LONG TERM PEAK ELECTRICITY DEMAND FORECASTING IN SOUTH AFRICA USING QUANTILE REGRESSION

By

Norman Maswanganyi
9300387

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy in STATISTICS

in the

Department of Mathematical and Computational Sciences
Faculty of Science, Engineering and Agriculture
University of Venda, Thohoyandou, Limpopo
South Africa

Supervisor: **Prof Caston Sigauke**

Co-Supervisor: **Prof Edmore Ranganai**

Abstract

It is widely accepted that South Africa needs to maximise sustainable electricity supply growth to meet the new and growing demand for higher economic growth rates, especially in energy-intensive sectors. To diversify the energy mix, the country also needs to take urgent actions to ensure the sustainability of renewable energy and energy efficiency by 2030. Hence, it is important to provide a modelling framework for forecasting long-term peak electricity demand and quantifying uncertainty of future electricity demand for better electricity security management. In order to estimate and capture changes in long-term peak electricity demand, the study employed quantile regression (QR) based models, including hybrid models for assessing and managing electricity demand using South African data. The changes in long-term electricity demand depend on network location areas and the uncertainties within the energy sectors. Long-term peak electricity demand forecasting using QR models seems scarce in South Africa. The current study closes a gap by developing a modelling framework that can be used for future electricity demand forecasting. Although many studies have been done on short-, medium and long-term peak electricity demand forecasting, an investigation of the extremal quantile regression (EQR) model for forecasting electricity demand (based on combined economic and weather conditions) still needs to be explored as far as we know. Accurately predicting extreme electricity demand distributions would significantly mitigate load shedding and overloading and allow energy-efficient storage. This thesis identifies weather-related and non-weather-related factors using the EQR approach to modelling and estimating the error of extremely low and high quantiles of peak electricity demand. Results from the thesis show that EQR provides a higher level of detail and can model the non-central behaviour of electricity demand than the other models used in the study. The study has shown how the additive quantile regression (AQR) model can provide the highest predictive ability and create superior accuracy of the forecast results. Power systems reliability requires a probabilistic characterisation of extreme peak loads, which results in severe system stress and causes grid problems. Accurate predictions of long-term electricity demand are very important as such forecasts can be used in the timing and rate of occurrence of such extreme peak loads. The study used hybrid additive quantile regression coupled with autoregressive models and variable selection using Lasso for hierarchical interactions to examine the power system's reliability in random extreme peak loads.

Keywords: *Extreme quantile regression, Forecasting, Generalised additive model, Long-term peak electricity demand, Quantile regression.*

Declaration

I, NORMAN MASWANGANYI, [9300387], hereby declare that the thesis titled: "Long term peak electricity demand forecasting in South Africa using quantile regression for the Ph.D. degree in Statistics at the University of Venda, hereby submitted by me, has not been submitted for any degree at this or any other university, that it is my own work in design and in execution, and that all reference material contained therein has been duly acknowledged.

Signed (Student): .... Date: 08 july 2024.

Dedication

This work is dedicated to my family. Their patience and understanding throughout the completion of this thesis is highly appreciated.

Acknowledgment

I want to express my sincere gratitude to my academic supervisor Prof. Caston Sigauke and co-supervisor, Prof Edmore Ranganai, for their insights and inspirations. The information and positive feedback I received were always excellent, helpful, and encouraged me to update and improve my writing. I am particularly greatly indebted to my brother academically, Prof Daniel Maposa, for providing great advice that helped determine the thesis work. Finally, I would like to recognize all staff members and students in the Department of Statistics and Operation Research at the University of Limpopo for developing my career.

Table of contents

List of Figures

List of Tables

Abbreviations

AATCI	Average of average coastal and interior temperature
AFs	Availability factors
AmaxTCI	Average maximum of coastal and interior temperature
AminTCI	Average minimum of coastal and interior temperature
AMTC	Average monthly coastal temperature
AMTI	Average monthly interior temperature
ANN	Artificial neural networks
AQR	Additive quantile regression
AQTC	Average quarterly coastal temperature
AQTI	Average quarterly interior temperature
AR	Autoregressive
ARIMA	Autoregressive integrated moving average
ARIMA-GBR	Autoregressive integrated moving average-gradient boosting regression
ARIMA-SVR	Autoregressive integrated moving average-support vector regression
BAU	Business-as-usual
CDD	Cooling degree days
CI	Confinement index
COD	Coefficient of determination
CP	Coverage probability
CRPS	Continuous rank probability score
CSIR	Council for scientific and industrial research
CV	Cross validation
DAAMTCI	Difference between average of average monthly coastal and interior temperatures
DAAQTCI	Difference between average of average quarterly coastal and interior temperatures
DAH	Day after holiday
DBH	Day before holiday
DD	Daily demand
DH	Day Holiday

DmaxTCI	Difference between average maximum of coastal and interior temperatures
DminTCI	Difference between average minimum of coastal and interior temperatures
DoE	Department of energy
DPD	Daily peak demand
DPED	Daily peak electricity demand
DSS	Dawid–Sebastiani score
ELC	Electric power consumption
EM	Extremal mixture
EQR	Extreme quantile regression
ES	Element set
EPLF	Extreme peak load frequency
EVT	Extreme value theory
FCEH	Final consumption expenditure of household
FORs	Forced outage rates
fQR	Quantie regression forecasting
fQRA	Quantie regression averaging forecasting
fGAM	Generalised additive model forecasting
GA	Generic algorithms
GAM	Generalized additive model
GARCH	Generalized autoregressive conditional heteroskedasticity
GCV	Generalised cross-validation
GDP	Gross domestic product
GEV	Generalised extreme value
GEVD	Generalised extreme value distribution
GJR	Glosten–Jagannathan–Runkle
GLAD	Generalised lambda distribution
GLD	Generalised logistic distribution
GPD	Generalised Pareto distribution
GRM	Generation reserve margin
GRU	Gated recument unit
GSP	Generalised single Pareto distribution
GW	Gigawatt
HDD	Heating degree days
HEG	High economic growth
IEE	Improved energy efficiency
IID	Independent and identically distributed
IW	Interval width
KBES	Knowledge-based expert system

