Models for short-term forecasting of spike occurrences in Australian electricity markets: a comparative study

Michael Eichler[a], Oliver Grothe[b], Hans Manner[b], Dennis Türk[a,*]

[a]*Department of Quantitative Economics, Maastricht University, P.O.Box 616, 6200 MD Maastricht, The Netherlands*

[b]*Seminar für Wirtschafts- und Sozialstatistik, Universität zu Köln, Albertus-Magnus-Platz, 50923 Köln, Germany*

Abstract

Understanding the dynamics of extreme observations, so called spikes, in real-time electricity prices has a crucial role in risk-management and trading. Yet the contemporaneous literature appears to be at the beginning of understanding the different mechanisms that drive spike probabilities. We reconsider the problem of short-term, i.e., half hourly, forecasts of spike occurrence in the Australian electricity market and develop models, tailored to capture the data properties. These models are variations of a dynamic binary response model, extended to allow for regime specific effects and an asymmetric link function. Furthermore, we study a recently proposed approach based on the autoregressive conditional hazard model. The proposed models use load forecasts and lagged log-prices as exogenous variables. Our in- and out-of-sample results suggest that some specifications dominate and can therefore be recommended for the problem of spike forecasting.

Keywords: intraday electricity spot prices, price spikes, forecasting binary variables, autoregressive conditional hazard, dynamic binary response

1. Introduction

One of the most important stylized features of electricity spot prices – besides mean-reversion and strong seasonality – is the infrequent occurrence of spikes. These are periods of extreme prices that are typically short-lived and during which the spot price exceeds its normal level many times over. As pointed out by, e.g., Knittel and Roberts (2005) and Cartea and Villaplana (2008), spikes occur due to price inelastic demand for electricity when either the demand varies (often as a result of weather conditions) or the supply decreases, e.g., due to outages of generators or transmission lines. These effects can be amplified by generator bidding behavior, to be described later, as the central dispatch process does sometimes require to call generators of high marginal costs into production to satisfy demand. Thomas et al. (2011) state that extreme spikes are even more prevalent in markets with compulsory participation, as is the case in Australia's National Electricity Market (NEM).

Price spikes constitute a major source of risk for market participants such as retailers, who buy electricity for a flexible price from the NEM but sell for a fixed price to consumers. Advanced knowledge about the occurrence of spikes can help to reduce this risk due to demand side participation. Demand side participation corresponds to the situation when consumers reduce their consumption of electricity in response to a change in market conditions, such as high spot prices. For instance, retailers could incur contracts with industrial consumers that allow them to interrupt supply for a certain amount of hours a day whenever prices exceed a specified threshold (see

[☆]We thank Marco Avarucci, Franz Palm, Stephan Smeekes and Francesco Violante for helpful comments. The usual disclaimer applies.

[*]Corresponding author

Email addresses: m.eichler@maastrichtuniversity.nl (Michael Eichler), grothe@statistik.uni-koeln.de (Oliver Grothe), manner@statistik.uni-koeln.de (Hans Manner), d.tuerk@maastrichtuniversity.nl (Dennis Türk)

AEMO, 2010). Furthermore they could close bilateral contracts with producers, paying a premium for producers to insure them against extreme prices. At the same time producers could make use of this knowledge to set strategic bids. Moreover, industries that have direct market access and are able to shift energy consumption over short periods of time such as refrigerated warehouses or aluminium smelters could make use of spike forecasts to reduce spending on electricity. A good overview concerning the different possibilities of demand side participation is given in AEMO (2013). The document also states that (and describes how) AEMO is currently introducing a demand response mechanism and ancillary services unbundling. The demand response mechanism aims at facilitating consumers the decision to continue consumption, or reduce their consumption by a certain amount, for which they would be paid the prevailing spot price. The ancillary services unbundling changes will enable third parties to register and sell frequency control ancillary service using aggregated loads independently of the retailer. In the light of these developments reliable forecasts of spike probabilities could provide valuable information for market participants.

The relevance of modeling price spikes has been acknowledged in the literature, and former studies have already successfully treated the subject of time-varying transition probabilities between extreme and moderate price regimes. For daily spot prices Mount et al. (2006) extended the literature on regime-switching models by allowing the transition probabilities to depend on the load or the implicit reserve margin, i.e. excess supply capacity relative to demand level, using a logistic function. Eichler and Türk (2013) implemented the same idea for a semiparametric framework, leaving the distribution of spikes unspecified. Kanamura and Ohashi (2008) showed that not only the reserve margin, but also the deterministic and stochastic trends of demand fluctuation are important factors to determine the transition probabilities of regime-switching.

Alternatively to these approaches, which jointly model the base and spike regime for spot prices, Lu et al. (2005), Zhao et al. (2007a), Zhao et al. (2007b), Christensen et al. (2009) and Christensen et al. (2012) exclusively focus on modeling spike occurrences. The first three contributions use a statistical approach to identify spikes and then propose data mining techniques combined with market information as, e.g., demand, supply, and net interchange between regions in order to conduct spike forecasting for Australian intraday spot prices. Christensen et al. (2009) propose to identify spikes as threshold exceedances and to subsequently use a Poisson autoregressive framework (further using temperature and loads as exogenous variables) in which price spikes occur as a result of the latent arrival and survival of system stresses. The approach is thus different from former research in two aspects: First the authors use the concept of economic price spikes (or extremes) instead of statistical concepts. 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. Evidence for the importance of such economic spikes is, e.g., given by the existence of cap options which protect their owners against prices exceeding A$300/MWh for the Australian electricity markets. Second, the before mentioned models essentially regard price spikes as a memoryless process with an intensity that is independent of its history. However, there is evidence to suggest that the intensity of the spiking process is neither homogeneous nor driven solely by deterministic factors. The authors thus show that there appears to be a significant historical component which is important in explaining the intensity of the spiking process.

Furthermore Christensen et al. (2012) recently proposed the autoregressive conditional hazard (ACH) model to forecast the occurrence of economic price spikes in the Australian electricity market for a half hourly lead time. A drawback of the ACH model, as applied by the authors, is the implicit assumption that the rate of spike occurrence, i.e., the instantaneous probability of spikes, depends only on exogenous factors, such as loads and temperature. The underlying spike durations on the other hand follow an ARMA type process which without the effect of the exogenous variables would only adapt slowly to changing dynamics in spike occurrences. Although the ACH model is an interesting approach, it is thus not capable of capturing some important features of spike occurrences in electricity spot prices for the Australian markets.

The contribution of the paper at hand is thus to first find characteristic patterns in the occurrence of economic spikes. In particular, we show that periods of extreme prices often occur in blocks of one hour and longer. Furthermore spikes tend to occur in clusters and to reappear after 24 hours. Moreover, such periods are mostly preceded by periods of increasing prices. Second, in order to capture these features, we make use of the recently proposed dynamic logit model by Kauppi and Saikkonen (2008) and tailor this model to the problem at hand. The simplest of the logit models results in forecasting a spike for the next period if we observe a spike in the current period and to not forecast a spike if we observe no spike. This model serves as a baseline model. The remaining models incorporate additional past information, decaying influence of spikes with growing durations and forecasts of electricity loads as exogenous variables, as well as simple regime-switching mechanisms distinguishing sequences of consecutive spikes from more tranquil times. Additionally, we suggest replacing the logistic function by an asymmetric link function and find that this can lead to significant improvements of the model fit. Furthermore, we consider a simple extension of the ACH model employed by Christensen et al. (2012) that includes additional information about the past of the price process and thus considerably improves the performance of the model. The proposed models are compared based on their in-sample and out-of-sample performance and model confidence sets are computed (when sensible) to distinguish the best performing ones.

The rest of the paper is structured as follows. In Section 2 we describe the Australian electricity market and analyze the features of the data. In Section 3 we present the statistical models studied in this paper. Additionally we briefly discuss forecasting electricity loads, which are known to drive a large part of the observed price movements and are required by all but one of the considered models. Section 4 gives a comparative forecast evaluation of the models, while Section 5 concludes.

2. Market structure and data description

In this section, we provide a detailed analysis of the occurrence patterns of price spikes. While most models for electricity spot prices assume that spikes occur more or less randomly over time, our results show that there are clear dependence patterns, which are neglected in these traditional approaches. Similar observations can be found in Christensen et al. (2009, 2012). We start the analysis by a brief description of the structure of the Australian electricity market. More extensive details can be found in Anderson et al. (2007), Chan et al. (2008), or AEMO (2010).

