniform distributions using the vine-copula method', *Computational Statistics and Data Analysis*, 51, 2889-2906.

Lambert, P. and Vandenhende, F., (2002), 'A copula-based model for multivariate non-normal longitudinal data: analysis of a dose titration safety study on a new antidepressant', *Statistics in Medicine*, 21, 3197-3217.

Levina, E., Rothman, A. and J. Zhu, (2008), 'Sparse Estimation of Large Covariance Matrices via a Nested Lasso Penalty', *Annals of Applied Statistics*, 1, 245-263.

Liu, X., Daniels, M. and B. Marcus, (2009), 'Joint Models for the Association of Longitudinal Binary and Continuous Processes with Application to a Smoking Cessation Trial', *Journal of the Americian Statistical Association*, 104, 429-438.

Meester, S. and MacKay, J., (1994), 'A Parametric Model for Cluster Correlated Categorical Data', *Biometrics*, 50, 954-963.

Min, A., and Czado, C., (2008), 'Bayesian inference for multivariate copulas using pair-copula constructions', working paper.

Nelsen, R., (2006), *An Introduction to Copulas*, 2nd ed., New York: NY: Springer.

Nott, D. and Kohn, R., (2005), 'Adaptive sampling for Bayesian variable selection', *Biometrika*, 92, 747-763.

Panagiotelis, A. and Smith, M., (2008), 'Bayesian identification, selection and estimation of semiparametric functions in high-dimensional additive models', *Journal of Econometrics*, 143, 291-316.

Pardo, A., Meneu, V. and Valor, E., (2002). 'Temperature and seasonality influences on Spanish electricity load', *Energy Economics*, 24, 55-70.

Peirson, J. and Henley, A., (1994), 'Electricity load and temperature', *Energy Economics*, 16, 235-24.

Pitt, M., Chan, D. and R. Kohn (2006), 'Efficient Bayesian Inference for Gaussian Copula Regression Models', *Biometrika*, 93, 537-554.

Pourahmadi, M., (1999), 'Joint mean-covariance models with applications to longitudinal data: Uncontrained parameterisation', *Biometrika*, 86, 677-69.

Ramanathan, R., Engle, R., Granger, C., Vahid-Araghi, F. and Brace, C., (1997), 'Short-run forecasts of electricity loads and peaks', *International Journal of Forecasting*, 13, 161-174.

Silva, R. and Lopes, H., (2008), 'Copula, marginal distributions and model selection: a Bayesian note', *Statistics and Computing*, 18, 313-320.

Sklar, A. (1959), 'Fonctions de répartition à n dimensions et leurs marges', *Publications de l'Institut de Statistique de L'Universit de Paris*, 8, 229-231.

Smith, M. and Kohn, R., (2002), 'Parsimonious Covariance Matrix Estimation for Longitudinal Data', *Journal of the American Statistical Association*, 91, 460, 1141-1153

Soares, L. and Medeiros, M., (2008), 'Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data', *International Journal of Forecasting*, 24, 630-644.

Song, P. (2000), 'Multivariate Dispersion Models Generated from Gaussian Copula', *Scandinavian Journal of Statistics*, 27, 305-320.

Sun, J., Frees, E. and Rosenberg, M., (2008), 'Heavy-tailed longitudinal data modeling using copulas', *Insurance Mathematics and Economics*, 42, 817-830.

Yuan, M. and Lin, Y., (2007), ‘Model Selection and Estimation in the Gaussian Graphical Model’, *Biometrika*, 94, 19-35.

	Model A			Model B				Model C			
Estimator	E1	E2/E4	E3	E1	E2	E3	E4	E1	E2	E3	E4
Mean Bias $\hat{b}(\tau_{t,s})$											
Dependent	-0.0051	-0.0121	-0.1204	0.0001	-0.0045	-0.1554	-0.0812	-0.0039	-0.0065	-0.0414	-0.0422
Independent	-0.0009	-0.0020	0.0024	0.0002	0.0107	0.0147	0.0508	0.0010	0.0446	0.0051	0.0092
Mean Width $\hat{w}(\tau_{t,s})$											
Dependent	0.0872	0.0899	0.1084	0.0734	0.0743	0.1322	0.1081	0.0861	0.0871	0.1015	0.1024
Independent	0.0658	0.2088	0.0098	0.0103	0.1619	0.0343	0.2109	0.0044	0.0995	0.1230	0.1938

Table 1: Means for the estimated bias $\hat{b}(\tau_{t,s})$ and the width $\hat{w}(\tau_{t,s})$ of the posterior probability intervals for Kendall's tau in the simulation study in Section 4. Results are given for all combinations of model and estimator. In each case the mean is reported for results over the 50 replicates in the simulation, and $N = 21$ pair-copulas broken down by type (dependent or independence).

