Proceedings of the International Conference on Industrial Engineering and Operations Management
Pretoria / Johannesburg, South Africa, October 29 – November 1, 2018

Long-term electricity demand forecasting using a generalised additive mixed quantile averaging (GAMMQV) model

Paul Mokilane and Pravesh Debba
Council for Scientific and Industrial Research, Pretoria, South Africa and
School of Statistics and Actuarial Science, University of the Witwatersrand, South Africa.
PMokilane@csir.co.za, PDebba@csir.co.za

V.S.S. Yadavalli
Department of Industrial and Systems Engineering, School of Engineering, University of Pretoria, South Africa
Sarma.Yadavalli@up.ac.za

Caston Sigauke
University of Venda, University Road, Thohoyandou, Limpopo Department of Statistics Limpopo, South Africa
caston.sigauke@univen.ac.za

Abstract

The paper discusses the development and application of GAMMQV in forecasting the long-term electricity demand in South Africa. The long-term hourly demand from 2007 to 2023 with in-sample forecasts from 2007 to 2012 and out-of-sample forecasts from 2013 to 2023 were done. The actual and forecasted demand distributions closely matched between 2013 and 2015. Therefore, the forecasted demand distribution is expected to represent the actual demand distribution until 2023. The findings are that (a) the expected demand and daily demand profiles are well forecasted and (b) future distributions of hourly demand and peak daily demand are likely to shift towards lower demand over the years until 2023. The contributions of the paper are (a) the development of GAMM with trend model in forecasting long-term electricity demand, harnessing the correlation structures within different hours (c) inclusion of a nonlinear trend with forecasted values from quantile regression (QR) and (d) the development and application of GAMMQV to the South African data.

Keywords
GAMM, probabilistic forecasts, MAPE, density function, distributions

1. Introduction

The economic growth of any country is dependent on its energy security. Long-term electricity demand forecasts are important to those who manage power utility companies including political principals making decisions pertaining to the countries energy security. The decisions include generation capacity, maintenance planning, decisions on the optimal energy mix and plans for future infrastructure expansions. Electricity demand trend in South Africa is unpredictable, it showed an upward trajectory between 1997 and 2007, and it then stabilized between 2008 and 2011 and started declining in the last four years until 2015, so it is uncertain whether the latest downward trend will continue or will revert to the earlier upward trajectory. The accuracy of the demand trend is important otherwise long-term forecasts could vastly deviate from the future actual demand which could have serious planning implications. For planning purposes, the country must assess its capacity to continue meeting its electricity demand in future with the existing generation infrastructure and whether there will be a need for infrastructure expansion. These decisions are made under uncertainties which among others emanate from the changing weather conditions, market penetration of renewable sources of electricity whose data are not adequately collected in South Africa, the growing market of electric vehicles, market penetration of power saving appliances, the escalating cost of electricity and unpredictable long-term economic growth. The decisions could have far reaching consequences because the decision for infrastructure expansion could result in the construction of unnecessary power generating facilities while the decision not to expand on the current infrastructure could result in failure to meet future electricity demand which could hurt the economy and result in an unintended loss of jobs. Hyndman & Fan (2010) argue that an overestimate of long-term electricity demand will result in substantial wasted investment in the construction of excess power facilities, while an underestimate of demand will result in insufficient generation and unmet future demand.

In the literature to date, short-term electricity demand forecasting has attracted substantial attention due to its importance for power system control, unit commitment, economic dispatch and electricity markets while on the other hand the medium and long-term forecasting has not received as much attention, despite their value for system planning and budget allocation (Fasiolo, Goude, Nedellec, & Wood, 2017; Gaillard, Goude, & Nedellec, 2016; Hyndman & Fan, 2010). Sigauke (2017) indicated that medium-term electricity demand is important to decision-makers in power utility companies for planning power generation, maintenance planning and for risk assessment. Long-term demand forecasting according to Hyndman & Fan (2010) corresponds to the forecast horizon from several months to several years ahead. Hong, Wilson, & Xie (2014) argue that forecasting by nature is a stochastic problem, but most of the utilities are still developing and using point forecasts instead of probabilistic forecasts. One of the advantages of probabilistic forecasts is that they provide estimates of the full probability distributions of the possible future values of electricity demand and most importantly the uncertainties in the forecasts are quantifiable (Hong et al., 2016, 2014; Hong & Fan, 2016; Hyndman & Fan, 2010; McSharry, Bouwman, & Bloemhof, 2005; Mokilane et al., 2018; Sigauke, 2014, 2017). The literature on statistical models used in electricity demand forecasting is dominated by parametric modelling approaches. In parametric modelling, the model is completely defined by a small set of parameters. There is limited literature on the application of non-parametric models in long-term electricity demand forecasting. In non-parametric modelling, the relationships between the outcome variable and the covariates are defined by functions whose shapes are fully determined by the data, and the number of parameters is determined by the size of the data. A semi-parametric model is the hybrid of parametric and non-parametric models. Semi-parametric models have been applied in electricity demand forecasting (Farland, 2013; Hyndman & Fan, 2010; Ruppert, Wand, & Carroll, 2003; Sigauke, 2017). Sigauke (2017) indicated that a generalised additive model (GAM) is classified as a statistical learning technique in forecasting electricity demand which is not discussed in the literature in South Africa. A generalised additive mixed model (GAMM) is the extension of the GAM model.

The critical review of literature on other approaches to electricity demand forecasting such as computational intelligence-based techniques is discussed in Hippert, Pedreira, & Souza (2001). Some comprehensive reviews of demand forecasting models which are commonly used in the energy sector are given by Suganthi & Samuel (2012); (Hong & Fan (2016); (Jebaraj & Iniyan (2006); Alfares & Nazeeruddin (2002). There is a vast body of literature on the electricity demand forecasting in South Africa (Amusa, Amusa, & Mabugu, 2009; Rasuba et al., 2010; Inglesi-Lotz, 2011; Koen & Holloway, 2014; Koen, Magadla, & Mokilane, 2014; Mokilane et al., 2018; Sigauke & Chikobvu, 2011; Sigauke, 2014, 2017; Ziramba, 2008). This paper focuses on the development and application of the GAMM quantile averaging (GAMMQV) model in electricity demand forecasting which is an alternative approach to the GAM model developed by Hastie & Tibshirani (1986) applied to load forecasting and the one used by Sigauke (2017) applied to electricity demand forecasting in South Africa.