2.1. The Australian electricity market

Since December 1998 the Australian National Electricity Market (NEM) is operating as a wholesale market supplying electricity to retailers and end-users in Queensland (QLD), New South Wales (NSW), Victoria (VIC) and South Australia (SA). Tasmania joined the NEM in 2006. The Snowy region located in southern NSW was initially a separate market in the NEM, but was abolished on the 1st of July 2008 and its generation assets were incorporated into the NSW and VIC regions. Thus, today operations are based in five interconnected regions that largely follow state boundaries. The Australian Electricity Market Operator (AEMO) provides the needed infrastructure, allowing for more than 10 billion Australian dollars worth of electricity to be traded per year.

Exchange between producers and retailers/consumers is facilitated through a pool in which output of all generators is aggregated and scheduled to meet forecasted demand. In contrast to most other electricity markets, the wholesale trading is conducted as a real-time market where supply and demand are instantaneously matched through a centrally coordinated dispatch process. Each of the approximately 260 registered generators is allowed to submit up to ten price-volume combinations for each 5 minute interval of the next day (daily bids) before 12:30 pm. However, as opposed to pure day-ahead markets, the generator has the right to change the corresponding volume for each submitted price (re-bid) up to 5 minutes before execution. This is almost equivalent to a continuous trading setup. From all offers submitted, AEMO determines the generators required to produce electricity based on the principle of meeting prevailing demand in the most cost-efficient way and dispatches them into production. In accordance with the bidding structure such a dispatch price is determined every 5 minutes and represents the cost to supply the last megawatt of electricity to meet demand. The resulting price applies to all generators scheduled into production regardless of the level of their original offer. Six dispatch prices are averaged every half-hour to determine the spot price, yielding 48 trading intervals for each day and region. These spot prices are used as the basis for the settlement of financial transactions for all electricity related products traded in the NEM and are publicly available from AEMO.

Extreme positive price spikes can occur in case that demand exceeds moderate priced supply and offers of high pricing generators apply. This phenomena usually occurs through unexpected increases in demand and when generators or transmission lines fall out (see, e.g., the discussion about price formation in electricity markets in Knittel and Roberts, 2005 and the references therein). Given the quasi-continuous trading in the Australian market, intraday forecasts of price spike probabilities are particularly relevant for market participants.

2.2. Data-analysis

Our data set consists of half-hourly spot prices, i.e., the highest frequency available, for the four main Australian markets Victoria (VIC), New South Wales (NSW), Queensland (QLD) and South Australia (SA) starting 1 January 2003 and ending on 31 December 2012. Data from earlier periods were omitted from the analysis because the four markets became physically interconnected only in the year 2001.

For our analysis, we distinguish between moderate prices, price spikes, and extreme price spikes. Following Christensen et al. (2012), we define prices exceeding a threshold of A\$100/MWh as spikes whereas prices above A\$300/MWh are considered as extreme spikes. The reasoning for using economic price spikes was given in Section 1. Although these choices appear somewhat arbitrary, the lower threshold of A\$100/MWh seems to be widely accepted by market participants (see Christensen et al., 2012) while the higher threshold corresponds to the strike price of heavily traded derivatives. Needless to say that other definitions of spikes based on statistical arguments are possible, see, e.g., Chan et al. (2008), Korniichuk (2012), Janczura et al. (2013) and Eichler and Türk (2013). The last two publications propose approaches for daily data, but techniques suggested could be applied to high frequency data. Note that for convenience we restrict the graphical data analysis to spikes defined as prices greater A\$100/MWh while reporting the modeling results for both thresholds. Negative prices can potentially occur with a floor of -A\$1 000/MWh but are very rare, whereas positive prices are capped at a maximum of A\$12 500/MWh.[1]

Table 1: *Number of extreme prices above A\$100/MWh and above A\$300/MWh*

	VIC		NSW		QLD		SA	
Threshold	A\$100	A\$300	A\$100	A\$300	A\$100	A\$300	A\$100	A\$300
2003	89	18	105	39	81	30	246	18
2004	114	34	335	139	215	67	454	63
2005	117	24	197	68	104	28	337	41
2006	199	48	157	32	141	27	398	62
2007	1453	132	1530	213	1554	160	1442	78
2008	233	21	150	23	195	62	267	78
2009	197	36	304	88	262	42	328	128
2010	101	45	93	52	56	22	158	58
2011	55	11	201	38	184	37	107	29
2012	171	17	119	2	257	36	280	23

Note: The table exhibits the number of spikes (defined as extreme prices above A\$100/MWh and above A\$300/MWh) for each year between 2003 and 2012.

An analysis concerning the number of spikes is given in Table 1. We split up the sample into sub-periods of one year in order to identify changing behavior over time. Comparing the different sub-periods it can be seen that the spike occurrence was exceptionally high in 2007 showing around 7 times the average of the remaining years (being 200 spikes per year) when considering spikes

[1] The maximum price cap was increased from A\$5 000/MWh to A\$10 000/MWh in April 2002 and again from A\$10 000/MWh to A\$12 500/MWh in July 2010.

as prices exceeding A$100/MWh. When looking at spikes exceeding A$300/MWh it can be seen that year 2007 again produces overall higher numbers of exceedances than the remaining years. Furthermore it can be seen that for remaining years the average amount of large spikes per year tends to decline. We also find that prices exceed the threshold of A$100/MWh (excluding year 2007) around 4 times more often than threshold A$300/MWh. This fact indicates that most spikes are between A$100/MWh and A$300/MWh.

Table 2: *Descriptive statistics for extreme prices above A$100/MWh and above A$300/MWh*

	VIC		NSW		QLD		SA	
Threshold	A$100	A$300	A$100	A$300	A$100	A$300	A$100	A$300
Mean	438.53	2173.35	597.25	2177.79	503.70	2197.90	603.58	3277.07
Std dev	1207.76	2608.33	1503.09	2681.60	1236.39	2379.83	1773.55	3675.88
Kurtosis	39.40	5.23	22.88	4.46	26.94	4.21	23.97	2.51
Skewness	5.79	1.79	4.36	1.58	4.73	1.48	4.62	1.06

Note: The table exhibits the mean, standard deviation, kurtosis and skewness of spikes.

Descriptive statistics concerning price spikes are given in Table 2. As for the descriptive statistics no systematic pattern could be identified, we restrain ourselves from presenting a table of sub-periods here to provide a better exposition. Nonetheless it should be stated that the first four central moments appear to be varying over time and that an appropriate model for the distribution of spikes would have to account for this possibility. Thus parametric Markov regime-switching models which are often used to model the occurrence of spikes for daily data might be problematic in this context.

Concerning spikes greater than A$100/MWh the following observations can be made. First, spikes in general have very high means that are far from the chosen threshold of A$100/MWh and thus what could be interpreted as moderate prices. Second, the variation in price spikes is high compared to the mean due to the possibility of prices going up to A$12,500/MWh. This yields standard deviations which in general are around twice the size of the corresponding means. Furthermore, we observe very high kurtosis indicating the possibility of very large prices and, in this context not surprisingly, positive skewness.

When looking at spikes exceeding A$300/MWh it can be stated that first and second central moment are on average more than A$2,000/MWh. This relative increase is in accordance with the fact that only 23% of all spikes are extremes exceeding A$300/MWh. Kurtosis and skewness on the other hand are much closer to normal for this subset of spikes than the respective measures of all observations exceeding A$100/MWh.

In our analysis concerning the occurrence and nature of spikes it stood out that the occurrence of spikes was exceptionally high in 2007. The reason was the "millennial drought" during this period, 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. As this period can be seen as unrepresentative for the rest of the time series we decide to ignore this year and split the data set into two subsamples.[2] The first sample will thus include observations from 1 January 2003 until 31 December 2006 while the second sample will contain data from 1 January 2008 until 31 December 2011. The data for the year 2012 is set aside for the out-of-sample evaluation of our models.

It should be pointed out that the presented ideas in the paper at hand are applicable for any real-time electricity market. We choose the Australian market as its regional spot markets are - in difference to, e.g., the European markets (where only the balancing markets are in real time) - organized in real-time. Further countries that adopted real-time spot markets are, e.g., Singapore and New Zealand.

