Dynamic copula based multivariate discrete choice models with applications

Michael Eichler[1], Hans Manner[*2], and Dennis Türk[1]

[1]*Department of Quantitative Economics, Maastricht University, The Netherlands*
[2]*Institute of Econometrics and Statistics, University of Cologne, Germany*

Abstract

We propose a dynamic model for multivariate binary outcomes, which allows the latent variables driving these observed outcomes to follow a vector autoregressive process. Furthermore the model is constructed using a copula representation for the joint distribution of the resulting innovations. This has several advantages over the standard multivariate probit model. First, it allows for nonlinear dependence between the error terms. Second, the univariate link function can be chosen freely. Third, the computational burden is greatly reduced, making estimation feasible in higher dimensions and for large samples. Finally, the model can easily be extended to allow for ordered outcomes and one may even model ordered and binary outcomes jointly. Conditions, needed for the estimated dependence parameter to lie inside the permissible parameter space are given, and the construction of confidence intervals for (small sample) cases in which these conditions are not fullfilled is discussed. Furthermore, the computation of marginal effects and the choice of the copula function for dimensions larger than two are treated. Two applications are presented. First, we treat the problem of forecasting the probability of extreme price (co-) occurrences, so called (co-) spikes, in Australian high frequency electricity markets, and second, we jointly model recession probabilities for four major economies. The supplementary material to this paper contains Monte Carlo simulations and some details concerning the computation of the likelihood function.

Keywords: Electricity price spikes, hierarchical archimedean copulas, multivariate probit, recession forecasting,vector autoregression

*Corresponding author, email: manner@statistik.uni-koeln.de.

1 Introduction

Models for binary outcomes such as logit and probit models are standard tools in econometrics, but the basic specifications are not always suitable for problems that arise in applications. In particular, the analysis of time-series data and modeling more than one variable are often of interest. Several studies have extended the basic model framework to the dynamic time series setting by including past information in the model. Examples are Dueker (1997), Kauppi and Saikkonen (2008) and Nyberg (2010) who apply dynamic binary choice models to the problem of modeling and forecasting recessions, Eichengreen et al. (1985) who model bank rate policy, or Eichler et al. (2014) who use a variation of the model of Kauppi and Saikkonen (2008) to forecast the occurrence of spikes in Australian Electricity prices. Theoretical properties of dynamic binary choice models are treated in De Jong and Woutersen (2011).

The first study introducing a multivariate probit model was Ashford and Sowden (1970). Recently, Nyberg (2013) and Candelon et al. (2013) proposed multivariate extensions of the dynamic binary choice model of Kauppi and Saikkonen (2008). The former paper proposes the generalization to a bivariate autoregressive probit model to jointly forecast recession probabilities for the US and Germany. The latter shows how to estimate the multivariate dynamic probit model in three dimensions using an exact maximum likelihood approach and applies the model to the problem of financial crisis mutation. Furthermore, Winkelmann (2012) presents an alternative model specification based on the recursive static bivariate probit model. The idea is to maintain the probit assumption for the marginal distributions while introducing non-normal dependence using copulas. In an application of the proposed copula bivariate probit model to analyse the effect of insurance status on the absence of ambulatory health care expenditure, the author shows that a model based on the Frank copula outperforms the standard bivariate probit model. Another notable contribution is Dueker (2005). He proposes a vector autoregressive model that allows for the inclusion of qualitative variables and applies it to model U.S. recessions and monetary policy contractions.

This paper makes the following contributions to the literature on multivariate discrete choice modeling. We introduce a copula based approach which permits the estimation of multivariate dynamic discrete choice models in dimensions larger than two, thus extending the ideas of Kauppi and Saikkonen (2008), Candelon et al. (2013), Nyberg (2013) and Winkelmann (2012). Our model allows for a flexible choice of the link functions, which may even differ for the distinct dependent variables in the model. Furthermore, the use of copulas makes it possible to model asymmetric dependence as well as tail dependencies between the innovations in the model. In order to make the computation of estimates in high dimensional models feasible we propose to use the class of nested Archimedean copulas (see, e.g., Okhrin et al. 2013 and Savu and Trede 2010) which yields closed form expressions for the likelihood function that can be evaluated at low computational costs, while maintaining a certain degree of flexibility. The representation of the model via the latent processes driving the observed outcomes implies that it can easily be extended to allow for ordered outcomes and that it can even be applied in situations in which some of the outcomes are binary and others are ordered.

For small samples it is possible that estimation results imply perfect positive (or negative) dependence. We analyze the necessary and sufficient conditions, which are needed for the estimated dependence parameter to lie in the interior of the parameter space and present a solution for cases in which these conditions are not fullfilled. A similar problem was studied in Butler (1996) in the specific context of censored probit models. Furthermore, we derive general formulas for computing the marginal effects of the model, complementing several studies that have treated the estimation of marginal effects in binary choice models. In particular, the analytical derivation of marginal effects in the context of the bivariate probit model is given by Greene (1996) and Christofides et al. (1997). Hasebe (2013) further derived the marginal effects for the bivariate (recursive) copula model of Winkelmann (2012). A recent extension for multivariate probit models with dimension greater than two was given by Mullahy (2011).

We present two applications, which demonstrate the practical relevance of our model framework in rather distinct contexts. First, we analyze the dynamic dependence between extreme price occurrences on four interconnected electricity markets in Australia. This subject is highly relevant for market participants in order to properly schedule demand/production, execute risk management and perform statistical arbitrage. In a univariate setting this problem has been studied by Christensen et al. (2009), Christensen et al. (2012), Clements et al. (2012), Korniichuk (2012), and Eichler et al. (2014), whereas Clements et al. (2015) treat this problem in a bivariate setting using a self-exciting peaks-over-threshold model. In the second application we revisit the problem of modeling recession probabilities in a multivariate setting and apply our framework to model the joint recession dynamics of the US, Canada, Japan and Germany. For the US this problem has received considerable attention in the literature in, e.g., Dueker (1997), Estrella and Mishkin (1998), Dueker (2005), Kauppi and Saikkonen (2008), or Nyberg (2010). A bivariate analysis for the U.S. and Germany can be found in Nyberg (2013).

The paper is structured as follows. In Section 2 we present the model and elaborate on the estimation, in particular the identification of the dependence parameter. We further discuss, the computation of marginal effects, the specification of the copula and extensions of the model. The empirical applications are presented in Section 3, while Section 4 concludes and outlines future research questions.

2 The model

In this section we present the dynamic copula based multivariate discrete choice (DCMDC) model. The general model specification is given in Section 2.1. In Section 2.2 we discuss maximum likelihood estimation of the model. Furthermore the estimation of the dependence parameter is treated in detail. Inference for the case when the dependence parameter is estimated to lie on the boundary of the parameter space is discussed in Section 2.3. Section 2.4 treats the computation of marginal effects. The precise choice of the copula function for dimensions larger than two is discussed in Section 2.5 and in Section 2.6 we outline the extension of the model to ordered outcomes.

2.1 Multivariate dynamic discrete choice models

Let $y_t = (y_{1t}, \ldots, y_{dt})'$, for $t = 1, \ldots, T$, be a d-dimensional vector of binary variables. Assume that the outcomes are driven by a vector of latent variables $y_t^* = (y_{1t}^*, \ldots, y_{dt}^*)'$. The observable variables are defined as

$$\begin{aligned} y_{it} &= 1 \text{ if } y_{it}^* > 0 \\ y_{it} &= 0 \text{ otherwise,} \end{aligned}$$

for $i \in \{1, \ldots, d\}$. Furthermore, let $x_t = (x_{1t}, \ldots, x_{kt})'$ be a vector of exogenous variables. Then generalizing the model by Kauppi and Saikkonen (2008) the latent variables are modeled as

$$y_t^* = \pi_t + \varepsilon_t \tag{1}$$

and

$$\pi_t = \alpha + Bx_t + \Gamma\pi_{t-1} + Dy_{t-1}, \tag{2}$$

where $\pi_t = (\pi_{1t}, \ldots, \pi_{dt})'$, $\alpha = (\alpha_1, \ldots, \alpha_d)'$ is a vector of intercepts, B is a $d \times k$ matrix, Γ and D are $d \times d$ matrices, and $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{dt})'$ is a vector of zero mean i.i.d. innovations. The parameters in the matrix Γ need to satisfy the usual restrictions to ensure stationarity of a VAR process. In this model specification the latent processes driving the binary outcomes are autocorrelated and it permits (dynamic) spillover effects between the endogenous variables. More general specifications in which π_t depends nonlinearly on its past or on exogenous variables are possible, see Section 3.1.2 for an example. The joint probabilities of the possible outcomes at time t depend on the distribution of ε_t and are given by

$$\begin{array}{ccccccc} P(y_{1t} = 1, \ldots, y_{dt} = 1) & = & P(\varepsilon_{1t} > -\pi_{1t}, \ldots, \varepsilon_{dt} > -\pi_{dt}) \\ \vdots \quad \vdots \quad \vdots & = & \vdots \quad \vdots \quad \vdots \\ P(y_{1t} = 0, \ldots, y_{dt} = 0) & = & P(\varepsilon_{1t} \leq -\pi_{1t}, \ldots, \varepsilon_{dt} \leq -\pi_{dt}). \end{array} \tag{3}$$

If ε_t is assumed to follow a multivariate normal distribution then the model is the multivariate probit model proposed by Candelon et al. (2013) and Nyberg (2013). However, we assume a more general joint distribution for the innovations based on the copula decomposition of multivariate distributions. In particular, the joint distribution H of ε_t can be written as

$$H(\varepsilon_{1t}, \ldots, \varepsilon_{dt}) = C_\theta(F_1(\varepsilon_{1t}), \ldots, F_d(\varepsilon_{dt})), \tag{4}$$

where F_i is the marginal distribution for ε_{it} and C_θ is a d-dimensional copula with parameter vector θ that captures the contemporaneous dependence between the innovations. For the copula any parametric family can in principle be used, see e.g. the books Nelsen (2006) or Joe (1997). We come back to this issue in Section 2.5. Common choices for the marginal distributions F_i are the standard normal distribution

$$F_i(\varepsilon) = \Phi(\varepsilon) \tag{5}$$

corresponding to the probit model or the logistic distribution

$$F_i(\varepsilon) = \frac{1}{1 + \exp(-\varepsilon)} \tag{6}$$

leading to a logit model. However, besides these two popular choices there are several alternative specifications available such as the Gumbel or the complementary log-log model (e.g. Greene 2011). In this paper we consider a further alternative distribution, namely the Burr-10 distribution,

$$F_i(\varepsilon) = \frac{1}{[1 + \exp(-\varepsilon)]^a}, \quad a > 0, \tag{7}$$

which was used by Nagler (1994) and results in the so called skewed logit (scobit) model. For $a = 1$ it nests the logit model, but otherwise the Burr-10 distribution is asymmetric.

In contrast to the logit and probit models, the probability of observing $y_{it} = 1$ is thus allowed to be most sensitive to changes in the explanatory variables at probabilities smaller or larger than 0.5.

Our model specification nests a number of specifications proposed in the literature. For $\Gamma = D = 0$, $F(\cdot) = \Phi(\cdot)$ and C being the Gaussian copula the model reduces to the standard multivariate probit model, see e.g. Ashford and Sowden (1970). The dynamic specification of Candelon et al. (2013) and Nyberg (2013) is obtained if the probit link and the Gaussian copula are used, and Γ and D are unrestricted. A bivariate, static (recursive) version of our model using probit link functions and a general dependence structure has been introduced by Winkelmann (2012).