KSA	Kingdom of Saudi Arabia
LASSO	Least absolute shrinkage and selection operator
L_{Md}	Lower median
LAD	Least absolute deviation
LEAP	Long-range energy alternative planning
LEP-SMs	Long-term electrical power system models
LogS	Logarithmic score
LOLE	Loss of load expectation
LOLP	Loss of load probability
LQR	Linear quantile regression
LSE	Least squares estimation
LSTM	Long short-term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
maxTC	Average maximum coastal temperature
maxTI	Average maximum interior temperature
Md	Median
MILP	Mixed-integer linear programming
minTC	Average minimum coastal temperature
minTI	Average minimum interior temperature
MLE	Maximum likelihood estimation
MPED	Monthly peak electricity demand
MSE	Mean square error
MSRE	Mean-squared relative error
MW	Megawatt
NARANN	Nonlinear autoregressive artificial neural network
NDP	National development plan
NGM	Nonlinear grey model
NGM-ARIMA	Nonlinear grey model-autoregressive integrated moving average
NLQR	Nonlinear quantile regression
NML	Novel machine learning
NPOT	Nonparametric peaks-over-threshold
OFQR	Optimal forecasting quantile regression
OLS	Ordinary least squares
PDF	Probability density function
PINAD	Prediction interval normalised average deviation
PINAW	Prediction interval normalised average width
PL	Pinball loss

PLACQR	Partially linear additive conditional quantile regression
PLAQR	Partially linear additive quantile regression
POT	Peaks-over-threshold
PP	Point process
QQ	Quantile to quantile
QPED	Quarterly peak electricity demand
QRA	Quantile regression averaging
QR	Quantile regression
quantGAM	Quantile generalised additive model
REIPPPP	Renewable energy independent power producer program
RELG	Renewal electric power generation
RMSE	Root mean squared error
SEC	Saudi electricity company
SARIMA	Seasonal autoregressive integrated moving average
SGB	Stochastic gradient boosting
SVM	Support vector machine
SVR	Support vector regression
TP	Total populations
U_{Md}	Upper median
US	United states
USB	United states bancorp

List of Notation and Special Symbols

$Q_\tau(y_t \mid \boldsymbol{x}_t)$	Conditional quantile function.
$q_{Y\mid X}(\tau)$	Extreme conditional quantile function.
s	Smooth functions.
S	Seasonal component.
ρ_τ	Pinball loss function.
$\pi(t)$	Penalised cubic regression smoothing spline.
$\hat{\beta}(\tau)$	Regression coefficient for the quantile.
$y_{t,\tau}$	Peak electricity demand function.
$\varepsilon_{t,\tau}$	Quantile error term.
λ	Smoothing parameter.
$q_{\tau,t}$	Quantile function.
$PL(q_{\tau,t})$	Pinball loss function in extreme quantile.
$L(y_t, Q_\tau)$	Pinball loss function in quantile.
$L(y_t, q_\tau)$	Loss function.
$S(y, p)$	Scoring rule function.
p	Order of non-seasonal autoregressive process.
q	Order of non-seasonal moving average process.
P	Order of seasonal autoregressive process.
Q	Order of seasonal moving average process.
σ_a^2	Variance of a_t.
$a_{t,\tau}$	White noise with mean Zero and variance, σ_a^2.
m	Fixed length.
m_n	Maximum over n time units.
F_t	Domain of attraction of Fr$\acute{e}$chet distribution.
τ	Unknown threshold.
h	Bandwidth.

$\boldsymbol{I}$	Indicator function.
t_1	Elastic net turning parameter.
$noltrend$	Nonlinear trend.
$K(.)$	Kernel function.
$\pi(y\|\beta)$	Likelihood function.
ψ	Learning rate.
μ	Location parameter.
ξ	Shape parameter
σ	Scale parameter.
$fplLassoI$	Model with pairwise interactions.
$fplLasso$	Model without pairwise interactions.

Research Outputs

A list of research outputs from this thesis is given below.

Peer Reviewed Journal Publications

1. Maswanganyi, N., Ranganai, E. and Sigauke, C. (2019). Long-term peak electricity demand forecasting in South Africa: A quantile regression averaging approach, AIMS Energy, vol. 7, no. 6, pp. 857-882. http://dx.doi.org/10.3934/energy.2019.6.857

2. Maswanganyi, N., Sigauke, C. and Ranganai, E. (2017). Peak electricity demand forecasting using partially linear additive quantile regression models. South African Statistical Journal: Peer-reviewed Proceedings of the 59th Annual Conference of the South African Statistical Association for 2017, pp. 25-32. ISBN 978-1-86822-692-4. https://journals.co.za/doi/pdf/10.10520/EJC-c04fd0069

3. Maswanganyi, N., Sigauke, C. and Ranganai, E. (2021). Prediction of extreme conditional quantiles of electricity demand: An application using South African data, Energies 2021, vol. 14, no. 20, pp. 6704. https://doi.org/10.3390/en14206704

4. Sigauke, C., Kumar S., Maswanganyi N. and Ranganai E. (2018). Reliable Predictions of Peak Electricity Demand and Reliability of Power System Management. In: System Reliability Management: Solutions and Technologies. Edited by Anand A. and Ram M. CRC Press, Tay-

lor and Francis, 1st Edition, Chapter 10. ISBN 9780815360728, eBook ISBN 9781351117654. https://doi.org/10.1201/9781351117661-10

Conferences attended

1. Maswanganyi, N., Sigauke, C. and Ranganai, E. (2017), Peak electricity demand forecasting using partially linear additive quantile regression models, 59th Annual conference of the South African Statistical Association (SASA 2017), 27-30 November 2017, University of the free state (Bloemfontein).

2. Maswanganyi, N., Ranganai, E. and Sigauke, C. (2018), Long-term Peak electricity demand forecasting in South Africa: A quantile regression averaging approach, 60th Annual conference of the South African Statistical Association (SASA 2018), 26-29 November 2018, University of South Africa (UNISA campus).

3. Maswanganyi, N., Sigauke, C. and Ranganai, E. (2019), Forecasting extreme conditional quantiles of electricity demand in South Africa, 61st Annual conference of the South African Statistical Association (SASA 2019), 25-29 November 2019, Nelson Mandela University (PE).

Summary of main contributions from the papers

This thesis consists of four peer-reviewed journal publications.

1. Paper 1 improves and extends the linear quantile regression (LQR) models for long-term peak electricity demand by including heating and cooling degree days and hierarchical pairwise interactions. The paper discusses a comparative analysis of ordinary least squares (OLS), linear quantile regression (LQR), QRA and generalised additive models (GAMs) in monthly peak electricity demand (MPED) and quarterly peak electricity demand (QPED) forecasting. The paper considered the analyses of support vector regression (SVR) and stochastic gradient

boosting (SGB) as two consistent benchmark models, especially with current trends in electricity demand forecasting. The paper further explores the inclusions of monthly and quarterly temperature covariates in the long-term peak electricity demand parameters through heating and cooling degree days. This paper discusses the two variable selection methods, least absolute shrinkage and selection operator (Lasso) and elastic net. It has been shown in this paper that the QRA model performs better for long-term peak electricity demand in QPED data than in MPED data.

