Modeling and forecasting multivariate electricity price spikes

Hans Manner*[1], Dennis Türk[2], and Michael Eichler[3]

[1]Institute of Econometrics and Statistics, University of Cologne, Germany
[2]RCMA Group Pte Ltd, Singapore
[3]Department of Quantitative Economics, Maastricht University, The Netherlands

Abstract

We consider the problem of forecasting the occurrence of extreme prices in the Australian electricity markets from high frequency spot prices. In particular, we are interested in the simultaneous occurrence of such so-called spikes in two or more markets. Our approach is based on a novel 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 distribution of the latent errors can be chosen freely. Third, the computational burden can be greatly reduced making estimation feasible in higher dimensions and for large samples. The model is applied to spikes in half-hourly electricity prices in four interconnected Australian markets. The multivariate model provides a superior fit compared to single-equation models and generates better forecasts for spike probabilities. Furthermore, evidence of spillover dynamics between the markets is revealed.

Keywords: Electricity price spikes, multivariate binary choice models, copulas, vector autoregression

JEL Classification: C32, C35, C53, Q41, Q47.

*Corresponding author, Meister-Ekkehart-Str. 9, 50937 Cologne, Germany, tel.: +49 221 4704130, email: `manner@statistik.uni-koeln.de`.

1 Introduction

Electricity spot prices are known to be characterized by the occurrence of extreme spikes caused by inelastic demand in case of demand shocks (e.g. extreme weather conditions) or supply disruptions due to outages of generating capacity or transmission lines; see Knittel and Roberts (2005), Cartea and Villaplana (2008), or Lindström et al. (2015). Many electricity markets, e.g. continental European capacity markets, allow separate prices for electricity and capacity reserves. In contrast, markets which exhibit one price only, as is the case in Australia, typically are characterized by heavier and more frequently occurring spikes. Therefore, a proper understanding of spike occurrence is particularly important for such markets.

There are several recent papers which discuss the dependence of electricity prices between markets. For instance, Le Pen and Sevi (2010) use a VAR-BEKK model for a discussion of spill-over effects in both mean and volatility between European electricity forward markets. For the Australian energy markets, which are also the main focus of this paper, Ignatieva and Trueck (2016) fit univariate GARCH model to deseasonalized half-hourly logarithmized spot prices and model the cross-dependence between two markets by copula functions. A similar approach is taken by Aderounmu and Wolff (2014b,a), who focus on the estimation of the tail dependence parameter, and Smith et al. (2012) who model daily Australian spot prices using a skew t-Copula. Finally, Smith and coauthors (Smith et al. 2010, Smith 2015) propose a copula regression model in which also the serial dependence of the electricity demand and spot prices are modelled by copulas. In contrast to the above approaches, which all directly model the dynamics and dependences of the electricity spot prices, we focus in the current paper on the spikes themselves and their dynamics.

The aim of this paper is to propose an econometric model for the prediction of (co-) spike probabilities in four interconnected Australian electricity markets. 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. (2013), Korniichuk (2012), and Eichler et al. (2014). However, as the Australian markets are directly interconnected and many market participants are active in more than one of these markets, a purely univariate treatment of price spikes seems inefficient and inappropriate. The only study deviating from the univariate analysis is the paper by Clements et al. (2015), who treat this problem in a bivariate setting using a self-exciting peaks-over-threshold model.

The first contribution of this paper is to propose the methodology to model and forecast multivariate price spike probabilities. The model we propose extends the dynamic binary choice models tailored to forecast the probability of spike occurrence developed in Eichler et al. (2014). This requires dynamic multivariate discrete choice models.

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 computations feasible in high dimensional models 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 cost, while maintaining a certain degree of flexibility. The representation of the model via the latent processes driving the observed outcomes implies that, in principle, 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.

The second contribution of this paper is to adapt the model to the problem of modeling and forecasting price spike occurrences for half-hourly Australian electricity prices in four interconnected markets. Our model takes information concerning the presence of market interconnectors into account by setting appropriate restrictions for the spillover dynamics between the different markets. Additionally, we allow for an exponentially decaying effect of the duration between subsequent spikes to capture the time dependence in an appropriate way. Our results reveal that multivariate modeling of price spikes greatly improves the model fit. There is clear evidence for dynamic spillovers between different markets. Furthermore, there is strong evidence for contemporaneous dependence between the markets captured via copulas connecting the latent innovations. The improvements in model fit extend to the out-of-sample evaluation of our model where we find that the predictive ability of the model improves upon more restrictive specifications. Both univariate spikes and co-spikes are predicted with higher accuracy.

The paper is structured as follows. In Section 2 we present the general econometric model and elaborate on its precise specification. The data and the empirical application are presented in Section 3, while Section 4 concludes. The appendix contains details on the estimation of the model and the computation of marginal effects.

2 A copula-based multivariate dynamic binary choice 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. The precise choice of the copula function for dimensions larger than two is discussed in Section 2.2. In Appendix A.1 we discuss the estimation of the model and Appendix A.2 treats the computation of marginal effects.

2.1 Model specification

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

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

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

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

with

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

where $\pi_t = (\pi_{1t}, \dots, \pi_{dt})'$, $\alpha = (\alpha_1, \dots, \alpha_d)'$ is a vector of intercepts, B is a $d \times k$ matrix, Γ and D are $d \times d$ matrices, and $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{dt})'$ 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 the model 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. In Section 3 we adapt this model to the specific problem of modeling joint spike probabilities of the interconnected Australian electricity markets by adding a nonlinear term to equation (2) and by placing suitable restrictions on the coefficient matrices Γ and D that take the specific structure of the markets into account. Note that in practice one needs to specify an initial value π_0

of the latent process. We follow Kauppi and Saikkonen (2008) and set π_0 equal to its unconditional mean.