Our specification has three advantages over the classical multivariate probit model. First, more general types of dependence can be allowed for. This includes the possibility of tail dependence and asymmetric dependence, i.e., the strength of cross dependence in ε_t can be different for small and large values. This can potentially result in a better model fit as linear correlation may not be an appropriate dependence measure in all situations. Second, the link function F can be chosen to be of any form, thus allowing to use distinct marginal distributions for the different dependent variables included in the system. The third advantage is the availability of closed forms for the distribution function H whenever the chosen copula has a closed form CDF[1], which is not the case for the multivariate normal distribution. This makes the estimation computationally more efficient, in particular in higher dimensions, as one does not need to rely on numerical integration techniques. Although the computational aspect may seem irrelevant given the availability of fast computers, note that the estimation of a three or four dimensional model with a Gaussian copula for reasonably large samples can be extremely time consuming or even infeasible.

2.2 Maximum likelihood estimation

The model introduced above can be estimated via maximum likelihood estimation (MLE). We begin by presenting the log-likelihood for the two-dimensional case. Let $z_t = (1\ x_t'\ \pi_{t-1}'\ y_{t-1}')$ be the vector that stacks all explanatory variables and let γ_i for $i = 1, 2$ be the column vector that stacks parameters from α, B, Γ and D corresponding to the i^{th} equation. Furthermore we make use of the fact that $C(u_1, 1) = u_1$, $C(1, u_2) = u_2$, and $C(u_1, 0) = C(0, u_2) = 0$. Next we define $s_{it} = 2y_{it} - 1$ for $i = 1, 2$. Then the proba-

[1] Admittedly this restricts the number of choices for the copula.

bilities corresponding to the four possible outcomes of y_t which are the contributions to the likelihood function, can be written as:

$$\begin{aligned} L(y_{1t}, y_{2t}) &= P(y_{1t}, y_{2t}|z_t) \\ &= y_{1t}y_{2t} - y_{1t}s_{2t}F_2(-z_t\gamma_2) - y_{2t}s_{1t}F_1(-z_t\gamma_1) + s_{1t}s_{2t}C_\theta[F_1(-z_t\gamma_1), F_2(-z_t\gamma_2)] \end{aligned} \tag{8}$$

The corresponding log-likelihood function for the model is then given by

$$LL(\gamma, \theta) = \sum_{t=1}^{T} \ln L(y_{1t}, y_{2t}), \tag{9}$$

where $\gamma = (\gamma_1, \gamma_2)$. In order to derive the likelihood for higher dimensions one has to compute the probabilities corresponding to the 2^d possible outcomes. These can be computed from the general formulas for computing rectangle probabilities for multivariate distributions given in, e.g., Nelsen (2006). Consider $\mathbf{a} = (a_1, a_2, \ldots a_d)$ and $\mathbf{b} = (b_1, b_2, \ldots b_d)$ where $a_i < b_i$, for all $i = 1, 2, \ldots, d$ and let $[\mathbf{a}, \mathbf{b}]$ denote the *d-box* $B = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]$, the Cartesian product of n closed intervals. Then for a given cumulative distribution function $H : R^d \to [0, 1]$, the probability of $\mathbf{X}$ lying in the *d-box* B defined by $\mathbf{a}$ and $\mathbf{b}$ is given by

$$P_H(B) = P(\mathbf{a} < \mathbf{X} < \mathbf{b}) = \sum_{j_1=1}^{2} \cdots \sum_{j_d=1}^{2} (-1)^{j_1+\ldots+j_d} H(x_{j_1}, \ldots, x_{j_d}), \tag{10}$$

with $x_{j_i} = a_i$ if $j_i = 1$ and $x_{j_i} = b_i$ if $j_i = 2$. For our specific case one has to set $a_i = -z_t\gamma_i$ and $b_i = \infty$ for an event occurring in component i, $(y_{it} = 1)$ and $a_i = -\infty$ and $b_i = -z_t\gamma_i$ for the probability of no event occurring in component i, $(y_{it} = 0)$. Corresponding to the possible number of outcomes, 2^d probabilities need to be computed, each consisting of up to 2^m terms with $m = \sum_{i=0}^{d} \mathbb{1}_{\{y_{it}=1\}}$ being the number of occurrences for the probability of interest. The log-likelihood is then computed similarly to the bivariate case. In the supplementary material to this paper we have written out the probabilities in equation (10) for the three and four dimensional case.

In principle estimating the DCMDC model via the maximum likelihood approach is straightforward. First we note that the existence of the maximum likelihood estimator for the regression parameters γ_i is guaranteed by the condition that there exists no parameter vector γ_i such that

$$s_{it}\, z_t\gamma_i > 0 \tag{11}$$

for all $t = 1, \ldots, T$ (cf Albert and Anderson 1984), where $s_{it} = 2\, y_{it} - 1$ as in the previous section. In the following, we assume that the above condition holds for $i = 1, \ldots, d$.

As discussed in Nyberg (2013), consistency and asymptotic normality of the MLE can be expected to hold under suitable regularity conditions. The existence of a MLE in the interior of the parameter space is implied by the conditions in equation (11) and Proposition 1 below. Uniqueness does not necessarily hold when jointly estimating γ and θ as the log-likelihood function is not strictly log-concave; see also Lesaffre and Kaufmann (1992) on the existence and uniqueness results. Furthermore, no results concerning the time series properties of the model in its most general form are known that would allow for the application of a central limit theorem to ensure asymptotic normality of the MLE. De Jong and Woutersen (2011) study such properties for (univariate) dynamic binary choice models containing lags of y_t. However, even for the univariate model of Kauppi and Saikkonen (2008) with lags of π_t no such results are known. For these reasons we leave a formal proof of asymptotic normality for our class of models for future research and focus on conditions for the existence of a MLE for θ in the interior of the parameter space.

For small samples it is possible that the estimated dependence parameter lies on the boundary of the permissible parameter space implying perfect positive (or negative) dependence. A closely related issue has been studied by Butler (1996) in the context of censored probit models. For this purpose we first give necessary and sufficient conditions for the maximum likelihood estimator of θ to lie in the interior of the parameter space for the bivariate case. Then we show how the argument extends to larger dimensions. In the next section we discuss how to draw inference in situations when the estimated dependence parameter lies on the boundary of the parameter space.

Proposition 1 *Consider the model defined by equations (2.1) and (2.4) for $d = 2$. Let the copula C_θ be parameterized by θ (possibly vector valued) and assume the copula has a limiting parameter θ_∞ for which it corresponds to the comonotonicity copula $C_{\theta_\infty}(u_1, u_2) = \min(u_1, u_2)$.*

(a) A necessary condition for the maximum likelihood estimator to fulfill $\hat{\theta} < \theta_\infty$ is that there exists at least one pair of observations such that $y_{1t} + y_{2t} = 1$.

(b) A sufficient condition for estimating $\hat{\theta} < \theta_\infty$ is that for all values of γ_1 and γ_2 there is at least one pair of observations such that (i) $y_{1t} + y_{2t} = 1$ and (ii) $P(Y_{1t} = y_{1t}) + P(Y_{2t} = y_{2t}) \leq 1$.

Furthermore assume the existence of a limiting parameter $\theta_{-\infty}$ for which the copula corresponds to the countermonotonicity copula.

(c) A necessary condition for the maximum likelihood estimator to fulfill $\hat{\theta} > \theta_{-\infty}$ is that there exists at least one pair of observations s.t. $y_{1t} = y_{2t}$.

(d) A sufficient condition for estimating $\hat{\theta} > \theta_{-\infty}$ is that for all values of γ_1 and γ_2 there is at least one pair of observations such that (i) $y_{1t} + y_{2t} = 2$ and $P(Y_{1t} = 0) + P(Y_{2t} = 0) \geq 1$, or (ii) $y_{1t} + y_{2t} = 0$ and $P(Y_{1t} = 0) + P(Y_{2t} = 0) \leq 1$.

Proof. In the following we denote $P(Y_{it} = 0) = F_i(-z_t\gamma_i)$ by F_{it}.

(a) Consider

$$P(Y_{1t} = 0, Y_{2t} = 0) = C_\theta(F_{1t}, F_{2t}),$$

and

$$P(Y_{1t} = 1, Y_{2t} = 1) = 1 - F_{1t} - F_{2t} + C_\theta(F_{1t}, F_{2t}).$$

Both these terms are increasing in C. Due to the Fréchet-Hoeffding upper bound $C_\theta(u_1, u_2) \leq \min(u_1, u_2) = M(u_1, u_2)$ these probabilities are maximized for $\hat{\theta} = \theta_\infty$. Thus if the stated condition does not hold the copula M maximizes the overall likelihood.

(b) The condition stated implies one of two possible cases that are symmetric to each other, so only one case needs to be considered. Let γ_1 and γ_2 be valid parameters for the regression part of the model. Without loss of generality, we may assume that γ_1 and γ_2 lie in a compact subset of the parameter space and suppose that at least one pair of observations fulfils $y_{1t} = 1$ and $y_{2t} = 0$. Then (ii) implies

$$F_{2t} = P(Y_{2t} = 0) \leq 1 - P(Y_{1t} = 1) = P(Y_{1t} = 0) = F_{1t}.$$

Now consider the corresponding joint probability of the observation

$$P(Y_{1t} = 1, Y_{2t} = 0) = F_{2t} - C_\theta(F_{1t}, F_{2t}).$$

For $C_{\theta_\infty} = M$ this probability becomes zero and implies a log-likelihood contribution of $-\infty$ assuring that the maximizer $\hat{\theta}_\gamma$ of the log-likelihood $LL(\gamma, \theta)$ for fixed γ satisfies $\hat{\theta}_\gamma < \theta_\infty$. Consequently, we have for every compact subset Γ of the parameter space for γ

$$\sup_{\gamma \in \Gamma} \hat{\theta}_\gamma < \theta_\infty.$$

Since by our assumption on γ in (11), the marginal and hence also the joint likelihood vanishes for $|\gamma_1| + |\gamma_2| \to \infty$, this proves $\hat{\theta} < \theta_\infty$.

(c) Consider

$$P(Y_{1t} = 1, Y_{2t} = 0) = F_{2t} - C_\theta(F_{1t}, F_{2t}),$$

and

$$P(Y_{1t} = 0, Y_{2t} = 1) = F_{1t} - C_\theta(F_{1t}, F_{2t}).$$

Both these terms are decreasing in C. Due to the Fréchet-Hoeffding lower bound $W = \max(u_1 + u_2 - 1, 0) \leq C_\theta(u_1, u_2)$ these probabilities are maximized for $\hat{\theta} = \theta_{-\infty}$.

(d) Let γ_1 and γ_2 be valid parameters for the regression part of the model and suppose that the pair (y_{1t}, y_{2t}) fulfills one of the stated conditions.

First, suppose that $y_{1t} = 1$ and $y_{2t} = 1$. Then for $\hat{\theta}_\gamma$ being on the lower bound of the permissible parameter space, condition (i) gives

$$\begin{aligned} P(Y_{1t} = 1, Y_{2t} = 1) &= 1 - F_{1t} - F_{2t} + C_{\theta_{-\infty}}(F_{1t}, F_{2t}) \\ &= 1 - F_{1t} - F_{2t} + F_{1t} + F_{2t} - 1 = 0. \end{aligned}$$

Thus for $C_\theta = W$ the probability becomes zero and implies a log-likelihood contribution of $-\infty$ assuring that $\hat{\theta}_\gamma > \theta_{-\infty}$.

Next, suppose that $y_{1t} = 0$ and $y_{2t} = 0$. The corresponding joint probability of the observation is

$$P(Y_{1t} = 0, Y_{2t} = 0) = C_\theta(F_{1t}, F_{2t}).$$

Since in case of $C_\theta = W$ we have $C_\theta = \max((F_{1t} + F_{2t} - 1, 0)$, (ii) would yield $C_\theta = 0$ and thus a joint probability of zero implying a log-likelihood contribution of $-\infty$ assuring that $\hat{\theta}_\gamma > \theta_{-\infty}$.