2. Methodology

The GAMM is an additive and a functional modelling technique where the impact of predictive variables is not only captured through the fixed effects but also through smooth functions which could be nonlinear. The goal of GAMM in this study is to model electricity demand (outcome variable) using its drivers (covariates), which are expressed in the form of fixed effects and others in the form of some smooth functions (splines). Smoothing splines are real functions that are piecewise-defined by polynomial functions called basis functions. The places, where the polynomial pieces connect are called knots. In GAMM, penalized regression splines are used in order to regularize the smoothness of the spline. The GAMM model is generally written as;

$$y_{th} = \beta_{0h} + \sum_{j=1}^{p} \beta_{hj} x_{thj} + \sum_{j=1}^{p} s_{hj}(x_{thj}) + Z_{th} b_t + u_{th} \tag{1}$$

where y_{th} is an electricity demand on day *t* at hour *h*; $\beta_{hj}x_{thj}$ is the linear part of the model, β_{hj} is the corresponding fixed parameters, β_{0h} is the constant parameter of the linear component of the model; s_{hj} are smooth functions of covariates x_{thj}; Z_{th} is a row of random effects model matrix; b_t is a vector of random effects coefficients and u_{th} are residual errors. $t = 1, 2, \dots, n$; $h = 1, 2, \dots, 24$ and $j = 1, 2, \dots, p$ and $x_{th1}, x_{th2}, \dots, x_{thp}$ are p covariates. Each smooth, $s_{hj}(x_{thj})$ in (1) is treated as having a fixed effects (unpenalised) component, which can be absorbed into the linear part of the model, and a random effects (penalised) component, which can be absorbed into $Z_{th}b_t$. The random effects component of the smooth has an associated Gaussian distribution assumption. The "Month", "day of week" and "day of year" were included in (1) as smooth functions. The following variables "fribtwn", "monbtwn" and "lngwknd" in Table A.1 in the appendix were included in (1) as part of the linear component of the model together with the trend variables. The GAMM in (1) is estimated using penalised cubic splines (Goude, Nedellec, & Kong, 2014; Wood, 2006).

$$\min_{\beta, s_{hj}} [\sum_{t=1}^{n} [y_{th} - \sum_{j=1}^{p} \beta_{hj}x_{thj} - Z_{th}b_t - \sum_{j=1}^{p} s_{hj}(x_{thj})]^2 + \sum_{j=1}^{p} \lambda_j [\int (f''(x))^2 dx] \quad (2)$$

According to Wood (2017), the degree of smoothness is controlled by the penalty parameter $\Lambda = (\lambda_j,\ j = 1,2,\dots,p)$, determining the roughness of the function estimate to the data. The generalised cross-validation criterion (GCV) is used to optimise this function. The smooth function s_{hj} is the sum of basis functions, where; $b_{hi}(x)$ is the *i* th basis function at hour *h* and β_{hi} are unknown parameters. Therefore, s_{hj} can be written as;

$$s_{hj}(x) = \sum_{i=1}^{q} \beta_{hi}\, b_{hi}(x) \quad (3)$$

where, q in (3) denotes the number of bases functions (dimensions). GAMM assumes that the model errors are identically and independently distributed. This assumption is not fulfilled in the case of a time series regression. Present values of the time series are correlated with past values and hence the errors of the model are also correlated. In such cases, the errors are said to be autocorrelated. This implies that estimated regression coefficients and residuals of the model might be biased, which also implies that the confidence intervals would be incorrect. This problem is avoided by including the autoregressive moving average model (ARMA) for the errors in our GAMM models. Harnessing the correlation structures within hours could improve the forecasts. The correlation structures within hours were found to differ. Some hours have simple correlation structures (autoregressive (AR) process) while others have ARMA (*p*,*q*) correlation structures. Errors in equation (1) are therefore modelled using the ARMA (*p*, *q*) process given in (4);

$$u_t = \emptyset_1 u_{t-1} + \emptyset_2 u_{t-2} + \cdots + \emptyset_{t-p} u_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (4)$$

where; $u_{t-1}, u_{t-2}, \dots, u_{t-p}$ are the AR component with coefficients $\emptyset_1, \emptyset_2, \dots, \emptyset_{t-p}$ and $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ are the MA components with coefficients $\theta_1, \theta_2, \dots, \theta_q$. The *p* in (4) represents the number of AR terms while the *q* represents the number of MA terms. The GAMM model was extended in this paper by including the fitted trend to the fixed effects component of the model. The nonlinear trend was fitted to the electricity demand data as shown in Figure 1 (electricity demand at hour 18h00). The red line represents the fitted demand trend between 2007 and 2012. The fitted values were then extracted to form the derived variable referred to as "trendfitted". The challenge was that only trend values up to 2015 were available and in order to use trend to forecast electricity demand beyond 2015, the trend values needed to be forecasted. The "trendfitted" was then used as the outcome variable and various time-related variables were used as covariates, namely day, public holidays, months, weekends, December break and the Fourier series or harmonic terms were used to capture the cycles inherent in the "trendfitted" variable in order to forecast its future values using quantile regression. The trend was forecasted at 0.05, 0.10, 0.25, 0.50, 0.75, 90, 95 quantiles of its distribution. The same procedure was followed to get the "trendfitted" for other hours.