The probabilities of the 2^d possible outcomes at time t depend on the distribution of ε_t and are given by

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

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

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

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) for various examples of copulas. We come back to this issue in Section 2.2. 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

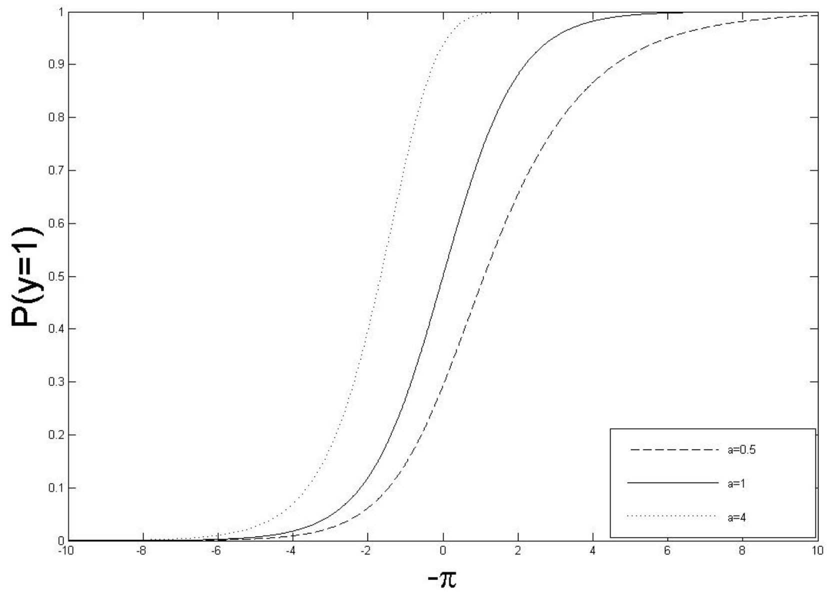

Figure 1: Scobit link function

allowed to be most sensitive to changes in the explanatory variables at probabilities smaller or larger than 0.5. This has the effect that, in contrast to the logit and probit models, marginal effects are maximized at $P \neq 0.5$, i.e. the probability of observing a one is allowed to be most sensitive to changes in the explanatory variables at probabilities smaller or larger than 0.5. Figure 1 shows the scobit link function. The solid line corresponds to the logit model for which $a = 1$.

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 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 are permitted. 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 some situations.

Second, the link function F can be of any form. This allows the use of 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 for 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. In our application below we have $d = 4$ and over 80.000 observations available for which the computations using a Gaussian copula are impossible.

The model can be estimated straightforwardly using maximum likelihood estimation. Details can be found in Appendix A.1. Furthermore, in Appendix A.2 we explain how marginal effects can be computed, which are typically reported due to the fact that the model coefficients cannot be interpreted directly.

2.2 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 too fast as the dimension increases.[2] 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

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

[2] 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.

with $\phi(1) = 0$ and $\phi(0) = \infty$. Then the function

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

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 single dependence parameter determines the dependence between all d variables. However, this class can be extended by relying on a nested dependence structure that allows for $d-1$ distinct generators to construct d-dimensional copulas that only have partial exchangeability. The fully nested Archimedean copula is given by

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

The dependence parameter of generator $j-1$, denoted by 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

$$C(u_1, u_2, u_3) = C_{\theta_2}(C_{\theta_1}(u_1, u_2), u_3), \tag{8}$$

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

$$C(u_1, u_2, u_3, u_4) = C_{\theta_3}(C_{\theta_2}(C_{\theta_1}(u_1, u_2), u_3), u_4), \tag{9}$$

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,

$$C(u_1, u_2, u_3, u_4) = C_{\theta_3}(C_{\theta_1}(u_1, u_2), C_{\theta_2}(u_3, u_4)), \tag{10}$$

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. As these copulas have the same number of parameters a comparison of the log-likelihood statistic is recommended.

The second issue is the precise choice of the nesting structure and the ordering of the variables. In principle one could test all possible orderings and chose the one that results in the highest log-likelihood. However, as there are $d!$ possible orderings this may be computationally very demanding and we propose a more elegant way to address this problem. 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 $P_{it} = P(Y_{it} = y_{it})$ 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. To illustrate this idea, consider an example for $d = 3$ variables and let $\hat{\rho}_{rank}(x, y)$ be the estimated rank correlation between x and y. If, e.g., $\hat{\rho}_{rank}(\hat{\varepsilon}_1, \hat{\varepsilon}_3) = 0.6$, $\hat{\rho}_{rank}(\hat{\varepsilon}_1, \hat{\varepsilon}_2) = 0.4$ and $\hat{\rho}_{rank}(\hat{\varepsilon}_2, \hat{\varepsilon}_3) = 0.35$ then for the hierarchical Archimedean copula in equation (8) one would chose the ordering (y_{1t}, y_{3t}, y_{2t}) because the pair $(\hat{\varepsilon}_1, \hat{\varepsilon}_3)$ has the highest degree of dependence, whereas the other two pairs have a lower degree of dependence that is of similar magnitude.

3 Data and empirical results

In this section we apply the proposed DCMDC model to 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. (2016). 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.[3]

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 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 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

[3]Recall that dynamic multivariate probit models as suggested in Nyberg (2013) and Candelon et al. (2013) cannot be estimated for the problem at hand due to the fact that this would require the evaluations of the CDF of a four-variate Normal distribution for a sample of over 80,000 observations, which is computationally infeasible.

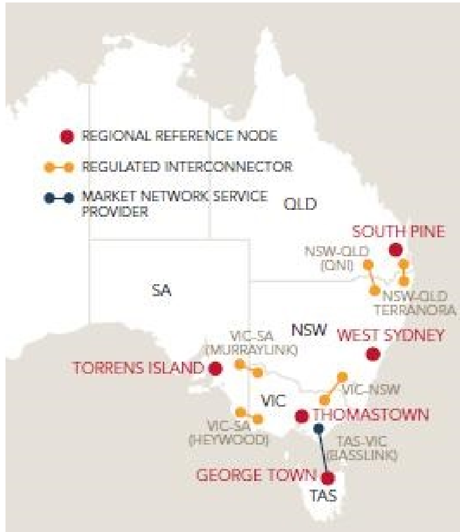

Figure 2: Different regions forming the NEM. The figure was taken from AEMO (2010).