t	Second Index s 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
2	0.56																					
3	-	0.69																				
4	-0.47	0.32	0.79																			
5	-	-	-	0.75																		
6	-	-	-	0.22	0.59																	
7	-	-	-0.43	-	0.67	0.86																
8	-	-	-	-	-	-	0.91															
9	-	-	0.29	-	0.27	-	0.49	0.87														
10	0.45	-	-	-	-	-	-	-0.16	0.81													
11	-	-	-	-	-0.20	-	-	-	0.55	0.90												
12	-	-	-	-	-	-	-0.29	-	-	-	0.85											
13	-	-	-	-	-	-	-	-	-	-	0.21	0.88										
14	0.20	-	-	-	-	-	-0.32	-	-	-	-	-	0.95									
15	-0.22	-	-	-	-0.52	-	-	0.18	-	-	-	-	-	0.93								
16	-	-	-	-	-	-	-	-	-	-	-0.55	-	-	0.47	0.85							
17	-	-	-	-	-	-	-	-	0.25	-	-	-	-	-	-	0.91						
18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.19	-	0.91					
19	-	-	-	-	-	-	-	-	-	-	-	-	-0.53	-	-	-	-	0.95				
20	-	-	-	0.67	0.20	-	-	-	-	-	-	-	-	-	-	-	-	-	0.94			
21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.50	-	-	-	-0.22	0.92		
22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.93	
23	-	-	-	-	-	-	-	-	-	0.29	-	-	-	-	-	-	-	-	-	-0.59	-	0.86

Table 2: Posterior means $E(\rho_{t,s}|\boldsymbol{x})$ for t_7 pair-copula fit to the cow liveweight data. Only those values which correspond to pair-copula with posterior proabilities $\text{pr}(\gamma_{t,s}=1|\boldsymbol{x}) > 0.5$ are reported.

Transition	SS1	SS2
$\gamma^{\text{old}} = 0 \rightarrow \gamma^{\text{new}} = 0$	64.3%	49.7%
$\gamma^{\text{old}} = 0 \rightarrow \gamma^{\text{new}} = 1$	4.3%	4.2%
$\gamma^{\text{old}} = 1 \rightarrow \gamma^{\text{new}} = 0$	4.3%	4.2%
$\gamma^{\text{old}} = 1 \rightarrow \gamma^{\text{new}} = 1$	27.1%	41.9%
Execution Time per 1000 Sweeps	180 mins	264 mins

Table 3: Percentage of times the four possible transitions occured in all the Metropolis-Hastings steps for the cow liveweight data. Results are reported for the differing proposal densities given by SS1 and SS2. Execution time (in serial on a standard Intel chip) per 1000 sweeps of the sampling scheme is also given.

$$
\begin{array}{cccccccccccc}
\begin{Bmatrix} u_{2|1} \\ u_{1|2} \end{Bmatrix} & & \begin{Bmatrix} u_{3|2} \\ u_{2|3} \end{Bmatrix} & & \begin{Bmatrix} u_{4|3} \\ u_{3|4} \end{Bmatrix} & \cdots \quad \cdots & & \begin{Bmatrix} u_{T-2|T-3} \\ u_{T-3|T-2} \end{Bmatrix} & & \begin{Bmatrix} u_{T-1|T-2} \\ u_{T-2|T-1} \end{Bmatrix} & & \begin{Bmatrix} u_{T|T-1} \\ u_{T-1|T} \end{Bmatrix} \\
\downarrow & \swarrow & \downarrow & \swarrow & \downarrow & & \swarrow & \downarrow & \swarrow & \downarrow & \swarrow & \\
\begin{Bmatrix} u_{3|1} \\ u_{1|3} \end{Bmatrix} & & \begin{Bmatrix} u_{4|2} \\ u_{2|4} \end{Bmatrix} & & \begin{Bmatrix} u_{5|3} \\ u_{3|5} \end{Bmatrix} & \cdots \quad \cdots & & \begin{Bmatrix} u_{T-1|T-3} \\ u_{T-3|T-1} \end{Bmatrix} & & \begin{Bmatrix} u_{T|T-2} \\ u_{T-2|T} \end{Bmatrix} & & \\
\downarrow & \swarrow & \downarrow & \swarrow & \downarrow & & \swarrow & \downarrow & \swarrow & & & \\
\begin{Bmatrix} u_{4|1} \\ u_{1|4} \end{Bmatrix} & & \begin{Bmatrix} u_{5|2} \\ u_{2|5} \end{Bmatrix} & & \begin{Bmatrix} u_{6|3} \\ u_{3|6} \end{Bmatrix} & \cdots \quad \cdots & & \begin{Bmatrix} u_{T|T-3} \\ u_{T-3|T} \end{Bmatrix} & & & & \\
\downarrow & \swarrow & \downarrow & \swarrow & \downarrow & & \swarrow & & & & & \\
\vdots & & \vdots & & & & & & & & & \\
\downarrow & \swarrow & & & & & & & & & & \\
\begin{Bmatrix} u_{T|1} \\ u_{1|T} \end{Bmatrix} & & & & & & & & & & &
\end{array}
$$