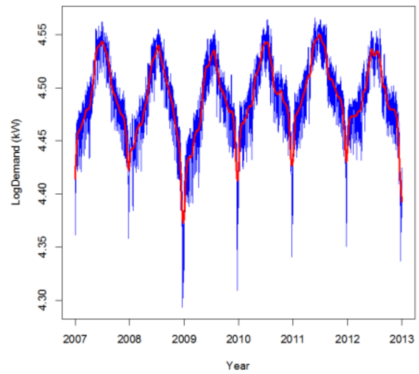


Figure 1: The nonlinear trend fitted to the demand data at hour 18

The GAMM model was then used to predict hourly electricity demand using calendar variables and the derived trend variable as drivers of electricity demand. A separate model for each hour was fitted and this is consistent with approaches of other researchers (Fan & Chen, 2006; Fay, Ringwood, Condon, & Kelly, 2003; Hyndman & Fan, 2010; Ramanathan, Engle, Granger, Vahid-Araghi, & Brace, 1997) who considered methods that fit separate models to the data from each half-hourly period in order to forecast future electricity demand. Their argument is that, the demand patterns change throughout the day and therefore, better estimates can be obtained if each half-hourly period is treated separately. Soares & Medeiros (2005) argue that considering separate models for each hour of the day, avoids modelling complicated intraday patterns in the hourly load, which is commonly called load profile. We argue that people carry out similar electricity demand influencing activities at specific hours across days, for example, at around 18h00 when people usually return home, some of the common activities include, cooking, watching television, bathing and ironing clothes for the next day. These activities are carried out every day at around the same time, hence the within hour correlation is expected to be stronger than the between hours correlation, therefore, modelling hours separately is likely to give better forecasts especially if within hour correlation structures are harnessed. The fitted model was referred to as a "GAMM with trend". The simple GAMM model with no trend was fitted and it was referred to as a "GAMM" model. The forecasts of the developed "GAMM with trend" and the "GAMM" models were combined to develop a model referred to as a "GAMM quantile averaging" model (GAMMQV). The GAMMQV model is given in (5);

$$y_i = \beta_0 + \beta_1 GAMM + \beta_2 GAMM_{trend} + \varepsilon_i \text{ and } i = 1, 2, \dots, n \tag{5}$$

where; $GAMM$ are forecasts from GAMM model and $GAMM_{trend}$ are forecasts from the GAMM model with trend and β_1 and β_2 are their corresponding respective regression coefficients. The τth quantile of ε_i is assumed to be zero and the corresponding quantile regression model is;

$$Q_\tau(Y|GAMM, GAMM_{trend}) = \beta_0 + \beta_1 GAMM + \beta_2 GAMM_{trend} \tag{6}$$

where;

$$Q_\tau(Y|GAMM, GAMM_{trend}) = \inf\{y: F_\tau(y|GAMM, GAMM_{trend}) \geq \tau\}, \tag{7}$$

is the conditional τ th quantile of the response (y_i) given the covariates ($GAMM \text{ and } GAMM_{trend}$) and $Q_\tau(Y|GAMM, GAMM_{trend})$ is a non-decreasing function of τ for any given covariate. $\beta = (\beta_1, \beta_2)$ is the vector of electricity demand parameters and is the marginal change in the quantile because of the marginal change in covariate. In estimating the QR model for a given quantile, the ideas of Koenker (2005) and Yue & Rue (2011) who used the standard approach of Koenker & Bassett Jr (1978) to estimate their QR model were used. QR minimises the tilted absolute function $\rho_\tau(.)$, which they called the check-function (Maistre, Lavergne, & Patilea, 2017), which asymmetrically weights residuals from the model to a degree that depends upon τ.

$$\rho_\tau(\varepsilon) = \begin{cases} (1-\tau)\varepsilon, & \varepsilon < 0 \\ \tau\varepsilon, & \varepsilon \geq 0, \end{cases} \qquad 0 < \tau < 1; \tag{8}$$

$\rho_\tau(\varepsilon)$ is a continuous piecewise linear function and non-differentiable at $\varepsilon = 0$ but differentiable everywhere else (Yue & Rue, 2011). This check-function ensures that all ρ_τ are positive and the scale is based on the probability τ.

The model parameters are estimated by;

$$\hat{\beta}(\tau) = argmin_{\beta \epsilon \mathcal{R}^p} \sum_{i}^{n} \rho_\tau[(y_i - (\beta_0 + \beta_1 GAMM + \beta_2 GAMM_{trend})]. \quad (9)$$

$\rho_\tau(\varepsilon)$ is a continuous piecewise linear function. The detailed formulation of quantile regression is available in Koenker & Bassett Jr (1978). The objectives of the study were; to forecast electricity demand distributions in South Africa in the long-term horizon (ten years), to forecast the distributions of the morning (demand at 06h00, 07h00 and 08h00) and the afternoon (demand at 18h00, 19h00 and 20h00) peak electricity demand over the years until 2023 in South Africa and to investigate the shifts in the distributions of electricity demand over the years until 2023.

3. Data

The South African total hourly electricity demand data from 1997 to 2015 is depicted in Figure 2. During this period the highest hourly electricity demand was 36 826 kW in 2011, while the minimum was 13 533 kW in 1998. Figure 2 shows an upward trend between 1997 and 2011 which could be attributed to the government's efforts to make electricity accessible to every South African household and consequently a lot of households were being connected to the grid and the growing economy during the same period (Mokilane et al., 2018). The electricity demand in Figure 2 took a downward trend in the latest four years until 2015 which could be attributed to the growing renewable sources of electricity, the sluggish economic growth, the steep increases in electricity tariffs and the market penetration of energy efficient appliances among others.