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. Similarly, in January 2013 there have been severe floods in Queensland which led to far more frequent occurrence of spikes than usual.

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.

Formally, let $S_{i,t}$ be a binary variable representing the occurence of a spike in market i and time t. Table 1 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.[4] 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 $P(S_{\mathrm{SA}}|S_{\mathrm{VIC}})$, $P(S_{\mathrm{SA}}|S_{\mathrm{QLD}})$ and $P(S_{\mathrm{SA}}|S_{\mathrm{NSW}})$ are, respectively, greater than $P(S_{\mathrm{VIC}}|S_{\mathrm{SA}})$, $P(S_{\mathrm{QLD}}|S_{\mathrm{SA}})$ and $P(S_{\mathrm{NSW}}|S_{\mathrm{SA}})$ 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 $P(S_{\mathrm{NSW}}|S_{\mathrm{VIC}})$ and $P(S_{\mathrm{QLD}}|S_{\mathrm{VIC}})$ are higher than the respective reverse conditional probabilities. Concerning the relationship between NSW and QLD no

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

Table 1: 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
$P(S_{2,t} \mid S_{1,t})$	0.5310	0.3804	0.8428	0.6817	0.4164	0.2746
$P(S_{1,t} \mid S_{2,t})$	0.4637	0.3019	0.5596	0.6195	0.3167	0.2298
$P(S_{2,t} \mid S_{1,t-1})$	0.4055	0.3025	0.7001	0.5479	0.3541	0.2327
$P(S_{1,t} \mid S_{2,t-1})$	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. $S_{1,t}$ denotes the first named market in the pair and $S_{2,t}$ denotes the second named market. 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.

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 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.

3.2 Model specification

Based on our findings in Section 3.1, on Eichler et al. (2014) where univariate spike probabilities are modeled, and on Lindström et al. (2015) who identify drivers for spikes in the Nord Pool market, we formulate the following model for the multivariate spike probabilities:

$$\pi_t = \alpha + Bl_t + \Gamma\pi_{t-1} + D_1 y_{t-1} + e^{\mathrm{diag}(-D_3 d_{t-t_n})} D_2. \tag{11}$$

Here $y_t = (y_{\mathrm{VIC},t}, y_{\mathrm{NSW},t}, y_{\mathrm{QLD},t}, y_{\mathrm{SA},t})'$ is the vector of observed spikes, $\pi_t = (\pi_{\mathrm{VIC},t}, \pi_{\mathrm{NSW},t}, \pi_{\mathrm{QLD},t}, \pi_{\mathrm{SA},t})'$ the vector latent process driving the spike probabilities, and α is a (4×1) column vector of constants. The vector of loads, $l_t = (l_{\mathrm{VIC},t}, l_{\mathrm{NSW},t}, l_{\mathrm{QLD},t}, l_{\mathrm{SA},t})'$, represents demand for electricity.[5] 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 (11).

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; see Hawkes (1971). The coefficients in D_2 measure the impact that newly occurring spikes have on the probability of further spikes, whereas the coefficients in D_3 measure the speed of the exponential decays in the absence of further spikes. 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 (11). First we restrict this matrix to be diagonal, in which case the evolution of π_t is described by four univariate models. We term this the 'single equation model' as it can be estimated one equation at a time. 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 more general specification, corresponding to the second possible restriction, 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 2.2. In order to decide on the ordering of the variables and the precise nesting structure we use the approach suggested

[5] Here we use the actual loads instead of predictions as argued by Christensen et al. (2012) because half-hourly load forecasts are typically very precise and the authors show that using true instead of predicted loads does not alter the results.

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 2. 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 proceed using $C(u_{\mathrm{VIC}}, u_{\mathrm{NSW}}, u_{\mathrm{QLD}}, u_{\mathrm{SA}}) = C_{\theta_3}(C_{\theta_1}(u_{\mathrm{VIC}}, u_{\mathrm{SA}}), C_{\theta_2}(u_{\mathrm{NSW}}, u_{\mathrm{QLD}}))$ with $\theta_1, \theta_2 > \theta_3$ as the nesting structure. The alternative nesting $C(u_{\mathrm{VIC}}, u_{\mathrm{NSW}}, u_{\mathrm{QLD}}, u_{\mathrm{SA}}) = C_{\theta_3}(C_{\theta_2}(C_{\theta_1}(u_{\mathrm{VIC}}, u_{\mathrm{SA}}), u_{\mathrm{NSW}}), u_{\mathrm{QLD}})$ 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. This implies a larger degree of dependence for large values of the innovations than for small ones. This model is termed 'fully multivariate' model as it allows for dependence between the innovations and allows for spillovers due to unrestricted coefficients in the coefficient matrices Γ and D_1 for directly connected markets. 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 based on Archimedean copulas 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 very costly in four dimensions and for such a large sample size.

To summarize, based on equation (11) 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. single equation model), the model allowing for spillovers via π_{t-1} but assuming an independence copula (intermediate model) and the same model but assuming a nested Gumbel copula (fully multivariate model). Note that the 'single equation' 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.

3.3 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 3. We only report the parameter es-

Table 2: 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 (11) 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.

timates for the 'single equation' and the 'fully multivariate' models, as the parameter estimates for the intermediate model allowing for spillovers but with the independence copula are quite similar to those of the most complex model. Standard errors computed using a robust 'sandwich' formula are given in parentheses.

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. Some estimates have high standard errors that can be explained by the high correlation between the regressors and resulting multicollinearity.