Finally, we get $\hat{\theta} > \theta_{-\infty}$ similarly as in (b) from the diverging behavior of the log-likelihood for large values of γ.

■

For illustration consider as a simple example the (static) bivariate probit model

$$\begin{aligned} y_{1t}^* &= \gamma_1 z_{1t} + \varepsilon_{1t} \\ y_{2t}^* &= \gamma_2 z_{2t} + \varepsilon_{2t}, \end{aligned} \tag{12}$$

where z_{1t} and z_{2t} are independent standard normal variables, $\gamma_1 = \gamma_2 = 1$, and ε_{1t} and ε_{2t} standard normal dependent through a Gumbel copula with $\theta = 4$. In this case, the critical region for the parameter vector γ where the log-likelihood $LL(\theta_\infty, \gamma)$ stays finite can be described by

$$\alpha_1 \gamma_1 < \gamma_2 < \alpha_2 \gamma_1$$

where

$$\alpha_1 = \max\{z_{1t}/z_{2t} | (y_{1t} > y_{2t} \text{ and } z_{2t} < 0) \text{ or } (y_{1t} < y_{2t} \text{ and } z_{2t} > 0)\}$$

and

$$\alpha_2 = \min\{z_{1t}/z_{2t} | (y_{1t} < y_{2t} \text{ and } z_{2t} < 0) \text{ or } (y_{1t} > y_{2t} \text{ and } z_{2t} > 0)\}.$$

If $\alpha_1 \geq \alpha_2$ the region is empty and the sufficient condition for the existence of the MLE is satisfied.

Consider the graphical illustration based on $T = 100$ random obervations from (12) in Figure 1, in which the copula was parametrized by $\rho = 1 - 1/\theta$ that lies in $(0, 1)$ and $\rho = 1$ corresponds to perfect dependence. The left panel shows $\hat{\rho}_\gamma = \text{argmax}_\rho LL(\rho|\gamma)$, the middle panel the corresponding value of the log-likelihood function and the right panel the value of the profile likelihood $LL(\hat{\gamma}_\theta|\theta)$. The critical region in which $LL(\theta_\infty, \gamma)$ stays finite is marked by the black lines and the MLE is represented by the black dot. For a large set of value for γ_1 and γ_2 the dependence parameter lies on the boundary of the parameter space. However, in this particular case the MLE exists. Furthermore, the profile likelihood is found to be bimodal indicating that the estimation of the model can become problematic.

An alternative way to state the sufficient conditions is as follows. Define the set

$$A_{it} = \begin{cases} [0, F_{it}] & \text{if } y_{it} = 0 \\ [F_{it}, 1] & \text{if } y_{it} = 1. \end{cases}$$

Then comonotonicity is possible only if

$$A_{1t} \cap A_{2t} \neq \emptyset, \quad \forall t \in \{1, \ldots, T\} \tag{13}$$

for some values of γ_1 and γ_2. Vice versa, if the intersection is empty for all possible values of γ_1 and γ_2 then the ML-estimator must satisfy $\hat{\theta} < \theta_\infty$. It is straightforward to check that the conditions are equivalent to the ones stated above. However, using this notation the extension to $d > 2$ follows directly. If the dependence of all pairs of variables depends on θ then comonotonicity is only possible if

$$A_{1t} \cap A_{2t} \cap \ldots \cap A_{dt} \neq \emptyset, \quad \forall t \in \{1, \ldots, T\}, \tag{14}$$

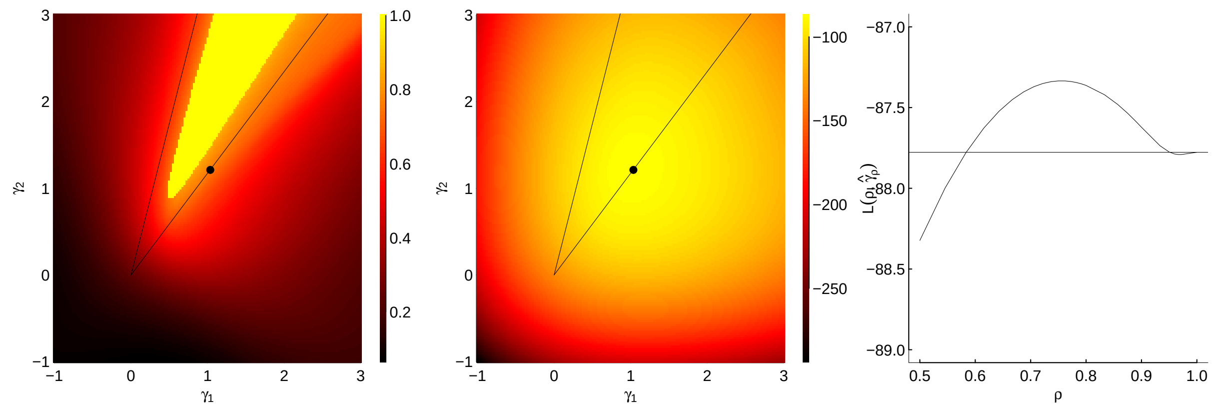

Figure 1: $\hat{\rho}_\gamma$ for artificial data from model (12) (left panel), the corresponding log-likelihood (middle panel) and the profile likelihood (right panel).

and $\hat{\theta} < \theta_\infty$ when the intersection is empty. In case there are multiple dependence parameters each characterizing the dependence for a subset of the variable pairs (as is the case in the hierarchical models we introduce in Section 2.5 below), the sufficient condition simply must hold for at least one of the pairs. Care must be taken when the copula is parametrized through a correlation matrix. In order to ensure that pairwise correlations lie in $(-1, 1)$ the above stated conditions must hold pairwise. However, this does not prevent degenerate behavior, as the correlation matrix may still not be positive definite.[2]

2.3 Inference when $\hat{\theta}$ lies on the boundary of the parameter space

In practice, the question arises of how inference on the copula parameter can be drawn in cases when $\hat{\theta} = \theta_\infty$. Due to the fact that the estimated parameter lies on the boundary of the parameter space standard inference is not possible. We suggest computing confidence intervals for $\hat{\theta}$ based on the profile likelihood method. In particular, for a fixed θ the profile likelihood estimator for γ is defined as

$$\hat{\gamma}_\theta = \underset{\gamma}{\operatorname{argmax}}\, LL(\gamma|\theta),$$

[2]Note that countermonotonicity is not treated for $d > 2$ as the Fréchet-Hoeffding lower bound is not a copula for dimensions larger than two.

where $LL(\gamma|\theta)$ denotes the likelihood function in (9) when holding θ constant. The maximum likelihood estimator for θ is then given by

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, LL(\hat{\gamma}_\theta|\theta)$$

and $(\hat{\gamma}_{\hat{\theta}}, \hat{\theta})$ is the ordinary maximum likelihood estimator. We can construct a confidence interval for θ by choosing (θ_l, θ_u) such that

$$2[LL(\hat{\gamma}_{\hat{\theta}}, \hat{\theta}) - LL(\hat{\gamma}_{\theta_l}, \theta_l)] = 2[LL(\hat{\gamma}_{\hat{\theta}}, \hat{\theta}) - LL(\hat{\gamma}_{\theta_u}, \theta_u)] = C_{1,1-\alpha}, \tag{15}$$

where $C_{1,1-\alpha}$ is the $(1-\alpha)$ quantile of the $\chi^2(1)$ distribution. In principle, this approach can also be used when $\hat{\theta} = \infty$, in which case the likelihood function is typically very flat with respect to θ. However, the approximation with a $\chi^2(1)$ distribution when $\hat{\theta}$ can lie on the boundary of the parameter space, but the LR statistics will follow a mixture of a $\chi^2(1)$ and a point mass at 0; see, e.g. Self and Liang (1987) or Andrews (2001). Nonetheless, the weights of the mixture are not known and can only be computed by simulations. We propose to approximate this distribution using a parametric bootstrap, which is implemented as follows.

1. Select a canditate value θ_l (θ_u) for the lower (upper) boundary of the CI. Obtain the profile likelihood estimator $\hat{\gamma}_{\theta_l}$. Furthermore, obtain the unrestricted MLE and compute the likelihood ratio statistic (LR) for testing $H_0 : \theta = \theta_l$ using (15).

2. Simulate from the model of interest using parameters θ_l and $\hat{\gamma}_{\theta_l}$ to obtain a bootstrap sample y_t^* for $t = 1, \ldots, T$. For the exogenous variables x_t use the observed values.

3. Based on the bootstrap sample y_t^* for $t = 1, \ldots, T$ compute the unrestricted MLE, as well as the profile MLE at θ_l, to obtain the bootstrap likelihood ratio statistic LR^* based on (2.15) for testing $H_0 : \theta = \theta_l$.

4. Repeat Steps 2 and 3 M times to obtain the bootstrap distribution of LR^*. Compute the empirical $1-\alpha$ quantile of the bootstrap distribution, $C^*_{1-\alpha}$. If $LR < C^*_{1-\alpha}$ the candidate value θ_l is inside the confindence interval, if $LR > C^*_{1-\alpha}$ it is outside the confidence interval.

5. Repeat Steps 1 to 4 over a grid of values for θ_l and θ_u until $LR \approx C^*_{1-\alpha}$ to find the upper and lower bounds of the confidence interval.

Note that although this approach is straightforward and works well, as our simulations below indicate, it is computationally very demanding due to the need to bootstrap the likelihood ratio statistic for all candidate values of (θ_l, θ_u). Clever choices for candidate values using confidence intervals based on the $\chi^2(1)$ distribution should be used to speed up the computations and we recommend the use of common random numbers to ensure convergence of the search for (θ_l, θ_u).

We illustrate the quality of the resulting confidence intervals through a small simulation study based on the model in (12) with sample size $T = 100$. We let the dependence parameter θ of the Gumbel copula vary from 1.5 to 6 in steps of size 0.5.[3] The model parameters are estimated by MLE and θ is constrained to lie in the interval $[1.01, 100]$. However, we explicitly allow for $\theta = \infty$ in the estimation and we estimate the remaining parameter using the profile likelihood in that case. Table 1 reports the fraction of samples for which $\hat{\theta} = \infty$. Furthermore, it compares the corresponding coverage probability of the 95% confidence intervals computed in the traditional way using estimated standard errors (Wald CI) with the ones estimated when applying the profile likelihood method, either by using the $\chi^2(1)$ distribution or the bootstrap method. To compute Wald confidence intervals we excluded the value $\theta = \infty$, but computed the confidence interval using the estimator restricted to $[1.01, 100]$. The results are based on 10,000 Monte Carlo replications. For the bootstrapped confidence intervals only 1,000 Monte Carlo simulations and 1,000 bootstrap replications were used. It stands out that for large values of θ the dependence parameter is estimated on the boundary of the parameter space in a large fraction of the cases. In these cases the coverage rates of the classical confidence intervals is well below its nominal level. The confidence intervals based on the profile likelihood method, on the other hand, have coverage rates that are close to the nominal level, but that are too small for low values of θ and too large when θ is large. It appears that whenever the sufficient condition for a finite $\hat{\theta}$ is not fulfilled the likelihood function is extremely flat resulting in a coverage rate above 95%. Furthermore, the $\chi^2(1)$ approximation of the distribution of the likelihood ratio statistic is likely imprecise near the boundary of the parameter space. The bootstrap appears to fix this problem, resulting in coverage rates that are close to 95%.

[3]Note that in the case of the Gumbel copula $\theta = 1$ corresponds to the independence copula, whereas $\theta = \infty$ corresponds to the comonotonicity copula. Values of θ larger than 3 indicate very strong dependence.