2. Paper 2 extends the partially linear additive quantile regression (PLAQR) model by coupling it with an autoregressive (AR) model and including a nonlinear trend variable as one of the covariates. The results of this paper are presented with a comparative analysis of the developed models, one with pairwise interactions and another without interactions. The forecast results from the models are combined using a forecasting combination algorithm, which improves the forecast accuracy. This paper analyses the average loss suffered by PLAQR with pairwise interaction and without interaction based on the pinball loss function. The paper includes a nonlinear trend covariate using a penalised cubic smoothing spline.

3. Paper 3 models the upper tail of the distribution of extreme peak electricity demand using regression quantile and extremal mixture models The paper uses AQR, extremal mixture (EM) and nonlinear quantile regression (NLQR) models to predict extreme conditional quantiles. The model performances are compared based on the probability scores to select the best model. The comparative prediction evaluations in this study are of interest in comparing models. The comparison addresses the number of uncertainties within the energy sector and quantifies the uncertainties in the estimated distribution parameters. Predicting extremely high quantiles of daily peak electricity demand could help system operators know the greatest demand that will enable them to supply adequate electricity to consumers and shift demand to off-peak periods. They are also necessary for planning power systems and assessing investment projects in South Africa. To the best of our knowledge,

the prediction of extreme conditional quantiles of daily peak electricity demand in South Africa using hybrid additive quantile regression, non-linear quantile regression, and extremal mixture models has not been performed previously .

4. Paper 4 examines the power systems reliability in the presence of random extreme peak loads using additive quantile regression (AQR) models. It summarises the forecast accuracies of quantiles and extends the significance of AQR models on long-term peak electricity demand forecasting. It presents the system reliability analysis and applicability of regression models for daily peak electricity demand.

Chapter 1

General introduction

This chapter gives the general framework of the study. The long-term electricity demand forecasting as a comprehensive solution for electricity demand and how it deals with long time horizons. The study elaborate challenges that South Africa faces in implementation of unified approach to electricity demand forecasting. The study methods such as QR models are well discussed and indicated as most useful in electricity demand forecasting. Moreover, the background , statement of the problem, aim and objectives of the study, significance of the study, contributions and outline of the thesis are also discussed.

1.1 Background

The South African National Electricity Control Centre at Germiston controls the power transmission network throughout the country. South Africa's public power utility (Eskom) and the government's Department of Energy (DoE) are aware of the basic demand for the country and its neighbouring countries, to which they also supply electricity. The country's unpredictable economic growth is increasingly focusing on industrialisation and its universal electrification programme and taking electricity into the length and breadth of the country. In order to meet the demand for electricity in South Africa, long-term electricity demand forecasting should be conducted to cover the total requirement for electricity. Forecasting long-term electricity demand helps plan and allocate resources. Furthermore, it helps diversify its output to stabilise its income over time. In addition, driven by population and economic growth, South Africa's electricity demand is expected to increase in the next 20 to 30 years. Forecasting electricity demand in a developing country such as South Africa is challenging due to weather-related factors such as temperature, economic, and calendar effects. Generally, electricity demand is non-stationary and exhibits an upward trend due to economic and population growth. However, forecasting demand is becoming the lifetime business in the world, where the tidal waves of change are sweeping the most established structures inherited by human society. Electricity demand forecasting is a systematic process that involves anticipating an organisation's future demand for electricity and services under a set of uncontrollable and competitive forces. Hyndman and Fan (2010) describe comparatively that electricity demand forecasting is divided into three groups, namely, short, medium and long-term electricity demand forecasting. Short-term electricity demand forecasting works nicely with hourly, daily, weekly, or monthly forecasts. Electricity demand forecasting is well covered in much literature, and electricity demand studies are very important in South Africa. Short-term electricity demand forecasting received considerable attention due to its significance for power system planning, electricity market, unit commitment and economic dispatch control than medium and long-term forecasting. Medium-term and long-term forecasting are less explored as compared to short-term forecasting. This study considers long-term electricity demand forecasting as a comprehensive solution for electricity demand. Long-term electricity demand forecasting normally deals with long time horizons. It also covers forecasting for one year or ten years and sometimes up to three decades. Furthermore, it is sometimes characterised by large uncertainties in the distant future. The uncertainty causes long-term electricity demand forecasting to be challenging to forecast more accurately over a planning period. This study focuses on analysing electricity demand using extreme Quantile regression (EQR) as an approach that uses Quantile Regression (QR) models in forecasting long-term peak demand. Essentially, the QR approach estimates error quantiles derived from past verification of electricity demand forecasting and applies them to real-time forecasts. As introduced by Koenker and Bassett Jr (1978), QR is a statistical technique intended to identify more effects than linear regression. It does not restrict recognition to the conditional mean and therefore allows approximation of the full conditional distribution of a response variable (Davino et al., 2013). It also provides an alternative approach to the conditional mean method by inserting the estimation of conditional mean functions with procedures for estimating a whole family of conditional quantile functions. Furthermore, QR can provide a complete statistical analysis of the stochastic relationships among random variables. QR has also been applied operationally in other areas, particularly for assessing quantile forecasts in probabilistic wind power forecasts (Bremnes, 2006). Electricity demand forecasting methodology is divided into two broad categories, such as statistical and artificial intelligence techniques (Hong and Fan, 2016). Artificial intelligence techniques include knowledge-based expert systems (KBES), artificial neural networks (ANNs), and a fuzzy inference system model. These models are useful in electricity demand forecasting. The ANN models have received the most attention and are the most popular intelligent system technique. Statistical methods such as QR models link electricity demand forecasting to its driving factors by a statistical approach. The model parameters are estimated using QR models on historical electricity demand data and its components. Parameter-based electricity demand forecasting methods can be classified under three categories: regression methods, time series prediction methods, and grey dynamic methods. This study focuses on QR techniques with generalised additive, semi-parametric and non-parametric models for long-term peak electricity demand.

As Davino et al. (2013) put it in their study, the idea behind the QR model is that it estimates the entire conditional distribution of a dependent variable. It is, therefore, gradually advancing into a comprehensive strategy for accomplishing the regression picture compared to standard regression that gives an incomplete picture for a set of distributions (Koenker and Hallock, 2001).

1.2 Statement of the problem

This research addresses the critical issue of electricity demand forecasting, which has profound implications for industrial management decisions, capital investments, operations, tariffs, and more. The growth of heavy industry, particularly mining in South Africa, significantly impacts electricity demand.

Population growth, GDP, and living standards are closely linked to electricity demand. However, South Africa faces the challenge of implementing a unified approach to electricity demand forecasting, particularly using Quantile Regression (QR) models, which have shown promise but need to be more explored in this context. This research uses South African data to create QR-based models for long-term peak electricity demand forecasting, thereby enhancing our comprehension of this intricate matter

1.3 Aim and objectives of the study

Aim

The study aims to develop a modelling framework which can be used in the electricity sector for long-term peak electricity demand forecasting.