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; compare Figure 1 for the corresponding shape of the link function. The parameters of the copula indicate strong dependence between the innovations of the model. The strongest 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 'single equation', 'intermediate' and 'fully multivariate' model. The differences in log-likelihoods, taking the values 5,995, 5,788 and 4,422, respectively, are statistically sig-

Table 3: Parameter estimates for price spike models

	Single equation				Fully multivariate			
	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
	(0.0064)	(0.0141)	(0.0060)	(0.0138)	(1.2817)	(0.5543)	(0.6270)	(1.3838)
$\pi_{\mathrm{VIC},t-1}$	0.2168				-0.1480	-0.0357		0.1125
	(0.0001)				(0.2562)	(0.0711)		(0.1361)
$\pi_{\mathrm{NSW},t-1}$		0.1950			0.0304	-0.1139	0.1724	
		(0.0001)			(0.0153)	(0.0631)	(0.0329)	
$\pi_{\mathrm{QLD},t-1}$			0.3949			0.1481	-0.0355	
			(0.0006)			(0.0558)	(0.0221)	
$\pi_{\mathrm{SA},t-1}$				-0.2782	0.0619			-0.2169
				(0.0002)	(0.0543)			(0.3061)
D_2	4.8110	9.0512	7.1481	7.1043	9.9913	9.9951	7.6561	9.9783
	(6.5214)	(0.0449)	(0.0981)	(5.2847)	(1.6280)	(1.2739)	(2.6854)	(4.5741)
D_3	0.5199	0.8650	0.6095	0.5331	0.5982	0.7154	0.6696	0.7959
	(0.2000)	(0.0064)	(0.0062)	(0.0666)	(0.2083)	(0.1022)	(0.2442)	(0.4803)
$y_{\mathrm{VIC},t-1}$		0.8827		1.1288		0.6264		1.5223
		(0.9185)		(1.6636)		(0.7080)		(0.4914)
$y_{\mathrm{NSW},t-1}$	0.7031		1.1297		0.9951		1.2406	
	(0.2531)		(0.8959)		(0.2705)		(0.3502)	
$y_{\mathrm{QLD},t-1}$		0.6263				0.8217		
		(0.1183)				(0.2805)		
$y_{\mathrm{SA},t-1}$	1.8219				1.1953			
	(1.8180)				(0.3773)			
Load	1.2283	1.4270	1.0050	1.7314	1.3776	1.8001	1.5556	1.5371
	(0.0017)	(0.0013)	(0.0010)	(0.0008)	(0.3265)	(0.2192)	(0.1176)	(0.2430)
scobit	0.4382	0.4873	0.5532	0.7918	0.2851	0.5082	0.5569	0.4808
	(0.0013)	(0.0006)	(0.0002)	(0.0004)	(0.0745)	(0.1031)	(0.3285)	(0.1730)
$[\theta_1, \theta_2, \theta_3]$					[2.5664,3.0475,1.8472]			
					[(0.1983),(0.2934),(0.1088)]			

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

nificant using likelihood ratio tests[6]. 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 fully 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 4. 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 of a market's own lag, 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% (see Table 1), 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 $y_{\mathrm{VIC},t-1}$ on the spike probability in VIC at time t is 62%, while the marginal effects of $y_{\mathrm{SA},t-1}$ and $y_{\mathrm{NSW},t-1}$ 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

[6]Note that the models are nested.

Table 4: Average marginal effects of lagged spikes

	Univariate probabilities				Co-spike probabilities		
	VIC	NSW	QLD	SA	VIC-NSW	VIC-SA	NSW-QLD
$y_{\text{VIC},t-1}$	0.1909	0.0028	-	0.0126	0.0082	0.0188	0.0006
$y_{\text{NSW},t-1}$	0.0047	0.1385	0.0085	-	0.0088	0.0016	0.0131
$y_{\text{QLD},t-1}$	-	0.0040	0.0866	-	0.0007	-	0.0098
$y_{\text{SA},t-1}$	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 endogenous variables to be one (left panel) and for pairs of endogenous 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.

system as a whole is already under stress.

The right panel of Table 4 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. 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 documented for marginal probabilities 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 for co-spikes in VIC and SA. It turns out, that the lagged marginal effect of $y_{\text{VIC},t-1}$ is 25% while the ones for $y_{\text{SA},t-1}$ and $y_{\text{NSW},t-1}$ are 12% and 7%, respectively. Again, we find that the size of marginal effects will heavily depend on the overall state of the system and that average marginal effects alone do not convey the whole picture.

3.4 Out-of-sample results

The predictive ability of our proposed models is compared to a simple benchmark model that is based on the observations that spikes tend to occur in blocks of consecutive spikes. Thus a natural benchmark is the naive model that predicts a spike in market i in period t whenever a spike has occurred in period $t-1$. This model has been used as a benchmark by Clements et al. (2013) and Eichler et al. (2014) where it was shown to provide surprisingly good forecasts. The clear disadvantage is the this model can never predict the first spike in a block, nor the first non-spike after a block. In fact, this model which we term 'naive model' is nested in our specification by setting

$$\pi_{it} = \alpha_i + d_i y_{i,t-1}$$

and assuming an independence copula.

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

$$R^2_{\text{pseudo}} = 1 - L/L_0, \tag{12}$$

with L giving the log-likelihood value that corresponds to the model under consideration, while L_0 is log-likelihood assuming constant probabilities. The predictive log-likelihood is the out-of-sample value of the log-likelihood function and the Cramer statistic is computed as

$$\mathrm{CR}_i = \frac{1}{K_i}\sum_{t=T+1}^{T+K} \hat{P}(y_{it}=1)\, y_{it} - \frac{1}{K-K_i}\sum_{t=T+1}^{T+K} \hat{P}(y_{it}=1)(1-y_{it}), \tag{13}$$

where $K_i = \sum_{t=T+1}^{T+K} y_{it}$ is the number of spike occurrences in market i in the out-of-sample period $\{T+1,\dots,T+K\}$. 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

Table 5: Out-of-sample fit

	R^2	L	$\text{CR}_{stacked}$	CR_{VIC}	CR_{NSW}	CR_{QLD}	CR_{SA}
Naive	0.4723	-2356	0.4599	0.4862	0.5533	0.3384	0.5144
Single equation	0.5001	-2232	0.3640	0.4638	0.5215	0.2708	0.3209
Intermediate	0.5496	-2011	0.4021	0.4849	0.5188	0.3023	0.3926
Fully multivariate	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.

evaluate the joint fit of the model. For this purpose we adapt it as

$$\text{CR}_{stacked} = \frac{\sum_{i=1}^{d}\sum_{t=T+1}^{T+K}\hat{P}(y_{it}=1)\,y_{it}}{\sum_{i=1}^{d}\sum_{t=T+1}^{T+K}y_{it}} - \frac{\sum_{i=1}^{d}\sum_{t=T+1}^{T+K}\hat{P}(y_{it}=1)(1-y_{it})}{K-\sum_{i=1}^{d}\sum_{t=T+1}^{T+K}y_{it}}. \tag{14}$$

The results are documented in Table 5. The values of the pseudo R^2 statistic indicate that the naive model gives a 47% improvement compared to a model only including a constant term. The single equation model obtains a value of 50%. 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) log-likelihood 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.