Figure 1: The dependencies between $\{u_{t|j}, u_{j|t}\}$ values resulting from both the forward and backwards recursions of Algorithm 1. Directed arcs indicate that perturbing the values at the origin node affects the values at the terminal node.

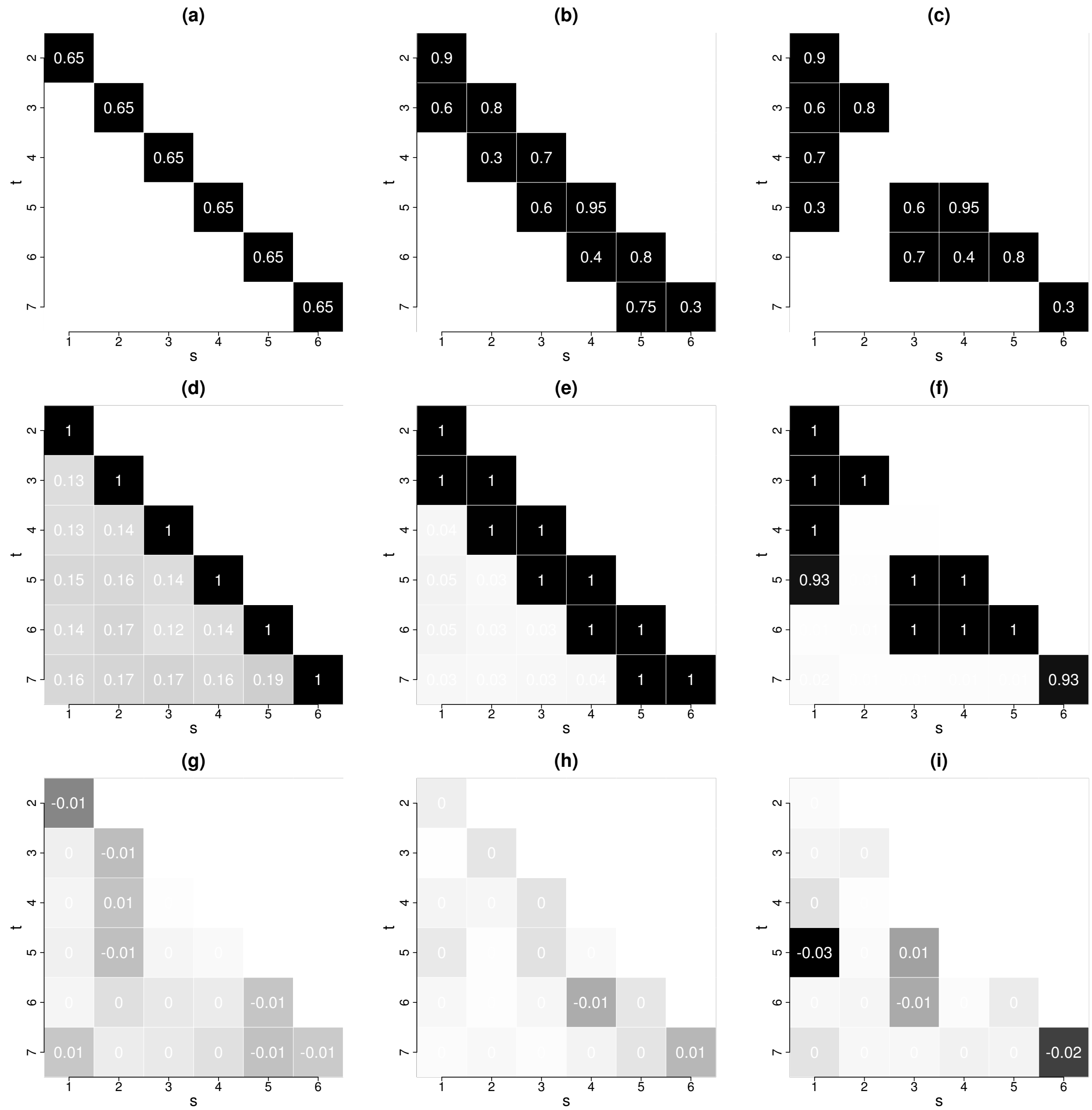

Figure 2: Simulation design and results for estimator E1, with the three columns of panels corresponding to models A, B and C, respectively. Panels (a)-(c) plot the true values of $\tau_{t,s}$ for the models in each row t and column