[2]Initially, the analysis was performed including the year 2007 and our models also performed well for forecasting during this uncommon year.

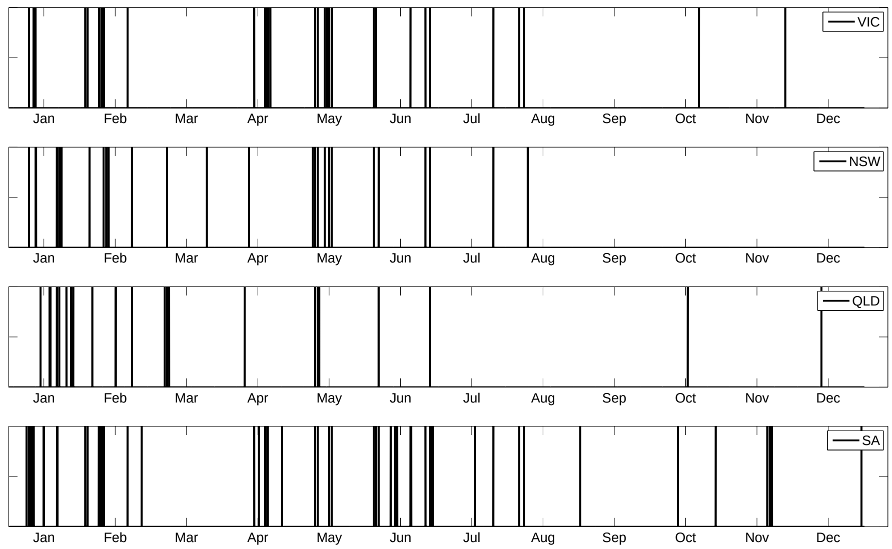

Figure 1: Spike occurrences during the year 2010.

Figure 1 depicts the occurrence of spikes for the year 2010 for all markets.[3] It can be seen from the graph that spikes do not only occur in clusters, i.e., that there are times of high probabilities of spikes and times of low probabilities, but that in fact spikes occur in blocks of varying length, i.e., that there are consecutive prices above the threshold. Figure 2 shows histograms of the block length for all four markets for the period 01.01.2008 to 31.12.2011. The histogram shows that a characteristic property of spikes is that once a spike has occurred it is very likely to be followed by further spikes. This observations is in accordance with what one would expect and a justification of regime-switching models for electricity prices. When demand goes up unexpectedly or supply decreases, e.g., due to outages of generators or transmission lines, extreme prices will occur. These extreme prices will prevail until demand decreases (e.g., due to bilateral agreements between retailers and industrial clients) or until the supply side problem is solved. For daily data such behavior is considered a stylized fact of electricity prices (Janczura and Weron, 2010). For intraday data, however, this has not yet been explicitly acknowledged.

Finally, Figure 3 shows the probabilities of spike occurrences conditional on a previous spike for each of the four electricity markets between 01.01.2008 and 31.12.2011 using a correlogram of spike occurrences. One can see a clear pattern that is identical for all four markets. The pattern is in accordance with our previous finding that spikes are often followed by further spikes. Furthermore, spikes are likely to be followed by spikes at the same time on the following day, confirming the aforementioned stylized facts that daily price spikes tend to occur in blocks as well. One may suspect that due to intraday seasonality spikes always occur at the same time of the day when demand is high. However, Figure 4 showing the distribution of spikes and loads over all half-hour intervals of the week suggests that this is only partly true. Although most spikes occur close to midday and the early evening, one can see that spikes are distributed over most parts of the day except night hours. A complementary explanation might be that shortages can be triggered by reparation or maintenance which might be happening over several days during the same time. Nonetheless it can be seen from the figure that spikes are more likely in times of high demand, i.e., high loads.

Based on previous studies (see e.g. Rambharat et al., 2005; Mount et al., 2006 and Christensen et al., 2012) we use load forecasts as exogenous variables. Loads are available from AEMO on a half hourly basis. Including this variable allows to capture a large part of the seasonal patterns in

[3]For other time periods the graphs look very similar and are available upon request.

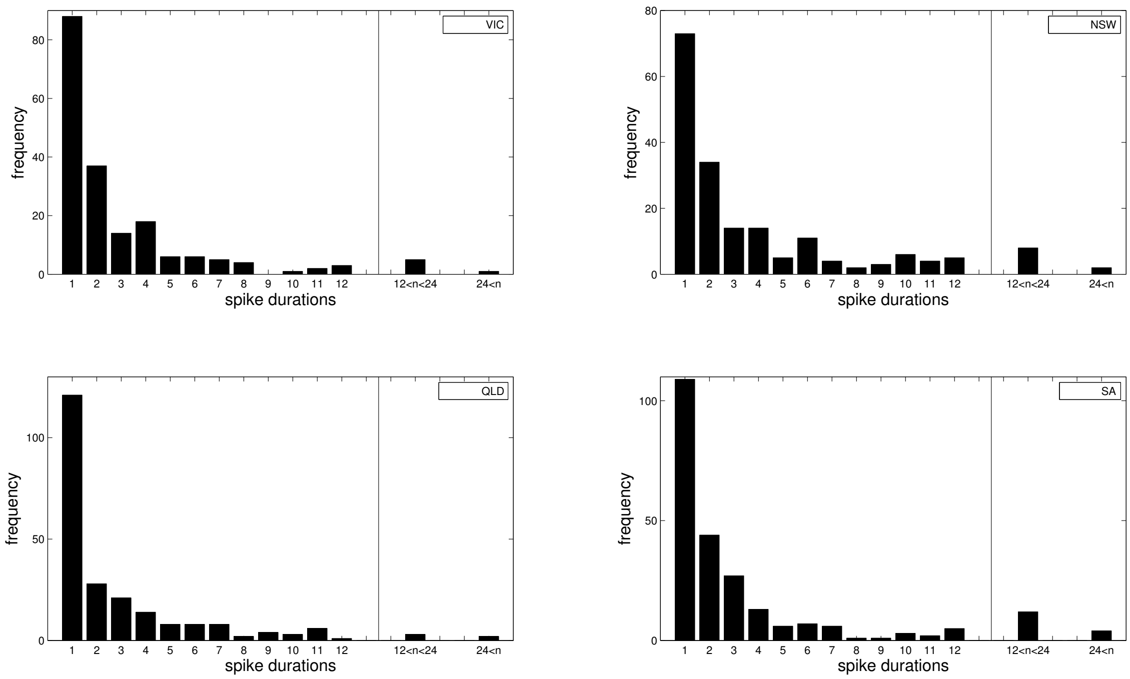

Figure 2: Histograms of the durations of spikes for VIC, NSW, QLD, and SA for the period 01.01.2008 to 31.12.2011. The duration of a spike is given by the number of consecutive spike events.

spike occurrences. This applies to daily and weekly seasonality as well as seasonality over the year. Consequently, directly modeling the seasonality is not necessary and does not lead to better model performance. We initially also considered using temperature variables which can be obtained from the Australian Bureau of Meteorology on a daily basis (see Christensen et al., 2012). As they turned out to be insignificant in most cases and did not improve the forecasting performance of our models we ultimately omitted them from our analysis. AEMO further provides short term projected assessment of system adequacy data (see AEMO, 2012). We also considered the use of these forecasts. In order to do so we replaced half hourly load forecasts with some of the measures provided by AEMO or combinations of them, which give proxies for the reserve margin.[4] Similar ideas on combining reserve capacity, supply and demand can e.g. be found in Lu et al. (2005). Nonetheless none of the used measures yielded a better fit in terms of BIC than half hourly load forecasts. The reason for this slightly worse performance might be the fact that the used measures have forecast horizons of between 16.5 to 40 hours. The data provided by AEMO could thus still be favorable in case researchers want to conduct multi-step forecasting.

3. Models to forecast spikes

In this section we present various approaches for modeling and forecasting spikes in half-hourly electricity spot prices. The models we consider use load forecasts as an exogenous variable. Therefore, we first have to address the problem of forecasting loads. The forecasting method used to produce our half hourly load forecasts is presented in Section 3.1. In Section 3.2 we present variations of a dynamic logit model proposed in Kauppi and Saikkonen (2008) that have been tailored to the problem of short-term spike forecasting by including past price information. Christensen et al. (2012) used a somewhat similar approach, i.e., the ACH model, which we present in Section 3.3.