Looking at the Cramer measure it stands out that the naive model performs surprisingly well and better than the more complex alternatives, which is in line with the findings in Eichler et al. (2014). However, note again the obvious disadvantage of this trivial forecasting approach of not having any value for forecasting the first spike in a sequence of spikes, nor the first non-spike after such a sequence. Focusing on the other models, the overall Cramer measure, 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 pseudo R^2 and predictive log-likelihood. This may be due to the fact that this measure only considers univariate probabilities. When looking at the Cramer statistic for each time series we can see that the same argument as for the overall Cramer statistic applies. Only for

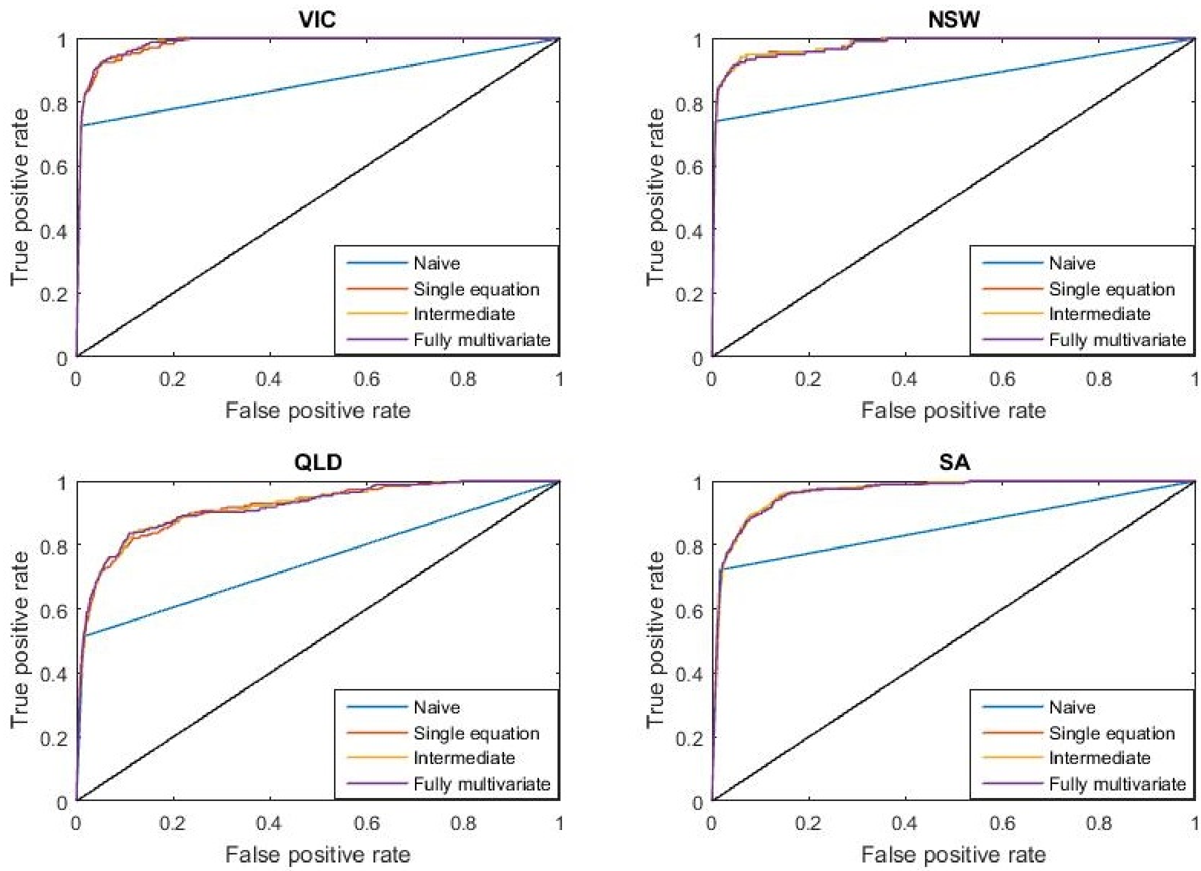

Figure 3: Receiver operating characteristic (ROC)

NSW the single equation 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. A graphical illustration of the predictive ability of the competing models is the receiver operating characteristic (ROC) in Figure 3. It varies the probability threshold τ at which a spike is predicted and plots the false positive rate (FPR) against the true positive rate (TPR) for all $\tau \in (0,1)$. The diagonal line shows the pairs that would arise for completely random guesses and a curve above the diagonal indicates predictions that are better than random guesses. The figure shows our models perform very similarly, whereas the naive model performs worse than the more sophisticated approaches.

As noted above, the naive model is not able to predict the first spike in block of spikes, nor the first non-spike after such a block. In order to assess how our proposed models perform for predicting such events we recomputed the ROC for two subsamples