Objectives

The specific objectives are to:

- (Chapter 4): improve and extend the linear quantile regression (LQR) models for long-term peak electricity demand by including heating and cooling degree days and inclusion of hierarchical pairwise interactions.
- (Chapter 5): extend the partially linear additive quantile regression (PLAQR) model by coupling it with an autoregressive (AR) model and including a nonlinear trend variable as one of the covariates.
- (Chapter 6): model the upper tail of the distribution of extreme peak electricity demand using regression quantile and extremal mixture models,

- (Chapter 7): examine the power systems reliability in the presence of random extreme peak loads using additive quantile regression (AQR) models,
- improve and extend AQR models on long-term peak electricity demand by using hybrid models that combine generalised additive models (GAMS) with QR models,

1.4 Significance of the Study

Energy problems are complicated and require time or internal and external expertise to improve forecasting techniques' performance and efficiency. The study compares the developed quantile regression-based models using density forecasting for long-term peak electricity demand in South Africa. Density forecasting gives a complete description of the uncertainty associated with a prediction. Density forecasts are more helpful and are necessary to evaluate and limit the financial risk produced by demand variability and forecasting uncertainty (Hyndman and Fan, 2010). It represents a statistical generalisation of QR. The density forecasts offer the possibility to implement policies because they exploit the precise structure of the probabilities estimated for future demand. The aspect of density forecasting is only one aspect that assesses the performance of the forecast.

The study presents density forecasts and discusses construction, presentation, and evaluation issues. Density forecasts contain more valuable information than point forecasts, which only estimate the future demand's expected value. The density forecasts also improve decision-making. Moreover, they minimise costs and risks in electricity markets due to their representation of uncertainty. Despite their value for grid companies, density forecasts of the long-term peak electricity demand still need adequate recognition in the literature McSharry et al. (2005). In particular, this study suggests that the QR model is one approach that can be used for density or probabilistic forecasting and for predicting extreme peak electricity demand in South Africa. QR is a robust method that can model the full distribution ignored by ordinary least squares and other approaches that focus on point forecasting. For example, the QRA and PLAQR models in Chapters 4 and 5 give the most accurate predictions compared to the OLS regression model.

1.5 Contributions

Considering an overview in Chapter 2, the main contribution of this study is to show the application of QR models for long-term peak electricity demand in South Africa by estimating the probability distribution of the possible future values of that demand. The process compares the semi-parametric extremal mixture model, QR and two benchmark models. Using South African data, QR models are considered in predicting the extremely high and extremely low quantiles of electricity demand. When modelling extreme high and low quantiles of long-term peak electricity demand, the AQR model can be considered the best model compared to other models. In the case of the possible largest electricity demand, predicting extremely high quantiles of DPED using the AQR model could assist the South African government and other operators in knowing when to supply adequate electricity to consumers. The contributions of the study are as follows:

- Many statistical applications focus on the lower or upper quantiles of the distributions. In this study, the QR and extremal mixture models estimate extremely low and extremely high quantiles of electricity demand using South African data.

- There have been many empirical studies on electricity demand. However, the peak electricity demand literature seems scarce in South Africa on the issue of long-term electricity demand, and this study closes a gap by introducing new QR approaches to forecast long-term peak electricity demand.

- The study introduces nonlinear trends as a covariate for QR models using penalised cubic smoothing splines to assess forecasting peak electricity demand. The proposed models have been shown to provide more accurate results than current methods.

- In this challenging time of load-shedding, the problem with South African electricity supply is a confluence of factors such as the absence of adequate planning, resources and scheduled maintenance. The study proposes new QR models for electricity demand forecasting to assist the system operators in knowing the possible largest demand, enabling them to supply adequate electricity to consumers and shift load to off-peak periods. The electricity demand faces challenges because of growing customer demand and new technologies in South Africa. Hence, selecting appropriate models on electricity demand is key in long-term electricity forecasting.

- The two drivers, such as heating degree days (HDD) and cooling degree days (CDD), are reviewed in Section 2.2 by Buechler et al. (2020) and analysed as detailed in Chapter 4 (quantile regression averaging approach paper). Including HDD and CDD drivers in this study is necessary in the current context of electricity demand analysis. Very few authors, including Buechler et al. (2020) and Bazmia et al. (2012) among others, have considered including HDD and CDD drivers in their studies. Importantly, the two drivers have been included in the study as they are the strongest drivers of long-term peak electricity demand.

- The long-term peak electricity demand forecasting, especially for extreme conditions, would be important for policy-making regarding the utilisation of renewable energies. The quantile regression methods proposed in this study will be useful for such a purpose.

1.6 Outline of thesis

The thesis is organised as follows: Chapter 1 introduces the background, aim and objectives of the study. This chapter focuses more on the QR techniques for predicting long-term peak electricity demand. Moreover, the summary of the problem statement, the significance of the study, the list of the main contributions, and the thesis outline are also provided in Chapter 1.

Chapter 2 reviews the literature on long-term peak electricity demand forecasting using different models. It demonstrates how QR models are used as compared to other models. It discusses past forecasts' strengths, weaknesses, successes, and challenges obtained from individual methods. Moreover, the chapter shows how the empirical results of each study are significant for the role of electricity demand predictions, portfolio management and risk modelling. Chapter 2 discusses a detailed review of articles on long-term peak electricity demand in South Africa and globally. The latest reviewed articles provided in Chapter 2 on long-term peak electricity demand globally include Lai et al. (2021); Alkhraijah et al. (2021); Gebremeskel et al. (2021); Jang et al. (2020); Guo et al. (2021); Agyei-Sakyi et al. (2021); Buechler et al. (2020); Alhajeri et al. (2020); Saba and Elsheikh (2020); Lu et al. (2021) and Norouzi et al. (2020) among others. Likewise, the recent articles reviewed on long-term peak electricity demand in South Africa are Mokilane (2018), Mokilane et al. (2018), Wright et al. (2019); Maswanganyi et al. (2019) and Ma and Wang (2020). The articles mentioned above are considered key to the theories used in this study. Their contributions are very important for this study and similar research areas in South Africa. In summary, the reviewed articles in Chapter 2 show the gaps in including weather variables or using temperature as explanatory variables in applying QR methods. The chapter also revealed another gap in the theory of long-term electricity demand predictions as advocated by the scarcest of literature in South Africa. To achieve the study's objectives, Chapter 3 explores in more detail the methodological approach, including how methods have been used for similar data or quantitative empirical analysis. This chapter thoroughly evaluates quantile regression-based models for long-term peak electricity demand in South Africa. In addition, the chapter concludes with EQR models used to predict extreme high and extreme low conditional quantiles. The chapter explores the different techniques largely preferred in electricity demand data analysis. This chapter is concerned with authors relying on a single methodology for determining the electricity demand of such a fast-growing and changing system. The models in this chapter are discussed and explained based on their performances.