[4]In order to check for the validity of combining one or several of these measures with our load forecasts we calculated the variance inflation factor for the load forecasts being regressed on a constant and each of the corresponding measures. As the smallest value of the variance inflation factor was still greater 12, indicating multicollinearity, we decided to restrain from using combinations.

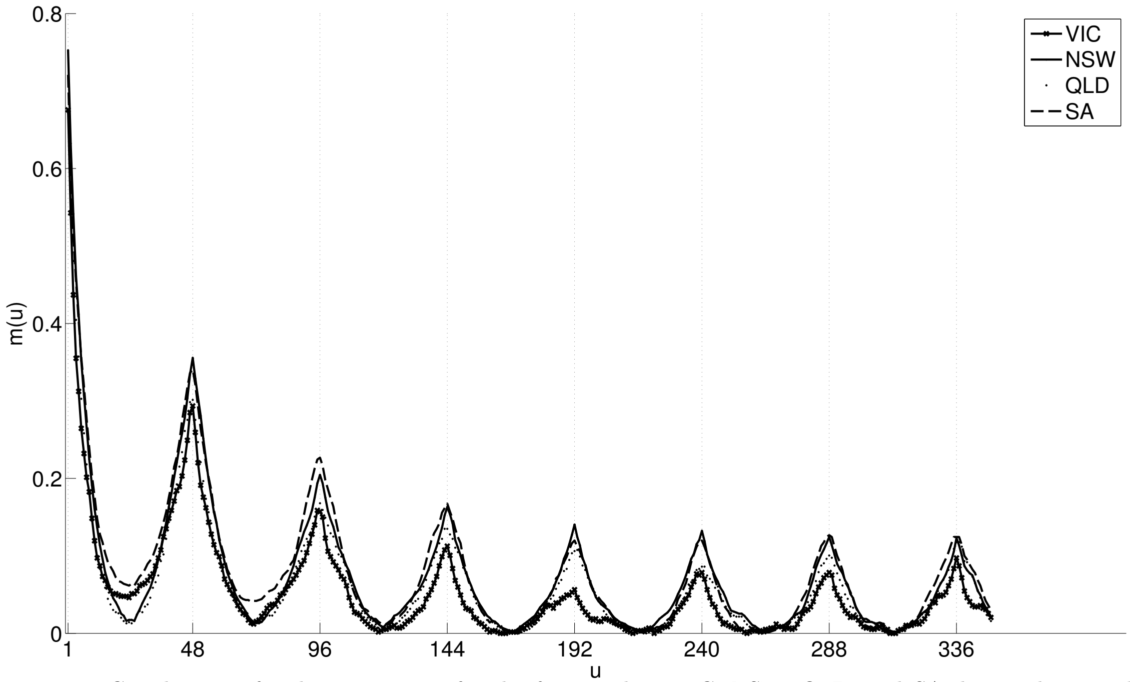

Figure 3: Correlogram of spike occurrences for the four markets VIC, NSW, QLD, and SA during the period 01.01.2008 to 31.12.2011. The correlogram gives the conditional probability $m(u)$ for a spike at time u given that a spike has occurred at time $t = 0$.

In order to make it comparable we further also include information of previous prices as exogenous variables in the ACH model. Note that AEMO reports loads and prices with a delay of less than ten minutes. For this reason it is possible to apply a model for half hourly load forecasts as well as the half hourly spike forecasting models presented in the following.

3.1. Forecasting loads

With exception of the benchmark model, all of the models that we consider in this paper include electricity load forecasts as an exogenous variable since loads are known to drive the movements in the electricity spot price to a large extent. Therefore, application of these models for forecasting the occurrence of spikes requires reliable forecasts of the loads themselves.

Short-term load forecasts are of independent interest as they form the basis for scheduling generation and transmission of electricity (e.g., Taylor et al., 2006). Consequently, a wide range of literature concerning short-term load forecasting has emerged during the last years. Distinct contributions to load forecasting based on univariate methods were made by Taylor (2003) and more recently Taylor (2012), who applied different techniques that were presented in a more general context in Taylor (2010). Furthermore Fan and Hyndman (2012) propose a new statistical methodology to forecast the short-term demand for two regions of the Australian electricity market including exogenous variables.

Despite its general importance, load forecasting will not be the focus of the present paper. Therefore, load forecasts were obtained based on a single method, namely the double seasonal exponential smoothing method by Taylor (2003, 2012). This univariate approach is straightforward to implement and has been shown to perform well (Taylor, 2012). Like Taylor (2012) we apply the natural log transformation to the loads in order to stabilize the variance of each series. As we are not interested in the unconditional levels we further demean each time series before fitting an additive version of the model.

The double seasonal exponential smoothing method decomposes the demeaned log-loads L_t (to which we will further refer as loads) into three components: the level l_t, one seasonal factor d_t for the intraday cycles, and one seasonal factor w_t for the intraweek cycles. The three components are

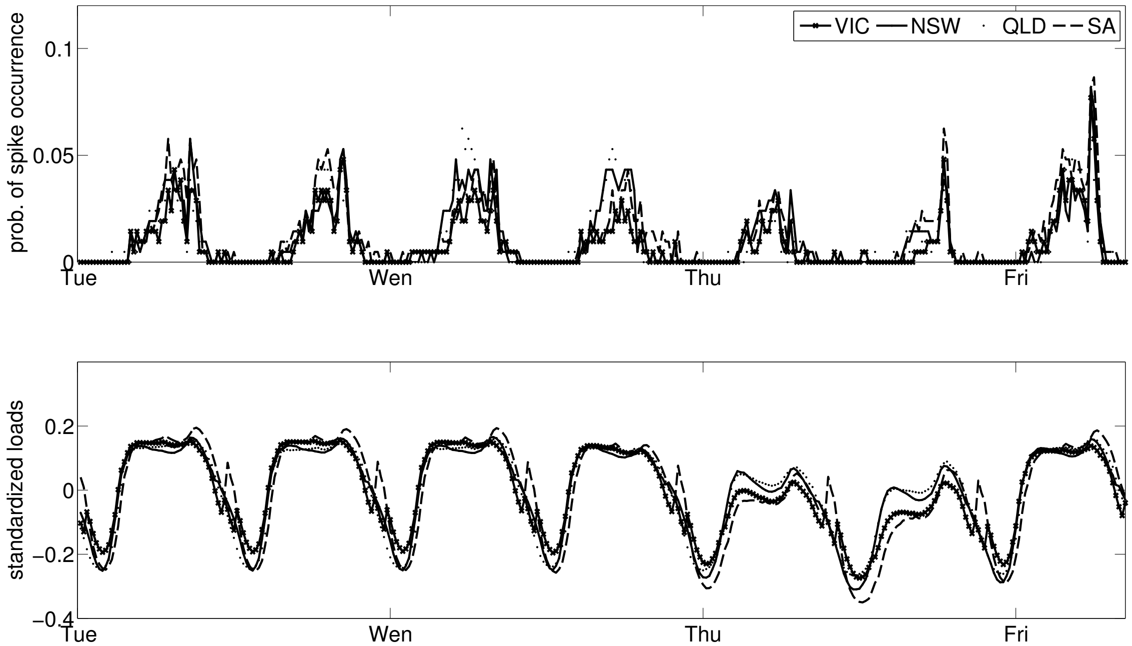

Figure 4: Weekly seasonality of spikes (top) and demeaned log-loads (bottom) for the period 01.01.2008 to 31.12.2011.

updated by the following smoothing equations

$$\begin{aligned} l_t &= (1-\alpha)\, l_{t-1} + \alpha\, (L_t - d_{t-m_1} - w_{t-m_2}) \\ d_t &= (1-\beta)\, d_{t-m_1} + \beta\, (L_t - l_{t-1} - w_{t-m_2}) \\ w_t &= (1-\gamma)\, w_{t-m_2} + \gamma\, (L_t - l_{t-1} - d_{t-m_1}), \end{aligned}$$

where α, β, and γ are smoothing parameters determining the smoothness of the three components. Corresponding to the number of periods in a day and a week, m_1 and m_2 are chosen to be 48 and 336. The load forecast $\hat{L}_{t+1}$ is computed from the smoothed components by

$$\hat{L}_{t+1} = (1-\phi)\, (l_t + d_{t+1-m_1} + w_{t+1-m_2}) + \phi\, L_t.$$

Here, the additional smoothing parameter ϕ allows for autocorrelation in the forecast errors. The values for the smoothing parameters are chosen to minimize the in-sample mean-squared-error, and the algorithm is initialized by appropriate averages over two weeks (Taylor, 2003).