Table 1: Coverage rates of confidence intervals for θ

θ	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
Wald CI	0.942	0.923	0.902	0.889	0.881	0.880	0.871	0.872	0.872	0.869
Profile CI	0.945	0.933	0.920	0.933	0.960	0.974	0.981	0.986	0.987	0.988
Bootstrap CI	0.946	0.958	0.954	0.951	0.952	0.947	0.96	0.941	0.954	0.965
Fraction $\hat{\theta} = \infty$	0.005	0.037	0.114	0.208	0.298	0.381	0.453	0.513	0.567	0.610

Note: Coverage probabilities of confidence intervals (CI) based on the standard approach and based on the profile likelihood for the Gumbel copula parameter in model (12) with $\gamma_1 = \gamma_2 = 1$ based on 10,000 Monte Carlo simulations (1,000 for the bootstrapped intervals). The dependence parameter is constrained to lie in the interval [1,100] and the value $\theta = \infty$ is permitted. The last row reports the fraction of samples for which $\hat{\theta} = \infty$.

2.4 Marginal effects

Due to the fact that the coefficients of discrete choice models do not have a direct interpretation, one commonly computes the marginal effects of changes in the exogenous variables on the probabilities of interest based on the estimated model parameters. This problem has been addressed for certain specifications of multivariate discrete choice models in Greene (1996), Christofides et al. (1997), Hasebe (2013) and Mullahy (2011). In the bivariate case marginal effects for joint conditional probabilities are obtained by calculating the derivative of (8):

$$\begin{aligned}\frac{\partial P(Y_{1t} = y_{1t}, Y_{2t} = y_{2t}|z_t)}{\partial z_t} &= y_{1t}s_{2t}f_1(-z_t\gamma_1)\gamma_1 + y_{2t}s_{1t}f_2(-z_t\gamma_2)\gamma_2 \\ &- s_{1t}s_{2t}\left[\frac{\partial C_\theta}{\partial u_1}\cdot f_1(-z_t\gamma_1)\cdot\gamma_1 + \frac{\partial C_\theta}{\partial u_2}\cdot f_2(-z_t\gamma_2)\cdot\gamma_2\right], \qquad (16)\end{aligned}$$

where f_1 and f_2 are the density functions corresponding to F_1 and F_2, respectively, and s_{it} has been defined in Section 2.2. We use $(\partial C_\theta)/(\partial u_i)$ instead of $\partial C_\theta(F_1(-z_t\gamma_1), F_2(-z_t\gamma_2))/\partial F_i(-z_t\gamma_i)$ to simplify notation.

For the d-variate case marginal effects for joint conditional probabilities are given as the derivative of (10):

$$\frac{\partial P(Y_{1t} = y_{1t}, \ldots, Y_{dt} = y_{dt}|z_t)}{\partial z_t} = \frac{\sum_{j_1=1}^{2}\cdots\sum_{j_d=1}^{2}(-1)^{j_1+\ldots+j_d}\partial H(x_{j_1}, \ldots, x_{j_d})}{\partial z_t}, \qquad (17)$$

with the corresponding x_{j_i} given below equation (10). The number of marginal effects which need to be calculated is 2^d while each equation itself will again be a function of the number of corresponding events with 2^m terms, where m is the number of ones in the event of interest. Furthermore the number of needed derivatives for each equation is a function of d and m equal to $\sum_{i=0}^{m} \binom{m}{i}(d-m+i)$. Thus computation of marginal effects for very high dimensions will be tedious. Nonetheless, for dimensions that are relevant in practice this is generally manageable.

In case one is interested in marginal effects with respect to a binary variable, say $z_{k,t}$, they can be computed using the difference, $P(Y_{1t} = y_{1t}, \ldots, Y_{dt} = y_{dt} | z_{k,t} = 1, z_{-k,t}) - P(Y_{1t} = y_{1t}, \ldots, Y_{dt} = y_{dt} | z_{kt} = 0, z_{-k,t})$. Here $z_{-k,t}$ denotes the vector z_t excluding $z_{k,t}$.

2.5 Specification of the copula

Until now we have left the specification of the copula function open. In principle any parametric copula C: $[0,1]^d \to [0,1]$ with unrestricted domain can be considered, see Nelsen (2006) or Joe (1997) for a large number of possibilities. In our situation the copula should ideally satisfy some properties to be useful. First, it must be available in dimensions larger than two. Second, while flexibility is generally speaking a nice feature the number of parameters should not grow to fast as the dimension increases.[4] Finally, the distribution function of the copula should be available in closed form. This precludes elliptical copulas such as the Gaussian and Student copulas, as their distribution functions are only defined implicitly and have to be evaluated using numerical integration techniques. This also precludes the popular class of vine copulas to handle dependence dimensions larger than two, see Aas et al. (2009) or Czado (2010) for an introduction. One class of copulas that satisfies the stated requirements are Archimedean copulas. They are defined through a generator function $\phi : [0,1] \to [0,\infty]$ that is continuous, strictly decreasing and convex with $\phi(1) = 0$ and $\phi(0) = \infty$. Then the function

$$C(u_1, \ldots, u_d) = \phi^{-1}(\phi(u_1) + \ldots + \phi(u_d))$$

is called an Archimedean copula. Some popular examples are the Clayton copula with $\phi(t) = (t^{-\theta} - 1)/\theta$, the Gumbel copula with $\phi(t) = (-\ln(t))^{\theta}$, or the Frank copula with $\phi(t) = -\ln((e^{-\theta t} - 1)/(e^{-\theta} - 1))$. A clear disadvantage of simple Archimedean copulas is that a dependence parameter determines the dependence between all d variables. However, they can be extended by relying on a nested dependence structure that allows

[4] This is not as important as in other applications of copula models where one may consider dimensions of 10 or higher. Here the number of parameters in the VAR-type model driving the latent process restricts the dimension of the model to a certain degree.

for $d-1$ distinct generators to construct d-dimensional copulas that only have partial exchangeability. The fully nested Archimedean copula is given by

$$C(u_1,\ldots,u_d) = \phi_{d-1}^{-1}(\phi_{d-1}\circ\phi_{d-2}^{-1}[\ldots(\phi_2\circ\phi_1^{-1}[\phi_1(u_1)+\phi_1(u_2)]+\phi_2(u_3))+\ldots \\ +\phi_{d-2}(u_{d-1})]+\phi_{d-1}(u_d)).$$

The dependence parameter of generator $j-1$, θ_{j-1} always has to be larger or equal to θ_j in order for C to be a copula. In the trivariate case this gives the copula

$$C(u_1,u_2,u_3) = C_{\theta_2}(C_{\theta_1}(u_1,u_2),u_3),$$

with $\theta_1 \geq \theta_2$. The bivariate dependence between u_1 and u_2 is characterized by θ_1, whereas the dependence of the pairs (u_1,u_3) and (u_2,u_3) is described by θ_2. In the four dimensional case this gives the fully nested structure

$$C(u_1,u_2,u_3,u_4) = C_{\theta_3}(C_{\theta_2}(C_{\theta_1}(u_1,u_2),u_3),u_4),$$

with $\theta_1 \geq \theta_2 \geq \theta_3$. Here θ_1 is the dependence parameter for the pair (u_1,u_2), θ_2 for the pairs (u_1,u_3) and (u_2,u_3), and θ_3 for the pairs (u_1,u_4), (u_2,u_4) and (u_3,u_4).

The nesting can also be done in different ways, see Savu and Trede (2010) or Okhrin et al. (2013) for a general treatment of hierarchical Archimedean copulas. In higher dimensions a large number of nesting structures are possible. In the four dimensional case only one alternative nesting structure is possible,

$$C(u_1,u_2,u_3,u_4) = C_{\theta_3}(C_{\theta_1}(u_1,u_2),C_{\theta_2}(u_3,u_4)),$$

with $\theta_1,\theta_2 \geq \theta_3$. Now θ_1 characterizes the dependence of the pair (u_1,u_2), θ_2 corresponds to the pair (u_3,u_4), and θ_3 is the dependence parameter of the pairs (u_1,u_3), (u_1,u_4), (u_2,u_3) and (u_2,u_4).

Two issues concerning the use of nested Archimedean copulas arise in practice. The first is the selection of the type of Archimedean copula. The three families mentioned above cover the possible types of dependence one typically encounters in applications. The Clayton family is characterized by a stronger dependence for small values of the random variable than for large realizations and it has positive lower tail dependence. The Gumbel copula implies upper tail dependence and stronger dependence of large values. Finally, the Frank copula is the only Archimedean copula that is rotationally symmetric (Frank 1979) and it has independent tails. In this sense it is similar to a Gaussian copula, although it has slightly lighter tails. Again, the question of which copula should be used is an empirical issue and should be determined by the model fit.

The second issues is the precise choice of the nesting structure and the ordering of the variables. Inspired by Hafner and Rombouts (2007) we suggest the following approach. Start with an arbitrary ordering of the data and estimate the multivariate discrete choice model using the independence copula $C(u_1, \ldots, u_d) = u_1 \cdot u_2 \cdot \ldots u_d$. Denote the resulting marginal probabilities $P_{it} = P(Y_{it} = y_{it}|z_t)$ for $t = 1, \ldots, T$ and $i = 1, \ldots, d$ and compute the Pearson residuals

$$\hat{\varepsilon}_{it} = \frac{y_{it} - P_{it}}{\sqrt{P_{it}(1 - P_{it})}}.$$

For these compute the Spearman rank correlation matrix. As the rank correlation is a copula based dependence measure this should give a good impression about the dependence structure of the true error terms ε_{it}. The estimated rank correlation matrix can then be used to decide upon a useful nesting structure and ordering of the variables.

2.6 Multivariate dynamic ordered discrete choice models

The models suggested above can easily be extended to allow for more than two outcomes, therefore generalizing ordered discrete choice models such as the ordered logit and probit models. Now we consider a vector of observed outcomes y_t that takes on the J distinct outcomes $j = 1, \ldots, J$. The latent variable y_t^* is again defined by equation (1) and (2). However, it is unclear how to include the past observed variable in the model. One possibility would to be include the indicator $y_{t-1} > 1$ whether the observed outcome was larger than one, but this depends strongly on the application, in particular on the number of outcomes that are possible. Alternatively, one can remove the lagged indicators from the model. Additionally, define the thresholds $\mu_{i,j}$ such that the observable outcomes relate to the latent variable as

$$y_{it} = j \text{ if } \mu_{i,j-1} < y_{it}^* < \mu_{i,j}, \tag{18}$$

with $\mu_{i,0} = -\infty$, $\mu_{i,1} = 0$ and $\mu_{i,J} = \infty$. The probabilities for the possible outcomes are computed similarly as for the binary case. For the bivariate model we have

$$\begin{aligned} P(Y_{1t} = j, Y_{2t} = k) = {} & C(F_1(\mu_{1,j} - z_t\gamma_1), F_2(\mu_{2,k} - z_t\gamma_2)) - C(F_1(\mu_{1,j-1} - z_t\gamma_1), F_2(\mu_{2,k} - z_t\gamma_2)) \\ & - C(F_1(\mu_{1,j} - z_t\gamma_1), F_2(\mu_{2,k-1} - z_t\gamma_2)) + C(F_1(\mu_{1,j-1} - z_t\gamma_1), F_2(\mu_{1,k-1} - z_t\gamma_2)). \end{aligned}$$

Again, for some probabilities this expression simplifies as several terms drop out. For larger dimensions the corresponding formula for multivariate interval probabilities given in equation (10) have to be used where now the $\mu_{i,j}$ determine the boundaries. The log-likelihood is easily obtained by multiplying indicators for the actual outcomes with their corresponding probabilities.

A notable possibility that arises is that one can combine, in a single multivariate model, variables with binary outcomes and variables with ordered multiple outcomes. This may prove useful in applications.