Chapters 4 to 7 have been submitted individually for publication; some are independent. However, the notations are consistent throughout the study. The detailed presentations for chapters 4 to 8 are given below.

In Chapter 4, the generalised additive and quantile regression averaging models are used in comparing the forecasting accuracy of monthly peak electricity demand and quarterly peak electricity demand data. The ridge, least absolute shrinkage and elastic net coefficients are estimated and compared using monthly and quarterly peak electricity demand data.

In Chapter 5, partially linear additive quantile regression models are applied to model and forecast daily peak electricity demand using South African electricity demand data sets. The findings in Chapter 5 show that the electricity demand in South Africa is highly sensitive to cold temperatures. The variable selection in Chapter 5 is made using the least absolute shrinkage and selection operator via hierarchical pairwise interactions. The main effects must be in the model if higher-order interactions are concluded. The results in Chapter 5 have revealed the usefulness of partially linear additive quantile regression models. The inclusion of a nonlinear trend in this chapter is determined using a penalised cubic smoothing spline.

Chapter 6 focuses on predicting extreme high and extreme low conditional quantiles of electricity demand using South African data. It extends Chapters 4 and 5 regarding the existing quantile regression forecasting models by including the extremal mixture models in estimating the extremely high and low electricity demand quantiles. In Chapter 6, the scoring rules, namely continuous ranked probability score, logarithmic score, Dawid-Sebastiani score, pinball loss and interval width, are used to compare the additive quantile regression, extremal mixture and nonlinear quantile regression models. Chapter 6 revealed that the distribution of daily peak electricity demand data during $2007-2011$ and $2017-2021$ are the same and change rapidly for the other three distributions. The three models mentioned in Chapter 6 are very important in predicting the extreme high and low electricity demand peaks as they yield accurate predictions. Accurate prediction of extreme electricity demand distribution is important to decision-makers in the electricity sector and should be monitored regularly.

To have accurate and reliable daily peak electricity demand forecasts, Chapter 7 presented two reliability indices: generation reserve margin and load probability. The generation reserve margin is the reliability index that measures how a power system exceeds peak demand. Chapter 7 revealed that the power system reliability requires a probabilistic characterisation of extreme peak loads, which produces severe stress to the system and causes problems to the grid. Chapter 8 summarises modelling by chapters, followed by a modelling discussion and summary of key findings. The chapter finalises with an overview of future research studies.

Chapter 2

Literature review

2.1 Introduction

A wide range of statistical methods has been suggested in the literature on forecasting electricity demand. This chapter reviews the main concepts used as the basis of this thesis. This study is concerned primarily with the performance of QR models, which are in and out of sample to forecast long-term peak electricity demand. In order to deal with long-term uncertainties within the energy sector, long-term peak electricity demand forecasting models which produce accurate forecasts are needed in decision-making and for planning purposes. Inaccurate forecasting may result in poor decision-making and bad policy establishment. For instance, the South African government, electricity consumers, producers and industries need accurate forecasts for future events to assist in long-term development plans for the country to minimise risks. The long-term electricity demand forecasting methods depend on the choice of variable selections and, very importantly, assumptions made for the long run.

The following demonstrates how the methods are reviewed and are given in separate sections developmentally: Section 2.2 reviews several approaches in the relevant literature on long-term electricity demand globally. It discusses the previous and most recent studies of worldwide electricity demand forecasting. The section also highlights the limitations of some traditional statistical models in predicting electricity demand. In contrast, Section 2.3 reviews the current literature and methodologies to forecast long-term peak electricity demand in South Africa. The section introduces electricity demand forecasting methodologies and discusses the most important characteristics of electricity demand, consumption and distributions. The limitations and shortcomings are identified and also presented. Section 2.4 addresses the techniques of long-term electricity demand. Moreover, Section 2.4 reviews QR methods and discusses the gaps in statistical techniques, and Section 2.5 concludes.

2.2 Review of articles on long-term peak electricity demand (global review)

Many researchers have dealt with long-term electricity demand forecasting globally using different models. For example, long-range energy alternative planning (LEAP) model, long-term electrical power system models (LEP-SMs), multilevel models, artificial neural networks (ANN), and time series analysis models have been used for long-term electricity demand forecasting (Perwez et al. (2015); Mir et al. (2020); Lai et al. (2021); Küçükdeniz (2010) Koen et al. (2014); Al-Saba and El-Amin (1999) and Ringwood et al. (2001)) among others.

In order to review the expected short-and long-term impacts of social distancing on the electricity demand in the Kingdom of Saudi Arabia (KSA), this study follows the work done by Alkhraijah et al. (2021) who use a case study to observe a reduction in the time required for the electricity demand to respond to temperature changes. Comparing 2020 and 2019, the study's results by Alkhraijah et al. (2021) clearly show that the peak electricity demand and energy consumption in April have increased by 3.53% and 3.15%. The results also show that the KSA has extremely high temperatures, and the annual peak demand typically coincides with hot weather in the Northern Hemisphere from June to August. Moreover, the results show that the correlation between the temperature and demand has increased dramatically during the full curfew period in Saudi Arabia.

To provide the information for a better understanding in interpreting long-term scenarios and the expectations of bioenergy's role in 2050, Szarka et al. (2017) review eight studies carried out in Germany. The authors describe, analyse and explain the role of bioenergy within the national energy system. A comparative analysis of the scenarios is performed by selecting, defining, quantifying, and explaining bioenergy-related indicators that characterise the energy system as a whole, considering the availability of high-quality data. Examples of these scenarios are:

- target scenarios, strategies, or pathways for reaching the set goals.
- explorative scenarios, which show the consequences of certain policy decisions.

In their findings, it is recommended to introduce standards for definitions and methodologies or minimum requirements for data quality and documentation.

A study by Gebremeskel et al. (2021) proposes the LEAP model to forecast the long-term evolution of energy and electricity demand in Ethiopia. The authors explore six different scenarios to unfold the future change, namely, business-as-usual (BAU), growth in electrification and urbanisation (E and U), high economic growth (HEG) and three policy-driven. The improved energy efficiency results (IEE-1, IEE-2 and IEE-3) show that applying energy efficiency policies and efficiency measures would only have a minor impact on the electricity demand. Hence, the application of energy efficiency policies is an important measure to fight the electricity demand mismatch causing power shortages and blackouts.

Cloud-based forecasting using a neural network (NN) model is proposed by Altinoz and Mengusoglu (2015) for long-term electricity demand forecasting in Turkey. The artificial neural fuzzy and neural network models for long-term electricity demand forecasting using historical temperature and demand data are considered. The results show that temperature and economic data are preferred as inputs for the proposed model. However, the study observed the following outcomes:

- Long-term forecasting using the proposed model is inaccurate and increasingly case-dependent.
- hourly electricity demand is affected by temperature.