3.2. Dynamic logit models

We now turn to the models for the occurrence of price spikes. These occurrences can be modelled by the binary process S_t that takes value 1 whenever the spot price exceeds the specified threshold (here A\$100/MWh or A\$300/MWh) and zero otherwise. Such a binary process is characterized by its hazard rate h_t, that is, by the probability of a spike occurring at time t conditional on past information up to time $t-1$. Thus the hazard rate is given by

$$h_t = \text{Prob}\big(S_t = 1 | \mathcal{H}_{t-1}\big),$$

where $\mathcal{H}_{t-1}$ denotes all relevant information observed up to time $t-1$.

In Section 2.2 we observed that the majority of spikes occur in blocks of more than one observation. A natural approach to model the occurrence of such clustered blocks is a logit model using exogenous information, such as loads, and past prices that can capture the persistence of spike regimes. In the context of daily spot price modeling, similar ideas have been used by Mount et al.

(2006), Huisman (2008) and Eichler and Türk (2013). Using a logistic regression, in this section, we then model h_t as

$$h_t = \frac{1}{1+\exp(-\pi_t)}, \tag{3.1}$$

with the precise form of π_t defining the different model specifications.

A weakness of the logit link function in (3.1) is that it is most sensitive to changes in π_t at a probability of 0.5 and that it is a symmetric function. The scobit (or skewed-logit), which was proposed by Nagler (1994), is an alternative link function that can be more sensitive to changes of π_t at values corresponding to probabilities different from 0.5. It makes use of the Burr-10 distribution in order to allow for the needed flexibility,

$$h_t = 1 - \frac{1}{[1+\exp(\pi_t)]^a}, \tag{3.2}$$

where $a > 0$. The Burr-10 distribution thus nests the logistic distribution (at $a = 1$) while allowing for a skewed response curve, with a measuring the skewness.

As a first logit model, called benchmark logit *(B-logit)* model, we propose a simple forecasting method that solely uses the fact that spikes tend to occur in blocks, i.e., that there are many sequences of consecutive spikes in the data:

$$\pi_t = \beta_0 + \beta_1 S_{t-1}, \tag{3.3}$$

where S_{t-1} is a spike indicator for period $t-1$. With this logit model we obtain probability forecasts that correspond to the conditional probability of a spike following a spike, $P(S_t = 1|S_{t-1} = 1)$ and the conditional probability of a spike occurring after a moderate price, $P(S_t = 1|S_{t-1} = 0)$. The model thus is a rudimentary version of the regime-switching models, which are used as work horses in daily spot price modeling (see Janczura and Weron, 2010 for an overview). Despite the fact that this approach immediately adapts after entering or leaving a spike block, it also has its disadvantages. In case of $P(S_t = 1|S_{t-1} = 1) \geq 0.5$ and $P(S_t = 1|S_{t-1} = 0) < 0.5$ one will, by construction, never be able to forecast the first spike of a block and will make one false prediction whenever a spike-block ends if a probability of 0.5 is taken as basis for this decision. As for periods in which spikes tend to occur in long blocks, this model still should yield a good fit, and we decide to use it as a benchmark for further model specifications.

Furthermore, we saw in Figure 1 that apart from occurring in blocks, spikes tend to occur in clusters. Another characteristic that stood out was the fact that spikes (high prices) in $t-1$ and $t-48$ appear to have the strongest effect on the probability to observe a spike at t. Thus a good model specification should account for these findings and at the same time allow to test if the lag of 24 hours might be caused by seasonality, i.e. spikes only occurring during certain hours of the day. We thus propose the following model including a richer amount of exogenous variables[5],

$$\pi_t = \beta_0 + \beta_1 \exp(-\beta_2 d_{t-t_n}) + \beta_3 \pi_{t-1} + \beta_4 \hat{L}_t + \beta_5 p_{t-1} + \beta_6 p_{t-48}, \tag{3.4}$$

which we call the Dynamic Hawkes logit/scobit *(DH-logit/scobit)* model, depending on the specification of the loss function. Again, β_0 corresponds to the constant term of the model while $\beta_1 \exp(-\beta_3 d_{t-t_n})$ is the simplest specification of a Hawkes process. It allows to model an exponential decay in the probability of spike occurrence as a function of the duration between the last spike (marked by t_n) and t, which is given by d_{t-t_n}. In contrast to the ACH model, which bases its dynamics on the duration between the last two observed spikes, we now allow the dynamics to adapt in correspondence with d_{t-t_n}. Thus the model yields faster adaption to changes in spike dynamics when entering or leaving times of turmoil (spike clustering). The lagged value of π_t is an autoregressive component, allowing for richer dynamics in the process π_t and thus the conditional probability of observing a spike at t. The parts, that are so far included in the model combine a

[5]Note that we will restrain ourselves to write down only the equations concerning π_t for the following models, in order not to exceed one row for an equation.

basic Hawkes model with the autoregressive logit model based on Kauppi and Saikkonen (2008). Furthermore, in order to account for seasonality and demand fluctuation we include half-hourly forecasts of demeaned log-loads, $\hat{L}_t$. Finally, we add lagged log-prices, p_{t-1} and p_{t-48}, to capture the feature of spike blocks in a similar way as the *B-logit* model with the difference that state-continuous lagged log-prices are used instead of the binary spike indicator. In this sense the model deviates from the specification proposed in Kauppi and Saikkonen (2008), which includes lags of the binary variable. Usually one does not have access to the variable driving the binary process in applications of binary choice models. This is different here, since we have full knowledge of the price process and ignoring the size of past prices would mean disregarding potentially important information. The *DH-logit/scobit* model is thus capable of capturing sequentially occurring spikes, while allowing for further information to indicate spikes that have not been preceded by another spike and to announce the end of a spike block before it is actually over. Furthermore dynamics will depend on the duration since the last spike has occurred.

Next, based on ideas of the literature on regime-switching models for daily spot prices we use the logit framework in order to build a dynamic threshold logit model which explicitly acknowledges the fact that price spikes are fundamentally different from moderate prices. This is typically done when modeling daily energy prices by distinguishing spike regimes and non-spike regimes. To this end we propose a simple regime-switching model, called regime-switching logit/scobit *(RS-logit/scobit)*. Recall that the difference between logit and scobit model depends again on the specification of the link function. The model is defined as

$$\pi_t = c_1 + \beta_1 p_{t-1} + \beta_2 p_{t-48} + (1 - S_{t-1})(c_2 + \beta_3 \hat{L}_t + \beta_4 p_{t-1} + \beta_5 p_{t-48}). \tag{3.5}$$

This model allows to isolate the effect of load forecasts to exclusively explain spike probabilities when observing moderate prices. If in contrast a spike has occurred at $t-1$, this exogenous information is disregarded and only past log-prices are assumed to be driving the probability of spikes. Thus we expect higher explanatory power of the exogenous variable loads when no spikes are prevailing. The reason is that the seasonal pattern, driving the probability of spike occurrence, should be mainly of strong influence when prices are not spiking. Once a spike has occurred in the previous observations the corresponding price can be assumed to dominate the probability which drives the occurrence of another spike. It should be mentioned that we also tried to combine the *DH-* and *RS-logit/scobit* specification. The parameters of the *RS-logit/scobit* model appear to dominate and the parameters for Hawkes and autoregressive part where not significant at a 5% level indicating multicollinearity in several cases. For this reason we did not pursue a combination of the two model frameworks. In the following, the question which of the models presented above is more appropriate is an empirical issue that we try to answer below.