3 Application

In this section we illustrate the proposed DCMDC model by two applications to real data. In Section 3.1 we consider a large data set of Australian intra-day electricity spot prices and we model the probability of extreme price occurrences, so called spikes, across four markets. Univariate treatments of this problem can be found in, e.g., Christensen et al. (2012), Eichler et al. (2014) and Hurn et al. (2014). To our knowledge, the only study considering the multivariate case is Clements et al. (2015), which relies on a bivariate self-exciting point process model to analyze inter-regional links between the probability of spike occurrence.

The second application presented in Section 3.2 considers the problem of modeling and forecasting recession probabilities as studied in, e.g., Kauppi and Saikkonen (2008) and Nyberg (2013).

3.1 Dynamics in the dependence of extreme price occurrence between real-time electricity markets

Understanding the co-dependence of spikes in real-time electricity prices between interconnected markets plays a crucial role in risk-management. Furthermore it is of great importance when pricing and hedging inter-regional spreads and/or to value interconnectors between these markets. In this context the Australian power exchange provides an ideal framework. It consists of five physically interconnected markets for which individual electricity prices are settled in a continuous trading scheme. Nowadays more than A$10 billion worth of electricity is traded annually. In each market, cap products are traded based on half-hourly electricity prices for which inter-regional spreads can be priced once the co-dependence of extremes is properly analyzed. Insights concerning the dynamics and dependence of co-spikes, i.e. simultaneous spikes in different markets, will further help to price and hedge Settlements Residue Auction products that are available in the

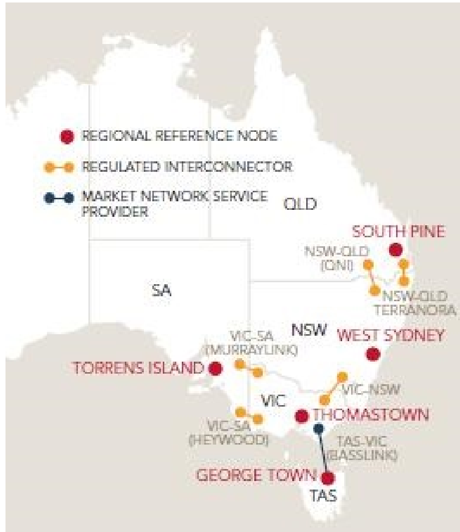

Figure 2: Different regions forming the NEM. The figure was taken from AEMO (2010).

Australian market. These products give the owner a share of the surplus generated on interconnectors which transfer electricity between two regions. While univariate models for spikes have recently been proposed for intradaily data in these markets, the nature of co-spikes remains still unexplored in the existing literature. Therefore, we address this topic and analyze the occurrence of such co-spikes and their dependence by using a variation of the DCMDC model.

3.1.1 Data description

Our data set consists of half-hourly spot prices, i.e. the highest frequency freely available, for the four main Australian markets Victoria (VIC), New South Wales (NSW), Queensland (QLD) and South Australia (SA). The region Tasmania was omitted as it is assumed to have a minor role being the smallest of the 5 National Electricity Market (NEM) regions. Figure 2 shows the regions in the market and the transmission lines that connect them. We use data between January 1, 2008 and December 31, 2012 for our analysis. This results in a total of 87,695 half-hourly observations. The reason to exclude earlier data is that in 2007 the occurrence of spikes was extremely high due to a "millennial drought", which not only reduced the amount of water available for hydro generation, but also limited the cooling water available for thermal (coal- and gas-fired) generators. This resulted in noticeably higher wholesale electricity prices as the cost of supply increased and the mix of generation sources changed.

Following Christensen et al. (2012), we define prices exceeding a threshold of A$100 per MWh as spikes. Although this choice appears somewhat arbitrary, the threshold of A$100/MWh is widely accepted by market participants (see Christensen et al. 2012). Other definitions of spikes based on statistical arguments have been proposed, by, e.g., Chan et al. (2008), Korniichuk (2012), Janczura et al. (2013) and Eichler and Türk (2013). Here we choose to use the concept of economic price spikes, in contrast to a statistical definition of price spikes. Our reasoning is that the latter is important and often used when it comes to disentangling the continuous from the discontinuous part of a time series. The former concept of economic spikes on the other hand is of importance when market participants can be expected to be interested in the probabilities of prices exceeding a certain threshold in order to adapt their behavior. Over the five year period under analysis there were 757 spikes in VIC, 867 in NSW, 954 in QLD, and 1,140 in SA. These spikes were shown to cluster strongly in time by, e.g., Eichler et al. (2014) who apply univariate dynamic binary choice models to analyze spike occurrences.

Table 2 presents some descriptive statistics for the occurrence of pairwise co-spikes. Directly interconnected markets are highlighted with bold letters. They exhibit a larger number of co-spikes than markets that are not directly connected. Among the not directly connected pairs NSW-SA can be seen to be more interdependent than the pairs VIC-QLD and QLD-SA. This is likely due to the physical proximity of the two markets and their indirect connection via VIC, see Figure 2. These results indicate that co-spikes are more likely between directly interconnected or locally close markets. A similar pattern can by seen when looking at the rank correlations between the sizes of the spikes.[5] The co-spikes of the directly interconnected regions, VIC-NSW, VIC-SA and NSW-QLD are characterized by a stronger dependence than those between the remaining markets. This finding indicates that interconnected markets exhibit stronger dependence in the magnitude of spikes than others.

Finally, the table reports the probabilities of observing a spike in one market conditional on a spike in another market during the same period and of observing a spike in one market conditional on a spike in another market during the previous period. As expected, these probabilities are larger for directly interconnected markets. Furthermore, it can be expected that the conditional probability of observing a spike in a small market (such as SA) conditional on observing a spike in a bigger directly or indirectly connected market should be greater than the probability of the larger market spiking conditional on a spike in the smaller market. The fact that $P(S_{\text{SA}}|S_{\text{VIC}})$, $P(S_{\text{SA}}|S_{\text{QLD}})$ and $P(S_{\text{SA}}|S_{\text{NSW}})$ are,

[5]Due to the fat tails of these extreme electricity prices (see Korniichuk 2012) we rely on Spearman rank correlations instead of linear correlations.

Table 2: Descriptive statistics for co-spikes

	VIC-NSW	VIC-QLD	**VIC-SA**	**NSW-QLD**	NSW-SA	QLD-SA
# of co-spikes	402	288	638	591	361	262
rank corr	0.7837	0.6559	0.8069	0.8837	0.6582	0.6021
$P(S_{2,t}\|S_{1,t})$	0.5310	0.3804	0.8428	0.6817	0.4164	0.2746
$P(S_{1,t}\|S_{2,t})$	0.4637	0.3019	0.5596	0.6195	0.3167	0.2298
$P(S_{2,t}\|S_{1,t-1})$	0.4055	0.3025	0.7001	0.5479	0.3541	0.2327
$P(S_{1,t}\|S_{2,t-1})$	0.3576	0.2442	0.4456	0.4979	0.2518	0.1886

Note: The table exhibits the number of co-spikes, the Spearman rank correlations, contemporaneous and lagged conditional probabilities for all six market combinations. Spikes are defined as prices greater A$100. The period under consideration goes from 01.01.2008 to 31.12.2012. Directly interconnected markets are indicated through the use of bold letters in row 1.

respectively, greater than $P(S_{\text{VIC}}|S_{\text{SA}})$, $P(S_{\text{QLD}}|S_{\text{SA}})$ and $P(S_{\text{NSW}}|S_{\text{SA}})$ for both contemporaneous and lagged conditional probabilties supports this assumption. It can also be seen that the conditional probability of a spike occurrence in SA is greatest for the directly interconnected market VIC, followed by the physically close market NSW which again exhibits higher values than the conditional probability to observe a spike in SA when a spike in QLD has occurred. For the conditional probabilities of the three larger markets, NSW, VIC and QLD, note that $P(S_{\text{NSW}}|S_{\text{VIC}})$ and $P(S_{\text{QLD}}|S_{\text{VIC}})$ are higher than the respective reverse conditional probabilities. Concerning the relationship between NSW and QLD no clear pattern can be observed. This is particularly interesting when considering that QLD and VIC are net exporters while NSW is a net importer of electricity. Nevertheless, we observe that for some pairs the dependence actually appears to be asymmetric. This may have been expected and is further in line with the findings of Lindstroem and Regland (2012) who also documented asymmetries in co-spike occurrences for daily spot prices between European electricity markets.

3.1.2 Model specification

Based on our findings in Section 3.1.1 and on Eichler et al. (2014), where univariate spike probabilities are modeled, we formulate the following model for the multivariate spike probabilities:

$$\pi_t = \alpha + Bl_t + \Gamma\pi_{t-1} + D_1 y_{t-1} + e^{\text{diag}(-D_3 d_{t-t_n})} D_2. \tag{19}$$

Here $y_t = (y_{VIC,t}, y_{NSW,t}, y_{QLD,t}, y_{SA,t})'$ is the vector of observed spikes, $\pi_t = (\pi_{VIC,t}, \pi_{NSW,t}, \pi_{QLD,t}, \pi_{SA,t})'$ the vector latent process driving the spike probabilities, and α is a (4×1) column vector of constants. The vector of loads, $l_t = (l_{VIC,t}, l_{NSW,t}, l_{QLD,t}, l_{SA,t})'$, represents demand for electricity. Finally, d_{t-t_n} is the (4×1) vector of durations between the last spike (denoted by t_n) and t in each market. Note that this model is a variation of model (2) introduced in Section 2.1, with π_t now being allowed to nonlinearly depend on past information through the last term in equation (19).

The matrices B, Γ, D_1, and D_3 are (4×4) coefficient matrices and D_2 is a (4×1) vector. Both B and D_3 are restricted to be diagonal. Thus electricity loads and durations are only allowed to affect the spike probabilities of the corresponding markets. The choice to include this last term is based on the findings of Eichler et al. (2014), where it was introduced in order to model an exponential decay in the probability of spike occurrence as a function of the duration which is given by d_{t-t_n}. It therefore resembles the dynamic structure of a Hawkes process. As this last term, i.e. the dynamic Hawkes term, already captures the effect of past spike occurrences in the corresponding market, we restrict the main diagonal of matrix D_1 to be zero in order to prevent multicollinearity. Furthermore the off-diagonal elements of D_1 are only allowed to differ from zero for markets that are directly connected to the corresponding market.

We consider two further restrictions on the coefficient matrix Γ in (19). First we restrict this matrix to be diagonal, in which case the evolution of π_t is described by four univariate models. Note that the fact that the matrix D_1 is not diagonal implies that spillover effects are nonetheless possible through the inclusion of lags of y_t from neighboring markets. The second specification leaves the coefficients corresponding to directly connected markets unrestricted, so that the model also allows for spillover effects via π_{t-1}.

The only part left to specify is the distribution of the error term ε_t in equation (1). For the marginal distribution we chose the flexible Burr-10 distribution given in equation (7) implying the scobit model of Nagler (1994). For the dependence between the innovations we consider two choices. The first one is the independence copula. The second one is the nested Archimedean copula from Section 2.5. In order to decide on the ordering of the variables and the precise nesting structure we use the approach suggested in that section and compute the rank correlation of the Pearson residuals based on the independence copula and using the second (more general) specification for Γ, to be found in Table 3. It can be seen, that the Spearman correlation between directly connected markets is higher than the one between the remaining market combinations. We decide to

Table 3: Spearman correlation for residuals

	VIC-NSW	VIC-QLD	**VIC-SA**	**NSW-QLD**	NSW-SA	QLD-SA
correlation	0.8722	0.7744	0.8849	0.8507	0.7837	0.6951

Note: The table exhibits Spearman correlations for residuals for all six market combinations. These residuals resulted from applying model (19) using the Independence copula and the Scobit approach for the link functions. The period under consideration goes from 01.01.2008 to 31.12.2011. Directly interconnected markets are indicated through the use of bold letters in row 1.

proceed using $C(u_{VIC}, u_{NSW}, u_{QLD}, u_{SA}) = C_{\theta_3}(C_{\theta_1}(u_{VIC}, u_{SA}), C_{\theta_2}(u_{NSW}, u_{QLD}))$ with $\theta_1, \theta_2 > \theta_3$ as the nesting structure. The alternative nesting $C(u_{VIC}, u_{NSW}, u_{QLD}, u_{SA}) = C_{\theta_3}(C_{\theta_2}(C_{\theta_1}(u_{VIC}, u_{SA}), u_{NSW}), u_{QLD})$ with $\theta_1 > \theta_2 > \theta_3$ also seems plausible, but resulted in an inferior fit. Finally, the chosen copula family is the Gumbel copula, which performed better in terms of the log-likelihood than alternative specifications that we considered. Note that it would be desirable to consider a Gaussian copula or even the standard multivariate probit model as a benchmark. However, estimating the proposed model takes more than one day. This implies that using the Gaussian copula for estimation can be expected to take several years as evaluating the CDF of the multivariate normal distribution is computationally costly in four dimensions and for such a large sample size.