According to Jang et al. (2020), long-term daily peak electricity demand forecasting plays a significant role in the planning of power systems. The study proposes a new approach: the nonhomogeneous generalised extreme value distribution (GEVD) model. The study uses the temperature and electricity demand data collected from 2002 to 2016 in Austin, Texas. The results show that the proposed approach quantifies the uncertainties in an integrative framework and provides useful insights into peak electricity demand's long-term evolution. The limitations of the study are:

- The authors use the Austin temperature data measurements and generalise them to represent the overall temperature patterns.
- GDP variable is not considered. Hence, it is one of the factors very important in determining electricity.

An early attempt is done by Guo et al. (2021) on the daily electricity consumption data for the building's full sample (n=117). The data is divided into a training dataset (n=89) and a testing dataset (n=28). The study proposes an autoregressive integrated moving average-support vector regression (ARIMA-SVR) method to predict electricity consumption from 1 to 28 days. The proposed model performances are done based on MSE, RMSE and MAPE. The model has been devised to be compared with ARIMA, autoregressive integrated moving average-gradient boosting regression (ARIMA-GBR), long short-term memory (LSTM) and gated recurrent unit (GRU) models. The strengths of the ARIMA-SVR model are as follows:

- It can be developed using a small training dataset while maintaining high accuracy.
- It does not require any additional variables; however, it is based on a value of its historical observations.
- It is very flexible and explanatory.

- It combines the advantages of both ARIMA and SVR models.
- It could extract both nonlinear and linear features.

The results show that the proposed model could improve the prediction performances of the ARIMA model to a certain degree when combined with nonlinear models that can capture nonlinear features in the dataset. This, of course, is assumed that using the ARIMA model alone could not capture the data characteristics adequately. However, the proposed model could improve the accuracy of forecasting efficiency in certain prediction horizons. The shortcomings of the study using the ARIMA-SVR model are:

- The proposed model depends on time series as it requires the univariate time series to be stationary or stationary after differencing the data.
- The study fails to use temperature and humidity in model construction.
- Estimating seasonal patterns in time series analysis using long-term electricity demand data could be a challenging task due to the short data history.

Recently, Agyei-Sakyi et al. (2021) analysed the determinants of electricity consumption and volatility-driven innovative roadmaps to one hundred per cent renewables for top consuming nations in Africa. The study proposes a novel machine learning (NML) model and volatility transition up to 2030. It utilises annual data from 1990 to 2019 on electric power consumption (ELC) and renewable electric power generation (RELG) extracted from energy data's Global Statistical Yearbook 2020 and total populations (TP) and

GDP from the World Bank's world development indicators. The study compares the electric power systems of the top four energy-consuming African countries: Egypt, Algeria, Nigeria and South Africa. It also illustrated that radical renewable electricity generation innovations are required in the post-2020 COVID-19 pandemic era to provide affordable and green electricity to all country populations. To ensure access to affordable, green and sustainable electric power for all by 2030, uncovering the determinants of domestic electric power demand would help decision-makers in top consuming African countries like South Africa to make an informed decision and implement sustainable policies (Agyei-Sakyi et al., 2021).

Using the hourly electricity consumption data for exploring global changes in electricity demand during the COVID-19 pandemic, Buechler et al. (2020) apply HDD and CDD for United States regions in the model. The linear regression model for predicting daily mean demand is used. The following factors are also considered: daily demand (DD), temperature variation, day of the week, seasonal factors, load growth and holiday effects. Other variables, such as economic factors affecting demand in specific regions, are also considered. OLS regression models the relationship between confinement index (CI) or mobility and electricity change. The linear regression model is fitted with dummy variables, and results show that a dummy variable improves model performance.

To address the inconsistent trend in long-term planning, Li and Jones (2019) uses the point process (PP) approach in extreme value theory to forecast long-term electricity demand. The method is proposed to model maximum substation demand as a function of trends in customer count, average demand and photovoltaic capacity. The proposed approach provides a way to predict maximum substation demand that leverages energy's stability and explanatory ability. The R package estimates trends and daily maximum load data from 2008 to 2018 at two substations in Western Australia. The study results show consistent outcomes between energy consumption and maximum demand forecasts, solving a long-term risk of inconsistent trends. However, the study's shortcomings are that the data was insufficient for the new technologies in applying the PP approach.

The long-term electricity demand forecasting and renewable energy sector can play a fundamental role in fighting the electricity crisis in many countries worldwide. The following studies have been reviewed in determining the impact of the COVID-19 pandemic on peak electricity demand. For example, regression and genetic algorithms (GA) models in the State of Kuwait (Alhajeri et al. (2020)); nonlinear autoregressive artificial neural network (NARANN) model in Egypt (Saba and Elsheikh (2020)); long-term scenarios in Saudi Arabia (Alkhraijah et al. (2021)); hybrid multi-objective optimiser-based models in United States (US) (Lu et al. (2021)); regressive and neural network models in China (Norouzi et al. (2020)); among others. Their results generally show that their methods will assist in planning plans to deal with the effects of the COVID-19 pandemic and minimise its impact on electricity demand.

To review the theory of extreme values in the electricity demand, Li and Jones (2019) address the shortcomings of long-term planning by proposing a point process approach and Hor et al. (2008) use GEV theory and block maxima approach to estimate the maximum load forecast errors to assess long-term electricity demand projection. However, the scarcity of extremes and characterisations of extreme value modelling are regarded as major shortcomings in extreme value analysis (Coles, 2001). A study by Beirlant et al. (2004) uses a two-step procedure based on two case studies in estimating extreme conditional quantiles. The procedure is evaluated using small sample simulation for both heavy-tailed and right-bounded distributions. The limitation of QR in estimating extreme conditional quantiles is that a one-step extreme conditional quantile procedure based on QR underestimates these conditional quantiles. The introduction of a two-step extreme conditional quantile procedure suggested overcoming this limitation (Beirlant et al., 2004). However, the results show that the two-step procedure could have been more useful for practical purposes Beirlant et al. (2004), which is not considered in this study. The work of Fisher and Tippett (1928) is very influential in presenting the background of extreme value theory to the extent that it discusses the limiting forms of the frequency distribution of the largest and smallest member of a sample. Although literature related to the development of a statistical theory of extreme values began in the 1920s, more information on the historical work of extreme value problems can be found on Berning (2010) followed by (Moroke, 2019). To build on the work of EVT to develop EQR models, Chernozhukov (2005) studied the asymptotic theory of extremal quantile regression. In this context, the study obtains the large sample properties of extremal (extreme order and intermediate order) quantile regression for the linear quantile regression models with conditional tails of the response variable. This is restricted to the domain of minimum attraction and is closed under the tail equivalence across conditioning values. Models such as parametric, semiparametric and nonparametric quantile regression are used to treat the asymptotic theory of extreme conditional quantile estimators. An example is work done by Daouia et al. (2013), who proposed kernel smoothing for extremal quantile regression. The paper uses nonparametric regression quantiles obtained by inverting a kernel estimator of the conditional distribution. The paper's main idea is to extend the asymptotics of the extreme conditional quantile estimator in a nonparametric regression model. Nonparametric models have also received considerable attention in papers such as Beirlant et al. (2004) and (Durrieu et al., 2015). Researchers such as d'Haultfoeuille et al. (2014) and Smith (1989) have developed semiparametric and parametric models by combining QRs in the tails and proposing some extensions based on the point process of high-level exceedances, respectively.