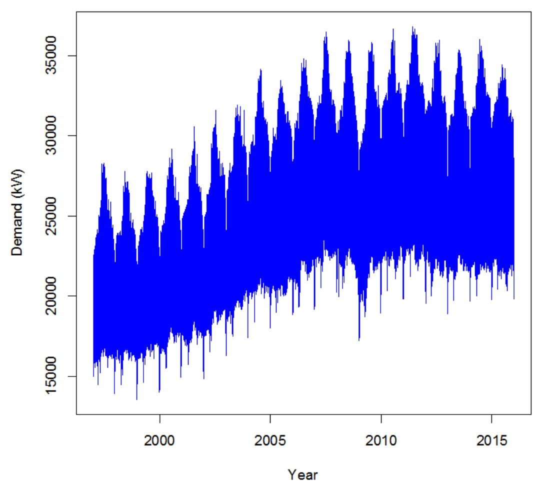


Figure 2: Hourly electricity demand between 2013 and 2015

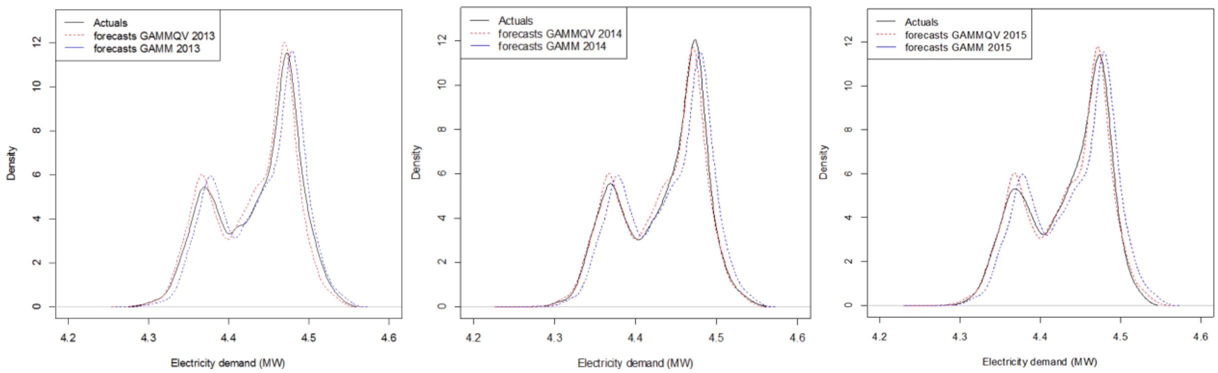
The model was built from the logarithmically transformed time series. Logarithmic transformations are convenient means of transforming a highly skewed variable into one that is approximately normal (Benoit, 2011). After carrying out a number of transformations of hourly demand from the Box & Cox (1964) family, we found that the logarithm is the best fit to the available data. We, therefore, modelled the logarithmic hourly demand data and this is consistent with Hyndman & Fan (2010) approach. The demand data between 2007 and 2012, inclusively, were used to train the model, while the data from 2013 to 2015 was withheld and used in model validation. The variables used in the modelling are given in Table A.1 in the appendix

4. Results and discussion

4.1 Sub-Headings

The electricity demand forecasts were validated; a) by comparing the distributions of the actual and the forecasted electricity demand (Figure 3), b) by comparing the daily profiles of the actual and the forecasted electricity demand (Figure 5), and c) by assessing the closeness of the point electricity demand forecasts to the actual electricity demand using the mean absolute percentage errors (MAPE) in Table 1. The electricity demand densities generated from the GAMMQV and the GAMM are compared to that of the actual electricity demand (Figure 3). Figures 3(a), 3(b) and

3(c) show that the forecasted electricity demand densities from the GAMMQV model (red graphs in Figure 3) better represented the densities of the actual electricity demand in 2013, 2014 and 2015 respectively.



(a) (b) (c)

Figure 3: Comparisons of actual and forecasted demand densities - GAMMQV vs GAMM: ((a) =2013 actual demand density against forecasted, (b) = 2014 actual demand density against forecasted density, (c) = 2015 actual demand density against forecasted)

The MAPE between the actual and the point electricity demand forecasts in 2013, 2014 and 2015 were below 4% in all hours, as shown in Table 1. Lewis (1982) indicates that a MAPE of less than 10% can be classified as highly accurate forecast. The GAMMQV model, therefore, provided better point forecasts compared to the GAMM model and the model produced accurate forecasts.

Table 1. Mean absolute percentage errors

	GAMM				**GAMMQV**			
Hours	**2013**	**2014**	**2015**	**Average**	**2013**	**2014**	**2015**	**Average**
0	2.013	2.600	3.190	**2.601**	1.706	1.914	2.633	**2.084**
1	2.105	2.686	3.093	**2.628**	1.701	1.942	2.657	**2.100**
2	2.115	2.635	3.088	**2.613**	1.688	1.896	2.653	**2.079**
3	1.880	2.327	2.852	**2.353**	1.901	1.927	2.587	**2.138**
4	1.847	1.937	2.535	**2.107**	2.294	2.217	2.641	**2.384**
5	2.648	2.601	2.931	**2.727**	3.183	3.152	3.181	**3.172**
6	3.371	3.258	3.970	**3.533**	3.146	3.124	3.490	**3.254**
7	2.684	3.089	3.714	**3.162**	2.018	2.172	2.886	**2.359**
8	2.434	2.760	3.331	**2.842**	1.742	2.029	2.681	**2.151**
9	2.380	2.618	2.783	**2.594**	1.791	2.031	2.284	**2.035**
10	2.290	2.666	3.124	**2.693**	1.826	1.969	2.507	**2.100**
11	2.245	2.610	3.121	**2.659**	1.893	2.022	2.592	**2.169**
12	2.223	2.593	3.075	**2.630**	1.934	2.041	2.608	**2.194**
13	2.148	2.507	2.939	**2.531**	2.028	2.102	2.583	**2.238**
14	2.140	2.481	2.916	**2.512**	2.156	2.207	2.646	**2.337**
15	2.056	2.297	2.799	**2.384**	2.381	2.226	2.649	**2.419**
16	2.054	2.077	2.811	**2.314**	2.513	2.308	2.709	**2.510**
17	2.033	2.025	3.245	**2.434**	2.397	2.510	3.013	**2.640**
18	2.128	1.958	3.205	**2.431**	2.111	2.677	2.777	**2.522**
19	2.346	1.962	3.176	**2.495**	1.872	2.304	2.504	**2.227**
20	2.300	2.395	3.742	**2.812**	1.900	1.946	2.925	**2.257**
21	2.290	2.670	3.806	**2.922**	1.806	1.806	2.907	**2.173**
22	2.214	2.574	3.401	**2.730**	1.664	1.835	2.718	**2.072**
23	2.117	2.639	3.150	**2.635**	1.726	1.873	2.659	**2.086**
Average	**2.253**	**2.498**	**3.167**	**2.639**	**2.057**	**2.176**	**2.729**	**2.321**

The MAPE and density functions in Table 1 and in Figure 3 respectively indicated that the GAMMQV gives better forecasts than the simple GAMM model. Therefore, the GAMMQV was used to forecast hourly electricity demand in this study.

4.2 Model diagnostics-GAMMQV

For illustration purposes, the model diagnostic results for hour 18 (corresponding to 18:00) are given in Figure 4. Figure 4 shows that a GAMMQV model is a good fit to the data. Figure 4(d) shows a plot of fitted values versus the response values which are scattered around a diagonal line showing a good fit. The clustering of residuals around zero in Figure 4(c) is the indication that the errors are independent of each other and there is no apparent relationship between them while Figure 4(b) shows that the residuals distribution was skewed to the left. The amount of variation explained by the GAMMQV model is 87.7% ($r^2 = 0.877$).