To summarize, based on equation (19) we have a total of three models which we will compare considering their ability to model and predict spike occurrences, namely the model with diagonal Γ matrix and an independence copula (i.e. univariate models), the model allowing for spillovers via π_{t-1} but assuming an independence copula and the same model but assuming a nested Gumbel copula. Note that the 'univariate' specification includes past information from neighboring markets. We have decided to allow for this because then all three models are based on the same information set and the comparison between them can be considered to be fair.

3.1.3 In-sample results

For our analysis we split the sample into an in-sample period to which we fit the model (2008-2011) and an out-of-sample period (2012) that we use to evaluate the forecasting performance of our model.

The estimation results can be found in Table 4. We only report the parameter estimates for the univariate and the Gumbel model, as the parameter estimates for the model

Table 4: Parameter estimates for price spike models

	Univariate				Gumbel			
	VIC	NSW	QLD	SA	VIC	NSW	QLD	SA
constant	-5.3902	-5.2451	-4.9162	-8.4559	-5.8689	-6.6440	-5.6830	-6.5014
$\pi_{VIC,t-1}$	0.2168				-0.1480	-0.0357		0.1125
$\pi_{NSW,t-1}$		0.1950			0.0304	-0.1139	0.1724	
$\pi_{QLD,t-1}$			0.3949			0.1481	-0.0355	
$\pi_{SA,t-1}$				-0.2782	0.0619			-0.2169
D_2	4.8110	9.0512	7.1481	7.1043	9.9913	9.9951	7.6561	9.9783
D_3	0.5199	0.8650	0.6095	0.5331	0.5982	0.7154	0.6696	0.7959
$y_{VIC,t-1}$		0.8827		1.1288		0.6264		1.5223
$y_{NSW,t-1}$	0.7031		1.1297		0.9951		1.2406	
$y_{QLD,t-1}$		0.6263				0.8217		
$y_{SA,t-1}$	1.8219				1.1953			
Load	1.2283	1.4270	1.0050	1.7314	1.3776	1.8001	1.5556	1.5371
scobit	0.4382	0.4873	0.5532	0.7918	0.2851	0.5082	0.5569	0.4808
$[\theta_1, \theta_2, \theta_3]$					[2.5664,3.0475,1.8472]			

Note: The table contains estimated parameters for the univariate model (left panel) and the DCMDC model, when using the Gumbel copula and $C_{\theta_3}(C_{\theta_1}(u_{VIC}, u_{SA}), C_{\theta_2}(u_{NSW}, u_{QLD}))$ as the nesting structure (right panel). The time-period under consideration is 01.01.2008 to 31.12.2011.

allowing for spillovers but with the independence copula are quite similar to those of the Gumbel model. We do not report standard errors, but note that all parameters were statistically significant at the 5% level.

The parameter estimates are in line with intuition. The loads have a positive effect on spike probabilities. Concerning the coefficients in D_2 and D_3 that characterize the exponentially decaying influence of the durations, we find that the estimated coefficients imply higher spike probabilities for shorter durations capturing the clustering of spikes. The coefficients on the lags of π_t and y_t provide evidence of spillover effects across markets. The interpretation of the individual parameters is difficult due to the interaction of the explanatory variables, resulting in some of the estimated coefficients for π_t to take on negative values.

The parameters of the scobit link function are all smaller than one. This implies that the slope of each individual link function takes on its maximum values at probability levels which are smaller than 0.5. The parameters of the copula indicate strong dependence between the innovations of the model. The strongest remaining dependence is reported between the error terms of NSW and QLD, with $\theta_2 = 3.05$. This value is followed by the dependence between the error terms of VIC and SA with $\theta_1 = 2.57$. The estimated parameter of the connecting copula is equal to $\theta_3 = 1.85$.

The in-sample fit of the models is compared using the Bayesian information criterion (BIC). The values of the BIC are equal to 12,325, 11,978 and 9,279, respectively for the univariate, independence and Gumbel model. Thus we can conclude that the additional flexibility of our model leads to notable improvements in the model fit. In particular, the size of the improvement when allowing for dependent errors is large, implying that the assumption of independence clearly cannot be maintained.

Based on the model that provides the best in-sample fit, i.e. the multivariate specification relying on the Gumbel copula, we compute the marginal effects caused by any of the lagged binary variables $y_{i,t-1}$. The corresponding results for the average marginal effects are documented in Table 5. The left panel presents average marginal effects that $y_{i,t-1}$ has on the probabilities that one of the four endogenous variables is equal to one, $P(Y_{j,t} = 1)$. It can be seen that the average marginal effect, caused by a market's own lag through the Hawkes term at a duration of one, is between 8.7% for QLD and 19% for VIC. Furthermore, the average marginal effects on neighboring markets are all positive, albeit - with values between 0.3% and 1.3% - far smaller than the average marginal effects caused by their own lagged dependent variables. In accordance with the descriptive statistics, lagged spike occurrences in the more influential market VIC do exert larger average marginal effects on the spike probability in SA than vice versa. Furthermore, lagged spike occurrences in NSW have stronger average marginal effects on QLD and VIC than lagged spikes occurrences in one of these two markets have on NSW. This is in accordance with the fact that NSW is the area with the highest electricity consumption while being a net importer and QLD and VIC being net exporters. Knowing that unconditionally $P(S_{i,t}|S_{i,t-1})$ is around 70%, it might be surprising that the largest average marginal effect is about 19%. Nonetheless one has to keep in mind that we are documenting the average marginal effect. This does not necessarily reflect the effect a lagged spike will have when, e.g., loads for the corresponding market are already high, as it is the case during day time when spikes generally occur. The same argument holds for lagged cross effects. For example, on the 4th of January 2008 VIC exhibited prices exceeding the threshold of 100 from 13:30 until 18:00. This occurred in combination with high loads and with spikes occurring in SA between 13:00 and 18:00. For these observations the average marginal effect of $y_{VIC,t-1} = 1$ on the spike probability in VIC at time t is 62%, while the marginal effects of $y_{SA,t-1} = 1$ and $y_{NSW,t-1} = 1$ are about 13% and 8%, respectively. Similar observations can be made for the lagged own and cross effects corresponding to the remaining three markets. This finding is reasonable as a spike in one market at time t can be expected to have larger impact on the probability of observing spikes at $t + 1$ in the same or connected markets when the system as a whole is already under stress.

The right panel of Table 5 reports the average marginal effects that $y_{i,t-1}$ has on the different co-spike probabilities for directly interconnected market pairs, $P(Y_{j,t} = 1, Y_{k,t} = 1)$. The average marginal effects are in accordance with the results from the descriptive data analysis in Section 3.1.1. As to be expected, lagged binary variables that correspond to one of the two markets of interest exhibit higher average marginal effects than those of other markets. Lagged spike occurrences in the more influential market VIC exert larger average marginal effects on the probability of co-spikes between VIC and SA than lagged spike occurrences in SA. Furthermore the marginal effect that a lagged spike occurrence in NSW has on co-spikes in NSW and QLD or NSW and VIC is higher than the corresponding marginal effect that is caused by a lagged spike in QLD or VIC. This can (as already for univariate probabilities documented in the left panel) be attributed to the fact that NSW is the area with the highest electricity consumption while being a net importer, and QLD and VIC being net exporters. We again report the average marginal effect caused by lagged spike occurrence in VIC, SA or NSW for January, 4th 2008 between 13:30 and 18:00. However, this time we look at the effects caused for co-spikes in VIC and SA. It turns out, that the lagged marginal effect caused by $y_{VIC,t-1}$ is 25% while the ones for $y_{SA,t-1}$ and $y_{NSW,t-1}$ are 12% and 7%, respectively. Again, we can state that the size of marginal effects will heavily depend on the overall state of the system.

3.1.4 Out-of-sample results

In order to compare the forecasting performance of the different models we compute McFadden's (1974) pseudo R^2 and the predictive log likelihood (LL) for the overall model, as well as the univariate Cramer (CR) statistic (see Cramer 1999) based on the out-of-sample data and the 1-step predictions of the underlying probabilities. The pseudo R^2 of McFadden (1974) is calculated as:

$$\text{pseudo } R^2 = 1 - LL/LL_0, \tag{20}$$

with LL giving the log-likelihood value that corresponds to the model under consideration, while LL_0 is log-likelihood assuming constant probabilities. The predictive log likelihood is

Table 5: Average marginal effects of lagged spikes

	Univariate probabilities				Co-spike probabilities		
	VIC	NSW	QLD	SA	VIC-NSW	VIC-SA	NSW-QLD
$y_{VIC,t-1}$	0.1909	0.0028	-	0.0126	0.0082	0.0188	0.0006
$y_{NSW,t-1}$	0.0047	0.1385	0.0085	-	0.0088	0.0016	0.0131
$y_{QLD,t-1}$	-	0.0040	0.0866	-	0.0007	-	0.0098
$y_{SA,t-1}$	0.0062	-	-	0.1399	0.0012	0.0094	-

Note: The table contains marginal effects that correspond to the estimated parameters on probabilites for only one of the four edongenous variables to be one (left panel) and for pairs of endongenous variables that belong to connected markets, to be jointly one (right panel). The time-period under consideration is 01.01.2008 to 31.12.2011.

the out-of-sample value of the log-likelihood function and the Cramer statistic is computed as

$$CR_i = \frac{\sum_{t=1}^{T} \hat{P}(y_{it} = 1) 1_{\{y_{it}=1\}}}{\sum_{t=1}^{T} 1_{\{y_{it}=1\}}} - \frac{\sum_{t=1}^{T} \hat{P}(y_{it} = 1) 1_{\{y_{it}=0\}}}{\sum_{t=1}^{T} 1_{\{y_{it}=0\}}}, \tag{21}$$

with $1_{\{y_{it}=1\}} = 1$ if $y_{it} = 1$ and 0 otherwise. The first term is the average of $\hat{P}(y_{it} = 1)$ conditional on $y_{it} = 1$, while the second term gives the average of $\hat{P}(y_t = 1)$ conditional on no spike having occurred at t. This measure heavily penalizes incorrect predictions. Furthermore, because each proportion is taken within the corresponding subsample, it is not unduly influenced by the large proportionate size of the group of more frequent outcomes. Apart from being considered for the marginal spike probabilities for each market separately, the Cramer statistic is further applied to the stacked marginal probabilities in order to evaluate the joint fit of the model. For this purpose we adapt it as

$$CR = \frac{\sum_{i=1}^{d} \sum_{t=1}^{T} \hat{P}(y_{it} = 1) 1_{\{y_{it}=1\}}}{\sum_{i=1}^{d} \sum_{t=1}^{T} 1_{\{y_{it}=1\}}} - \frac{\sum_{i=1}^{d} \sum_{t=1}^{T} \hat{P}(y_{it} = 1) 1_{\{y_{it}=0\}}}{\sum_{i=1}^{d} \sum_{t=1}^{T} 1_{\{y_{it}=0\}}}. \tag{22}$$