The extremal mixture models play a significant role in electricity demand data analysis. One of the advantages of using extremal mixture models is that it allows us to choose a flexible bulk model and tail model, which can fit non-extreme data (Hu, 2013). Even though this study is focused on semi-parametric, where semi-parametric additive models are used to establish the relationship between electricity demand and other variables (see Section 5.2.1), MacDonald et al. (2011) and MacDonald et al. (2013) make great contributions on extremal mixture models as they use nonparametric smoothing kernel density estimator for the bulk of the distribution for the tail model. The authors introduce a two-tailed extremal mixture model that overcomes the inconsistency in the cross-validation likelihood estimation of the bandwidth-heavy distributions. One of the advantages of using a nonparametric mixture model is that the tail model is more robust than other extremal mixture models. If the model is specified, the parametric extremal mixture model offers few flexibilities compared to the semi-parametric or non-parametric extremal mixture model (Hu, 2013). The works of MacDonald et al. (2011) and MacDonald et al. (2013) have combined the boundary-corrected kernel density estimator (boundary-corrected mixture model) to minimise the inherent bias of the kernel density estimator. However, the boundary-corrected mixture model can only be applied if a proper tail, pole and shoulder exist at the boundary. Furthermore, the boundary-corrected mixture model is more complex than the standard kernel generalised Pareto distribution (GPD) model, and the results are easy to summarise using sets of kernel density functions. To improve and extend the number of directions in extremal mixture models, Behrens et al. (2004) and do Nascimento et al. (2012) use parametric and semi-parametric mixture models to fit bulk distribution and GPD for the tail distributions, respectively. Both authors have applied the Bayesian approach, which takes advantage of prior information to compensate for the inherent sparse extreme data and assists in identifying the model components. considers GPD above the threshold and gamma distribution below the threshold citebehrens2004bayesian. In contrast, do Nascimento et al. (2012) included a weighted mixture of gamma densities for the bulk distribution and GPD as the tail distribution. In addition, Lee et al. (2012) extended the work done by Behrens et al. (2004) as they use the mixture of the exponential distribution below the threshold and GPD for the threshold excess. Their analysis shows that the exponential distribution mixture model is more flexible than others. However, the proposed model can not be appropriate for a process with a lower tail, a shoulder, or if it is zero at the boundary, and authors also support the idea by MacDonald et al. (2011), and MacDonald et al. (2013) of using boundary corrected mixture model. The other papers supported the use of extremal mixture models and have also been very instrumental in extremes literature, including Carreau and Bengio (2009); Frigessi et al. (2002); de Melo Mendes and Lopes (2004), and Zhao et al. (2010) among others.

2.3 Review of articles on long-term peak electricity demand in South Africa

Despite the extensive discussion of short-term electricity demand forecasting in literature, only some studies focus on long-term peak electricity demand forecasting in South Africa. To discuss a methodology for long-term electricity demand forecasting that is initially used for assisting with strategic planning for the South African branch of a multi-national company, Koen and Holloway (2014) have chosen multiple regression modelling as the forecasting technique for long-term forecasting. The paper discusses a multiple regression model and compares it with other approaches used in other studies. The results show that the forecasting methodology has a strong scientific basis and is suitable for supporting future strategic planning. Five covariates used include private consumption expenditure in Agriculture, rail freight and GDP, private consumption expenditure in domestic, population and GDP and coal and gold production volume indices. One of the paper's objectives is to fill a particular gap in energy research, such as predicting the future electricity demand in the South African economy and demography. However, the major conceptual limitation of this methodology is that it is not able to model the effect of variables that did not play a significant role in the historical data or of causal factors which could not be quantified (Koen and Holloway, 2014). On modelling and forecasting long-term daily peak electricity demand, Maswanganyi et al. (2017) apply the PLAQR model using South African data from January 2007 to December 2013. The 28 South African weather stations divided into coastal and inland regions are used. The variable selection is carried out using Lasso. The empirical results are presented with a comparative analysis of two developed models: pairwise and without interactions. Based on the pinball loss function, the average loss suffered by PLAQR with pairwise interaction is less than that of PLAQR without interaction. Moreover, based on root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE), PLAQR with pairwise interaction is better than PLAQR without interaction. The empirical results have also shown the usefulness of PLAQR models.

Rasuba et al. (2017) also propose a methodology for forecasting electricity demand in South Africa using the Council for Scientific and Industrial Research (CSIR) sectoral regression model. CSIR model is a collection of models that predict the expected electricity consumption in various electricity sectors to obtain a national forecast. The CSIR model mostly used in the study is a multiple linear regression, and it relates factors that can influence annual electricity consumption within individual sectors. The regression model has been interpreted as a purely statistical (empirical) relationship, not a cause-and-effect relationship. The long-term electricity demand forecasting of predictor variables is inserted into the regression models to create 25 years of forecasts of annual electricity demand per sector under different scenarios using historical data and forecasts and Eskom sectors. Predictor variables such as final consumption expenditure of household (FCEH), mining index, population and FCEH, among others, are used. CSIR model, Eskom and Econometric approaches are compared to assess the relative advantages and disadvantages. Rasuba et al. (2017) also indicate that it is not possible to model price elasticity successfully at the national level only, which is supported by Inglesi (2010), which provides a set of forecasts that attempt to incorporate price elasticity into the modelling of electricity forecasts.

In conclusion, although it is good to argue that price might affect demand, especially given the dramatic price increase proposed by Eskom, Rasuba et al. (2017), however, it fails to give historical data available to use as a basis for forecasting its effect quantitatively. One of the major areas for improvement of the CSIR model is that it does not include temperature as a covariate. However, the temperature is known to be a major driver of electricity demand. The recent papers that use long-term electricity demand forecasting in South Africa include Mokilane (2018); Hedden and Hedden (2015); Mokilane et al. (2018); Inglesi and Pouris (2010); Mokilane et al. (2019); among others. As effective measures to rapidly reduce the demand for electricity and deal with future electricity prices and government policy uncertainties within the energy industry in South Africa, Maswanganyi et al. (2019) compare the forecasting accuracy of MPED and QPED data using OLS, QR, GAM and QRA models. The paper uses South African data to present a QRA approach for long-term peak electricity demand forecasting. The empirical results show that the QRA model performed better in QPED data than in MPED data.