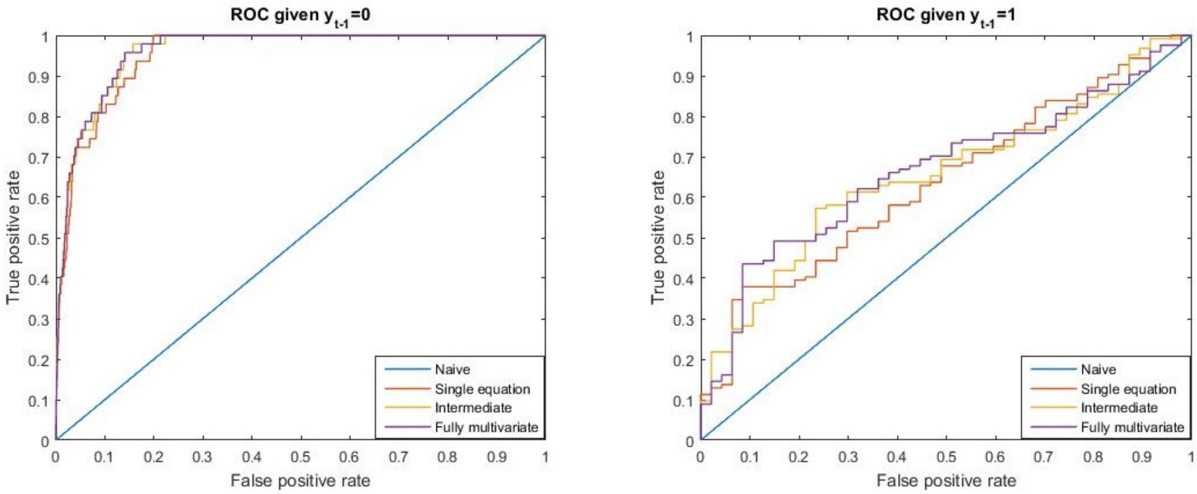

Figure 4: Conditional ROC for VIC

of the out-of-sample data. First, we consider only those periods t for which $y_{t-1} = 0$, i.e. we consider the problem of predicting spikes given that no spikes have occurred in the previous period. Second, we consider the observations for which $y_{t-1} = 1$, in which case a spike has occurred in the previous period and the market has entered a block of spikes. Figure 4 shows the curves for the Victoria (VIC) market. The results for the other markets look similar and are omitted to preserve space. It can be seen that our models perform well for both cases and that the fully multivariate model does slightly better than the less complex models. The high predictive power conditional on $y_{t-1} = 0$ stems from the fact that most non-spikes are predicted correctly and we thus have a fairly low false positive rate. Conditional on being in a block of spikes the predictive power is lower since there are naturally more false positives when a block of spikes ends. Note that the diagonal ROC for the naive model is an artifact from the fact that depending on the chosen probability threshold the FPR and TPR are both either 0 or 1.

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

Table 6: 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
Naive	0.3640	0.3771	0.3992	0.3560	0.1022	0.3794	0.3660	0.3355	0.2562
Single equation	0.3808	0.3530	0.1278	0.2633	0.1187	0.1378	0.3341	0.1636	0.1981
Intermediate	0.4089	0.3363	0.1203	0.3474	0.1167	0.1499	0.4148	0.1411	0.1856
Fully multivariate	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.

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:

$$CR_{co} = \frac{\sum_{t=T+1}^{T+K} \hat{P}(y_{it}=1, y_{kt}=0)\, y_{it}(1-y_{kt})}{\sum_{t=T+1}^{T+K} y_{it}(1-y_{kt})} - \frac{\sum_{t=T+1}^{T+K} \hat{P}(y_{it}=0, y_{kt}=1)(1-(1-y_{it})y_{kt})}{K - \sum_{t=T+1}^{T+K} y_{it}(1-y_{kt})}. \tag{15}$$

Here the first term gives the mean probability that the model yields for $\{y_{it} = 1, y_{kt} = 0\}$ 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 6 reports the results. Comparing the single equation and the intermediate models, we can conclude that the intermediate model performs better at forecasting co-spikes, whereas the single equation model might be slightly more reliable for predictions of the event that a single spike occurs. The fully multivariate model in contrast not only outperforms these two competitors for forecasting co-spikes, but also indicates a better performance when forecasting the event that only one of the two markets spikes. Again the naive model performs rather well for the event that only a single spike occurs, but it is always clearly outperformed by the fully multivariate model for predicting co-spikes. Ignoring the naive model, the fully multivariate model 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 fully multivariate model outperforms its less complex alternatives in terms of marginal spike forecasting, in terms of forecasting specific event combinations and regarding the overall in- and out-of-sample fit. Furthermore, we can conclude that the proposed models offer advantages over the naive forecasting scheme that predicts a spikes whenever a spike in the previous period has occurred.

4 Conclusion

This paper proposes a general model for multivariate binary outcomes in a time series setting, which is adapted to the problem of forecasting multivariate price spikes in Australian electricity prices. 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. Furthermore the computation of marginal effects for any probability of interest is straightforward.

The application of the paper shows that one can obtain considerable improvements in model fit and forecasting precision when applying the general multivariate model compared to single equation models or models assuming independent error terms. The model is applied to the problem of modeling and forecasting price spikes in half-hourly electricity prices in four interconnected markets in Australia. The basic model has been adapted to this specific problem by including an exponentially decaying effect of durations in order to better capture the clustering behavior of spikes and by allowing direct spillover effects only for markets that are directly interconnected. The model is compared to more restricted specifications and our results show that the full model complexity of allowing spillovers and dependent innovations leads to significant improvements in in-sample fit and in terms of forecasting performance. Advantages over a naive forecasting scheme are also pointed out.

Future research could aim at quantifying the potential economic gains from applying this complex model over simpler specifications. Furthermore, the stability of the dependence between the markets over time may be studied considering changes in the market infrastructure such as additional transmission capacities between markets. From a methodological perspective, the development of tools for an impulse response analysis could greatly improve the usefulness and interpretability of the model.

Appendix A Estimation and marginal effects

Appendix A.1 Maximum likelihood estimation

The model introduced in Section 2 can be estimated via maximum likelihood estimation (MLE). As discussed in Nyberg (2013), consistency and asymptotic normality of the MLE can be expected to hold under suitable regularity conditions, but a rigorous treatment of this poses some technical challenges and is beyond the scope of this paper.