s of each panel. Panels (d)-(f) plot the values of $\bar{P}_{t,s}$ defined in Section 4 in row t and column s of each panel. Panels (g)-(i) plot the estimated bias values $\hat{b}(\tau_{t,s})$ in row t and column s of each panel.

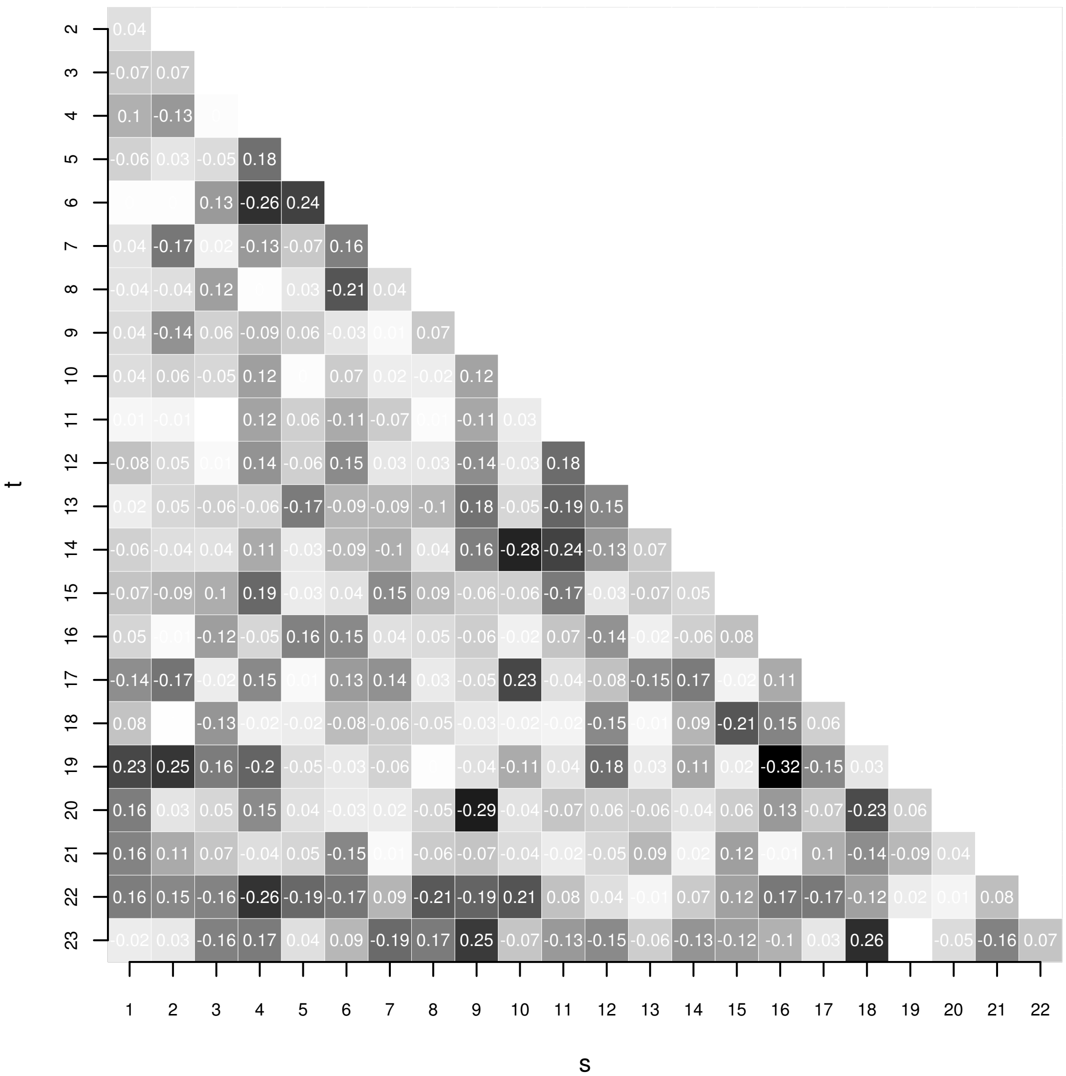

Figure 3: Difference in the posterior means of Kendall's tau measure under the full model minus the model with selection for the cow liveweight data fit with t_7 pair-copula. The number in row t and column s corresponds to pair-copula $C_{t,s}$.