3.3. The autoregressive conditional hazard model

In the ACH model originally proposed by Hamilton and Jorda (2002), the hazard rate depends only on the observed durations between (spike) events. Since the hazard rate is constant between the occurrence of two successive price spikes, the durations are geometrically distributed. More precisely, denoting the duration between the $(n-1)$th and nth spike by u_n and letting ψ_n be the conditional expectation of u_n given the previous durations $u_{n-1}, u_{n-2}, \ldots$, the hazard rate h_t at time t is given by

$$h_t = \frac{1}{\psi_{N(t-1)+1}}, \tag{3.6}$$

where $N(t-1)$ is the counting process giving the total number of spike events up to time $t-1$. The main idea of the ACH model and the underlying autoregressive conditional duration (ACD) model by Engle and Russel (1998) is that (similar to a GARCH specification) the intensities ψ_n depend only on the past durations and intensities by

$$\psi_n = \omega + \sum_{j=1}^{k} \alpha_j \, u_{n-j} + \sum_{j=1}^{l} \beta_j \, \psi_{n-j}. \tag{3.7}$$

As the distribution of the durations is extremely skewed, Christensen et al. (2012) apply a generalized version of the ACH model proposed by Fernandes and Grammig (2006), which makes use of a Box-Cox transformation to reduce the skewness of the durations. Thus the intensities ψ_n satisfy

$$\psi_n^{\nu} = \omega + \sum_{j=1}^{k} \alpha_j \, u_{n-j}^{\nu} + \sum_{j=1}^{l} \beta_j \, \psi_{n-j}^{\nu}, \tag{3.8}$$

where the parameter ν controls the Box-Cox transformation. The original ACH model is obtained for $\nu = 1$.

Equation (3.6) allows to include additional explanatory variables such as loads, which allow the hazard rate to vary without the occurrence of spikes. For instance, Christensen et al. (2012) consider the modified hazard rate

$$h_t = \frac{1}{\Lambda(\psi_{N(t-1)+1} + \exp(-\pi_t))}, \tag{3.9}$$

where the form of π_t again reflects the model specification when including explanatory variables. The function Λ was proposed by Hamilton and Jorda (2002)[6] to ensure that h_t satisfies the constraint $0 < h_t < 1$. The authors further propose to omit the intercept in (3.8) and add a constant to the equation driving π_t in (3.9). The parameters of the model are estimated by maximization of the likelihood function.

Using the model framework proposed by Christensen et al. (2012) we choose the regressors for π_t to consist of a constant and the forecasted demeaned log-load $\hat{L}_t$ as well as the lagged log-prices p_{t-1} and p_{t-48} as exogenous variables. For the linear regressor part in (3.9) this yields

$$\pi_t = \beta_0 + \beta_1 \, \hat{L}_t + \beta_2 \, p_{t-1} + \beta_3 \, p_{t-48}. \tag{3.10}$$

Because of its Box-Cox transformation in the reminder we will refer to it as *ACH-bc* model. We further replaced the temperatures variables used by Christensen et al. (2012) (which in most of the cases turned out to be insignificant) with p_{t-1} and p_{t-48}, thus adding valuable exogenous information to the model. The model can be seen as a direct competitor of the *DH-scobit* model. From (3.8) it can be inferred that (ignoring exogenous variables) this model bases its conditional probability of a spike occurrence on the duration between the last two observed spikes and updates its expectations concerning the duration only in case that a new event occurs. The *DH-scobit* on the other hand (again ignoring exogenous variables) updates its expectations based on the duration between the last observed spike and t. It thus can be expected to descent faster to low probability levels when the time since the last spikes increases and to increase the spike probability faster after a spike is observed as it adapts instantaneously. Recall in this context that electricity price spikes do not only appear in clusters, but usually in blocks of consecutive spikes. Although the ACH model includes exogenous information that allows to account for these features, it is not obvious that the dynamics implied by the model are entirely appropriate. The reason is the fact that – without accounting for the exogenous information – the model might need too much time to adapt whenever a block or cluster of spikes starts or ends. This is indeed what we find in the data and the inclusion of load and past price information alone does not mitigate this shortcoming of the model. We again considered combining the framework of the *RS-logit* with the *ACH-bc* model. As before the parameters of the *RS-scobit* appear to dominate the fit yielding the core parameters of the *ACH* model insignificant at the 5%-level in several occasions.

4. Forecast comparison

In this section we report the results of our empirical analysis. We consider spikes defined as prices exceeding both A\$100/MWh and A\$300/MWh. As pointed out in section 2.2 we use two different

[6] The precise form of the transformation is given in footnote 3 of Hamilton and Jorda (2002).

sample periods. The first one was chosen from 1st of January 2003 until 31st of December 2006 and can be thought of as a control sample. Data for the year 2007 was discarded because of the "millennial drought" and the resulting unrepresentatively high amount of spike occurrences. The second time window covers the period from 1st of January 2008 until 31st of December 2012, where the last year is considered for the out-of-sample evaluation. In Section 4.1 we report the in-sample fit for our models, whereas the out-of-sample results are reported in Section 4.2.

4.1. In-sample fit

We estimated all models presented in Section 3 with price spikes corresponding to thresholds of A$100/MWh and A$300/MWh for the two sample periods described above. In order to preserve space we do not report the parameter estimates here, but note that they are what is to be expected considering the data description in Section 2.2 and the estimates reported in Christensen et al. (2012).[7] As the models are partly non-nested we decided to compare the in-sample fit with help of the Bayesian information criterion (BIC). Given any two estimated models, the one with the lower BIC value is the one to be preferred. Compared to other information criteria as, e.g., the Akaike information criterion it penalizes additional parameters quite strongly, particularly with samples as large as ours, so it should favor parsimonious specifications. As we are ultimately interested in models with a good forecasting performance we expect less heavily parameterized models to be more suitable. Based on the BIC we also computed the model confidence set (MCS) of Hansen et al. (2011). The MCS is a set of models whose performance is not significantly different considering a certain goodness-of-fit criterion and it can be seen as an analogue to a confidence interval for competing (non-nested) models. Thus we acknowledge that it is unlikely that a single model outperforms all the others, but that there are multiple models that perform equally well. The MCS is determined using a sequence of hypothesis tests. It eliminates inferior models based on the criterion of interest. P-values for the sequential tests are determined by a block-bootstrap procedure as described in Hansen et al. (2011) and references therein. We used a size of 5%, 5000 bootstrap samples and a block length corresponding to 4 weeks of data. Table 3 reports the BIC for all estimated models, with bold numbers indicating that the MCS contains the corresponding model. The loss for the best performing model is highlighted with an asterisk.

Now consider the role of the link function, i.e., the choice between logit and scobit. Looking at the results it stands out that the *DH-scobit* yields smaller BIC statistics than the *DH-logit* in 13 out of 16 cases. This indicates that the resulting latent error terms of the DH model should not be assumed to be symmetrically distributed. In case of the RS model on the other hand, the *RS-logit* yields smaller values than the *RS-scobit* in 14 out of 16 cases, indicating that the latent error terms of the model at hand might be symmetrically distributed. This suggests that the regime-switching dynamics of the model capture the asymmetries in the innovations driving the spike probabilities. Furthermore when focusing on window I and a threshold of $A\$100$ it can be stated that the *DH-scobit* is contained in the MCS for all four markets while yielding the smallest BIC for three of them. The *RS-logit* still enters twice, yielding the smallest BIC for SA, while the remaining models do not make it into MCS at all. For the same window and a threshold of $A\$300$ the *DH-scobit* and the *RS-logit* both enter the MCS for all markets. The *DH-scobit* again yields the smallest BIC for the three biggest markets while *RS-logit* exhibits the smallest BIC of all models under consideration for SA. Furthermore now the *RS-scobit* enters the MCS three times while the *DH-logit* and the *ACH-bc* enter twice. Looking at window II the picture changes insofar, that now the *RS-logit* enters the MCS for all four markets and both thresholds, while the *DH-scobit* only enters twice for each threshold. The only remaining model that enters once for the threshold of $A\$100$ is the *RS-scobit* for NSW. In case of the threshold of $A\$300$ the only remaining model to enter the MCS twice is the *DH-logit*. For the threshold of $A\$100$ the *RS-scobit* further yields the smallest BIC for three out of the four markets while giving the smallest loss for two markets when looking at the threshold of $A\$300$. The *DH-scobit* only gives the smallest BIC for NSW and a threshold of $A\$100$ while *DH-logit* exhibits the smallest loss for QLD and SA when

[7] Detailed estimation results are available from the authors upon request.