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

$$\begin{aligned} L(y_{1t}, y_{2t}) &= P(y_{1t}, y_{2t}|z_t) \\ &= y_{1t}y_{2t} - y_{1t}s_{2t}F_2(-z_t\beta_2) - y_{2t}s_{1t}F_1(-z_t\beta_1) + s_{1t}s_{2t}C_\theta[F_1(-z_t\beta_1), F_2(-z_t\beta_2)]. \end{aligned} \tag{16}$$

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

$$LL(\beta, \theta) = \sum_{t=1}^{T} \ln L(y_{1t}, y_{2t}), \tag{17}$$

where $\beta = (\beta_1, \beta_2)$. 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 the random vector $\mathbf{X}$ to take values in $[\mathbf{a}, \mathbf{b}]$ is given by

$$P_H(B) = P(\mathbf{a} < \mathbf{X} < \mathbf{b}) = \sum_{j_1=1}^{2} \cdots \sum_{j_d=1}^{2} (-1)^{j_1+\ldots+j_d} H(x_{j_1}, \ldots, x_{j_d}), \tag{18}$$

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 $a_i = -z_t\beta_i$ and $b_i = \infty$ for an event occurring in component i, $(y_{it} = 1)$ and $a_i = -\infty$ and $b_i = -z_t\beta_i$ 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 $m = \sum_{i=1}^{d} \mathbb{1}_{\{y_{it}=1\}}$ being the number of occurrences for the probability of interest. The log-likelihood is then computed similarly to the bivariate case as the sum of the probabilities of the observed outcomes.

Here we give detailed expressions for the probabilities entering the likelihood function in three and four dimensions that are more user friendly than the general probabilities in equation (18). We start with an illustration for the three-dimensional case. Consider the vectors (a_1, a_2, a_3) and (b_1, b_2, b_3) with $a_i < b_i$ for all i. Then for the vector of random variables $X = (X_1, X_2, X_3)$ it holds in general that

$$\begin{aligned} P(a_1 < X_1 < b_1, a_2 < X_2 < b_2, a_3 < X_3 < b_3) &= H(b_1, b_2, b_3) - H(a_1, a_2, a_3) + H(b_1, a_2, a_3) \\ &-H(b_1, b_2, a_3) - H(b_1, a_2, b_3) + H(a_1, b_2, a_3) - H(a_1, b_2, b_3) + H(a_1, a_2, b_3). \end{aligned}$$

Since $F_i(-\infty) = 0$ and $F_i(\infty) = 1$ a number of terms drop out or simplify when evaluating the nine probabilities of interest. Using the shorthand notation $F_i = F_i(-z_t\beta_i)$ for $i = 1, 2, 3$ this leads to the following expressions for the probabilities.

$$\begin{aligned} P_t(1, 1, 1) &= 1 - C_\theta(F_1, F_2, F_3) + C_\theta(1, F_2, F_3) + C_\theta(F_1, 1, F_3) \\ &\quad + C_\theta(F_1, F_2, 1) - F_1 - F_2 - F_3 \\ P_t(1, 1, 0) &= F_3 - C_\theta(1, F_2, F_3) - C_\theta(F_1, 1, F_3) + C_\theta(F_1, F_2, F_3) \\ P_t(1, 0, 1) &= F_2 - C_\theta(1, F_2, F_3) - C_\theta(F_1, F_2, 1) + C_\theta(F_1, F_2, F_3) \\ P_t(0, 1, 1) &= F_1 - C_\theta(F_1, 1, F_3) - C_\theta(F_1, F_2, 1) + C_\theta(F_1, F_2, F_3) \\ P_t(1, 0, 0) &= C_\theta(1, F_2, F_3) - C_\theta(F_1, F_2, F_3) \\ P_t(0, 1, 0) &= C_\theta(F_1, 1, F_3) - C_\theta(F_1, F_2, F_3) \\ P_t(0, 0, 1) &= C_\theta(F_1, F_2, 1) - C_\theta(F_1, F_2, F_3) \\ P_t(0, 0, 0) &= C_\theta(F_1, F_2, F_3) \end{aligned}$$

The log-likelihood is obtained by multiplying the logs of these probabilities with indicators for the actual observations.

Using the same notation as for the three dimensional case, but omitting the dependence of the copula on θ we get the following statements for the 16 probabilities of interest in the four-dimensional case. Note that when the bivariate or trivariate margins of the copula are known most of the terms can be replaced by those. This may lead to a slight decrease in computation times in practice.

$$\begin{aligned}
P_t(1,1,1,1) &= 1 + C(F_1,F_2,F_3,F_4) - C(1,F_2,F_3,F_4) - C(F_1,1,F_3,F_4) - C(F_1,F_2,1,F_4) \\
&\quad - C(F_1,F_2,F_3,1) + C(1,1,F_3,F_4) + C(1,F_2,1,F_4) + C(1,F_2,F_3,1) \\
&\quad + C(F_1,1,1,F_4) + C(F_1,1,F_3,1) + C(F_1,F_2,1,1) - F_1 - F_2 - F_3 - F_4 \\
P_t(1,1,1,0) &= F_4 - C(F_1,F_2,F_3,F_4) - C(1,1,F_3,F_4) - C(1,F_2,1,F_4) - C(F_1,1,1,F_4) \\
&\quad + C(1,F_2,F_3,F_4) + C(F_1,1,F_3,F_4) + C(F_1,F_2,1,F_4) \\
P_t(1,1,0,1) &= F_3 - C(F_1,F_2,F_3,F_4) - C(F_1,1,F_3,1) - C(1,F_2,F_3,1) - C(1,1,F_3,F_4) \\
&\quad + C(1,F_2,F_3,F_4) + C(F_1,1,F_3,F_4) + C(F_1,F_2,F_3,1) \\
P_t(1,0,1,1) &= F_2 - C(F_1,F_2,F_3,F_4) - C(F_1,F_2,1,1) - C(1,F_2,F_3,1) - C(1,F_2,1,F_4) \\
&\quad + C(1,F_2,F_3,F_4) + C(F_1,F_2,1,F_4) + C(F_1,F_2,F_3,1) \\
P_t(0,1,1,1) &= F_1 - C(F_1,F_2,F_3,F_4) - C(F_1,F_2,1,1) - C(F_1,1,F_3,1) - C(F_1,1,1,F_4) \\
&\quad + C(F_1,1,F_3,F_4) + C(F_1,F_2,1,F_4) + C(F_1,F_2,F_3,1) \\
P_t(1,1,0,0) &= C(F_1,F_2,F_3,F_4) + C(1,1,F_3,F_4) - C(1,F_2,F_3,F_4) - C(F_1,1,F_3,F_4) \\
P_t(1,0,1,0) &= C(F_1,F_2,F_3,F_4) + C(1,F_2,1,F_4) - C(1,F_2,F_3,F_4) - C(F_1,F_2,1,F_4) \\
P_t(1,0,0,1) &= C(F_1,F_2,F_3,F_4) + C(1,F_2,F_3,1) - C(1,F_2,F_3,F_4) - C(F_1,F_2,F_3,1) \\
P_t(0,1,1,0) &= C(F_1,F_2,F_3,F_4) + C(F_1,1,1,F_4) - C(F_1,1,F_3,F_4) - C(F_1,F_2,1,F_4) \\
P_t(0,1,0,1) &= C(F_1,F_2,F_3,F_4) + C(F_1,1,F_3,1) - C(F_1,1,F_3,F_4) - C(F_1,F_2,F_3,1) \\
P_t(0,0,1,1) &= C(F_1,F_2,F_3,F_4) + C(F_1,F_2,1,1) - C(F_1,F_2,1,F_4) - C(F_1,F_2,F_3,1) \\
P_t(1,0,0,0) &= C(1,F_2,F_3,F_4) - C(F_1,F_2,F_3,F_4) \\
P_t(0,1,0,0) &= C(F_1,1,F_3,F_4) - C(F_1,F_2,F_3,F_4) \\
P_t(0,0,1,0) &= C(F_1,F_2,1,F_4) - C(F_1,F_2,F_3,F_4) \\
P_t(0,0,0,1) &= C(F_1,F_2,F_3,1) - C(F_1,F_2,F_3,F_4) \\
P_t(0,0,0,0) &= C(F_1,F_2,F_3,F_4).
\end{aligned}$$