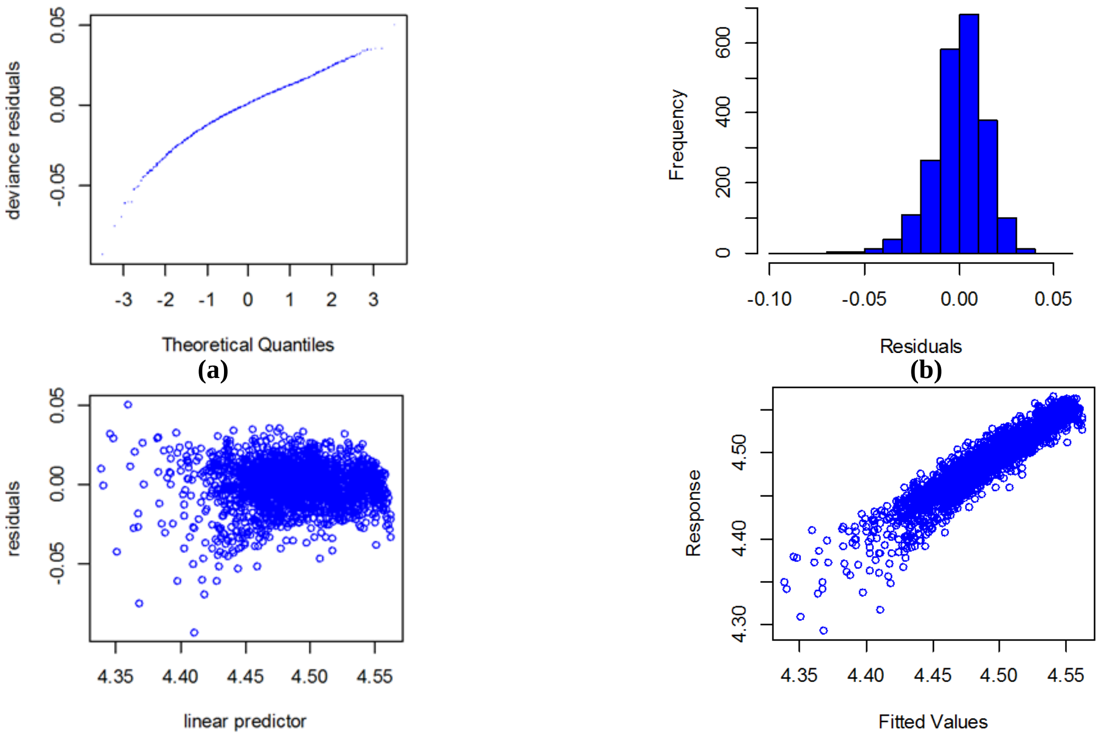
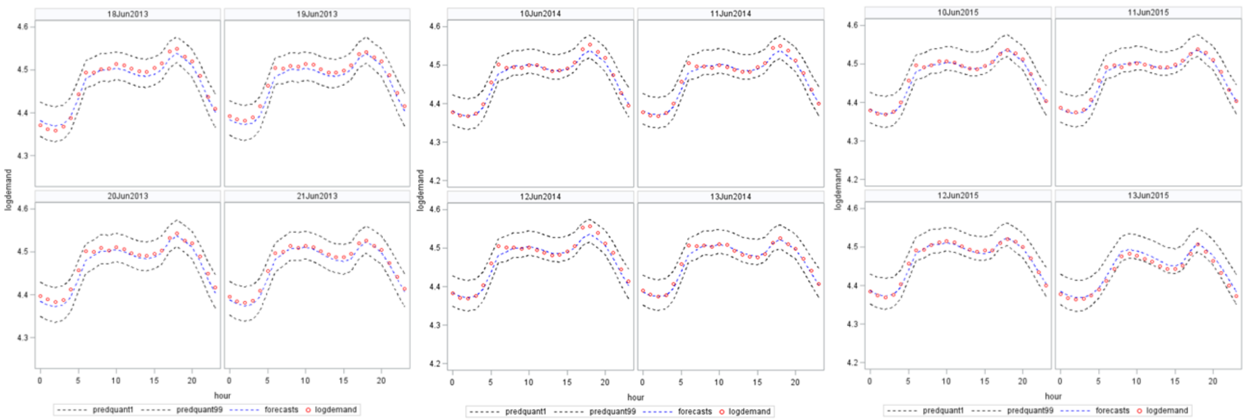


Figure 4: diagnostic plots (a = normal quantile-quantile, b = residuals histogram plot, c = linear predictor values against residuals, d = fitted values plotted against response values)

4.3 Forecasts

The three year out of sample forecasts were further validated by comparing the actual and forecasted daily electricity demand profiles. For discussion purposes, 4 days in June were selected in such a way that the day with the highest hourly electricity demand of the year was included. The highest electricity demand in 2013 was 35 393 kW on the 18 June, it was 36 039 kW in 2014 on the 12 June and 34 481 kW in 2015 on the 11 June and they were all at 18h00. Figures 5(a), 5(b) and 5(c) show that the forecasted (blue graph) electricity demand well represented the actual (red circles) electricity demand in 2013, 2014 and in 2015 respectively. The black graphs in Figure 5 represent the electricity demand forecasts at the 1st and 99th quantiles of the demand distribution. Considering the electricity demand forecasts on the 18th June 2013, at 18h00, the forecasted electricity demand was 34 602 kW (logdemand=4.5391) while the actual electricity demand was 35 393 kW (logdemand=4.5489). The forecasted electricity demand at the 1st quantile was 32 839 kW (logdemand=4.51639) while the forecast at the 99th quantile was 37 717 kW (logdemand=4.57654). Therefore, electricity demand at 18h00 on the 18th June 2013 is expected to fall between 32 839 kW and 37 717 kW with a 98% probability.