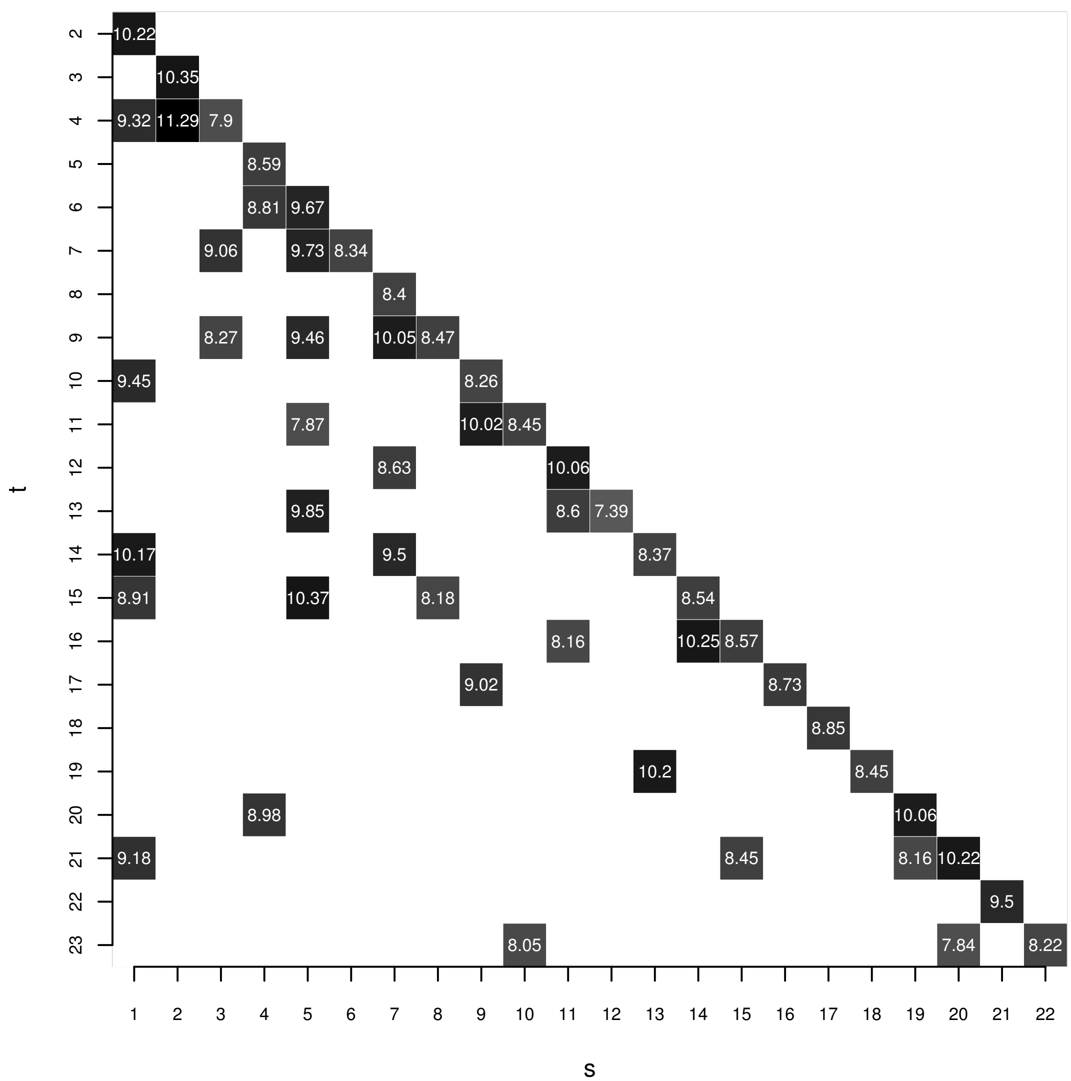

Figure 4: The posterior means $E(\nu_{t,s}|\gamma_{t,s} = 1, \boldsymbol{x})$ for the t pair-copula with varying degrees of freedom fitted to the cow liveweight data. Only those posteriors with $\mathrm{pr}(\gamma_{t,s} = 1|\boldsymbol{x}) > 0.5$ are reported, indicating the likely pattern of serial dependence. The number in row t and column s corresponds to pair-copula $C_{t,s}$.