Table 3: Bayesian information criterion

	VIC		NSW		QLD		SA	
Threshold	A$100	A$300	A$100	A$300	A$100	A$300	A$100	A$300
				Window I				
$B-logit$	2968	936	3615	1313	2984	1208	6935	1798
$DH-logit$	2084	**708**	2245	889	2163	**875**	5322	1418
$DH-scobit$	**1999***	**687***	**2126***	**837***	**2072***	**859***	**5209**	**1413**
$RS-logit$	2092	**691**	2290	**865**	**2107**	**874**	**5198***	**1401***
$RS-scobit$	2080	**693**	2274	**870**	2117	**881**	5209	1412
$ACH-bc$	2560	**820**	2551	1404	2391	**886**	5386	1534
				Window II				
$B-logit$	3383	781	3421	1038	3979	1519	4177	1577
$DH-logit$	2545	577	2362	796	3108	**1208***	3188	**1140***
$DH-scobit$	**2490**	579	**2246***	766	3003	**1213**	3097	**1142**
$RS-logit$	**2487***	**560***	**2269**	**751***	**2972***	**1228**	**3043***	**1152**
$RS-scobit$	2498	571	**2274**	762	2983	1239	3054	1163
$ACH-bc$	3104	619	2820	864	3741	1328	3555	1217

Note: The table contains the BIC for the 6 different model specifications with spike thresholds (thr) equal to A$100/MWh and A$300/MWh and time-periods 1st of January 2003 to 31st of December 2007 (window I), as well as 1st of January 2008 to 31st of December 2011 (window II). Values that correspond to models belonging to the model confidence set (MCS) with a size of 5% are printed in bold. To calculate the MCS a block-bootstrap procedure with 336*4 observations per block and a total of 5 000 simulations was used.

considering the threshold of $A\$300$. Overall, in terms of in-sample fit the results of Table 3 indicate that *RS-logit* and *DH-scobit* appear to be the best suited models for the data at hand in terms of BIC. Furthermore it can be stated that the *ACH-bc*, which in total only enters twice into the MCS, and the *B-logit*, which never enters, appear to be the least suited approaches to fit the data.

4.2. Out-of-sample evaluation

In order to compare the forecasting performance of our models we estimated each using the in-sample period defined above and performed 1-step ahead forecasts of the probability to observe a spike. Note that we did not re-estimate the model parameters each period, but that we only updated the information set.

Furthermore, we restrain from reporting hitting rates as the commonly used cut-off value of 0.5 might (as pointed out in Cramer, 2010) reflect the uneven sample composition, and in this case the prevalent outcome could be much better predicted than the rare alternative. In our case we certainly have one prevalent outcome (non-spikes) and a rather rare one (spikes). Cramer (1999, 2010) states that in case of unbalanced samples more sensible results are obtained by setting the cut-off value equal to the sample frequency. We state that the choice of an appropriate cut-off value strongly depends on how costly it is not to detect a spike and to forecast a spike when in fact none occurs and thus is a matter of the goal of the forecaster (see also Greene and Hensher, 2010 in this respect). Zhao et al. (2007b) even propose to directly integrate benefit-cost analysis into forecasting methods. Christensen et al. (2012) argued that not forecasting a spikes is in fact more costly than a false detection. This argument can be backed by the fact that electricity prices in Australia can be as high as $12 500.

While it is indeed important to be able to forecast the occurrence of spikes, it may very often be sufficient to have a good forecast for the probability of a spike. For this reason we present two types of measures in order to evaluate the predictive ability of the different models. The first measure is the predictive log-likelihood. It can be interpreted in the same way as the BIC. The difference is that now the number of parameters will not be taken into account as the out-of sample fit will not depend on it. This idea is in line with Kauppi and Saikkonen (2008), who used the the maximum value of the likelihood function as the main criteria for evaluating out-of-sample forecasts while facilitating interpretation by rescaling it into a pseudo R^2. We restrain from rescaling as the needed log-likelihood computed with only a constant term would not be the same for the ACH-bc model as for the remaining models. Nonetheless using the non-scaled log-likelihood functions allows us to make use of the MCS. Using the pseudo R^2 in contrast would yield loss functions that are bounded between zero and one at each point in time. This would yield differences between these loss functions which are then at each point in time bounded between -1 and 1. The use of significance tests, which are needed in order to decide if these differences are significantly different from zero, would be problematic in this case. Due to the shorter out-of-sample period the block length for the bootstrap is chosen to correspond to one week of data.

The second measure of forecasting performance is based on the forecasted spike probabilities, which we denote as $\hat{h}_t$, and the spike realizations S_t. Then the out-of-sample forecast performance can be evaluated by fit measures based on predicted probabilities (for an overview see e.g. Greene and Hensher, 2010) obtained from T^* out-of-sample observations. In the following, we use the statistic proposed by Cramer (1999) which is defined as

$$\mathrm{cr} = \mathrm{E}(\hat{h}_t|S_t = 1) - \mathrm{E}(\hat{h}_t|S_t = 0). \tag{4.1}$$

The first term on the right hand side gives the expectation of $\hat{h}_t$ conditional on a spike having occurred at t while the second term gives the expectation of $\hat{h}_t$ conditional on no spike having occurred at t. This measure heavily penalizes incorrect predictions. Furthermore, because each proportion is taken within the subsample, it is not unduly influenced by the large proportionate size of the group of more frequent outcomes. The resulting measure yields results that are bounded between -1 in case of not fit at all and 1 in case of a perfect fit. For this reason we restrain from using the MCS for this measure.

The results can be found in Table 4. Note that for the best performing model the loss is highlighted with an asterisk, whereas for the results concerning the predictive likelihood it is reported in bold whenever the corresponding model belongs to the MCS.

Concerning the predictive log-likelihood it can be stated that the *DH-scobit* yields the smallest value in four out of eight cases. Nonetheless the *RS-logit* and *RS-scobit* each are once the only model in the MCS. Furthermore all DH- and RS-specifications are equally often in the MCS with six times each. The *ACH-bc* only enters the MCS four out of eight times and never has the lowest predictive likelihood. Recall that the main difference between ACH- and DH-specification is that the latter is based on the duration since the last observed spike instead of the duration *between* the last two spikes. The results indicate that this simple adaption already strongly improves forecasting abilities of the model. The benchmark model only enters the MCS twice and yields the worst values in seven out of eight cases. This indicates that the model is not suited to appropriately fit the data. When shifting our attention to the Cramer statistic the picture changes in so far that the two *RS*-models together with the *B-logit* appear to dominate the results. The reason might be the fact that these models are changing their dynamics very strongly as soon as they enter a block of spikes and thus show high probabilities immediately after entering a spike block and low probabilities immediately after exiting it. The results acknowledge that a simple model assuming a spike in $t+1$ after observing a spike at t and none elsewise might be a hard to beat competitor depending on the purpose of the forecaster. Nonetheless it should be stated that the *DH* specifications (above all the *DH-scobit*) generally yield far better fits than the *ACH-bc*. The overall results thus encourage further investigation concerning the *RS*- and *DH*-specifications which appear to yield superior in- and out-of-sample fits to the formerly proposed *ACH-bc* model.

Table 4: Measures for predictive fit

	VIC		NSW		QLD		SA	
Threshold	A$100	A$300	A$100	A$300	A$100	A$300	A$100	A$300
# of spikes	171	17	119	2	257	36	280	23
				predictive log-likelihood				
$B-logit$	426.43	**82.21**	298.57	**29.50**	962.74	281.47	667.56	143.84
$DH-logit$	315.75	**47.02**	245.10	**14.49**	**797.25**	**217.97**	**634.29**	**107.33**
$DH-scobit$	294.37	**43.57***	206.53	**13.80***	**794.75***	**217.82***	**516.66**	**102.32**
$RS-logit$	**282.36***	**45.64**	199.33	**15.65**	**816.38**	231.27	**469.59***	**100.50**
$RS-scobit$	282.59	**45.78**	**197.73***	**15.59**	**816.32**	231.83	**469.62**	**100.46***
$ACH-bc$	364.85	**52.92**	293.22	**12.74**	978.13	223.45	**647.39**	**106.96**
				Cramer statistic				
$B-logit$	48.62	36.02	55.33*	-0.01	33.83	1.28	51.45*	25.39
$DH-logit$	44.76	32.39	37.30	0.72	30.77	2.87*	28.61	9.39
$DH-scobit$	46.61	33.90	44.69	1.32	32.53	2.80	39.08	14.41
$RS-logit$	52.16*	36.25	54.69	1.48	35.34*	2.39	50.90	25.80
$RS-scobit$	52.14	36.51*	54.67	1.55*	35.34*	2.35	50.89	25.83*
$ACH-bc$	45.80	12.39	27.86	1.06	25.98	2.23	30.14	4.40

Note: This table reports the predictive (negative) log-likelihood and Cramer statistic whereas the latter was multiplied by 100) for spike forecasts with thresholds (thr) equal to $100/MWh and $300/MWh for the out-of-sample forecasts from January 1 to December 31 2012. The measure for the best performing model is marked with an asterisk. In the case of the predictive log-likelihood for models belonging to the model confidence set (MCS) the loss is reported in bold. To get the MCS we used block-bootstrap with block-length of 336 and 5 000 simulations. The tests were based on a significance level of 5%.