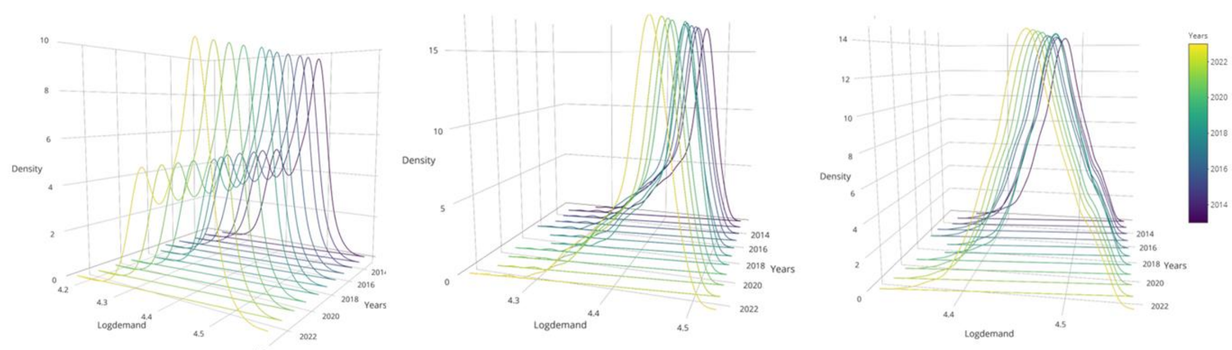


(a) **(b)** **(c)**

Figure 5: comparisons of actual and forecasted daily profiles: (a) = 2013 actual hourly demand against forecasted; (b) = 2014 actual hourly demand against forecasted; (c) = 2015 actual hourly demand against forecasted)

4.4 Comparing electricity demand distributions over the years

The electricity demand for each hour was forecasted at various quantiles of the demand distribution using the fitted GAMMQV model. The forecasted values were used to generate the density functions in Figures 6 (a), (b) and (c). By comparing the electricity demand density functions over the years, the insight into expected shifts in electricity demand patterns can be obtained, that is, whether the distribution of electricity demand in Figure 6 (a) is expected to shift towards higher or lower demand. The forecasted electricity demand distributions in Figure 6 (a) obtained from the GAMMQV model for the period investigated suggest that electricity demand from Eskom is likely to shift towards lower demand over the years until 2023.



(a) **(b)** **(c)**

Figure 6: Demand distribution between 2013 and 2023 - (a) Overall; (b) Morning peak; (c) Afternoon peak

The daily peak electricity demand is very important for planning purposes, as this represents the maximum that would need to be supplied in an hour and if the power generating company could meet the daily peak hourly electricity demand, it could meet any hourly demand. The morning and afternoon peak electricity demand distributions suggest a shift towards lower peak demand over the years until 2023 (Figures 6 (b) and (c)).

4.5 Conclusions

The forecasted electricity demand distributions closely match the actual electricity demand distribution between 2013 and 2015; therefore, there is no reason to doubt that the GAMMQV model will continue to forecast electricity demand accurately beyond 2015. The distribution of electricity demand from Eskom is expected to shift towards lower demand over the years. Both morning and afternoon peak electricity demand distributions are also expected to shift towards lower demand in future. The decline in electricity demand is apparent in the latest four years until 2015 which could be attributed to the growing renewable sources of electricity in South Africa, the sluggish economic growth, the steep increases in electricity tariffs and the market penetration of energy efficient appliances among others. The forecasted daily profiles from the GAMMQV model accurately capture the actual daily profiles in the long-term. The MAPE shows that the GAMMQV model gives good point forecasts. The probabilities of exceeding certain electricity demand can be obtained from the quantile forecasts of the GAMMQV model. The first contribution of the paper is the development and application of GAMMQV model in forecasting the long-term electricity demand in South Africa. The second contribution is the harnessing of the within hour correlation structures in forecasting the long-term hourly electricity demand using AR in some hours and ARMA in other hours. The third contribution is the probabilistic forecasting of long-term electricity demand using the South African data. The fourth contribution is the inclusion of the nonlinear trend with the future values forecasted using QR model. The trend values beyond 2012 were forecasted at 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95 quantiles of the distribution.

Acknowledgements

The research was conducted during a contract research project funded by Eskom and some funding from Thuthuka fund of the National Research Fund (NRF). I would therefore, like to thank Eskom and the NRF for funding the research. I would also like to thank the following Eskom staff members for supplying the data; Moonlight Mbata and Ntokozo Sigwebela. I would like to thank the following CSIR stuff members for their technical support; Nontembeko Dudeni-Tlhone, Nosizo Sebake and Renee Koen.

References

Alfares, H. K., and Nazeeruddin, M., Electric load forecasting: literature survey and classification of methods, *International Journal of Systems Science*, vol. 33, no. 1, pp.23-34, 2002.

Amusa, H., Amusa, K., and Mabugu, R., Aggregate demand for electricity in South Africa: An analysis using the bounds testing approach to cointegration, *Energy Policy*, vol. 37, no.10, pp. 4167-4175, 2009.

Benoit, K., Linear regression models with logarithmic transformations. *London School of Economics*, London, vol. 22, no. 1, pp. 23-36, 2011.

Box, G. E., and Cox, D. R., An analysis of transformations, *Journal of the Royal Statistical Society*. Series B (Methodological), pp. 211-252, 1964.

Rasuba, M., Khuluse, S. and Elphinstone, C. D., Forecasts for electricity demand in South Africa (2010-2035) using the CSIR sectoral regression model. Available: http://www.energy.gov.za/irp/irp%20files/CSIR_model_IRP%20forecasts%202010_final_v2.pdf, September 4, 2017.

Fan, S., and Chen, L., Short-term load forecasting based on an adaptive hybrid method, *IEEE Transactions on Power Systems*, vol. 21, no.1, pp.392-401, 2006.

Farland, J., Zonal and Regional Load Forecasting in The New England Wholesale Electricity Market: A Semiparametric Regression Approach, Masters thesis 1911 – February 2014, Available: http://scholarworks.umass.edu/theses/1120/?utm_source=scholarworks.umass.edu%2Ftheses%2F1120&utm_medium=PDF&utm_campaign=PDFCoverPages, August 16, 2017.

Fasiolo, M., Goude, Y., Nedellec, R., and Wood, S. N., Fast calibrated additive quantile regression, Available: https://arxiv.org/abs/1707.03307, August, 30, 2017.

Fay, D., Ringwood, J. V., Condon, M., and Kelly, M., 24-h electrical load data – A sequential or partitioned time series?, *Neurocomputing*, vol. 55, no. 3, pp. 469-498, 2003.

Gaillard, P., Goude, Y., and Nedellec, R., Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting, *International Journal of Forecasting*, vol. 32, no. 3, pp. 1038-1050, 2016.