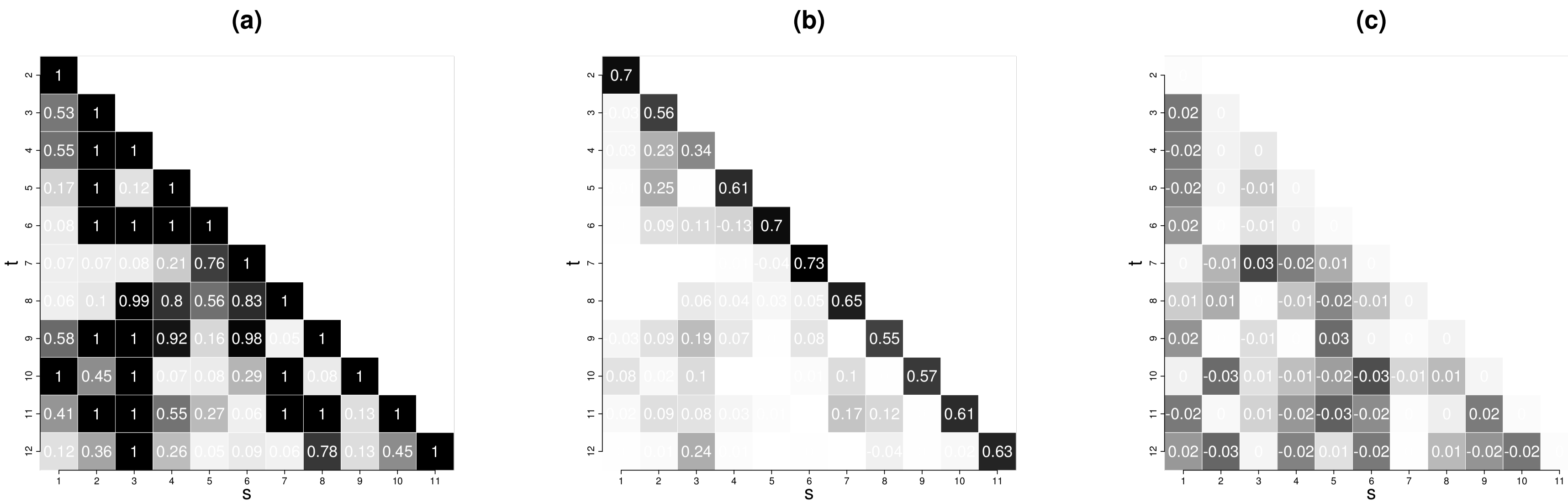

Figure 5: Gaussian pair-copula estimates for the NSW electricity example. Panel (a) depicts the posterior probabilities $\mathrm{pr}(\gamma_{t,s} = 1|\boldsymbol{x})$, while panel (b) depicts the corresponding posterior means of Kendall's tau $E(\tau_{t,s}|\boldsymbol{x})$. Panel (c) depicts the difference in the posterior means $E(\tau_{t,s}|\boldsymbol{x})$ obtained with and without selection. In all panels the (t, s)th entry corresponds to pair-copula $C_{t,s}$.