5. Conclusion

Accurately forecasting price spikes is essential for all market participants. Retailers could better hedge their positions while producers could include the information into their bidding strategies. Even industrial consumers that are capable of switching energy consumption over short periods of time could potentially profit from accurate spike forecasts. Furthermore the introduction of the demand response mechanism which aims at facilitating consumers the decision to continue consumption, or reduce their consumption by a certain amount, for which they would be paid the prevailing spot price makes the need for proper spike forecasting approaches economically even more eminent. The importance of forecasting spikes has already been pointed out by, e.g., Zhao et al. (2007a), Zhao et al. (2007b), Christensen et al. (2009) and Christensen et al. (2012), who proposed sophisticated models to achieve that goal.

In the paper at hand we reconsider this problem by first providing a detailed analysis of the data under consideration. Most importantly we show that spikes often occur in blocks, i.e., in a consecutive sequence, and that they exhibit a certain autocorrelation structure. Based on these characteristics we suggest a refinement (using alternative exogenous variables) of the model proposed by Christensen et al. (2012) and alternatives in form of logit specifications relying on the proposals of Kauppi and Saikkonen (2008) as well as ideas from the regime-switching literature. Furthermore instead of solely comparing the models in terms of certain loss functions we compute the model confidence set proposed by Hansen et al. (2011) in order to identify models that are equivalent in their performance.

The dynamic logit specifications, i.e., *DH-scobit* and *RS-logit* are shown to have a superior fit when evaluated both in-sample and out-of-sample. The reason is that these models manage to adequately respond to the fact that spikes tend to occur in blocks and clusters. The DH specification does so by allowing the dynamics to depend on the duration since the last spike has occurred. The RS framework on the other hand applies different dynamics whenever a spike has been observed as the most recent observation. Furthermore we allowed past log-prices to drive the dynamics of spike occurrence by including them, apart from load forecasts, as exogenous parameters. Load forecasts already have been used earlier as covariates and help capturing seasonalities. Although following Christensen et al. (2012) we solely consider half hourly one-step forecasts, practitioners could easily adapt the presented models to forecast spike probabilities over horizons longer than 30 minutes.

For future research it may be useful to study and forecast the multivariate behavior of price spikes. Such information would be valuable for market participants that are active on more than one market and try to minimize their overall price risk exposure. In this regard Cartea and González-Pedraz (2012) show for European electricity markets and daily spot prices that spikes in the price spread between two markets can make up to 40% of the total value of an interconnector. Further contributions for daily data and on the complete price distribution have already been made by, e.g., Worthington et al. (2005), Higgs (2009) and Ignatieva and Trueck (2011).

References

AEMO, 2010. An Introduction to Australia's National Electricity Market. Australian Energy Market Operator, Melbourne.

AEMO, 2012. Short Term PASA Process Description. Australian Energy Market Operator, Melbourne.

AEMO, 2013. Demand Response Mechanism and Ancillary Services Unbundling - High Level Market Design. Australian Energy Market Operator, Melbourne.

Anderson, E. J., Hu, X., Winchester, D., 2007. Forward contracts in electricity markets: the Australian experience. Energy Policy 35, 3089–3103.

Cartea, Á., González-Pedraz, 2012. How much should we pay for interconnecting electricity markets? a real options approach. Energy Economics 23, 14–30.

Cartea, Á., Villaplana, P., 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., Gray, P., Van Campen, B., 2008. A new approach to characterizing and forecasting electricity price volatility. International Journal of Forecasting 28, 728–743.

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

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

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.

Cramer, J. S., 2010. Logit Models from Economics and Other Fields. Cambridge University Press, New York.

Eichler, M., Türk, D., 2013. Fitting semiparametric Markov regime-switching models to electricity spot prices. Energy Economics 36, 614–624.

Engle, R. F., Russel, J. R., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66, 1127–1162.

Fan, S., Hyndman, R. J., 2012. Short-term load forecasting based on a semi-parametric additive model. IEEE Transactions on Power Systems 27, 134–141.

Fernandes, M., Grammig, J., 2006. A family of autoregressive conditional duration models. Journal of Econometrics 130, 1–22.

Greene, W. H., Hensher, D. A., 2010. Modeling Ordered Choices: A Primer. Cambridge University Press, New York.

Hamilton, J. D., Jorda, O., 2002. A model of the federal funds rate target. Journal of Political Economy 110, 1135–1167.

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. Econometrica 79, 453–497.

Higgs, H., 2009. Modelling price and volatility inter-relationships in the Australian wholesale spot electricity markets. Energy Economics 31, 748–756.

Huisman, R., 2008. The influence of temperature on spike probability in day-ahead power prices. Energy Economics 30, 2697–2704.

Ignatieva, K., Trueck, S., 2011. Modeling spot price dependence in Australian electricity markets with applications to risk management. Working paper, Macquarie University Sydney.

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

Janczura, J., Weron, R., 2010. An empirical comparison of alternate regime-switching models for electricity spot prices. Energy Economics 32, 1059–1073.

Kanamura, T., Ohashi, K., 2008. On transition probabilities of regime switching in electricity price. Energy Economics 30, 1158–1172.

Kauppi, H., Saikkonen, P., 2008. Predicting U.S. recessions with dynamic binary response models. The Review of Economics and Statistics 90, 777–791.

Knittel, C. R., Roberts, M. R., 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.

Lu, X., Dong, Z. Y., Li, X., 2005. Electricity market price spike forecast with data mining techniques. Electric Power Systems Research 73, 19–29.

Mount, T. D., Ning, Y., Cai, X., 2006. Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters. Energy Economics 28, 62–80.

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

Rambharat, R. B., Brockwell, A. E., Seppi, D. J., 2005. A threshold autoregressive model for wholesale electricity prices. Journal of the Royal Statistical Society. Series C (Applied Statistics) 54, 287–299.

Taylor, J. W., 2003. Short-term electricity demand forecasting using double seasonal exponential smoothing. Journal of the Operational Research Society 54, 799–805.

Taylor, J. W., 2010. Exponentially weighted methods for forecasting intraday time series with multiple seasonal cycles. International Journal of Forecasting 26, 627–646.

Taylor, J. W., 2012. Short-term load forecasting with exponentially weighted methods. IEEE Transactions on Power Systems 27, 458–464.

Taylor, J. W., de Menezes, L. M., McSharry, P. E., 2006. A comparison of univariate methods for forecasting electricity demand up to a day ahead. International Journal of Forecasting 22, 1–16.

Thomas, S., Ramiah, V., Mitchell, H., Heaney, R., 2011. Seasonal factors and outlier effects in rate of return on electricity spot prices in Australia's national electricity market. Applied Economics 43, 355–369.

Worthington, A., Kay-Spratley, A., Higgs, H., 2005. Transmission of prices and price volatility in Australian electricity spot markets: a multivariate GARCH analysis. Energy Economics 27, 337–350.

Zhao, J. H., Dong, Z. Y., Li, X., 2007a. Electricity market price spike forecasting and decision making. IET Generation, Transmission and Distribution 1, 647–654.

Zhao, J. H., Dong, Z. Y., Li, X., P, W. K., 2007b. A framework for electricity price spike analysis with advanced data mining methods. IEEE Transactions on Power Systems 22, 376–385.