Goude, Y., Nedellec, R., and Kong, N., Local short and middle term electricity load forecasting with semi-parametric additive models, *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 440-446, 2014.

Hastie, T., and Tibshirani, R., Generalized Additive Models, *Statistical Science*, vol. 1, no. 3, pp. 297-310, 1986.

Hippert, H. S., Pedreira, C. E., and Souza, R. C., Neural networks for short-term load forecasting: A review and evaluation, *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp. 44-55, 2001.

Hong, T., and Fan, S., Probabilistic electric load forecasting: A tutorial review, *International Journal of Forecasting*, vol. 32, no. 3, pp. 914-938, 2016.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., and Hyndman, R. J., Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond, vol. 32, no. 3, pp. 896-913, 2016.

Hong, T., Wilson, J., and Xie, J., Long term probabilistic load forecasting and normalization with hourly information, *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 456-462, 2014.

Hyndman, R. J., and Fan, S., Density forecasting for long-term peak electricity demand, *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 1142-1153, 2010.

Inglesi-Lotz, R., The evolution of price elasticity of electricity demand in South Africa: A Kalman filter application, *Energy Policy*, vol. 39, no. 6, pp. 3690-3696, 2011.

Jebaraj, S., and Iniyan, S., A review of energy models, *Renewable and Sustainable Energy Reviews*, vol. 10, no. 4, pp. 281-311, 2006.

Koen, R., and Holloway, J., Application of multiple regression analysis to forecasting South Africa's electricity demand, *Journal of Energy in Southern Africa*, vol. 25, no. 4, pp. 48-58, 2014.

Koen, R., Magadla, T., and Mokilane, P., Developing long-term scenario forecasts to support electricity generation investment decisions. *Proceedings of the 43rd Annual Conference of the Operations Research Society of South Africa*, p. 9, Pretoria, South Africa, September 14 - 17, 2014.

Koenker, R., *Quantile regression*. Cambridge university press, New York, NY, 2005.

Koenker, R., and Bassett Jr, G., Regression quantiles, *Econometrica: Journal of the Econometric Society*, vol. 46, no. 1, pp. 33-50, 1978.

Lewis, C. D., *Industrial and business forecasting methods: A Radical guide to exponential smoothing and curve fitting*, Colin David, London, Butterworth Scientific, 1982.

Maistre, S., Lavergne, P., and Patilea, V., Powerful nonparametric checks for quantile regression, *Journal of Statistical Planning and Inference*, vol. 180, pp. 13-29, 2017.

McSharry, P. E., Bouwman, S., and Bloemhof, G., Probabilistic forecasts of the magnitude and timing of peak electricity demand, *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1166-1172, 2005.

Mokilane, P., Galpin, J., Yadavalli, V. S. S., Debba, P., Koen, R., and Sibiya, S., Density forecasting for long-term electricity demand in South Africa using quantile regression, *South African Journal of Economic and Management Sciences*, vol. 21, no. 1, 2018.

Ramanathan, R., Engle, R., Granger, C. W., Vahid-Araghi, F., and Brace, C., Short-run forecasts of electricity loads and peaks, *International Journal of Forecasting*, vol. 13, no. 2, pp. 161-174, 1997.

Ruppert, D., Wand, M. P., and Carroll, R. J., *Semiparametric Regression*, Cambridge University Press, edition 1, Cambridge, New York, 2003.

Sigauke, C., and Chikobvu, D., Prediction of daily peak electricity demand in South Africa using volatility forecasting models, *Energy Economics*, vol. 33, no. 5, pp. 882-888, 2011.

Sigauke, C., Modelling electricity demand in South Africa, University of the Free State, Doctoral dissertation, University of the Free State, Available: http://scholar.ufs.ac.za:8080/xmlui/handle/11660/1569, 2017.

Sigauke, C., Forecasting medium-term electricity demand in a South African electric power supply system, *Journal of Energy in Southern Africa*, vol. 28, no. 4, pp. 54-67, 2017.

Soares, L. J., and Medeiros, M. C., Modeling and forecasting short-term electric load demand: a two-step methodology. Working Paper 495, Department of Economics, Pontifical Catholic University of Rio de Janeiro, Available: https://pdfs.semanticscholar.org/734b/3f6565243912784ad7b1a7421acb7188c9ca.pdf., February 15, 2018.

Suganthi, L., and Samuel, A. A., Energy models for demand forecasting-A review, *Renewable and Sustainable Energy Reviews*, vol. 16, no. 2, pp. 1223-1240, 2012.

Wood, S. N., *Generalized additive models: an introduction with R*, Chapman and Hall/CRC, London, 2006.

Wood, S. N., P-splines with derivative based penalties and tensor product smoothing of unevenly distributed data, *Statistics and Computing*, vol. 27, no. 4, pp. 985-989, 2017.

Yue, Y. R., and Rue, H., Bayesian inference for additive mixed quantile regression models, *Computational Statistics and Data Analysis*, vol. 55, no. 1, pp. 84-96, 2011.

Ziramba, E., The demand for residential electricity in South Africa, *Energy Policy*, vol. 36, no. 9, pp. 3460-3466, 2008.

Biographies

Paul Mokilane is a Senior Statistician in Spatial Planning and Systems within the Built Environment at the Council for Scientific and Industrial Research (CSIR). He is a PhD candidate at the University of the Witwatersrand in the School of Statistics and Actuarial Sciences. He serves in the board of the Institute of Certificated and Chartered Statisticians of South Africa (ICCSSA) and he is a member of the South African Statistical Association (SASA).

Pravesh Debba is a Manager of Spatial Planning and Systems within the Built Environment at the Council for Scientific and Industrial Research. He was an Associate Professor in Statistics at the University of South Africa. He holds a PhD in Statistics from the International Institute for Geo-Information Science and Earth Observation (ITC) and Wageningen University in the Netherlands and an MSc degree in Biostatistics from the University of Hasselt in Belgium. He has been visiting professor in the School of Statistics and Actuarial Sciences at the University of the Witwatersrand (Wits) since 2015 and has successfully co-supervised several masters and currently co-supervising several doctoral students in Statistics. He has served as president of the South African Statistical Association (SASA), and as chairperson of the Institute of Certificated and Chartered Statisticians of South Africa (ICCSSA), elected member of the International Statistical Association (ISI) and editorial board member of the Spatial Statistics Journal published by Elsevier. He has published around 40 research papers in journals, conference proceedings or book chapters.