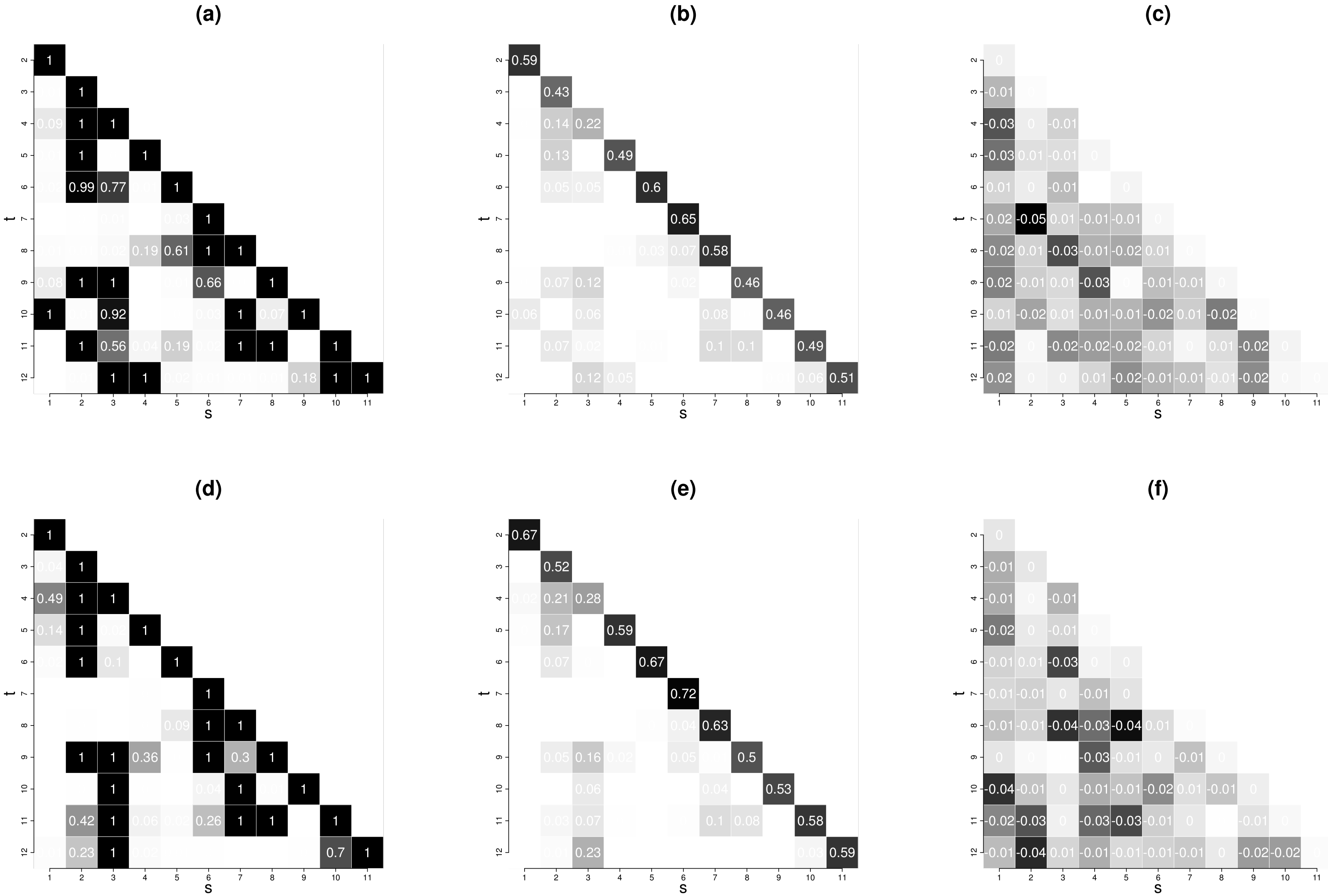

Figure 6: Gumbel and Clayton pairwise copula estimates for the NSW electricity example. Panels (a) and (b) depict the probabilities $\mathrm{pr}(\gamma_{t,s} = 1|\boldsymbol{x})$ and the corresponding means $E(\tau_{t,s}|\boldsymbol{x})$ for the Gumbel pairwise copula. Panel (c) depicts the difference between the posterior means of $\tau_{t,s}$ with and without selection. Panels (d), (e) and (f) depict the same, but for the Clayton pairwise copula.