The results are documented in Table 6. The pseudo R^2 indicate that the univariate model gives a 50% improvement with respect to a model only including a constant term. When allowing for spillover via π_{t-1} the pseudo R^2 increases to about 55%, whereas the generalization to dependent error terms leads to a further improvement to a value of 62.5%. The predictive (negative) LL statistics give the same ranking as the pseudo R^2. We decided to include them nonetheless in order to give the reader an impression about their absolute magnitude as a measure of the quality of the density forecasts. The overall CR measure,

Table 6: Out-of-sample fit

	R^2	LL	CR	CR_{VIC}	CR_{NSW}	CR_{QLD}	CR_{SA}
Univariate	0.5001	2232	0.3640	0.4638	0.5215	0.2708	0.3209
Independence	0.5496	2011	0.4021	0.4849	0.5188	0.3023	0.3926
Gumbel	0.6250	1674	0.4169	0.4856	0.5170	0.3084	0.4307

Note: The table contains pseudo R^2, predictive log-likelihood and Cramer statistic results for the out-of-sample period from 01.01.2012 to 31.12.2012.

which stacks the univariate probabilities and the observed binary variables for each of the four markets, also ranks the models consistently with their complexity. However, in this case the generalization to dependent errors only results in a minor improvement compared to the improvements yielded for R^2 and predictive LL. This may be due to the fact that this measure only considers univariate probabilities. When looking at the CR for each time series we can see that the same argument as for the overall CR applies. Only for NSW the univariate model appears to perform slightly better than its generalizations. Altogether the results clearly illustrate that for the problem at hand multivariate modeling improves quality of the forecasts.

Next we analyze how well the models perform when forecasting co-events between directly interconnected markets. To be more precise, we evaluate their capability of forecasting probabilities for directly interconnected markets to co-spike. Furthermore, we look at the forecasts of the event that only one of the two markets spikes. Being able to forecast these events is important when pricing inter-regional settlements residue. Reliable forecasts for the probability of a single spike will help speculators to better price their bids for transmission rights in one direction or the other. Quantifying the expected probability of two connected markets to co-spike will allow one to bet on spreads between resulting spikes by bidding on transmission rights in both directions.

We compare the forecasting performance using a variation of the Cramer statistic (CR_{co}). The adaptation is straightforward. An example on how to calculate CR_{co} for the co-event that $y_{it} = 1$ and $y_{kt} = 0$ is given by:

$$CR_{co} = \frac{\sum_{t=1}^{T} \hat{P}(y_{it} = 1, y_{kt} = 0) 1_{\{y_{it}=1, y_{kt}=0\}}}{\sum_{t=1}^{T} 1_{\{y_{it}=1, y_{kt}=0\}}} - \frac{\sum_{t=1}^{T} \hat{P}(y_{it} = 0, y_{kt} = 1) 1_{\{y_{it}=0 \text{ or } y_{kt}=1\}}}{\sum_{t=1}^{T} 1_{\{y_{it}=0 \text{ or } y_{kt}=1\}}}. \tag{23}$$

Here the first term gives the mean probability that the model yields for $\{y_{it} = 1, y_{kt} = 0\}$

Table 7: Out-of-sample Cramer statistic for co-events

	VIC,NSW			VIC,SA			NSW,QLD		
	1,1	1,0	0,1	1,1	1,0	0,1	1,1	1,0	0,1
# of events	106	65	13	161	10	119	91	28	166
Univariate	0.3808	0.3530	0.1278	0.2633	0.1187	0.1378	0.3341	0.1636	0.1981
Independence	0.4089	0.3363	0.1203	0.3474	0.1167	0.1499	0.4148	0.1411	0.1856
Gumbel	0.4281	0.3612	0.1018	0.4352	0.0959	0.1730	0.4651	0.1253	0.2017

Note: Cramer statistic for bivariate coevents concerning directly interconnected markets. The out-of-sample period under consideration is 01.01.2012 to 31.12.2012.

when being calculated only for the observations for which this event really occurs. The second term gives the average probability of $\{y_{it} = 1, y_{kt} = 0\}$ for all t at which this event does not occur.

Table 7 reports the results. Comparing the univariate and the independence model, we can conclude that the independence model performs better at forecasting co-spikes, whereas the univariate model might be slightly more reliable for predictions of the event that a single spike occurs. The Gumbel model in contrast not only outperforms its two competitors for forecasting co-spikes, but also indicates a better performance when forecasting the event that only one of the two markets spikes. In total it gives the best predictions in 6 out of 9 cases. The three cases in which it does not outperform are the ones with very few occurrences. The fact that in these three cases we only have 10, 13 and 28 observations suggests that the rankings are not very reliable. Furthermore, these cases indicate a strong asymmetry concerning the occurrence of spikes in only a single market. A sensible extension of our model may allow for a non-exchangeable dependence structure, i.e. a Copula for which $C(u, v) \neq C(v, u)$. We leave this topic for future research.

Overall, it can be stated that the Gumbel model outperforms its less complex alternatives in terms of univariate spike forecasting, in terms of forecasting specific event combinations and regarding the overall in- and out-of-sample fit.

3.2 Recession dynamics

In this section we consider the problem of modeling the dynamics of business cycle recession and expansion periods. This problem has been addressed in various studies in a univariate context, e.g. Estrella and Mishkin (1998), Nyberg (2010) and Kauppi and Saikkonen (2008). However, the only study considering the problem in a multivariate context is Nyberg (2013) who applies a dynamic bivariate probit model to data for the US and Germany (Ger). We consider the same problem, but extend the analysis to a four dimensional setting additionally including Canada (Can) and the UK.

3.2.1 Data

For the US the commonly used recession indicators by the National Bureau of Economic Research (NBER) are considered. For the remaining countries the turning points provided by the Economic Cycle Research Institute (ECRI)[6] are used. We base our study on monthly data for the period January 1970 until May 2013. Following Nyberg (2013), two exogenous variables are considered, namely stock returns r_t^i on country i's MSCI stock index and long-short terms spread S_t^i. The latter are computed as the difference between the interest rates on 10-year and 3-month government bonds. The interest rate data has been downloaded from the FRED database by the Federal Reserve Bank of St. Louis.

3.2.2 Results

A number of choices need to be made when applying our proposed model to recession dynamics for the US, Canada, Germany and the UK. The first issue is the ordering of the variables concerning the hierarchical modeling with nested Archimedean copulas. Following our suggested approach in Section 2.5 we initially estimated various model specifications using an independence copula and looked at the rank correlation matrix to decide on reasonable orderings. The ordering that turned out to most appropriate $C(u_{US}, u_{Can}, u_{Ger}, u_{UK}) = C_{\theta_3}(C_{\theta_2}(C_{\theta_1}(u_{US}, u_{Can}), u_{UK}), u_{Ger})$. The copula that gave the best fit was the fully nested Gumbel copula. This implies that the first estimated dependence parameter describes the dependence between the innovations for the US and Canada, the second copula parameter the dependence for the pair US-UK and Can-UK and the third parameter the dependence for the remaining pairs. Recall the restriction $\theta_1 \geq \theta_2 \geq \theta_3$. Note that the ordering was robust across different model specifications.

The next issue to be considered is the precise model specification. As pointed out in previous studies, the information on business cycle turning points is only known with a significant delay. Therefore the following models are based on equation (2) without the lagged recession indicator y_{t-1}. We consider five model specifications:

[6]https://www.businesscycle.com/

Model 1: The coefficient matrix Γ is restricted to be diagonal, the matrix B is restricted such that only r_t^i and S_t^i can influence π_{it} and the copula is set to the independence copula. This corresponds to univariate modeling approaches of recession dynamics.

Model 2: The same as Model 1, but allowing for a fully nested Gumbel Copula to capture the dependence between the innovations.

Model 3: The same as Model 2, but the matrix Γ is unrestricted.

Model 4: The same as Model 2, but the matrix B is unrestricted.

Model 5: All parameters are unrestricted.

Besides the restrictions due to the model specification we restrict the matrix Γ to have some 0 entries. The restrictions were based on a general-to-specific model search. We do not report the full estimation results for all 5 model specifications. Note, however, that the model fit in terms of the BIC evolved from Model 1 to Model 5 as $1337 \rightarrow 1210 \rightarrow 1062 \rightarrow 891 \rightarrow 731$. Thus each generalization of the model specification leads to a significantly better model fit. In Table 8 we report the parameter estimates for Model 1 and Model 5. The parameter values of the autoregressive terms show high persistence in the recession regimes for all countries and for both models. The spillover effects via the lags of π_{it} indicate spillovers from the US to the other countries and including the lag of $\pi_{Ger,t}$ helps predicting recessions for the US and the UK (with a negative coefficient). One has to be careful when interpreting the exogenous variables for Model 5, as the regressors are highly correlated, making it hard to separate effects of individual variables. Nevertheless, these results are in line with the results in Nyberg (2013). Finally, two observations can be made concerning the estimated dependence parameters. First, θ_1 is estimated to be quite large, which may partially be explained by the discussion in Section 2.2. Second, the estimated standard errors are extremely large, most notably for θ_2. This is confirmed by large variation of the estimated dependence parameter across the other model specifications. This is not entirely surprising as we try to identify the dependence parameter in a complex model using binary data and, in contrast to the application to electricity price spikes, we have a relatively small sample.

Table 8: Parameter estimates for multivariate recession models

	Model 1				Model 5			
	US	Can	UK	Ger	US	Can	UK	Ger
constant	0.0796	-0.193	-0.0953	-0.0484	0.1657	0.7065	0.7776	0.151
	(0.0431)	(0.0266)	(0.0359)	(0.0218)	(0.0568)	(0.2727)	(0.1023)	(0.0547)
$\pi_{US,t-1}$	0.8771				0.9304	0.5782	0.1672	-0.0409
	(0.0313)				(0.0028)	(0.1379)	(0.0805)	(0.0129)
$\pi_{Can,t-1}$		0.9096				0.9087		
		(0.0091)				(0.0117)		
$\pi_{UK,t-1}$			0.8956				0.9999	
			(0.0332)				(0.0001)	
$\pi_{Ger,t-1}$				0.8652	-0.0738		-0.3251	0.9417
				(0.0125)	(0.0321)		(0.1954)	(0.0018)
Spread US	-0.2588				-0.1916	-0.3766	-0.2994	-0.1434
	(0.0525)				(0.0357)	(0.2388)	(0.0866)	(0.0414)
MSCI US	-0.0953				-0.213	-0.1769	-0.0494	0.0412
	(0.0355)				(0.0664)	(0.1483)	(0.2707)	(0.0361)
Spread Can		-0.1818			0.0751	-0.654	0.1544	0.0149
		(0.0149)			(0.0304)	(0.2262)	(0.1264)	(0.0367)
MSCI Can		-0.0196			0.0273	-0.1367	-0.046	0.0383
		(0.0163)			(0.0375)	(0.0932)	(0.0887)	(0.0358)
Spread UK			-0.0157		-0.0774	-1.0153	-0.1438	-0.0483
			(0.0164)		(0.0197)	(0.3435)	(0.0689)	(0.0217)
MSCI UK			-0.0619		0.0346	0.1795	-0.2741	-0.0367
			(0.0093)		(0.0305)	(0.1483)	(0.1828)	(0.0184)
Spread Ger				-0.1202	-0.3138	0.5676	-0.1115	-0.2103
				(0.0117)	(0.0603)	(0.1928)	(0.0768)	(0.0566)
MSCI Ger				-0.1202	0.068	-0.0077	0.2273	-0.0942
				(0.0282)	(0.0286)	(0.0772)	(0.0695)	(0.0253)
$[\theta_1, \theta_2, \theta_3]$					[9.6634, 2.7936, 1.6685]			
					(3.1127, 20.2940, 0.4045)			

Note: The table contains estimated parameters for the univariate model 1 (left panel) and the multi-avaraite model 5, for which all parameters are unrestricted (left panel). Corresponding standard errors are given in parentheses. The time-period under consideration is January 1970 to May 2013.