Venkata S.S. Yadavalli is a Professor and Head of Department of Industrial & Systems Engineering, at University of Pretoria, South Africa. He obtained his Ph D from Indian Institute of Technology, Chennai, in 1983. His main areas of interest include, Stochastic modeling towards, Reliability, Queuing, Inventory, Manpower planning systems and Finance. He has published over 150 research papers in these areas. He published articles in various national and international journals including, IEEE Transactions on Reliability, Applied Mathematics and Information Sciences, Computers & Industrial Engineering, Stochastic Analysis and Applications, International Journal of Systems Science, International Journal of Production Economics, International Journal of Production Research, Applied Mathematics & Computation, etc. He was past Vice-President and President of Operations Research Society of South Africa. He is in the editorial board of various national and international journals. He is a Fellow of the South African Statistical Association. He received various awards from various professional bodies like IEOM (Industrial Engineering & Operations Management, USA), South African Institute of Industrial Engineers, South African Institute of Mining and Metallurgy etc. for his contributions towards research, and ICQRIT.

Caston Sigauke holds a PhD in Statistics from the University of the Free State and an MSc in Operations from the National University of Science and Technology in Zimbabwe. He is a member of the: International Institute of Forecasters (IIF), Member ID: 7178; Operations Research Society of South Africa (ORSSA), Member ID 511 and South African Statistical Association (SASA). His fields of professional expertise are lecturing, research and consulting. His fields of expertise are forecasting and time series, statistics of extremes, statistical learning and modelling, exploratory data analysis with applications in probabilistic forecasting in energy and environmental systems. Current research is on probabilistic electricity demand forecasting and solar energy modelling.

Appendix

TABLE A.1: Variables used in the GAMM

Variable	Type of variable	Scale	Created variables
Demand	**Dependent**	**Continuous**	**Log transformed**
Newyear	**Independent**	**Dichotomous**	**1 if day is 01 January; 0 otherwise**
Humanrights	**Independent**	**Dichotomous**	**1 if day is 21 March; 0 otherwise**
FreedomDay	**Independent**	**Dichotomous**	**1 if day is 27 April; 0 otherwise**
WorkersDay	**Independent**	**Dichotomous**	**1 if day is 01 May; 0 otherwise**
YouthDay	**Independent**	**Dichotomous**	**1 if day is 16 June; 0 otherwise**
HeritageDay	**Independent**	**Dichotomous**	**1 if day is 24 September; 0 otherwise**
ReconciliationDay	**Independent**	**Dichotomous**	**1 if day is 16 December; 0 otherwise**
ChristmasDay	**Independent**	**Dichotomous**	**1 if day is 25 December; 0 otherwise**
GoodwillDay	**Independent**	**Dichotomous**	**1 if day is 26 December; 0 otherwise**
Month1	**Independent**	**Dichotomous**	**1 if Month is January; 0 otherwise**
Month2	**Independent**	**Dichotomous**	**1 if Month is February; 0 otherwise**
Month3	**Independent**	**Dichotomous**	**1 if Month is March; 0 otherwise**
Month4	**Independent**	**Dichotomous**	**1 if Month is April; 0 otherwise**
Month5	**Independent**	**Dichotomous**	**1 if Month is May; 0 otherwise**
Month6	**Independent**	**Dichotomous**	**1 if Month is June; 0 otherwise**
Month7	**Independent**	**Dichotomous**	**1 if Month is July; 0 otherwise**
Month8	**Independent**	**Dichotomous**	**1 if Month is August; 0 otherwise**
Month9	**Independent**	**Dichotomous**	**1 if Month is September; 0 otherwise**
Month10	**Independent**	**Dichotomous**	**1 if Month is October; 0 otherwise**
Month11	**Independent**	**Dichotomous**	**1 if Month is November; 0 otherwise**
Sin6	**Independent**	**Continuous**	**Sine term of Fourier series with Period 6**
Cos6	**Independent**	**Continuous**	**Cosine term of Fourier series with Period 6**
Sin12	**Independent**	**Continuous**	**Sine term of Fourier series with Period 12**
Cos12	**Independent**	**Continuous**	**Cosine term of Fourier series with Period 12**
Sin18	**Independent**	**Continuous**	**Sine term of Fourier series with Period 18**
Cos18	**Independent**	**Continuous**	**Cosine term of Fourier series with Period 18**
Sin24	**Independent**	**Continuous**	**Sine term of Fourier series with Period 24**
Cos24	**Independent**	**Continuous**	**Cosine term of Fourier series with Period 24**
Lag70128	**Independent**	**Continuous**	**The 1st time lag**
Lag70152	**Independent**	**Continuous**	**The 2nd time lag**
Lag70176	**Independent**	**Continuous**	**The 3rd time lag**
Lag70200	**Independent**	**Continuous**	**The 4th time lag**
Lag70224	**Independent**	**Continuous**	**The 5th time lag**
Lag70248	**Independent**	**Continuous**	**The 6th time lag**
Sun	**Independent**	**Dichotomous**	**1 if day is Sunday; 0 otherwise**
Mon	**Independent**	**Dichotomous**	**1 if day is Monday; 0 otherwise**
Tues	**Independent**	**Dichotomous**	**1 if day is Tuesday; 0 otherwise**
Wed	**Independent**	**Dichotomous**	**1 if day is Wednesday; 0 otherwise**
Thurs	**Independent**	**Dichotomous**	**1 if day is Thursday; 0 otherwise**
Fri	**Independent**	**Dichotomous**	**1 if day is Friday; 0 otherwise**
Fribtwn	**Independent**	**Dichotomous**	**1 if Friday preceded by a holiday; 0 otherwise**
Monbtwn	**Independent**	**Dichotomous**	**1 if Monday preceded a holiday; 0 otherwise**
Lngwknd	**Independent**	**Dichotomous**	**Long weekend**
Dec_closure	**Independent**	**Dichotomous**	**1 if period between 16 December and 01 January; 0 otherwise**
Winter_schoolholiDay	**Independent**	**Dichotomous**	**1 if period is during school closure in June/July; 0 otherwise**
Easter	**Independent**	**Dichotomous**	**1 if day is Easter; 0 otherwise**
Winter	**Independent**	**Dichotomous**	**1 if period is between June and August**