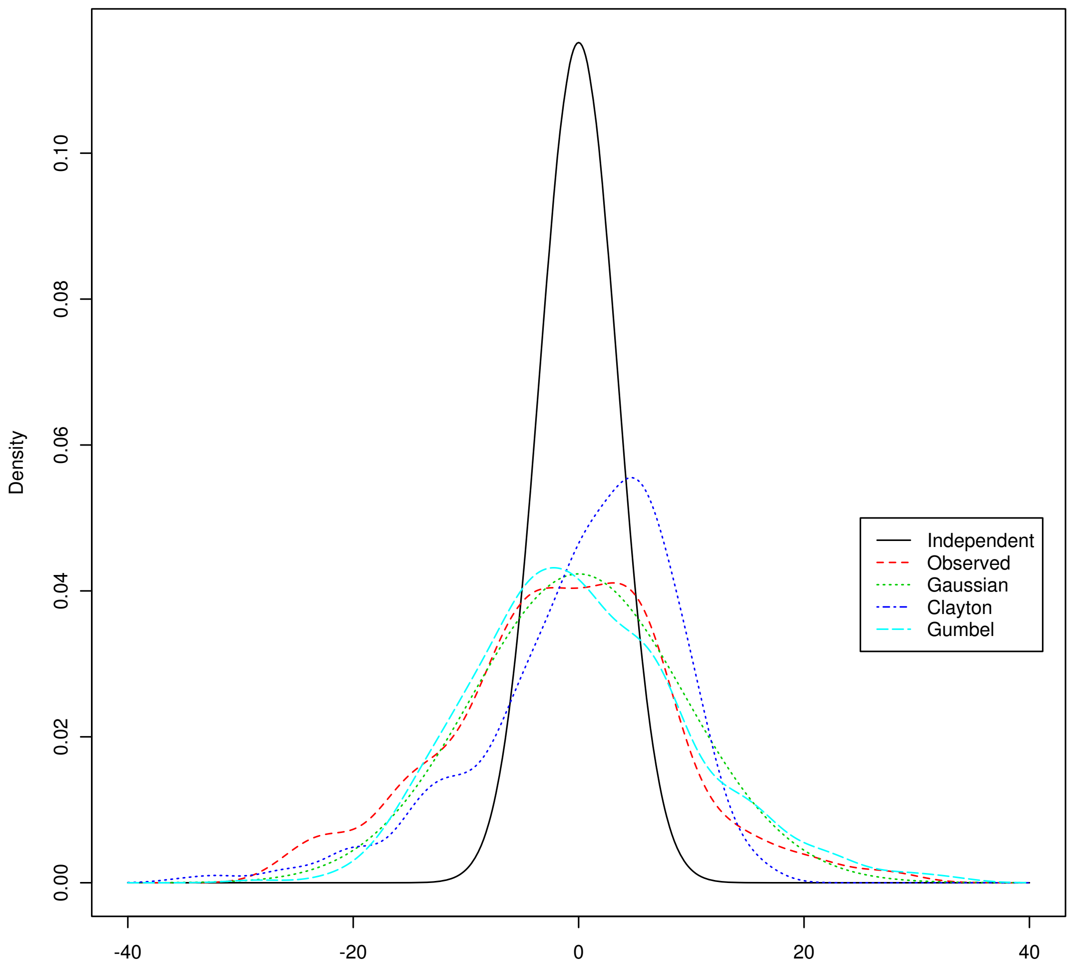

Figure 7: Distributions of S for the NSW electricity load example. The solid line corresponds to an assumption of intraday independence, while the distribution of the empirically observed data is given by the dashed (red) line. The distributions corresponding to the three parametric pair-copula models are also shown with line types (colours) as indicated.

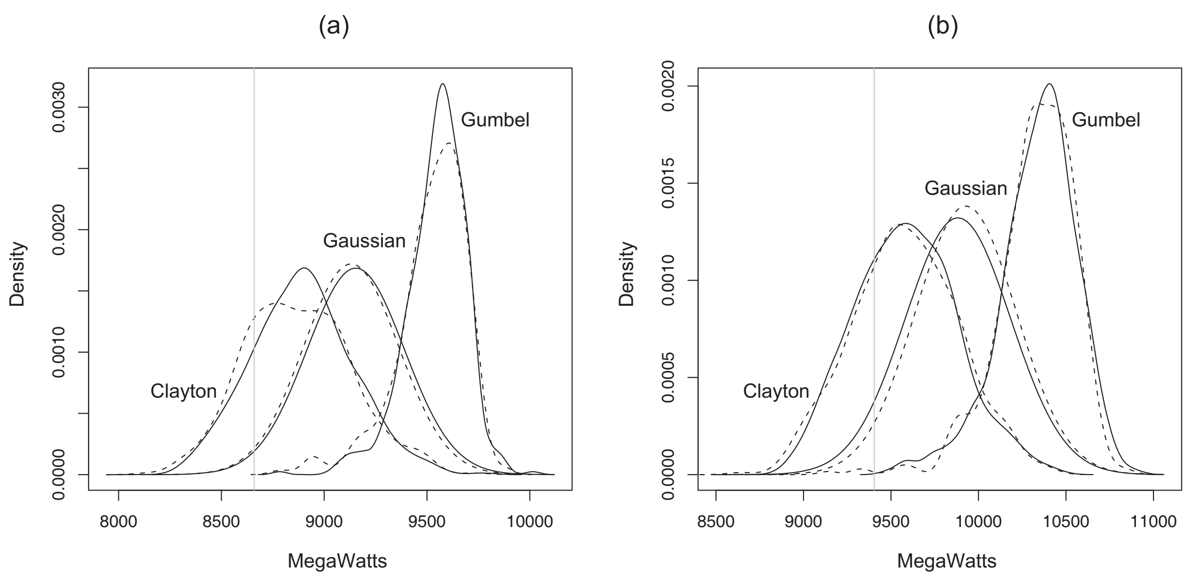

Figure 8: Intraday forecasts for electricity load on 3 January 2008. Panel (a) is for load at 13:30, while panel (b) is for load at 17:30, with both forecasts made conditional on load observed up to 09:30. Three different pair copulas were employed, and solid lines correspond to forecasts with selection and dashed lines without. The horizontal lines mark the actual load observed.