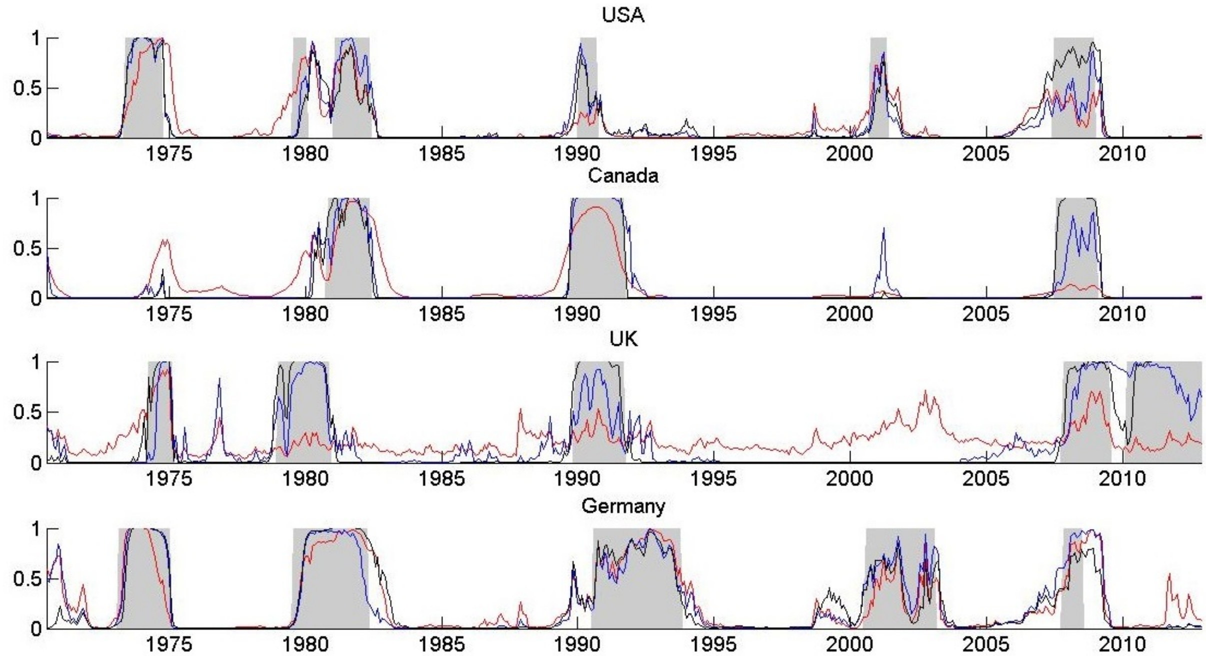

Figure 3: Estimated probabilities of recession based on Model 1 (red), Model 4 (blue) and Model 5 (black) with actual recessions shaded in gray.

In Figure 3 we provide plots of the estimated probabilities of a recession based on Models 1 (red), 4 (blue) and 5 (black) along with the actual recession dates (shaded in grey). One can clearly see that the multivariate models (4 and 5) are superior to the univariate model (Model 1) in capturing the recession regimes. Furthermore, Model 5 allowing for spillovers captures the recession regimes slightly better and does not lead to as many false alarms as the other specifications; see, e.g., the period between the last two recession for the UK.

Overall we conclude that multivariate modeling of recession probabilities is quite beneficial. We can achieve a significantly better fit than using univariate models due to spillovers, dependent innovations, but also by using exogenous information in the form of interest rate spread and stock market returns from other countries. One can therefore expect a better forecasting performance of multivariate models. However, we refrain from performing an out-of-sample evaluation of the model, because we do not have much data and an evaluation of the forecasting performance based on a single recession per country does not appear to be sensible. We leave this issue for future research using a larger cross section of countries to have enough out-of-sample observations to make a credible comparison of competing models.

4 Conclusion

This paper proposes a general model for multivariate binary and ordered outcomes in a time series setting. The model specification allows the researcher to freely choose the link function and copula, therefore nesting a large number of different models. Choosing a copula other than the Gaussian has the further advantage of drastically simplifying the computations needed to evaluate the likelihood function of the model at the price of some restrictions on the dependence structure. This makes the estimation of the model in dimension larger than two feasible even for very large sample sizes.

We discuss the identification of the dependence parameter and find that for small samples one might get estimates on the boundary of the parameter space implying perfect dependence. There is no simple answer on how to treat this issue when it occurs in an application, as one has to decide between a model with perfectly dependent or one with independent innovations. However, confidence intervals for the copula parameter can be computed using a bootstrap. The computation of marginal effects for any probability of interest is straightforward.

The applications of the paper show that one can obtain considerable improvements in model fit and forecasting precision when applying the general multivariate model compared to univariate models or models assuming independent error terms.

Finally, Monte Carlo simulations in the supplementary material to this paper suggest that the identification problem of the dependence parameter does not seem to affect the estimation of the remaining model parameters, while a misspecification of the link function or the copula can lead to biased estimates of marginal effects.

Several issues need to be addressed in future research. First of all, a formal proof for the consistency and asymptotic normality of the maximum likelihood estimator is needed. In particular, this requires the formulation of stationarity conditions for the model. Next, the the development of tools for an impulse response analysis could greatly improve the usefulness and interpretability of the model. Furthermore, the efficiency gains of estimators based on the multivariate model compared to univariate models, similar to the case of seemingly unrelated regressions in the linear case, could be studied. Finally, we adapted the univariate Cramer statistic in an adhoc manner in order to use it as a multivariate fit measure. A more thorough analysis of its properties, as well as the development of further goodness-of-fit measures for multivariate binary choice models, could be of interest.

References

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

AEMO (2010). *An Introduction to Australia's National Electricity Market*. Melbourne: Australian Energy Market Operator.

Albert, A. and J. A. Anderson (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika 71*, 1–10.

Andrews, D. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica 69*, 683–734.

Ashford, J. R. and R. R. Sowden (1970). Multivariate probit analysis. *Biometrics 26*, 535–546.

Butler, J. S. (1996). Estimating the correlation in censored probit models. *The Review of Economics and Statistics 78*, 356–358.

Candelon, B., E.-I. Dumitrescu, C. Hurlin, and F. C. Palm (2013). Multivariate dynamic probit models: An application to financial crises mutation. *Advances in Econometrics 32*, 395–427.

Chan, K. F., P. Gray, and B. Van Campen (2008). A new approach to characterizing and forecasting electricity price volatility. *International Journal of Forecasting 28*, 728–743.

Christensen, T., S. Hurn, and K. Lindsay (2009). It never rains but it pours: Modeling the persistence of spikes in electricity prices. *The Energy Journal 30*, 25–48.

Christensen, T., S. Hurn, and K. Lindsay (2012). Forecasting spikes in electricity prices. *International Journal of Forecasting 24*, 400–411.

Christofides, L. N., T. Stengos, and R. Swidinsky (1997). On the calculation of marginal effects in the bivariate probit model. *Economics Letters 54*, 203–208.

Clements, A. E., J. E. Fuller, and A. S. Hurn (2012). Semi-parametric forecasting of spikes in electricity prices. Working paper 82, NCER Working Paper Series.

Clements, A. E., R. Herrera, and A. S. Hurn (2015). Modeling interregional links in electricity price spikes. Working paper.

Cramer, J. S. (1999). Predictive performance of the binary logit model in unbalanced samples. *Journal of the Royal Statistical Society: Series D (The Statistician) 48*, 85–94.

Czado, C. (2010). Pair-copula constructions of multivariate copulas. In *F. Durante, W. Härdle, P. Jaworki, and T. Rychlik (Eds.), Workshop on Copula Theory and its Applications*.

De Jong, R. M. and T. Woutersen (2011). Dynamic time series binary choice. *Econometric Theory 27*, 673–702.

Dueker, M. J. (1997). Strengthening the case for the yield curve as a predictor of U.S. recessions. *Federal Reserve Bank of St. Louis Review in Business & Finance 2*, 41–51.

Dueker, M. J. (2005). Dynamic forecasts of qualitative variables: A qual VAR model of U.S. recessions. *Journal of Business and Economic Statistics 23*, 96–104.

Eichengreen, B., M. W. Watson, and R. S. Grossman (1985). Bank rate policy under the interwar gold standard: A dynamic probit model. *Economic Journal 95*, 725–745.

Eichler, M., O. Grothe, H. Manner, and D. Türk (2014). Models for short-term forecasting of spike occurrences in australian electricity markets: a comparative study. *The Journal of Energy Markets 7*, 55–81.

Eichler, M. and D. Türk (2013). Fitting semiparametric markov regime-switching models to electricity spot prices. *Energy Economics 36*, 614–624.

Estrella, A. and F. S. Mishkin (1998). Predicting U.S. recessions: Financial variables as leading indicators. *Review of Economics and Statistics 80*, 45–61.

Frank, M. (1979). On the simultaneous associativity of $f(x, y)$ and $x + y - f(x, y)$. *Aequationes mathematicae 19*, 194–226.

Greene, W. H. (1996). Marginal effects in the bivariate probit model. Working paper EC-96-11, NYU.

Greene, W. H. (2011). *Econometric Analysis* (7. ed.). Upper Saddle River, NJ: Prentice Hall.

Hafner, C. and J. Rombouts (2007). Semiparametric multivariate volatility models. *Econometric Theory 23*, 251–280.

Hasebe, T. (2013). Marginal effects of a bivariate binary choice model. *Economics Letters 121*, 298–301.

Hurn, S., A. Silvennoinen, and T. Teräsvirta (2014). A smooth transition logit model of the effects of deregulation in the electricity market. Working paper, School of Economics and Management, University of Aarhus.

Janczura, J., S. Trück, R. Weron, and R. C. Wolff (2013). Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling. *Energy Economics 38*, 96–110.

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

Kauppi, H. and P. Saikkonen (2008). Predicting u.s. receccions with dynamic binary response models. *The Review of Economics and Statistics 90*, 777–791.

Korniichuk, V. (2012). Forecasting extreme electricity spot prices. Working paper, Cologne Graduate School.

Lesaffre, E. and H. Kaufmann (1992). Existence and uniqueness of the maximum likelihood estimator for a multivariate probit model. *Journal of the American Statistical Association 87*, 805–811.

Lindstroem, E. and F. Regland (2012). Modeling extreme dependence between european electricity markets. *Energy Economics 34*, 899–904.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. Academic Press.

Mullahy, J. (2011). Marginal effects in multivariate probit and kindred discrete and count outcome models, with applications in health economics. Working paper w17588, NBER.

Nagler, J. (1994). Scobit: An alternative estimator to logit and probit. *American Journal of Political Science 38*, 230–255.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer.

Nyberg, H. (2010). Dynamic probit models and financial variables in recession forecasting. *Journal of Forecasting 29*, 215–230.

Nyberg, H. (2013). A bivariate autoregressive probit model: business cycle linkages and transmission of recession probabilities. *Macroeconomic Dynamics*.

Okhrin, O., Y. Okhrin, and W. Schmid (2013). On the structure and estimation of hierarchical archimedean copulas. *Journal of Econometrics 173*, 189–204.

Savu, C. and M. Trede (2010). Hierarchical archimedean copulas. *Quantitative Finance 10*, 295–304.

Self, S. G. and K.-Y. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association 82*, 605–610.

Winkelmann, R. (2012). Copula bivariate probit models: with an application to medical expenditures. *Health Economics*.