Appendix A.2 Computation of 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 (16):

$$\begin{aligned}\frac{\partial P(Y_{1t}=y_{1t},Y_{2t}=y_{2t}|z_t)}{\partial z_t} &= y_{1t}s_{2t}f_1(-z_t\beta_1)\beta_1 + y_{2t}s_{1t}f_2(-z_t\beta_2)\beta_2 \\ &- s_{1t}s_{2t}\left[\frac{\partial C_\theta}{\partial u_1}\cdot f_1(-z_t\beta_1)\cdot\beta_1 + \frac{\partial C_\theta}{\partial u_2}\cdot f_2(-z_t\beta_2)\cdot\beta_2\right], \qquad (19)\end{aligned}$$

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 Appendix A.1. We use $(\partial C_\theta)/(\partial u_i)$ instead of $\partial C_\theta(F_1(-z_t\beta_1),F_2(-z_t\beta_2))/\partial F_i(-z_t\beta_i)$ to simplify notation.

For the d-variate case marginal effects for joint conditional probabilities are given as the derivative of (18):

$$\frac{\partial P(Y_{1t}=y_{1t},\ldots,Y_{dt}=y_{dt}|z_t)}{\partial z_t} = \frac{\sum_{j_1=1}^{2}\cdots\sum_{j_d=1}^{2}(-1)^{j_1+\ldots+j_d}\partial H(x_{j_1},\ldots,x_{j_d})}{\partial z_t}, \qquad (20)$$

with the corresponding x_{j_i} given below equation (18). 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 $\sum_{i=1}^{m}\binom{m}{i}(d-m+i)$. 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 $z_{k,t}$.

References

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

Aderounmu, A. and R. Wolff (2014a). Assessing tail dependence in electricity markets. Working paper. Available at `http://dx.doi.org/10.2139/ssrn.2373591`.

Aderounmu, A. and R. Wolff (2014b). Modelling dependence of price spikes in Australian electricity markets. *Energy Risk 11*, 60–65.

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

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

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.

Cartea, Á. and P. Villaplana (2008). Spot price modeling and the valuation of electricity forward contracts: The role of demand and capacity. *Journal of Banking & Finance 32*, 2502–2519.

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 (2013). Semi-parametric forecasting of spikes in electricity prices. *Economic Record 89*, 508–521.

Clements, A. E., R. Herrera, and A. S. Hurn (2015). Modeling interregional links in electricity price spikes. *Energy Economics 51*, 383–393.

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. Springer, Dortrech.*, pp. 93–110.

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.

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.

Hawkes, A. G. (1971). Point Spectra of Some Mutually Exciting Point Processes. *Journal of the Royal Statistical Society B 33(3)*, 438–443.

Hurn, S., A. Silvennoinen, and T. Teräsvirta (2016). A smooth transition logit model of the effects of deregulation in the electricity market. *Journal of Applied Econometrics 31*, forthcoming.

Ignatieva, K. and S. Trueck (2016). Modeling spot price dependence in Australian electricity markets with applications to risk management. *Computers and Operations Research 66*, 415–433.

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. recessions with dynamic binary response models. *The Review of Economics and Statistics 90*, 777–791.

Knittel, C. R. and M. R. Roberts (2005). An empirical examination of restructured electricity prices. *Energy Economics 27*, 792–817.

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

Le Pen, Y. and B. Sevi (2010). Volatility transmission and volatility impulse response functions in European electricity forward markets. *Energy Economics 32*, 758–770.

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

Lindström, E., V. Norn, and H. Madsen (2015). Consumption management in the Nord Pool region: A stability analysis. *Applied Energy 146*, 239–246.

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 18*, 838–862.

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.

Smith, M. S. (2015). Copula modelling of dependence in multivariate time series. *International Journal of Forecasting 31*, 815–833.

Smith, M. S., Q. Gan, and R. J. Kohn (2012). Modelling dependence using skew t copulas: Bayesian inference and applications. *Journal of Applied Econometrics 27*, 500–522.

Smith, M. S., A. Min, C. Almeida, and C. Czado (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association 105*, 1467–1479.

Winkelmann, R. (2012). Copula bivariate probit models: with an application to medical expenditures. *Health Economics 21*, 1444–1455.