Moreover, Ma and Wang (2020) use South African energy consumption data from 1988 to 2016 for the prediction of energy consumption variation trends in South Africa based on the ARIMA model, nonlinear grey model (NGM), nonlinear grey model-autoregressive integrated moving average (NGM-ARIMA) models. The study uses 14 14-year steps ahead forecast to predict South African energy consumption. The results show that South African energy consumption prediction from 2017 to 2030 has a stable growth tendency. Moreover, the NGM-ARIMA model gives more accurate predictions compared to other models. The study provides scientific and reliable data support for predicting South African energy. The shortcomings of NGM, ARIMA and NGM-ARIMA models are:

- The models only forecast according to the historical change track of the data sequence itself.
- Construction of the models is a combination of prediction steps.
- They cannot accurately measure the data affected by multiple factors.

Wright et al. (2019) in their study proposes the mixed-integer linear programming (MILP) model for long-term electricity sector expansion planning to present a generation capacity expansion optimisation scenarios in South Africa. The three long-term generation capacity expansion planning scenarios considered in the South African electricity system are business-as-usual, least-cost and decarbonise scenarios. The study applies the MILP model to co-optimise energy and ancillary services from 2016 to 2050. The results show that South Africa has a unique opportunity to transition from an existing carbon dioxide (CO_2) and low water intensity electricity system in the long-term at a least-cost scenario. The decarbonised electricity system is feasible for South Africa and is more expensive than the least-cost scenarios. The limitations of the MILP model are explicit reserve classes with unique temporal characteristics and minimum stable levels.

Koen et al. (2014) use multilevel models to develop long-term scenario forecasts for South African load profiles. The multilevel modelling is then used to support decisions regarding the electricity generation capacity required. The results show that the proposed modelling is better at supporting electricity generation and expansion decisions for long-term forecasting. To show the quality of fit measurements in regression quantiles using the elemental set (ES) technique, the work of Ranganai (2016) uses the Halda data set to illustrate applications of ESs, QR and OLS methods. An example is when the author proposes a coefficient of determination (COD) measure and model selection indices based on the ES method. The author explains why OLS is still a preferred statistical technique in data analysis. Nevertheless, the OLS technique is susceptible to outliers under the heavy-tailed error distributions. In contrast to OLS modelling, QR performs better than the OLS technique. To a large extent, one of the quantile regression's most appealing features is that it helps describe the relationship between the covariates and the dependent variable, not only on the median but also on the tail of the conditional response variable. Furthermore, it also reveals the risk of immediate changes in the independent variable and their effect on the response variable.

Earlier studies applied the different QR models to forecast electricity demand. These examples are Gibbons and Faruqui (2014), which develop an optimal forecast quantile regression (OFQR) method and compare the results with OLS. The paper focuses on OFQR as an alternative proposed solution to OLS; Lancaster and Jae Jun (2010) explains how the Bayesian method can be applied to QR. The authors give an explicit form for the posterior density of quantiles and compare it with Jeffreys prior; Soyiri et al. (2013) use multistage quantile regression approach to predict excess demand for health care services.

2.4 Discussion of the identified gaps

To quantify the uncertainty of long-term peak electricity demand and address what type of model is required in the future, the study has to find the best-fitting model that provides the highest predictive accuracy. To achieve this or close the existing gaps in modelling, this study discusses the impact of weather and electricity drivers. It uses QR techniques to reduce uncertainties in long-term peak electricity demand. Based on the reviewed articles, the work done by some researchers, which include Rasuba et al. (2017), Guo et al. (2021) and Li and Jones (2019) among others, need to include weather or temperature as an explanatory variable. The long-term peak electricity demand is the highest over a specific period. It depends on the economy, weather, seasons, and days of the week. The temperature covariate is considered in this study based on its relationship with electricity demand, as detailed in Chapter 4 (Section 4.2.1). Some other papers that use temperature as a covariate include Alkhraijah et al. (2021); Altinoz and Mengusoglu (2015); Jang et al. (2020); Agyei-Sakyi et al. (2021); Hyndman and Fan

(2010); Kaza (2010); Blázquez et al. (2013); among others. However, those papers have yet to explore including temperature in monthly and quarterly temperature covariates in their parameters through heating and cooling degree days.

This study uses the QR model as an appropriate methodology for accomplishing this task to understand how the covariate affects the response variable. The study uses QR models to analyse the DPED data set as detailed in Chapters 5 and 6. Moreover, the PLAQR model in Chapter 5 and AQR model in Chapter 6 have produced the most accurate predictions compared to other models, such as the CSIR model by Rasuba et al. (2017) and artificial neural fuzzy and neural network models by Altinoz and Mengusoglu (2015) that have failed to give accurate predictions in long-term electricity demand forecasting.

Many statistical applications focus on the distributions' lower or upper quantiles. Consequently, the theory of extreme value techniques has been extensively used in recent years. In contrast to the diverse literature on the estimation of extreme quantiles, which includes those of Elamin (2018); Sigauke et al. (2013); Li and Jones (2019); Diriba et al. (2015); Sigauke and Nemukula (2020) among others, this study gives the comparative analysis of the performances of AQR, EM and NLQR models as detailed in Chapter 6. The two machine learning techniques, SVR and SGB, are benchmark models. Moreover, the study uses South African electricity demand data to focus on extremely high and low conditional quantiles. As such, it has yet to be carried out elsewhere to the best of our knowledge. To ensure an accurate, affordable, reliable and sustainable long-term electricity demand forecast, this study proposes a model that could assist the South African government to know how much electricity is needed during the heatwave period. Furthermore, the study recognises the need to assist the country in attaining long-term electricity demand sustainability by yielding accurate predictions. An example of this is the study carried out by Ma and Wang (2020). This study shows how QR methods could help overcome the limitations and pitfalls of methods mentioned in Sections 2.2 and 2.3.

2.5 Conclusion

In the literature review, different traditional regression methods (including quantile regression) have been applied to electricity demand. However, the methods did not consider the extremal quantile distributions for long-term peak electricity demand forecasting in greater detail. The different methods applied have limitations as they failed to prove they are useful for practical purposes, especially when modelling extreme peaks in electricity demand. The existing literature excludes important variables such as weather and other drivers of electricity demand. This study closes the gaps by considering the temperature and other weather variables. Moreover, the study has also considered QR techniques, namely AQR, EM, and NLQR, which are compared using the historical electricity demand data. These models are important in predicting extremely high peaks and low electricity demand as they yield accurate predictions. Therefore, the QR techniques currently used in this study fill a particular gap within the long-term peak electricity demand forecasting domain in South Africa. On the other hand, the accurate prediction of the extreme electricity demand distribution is important to decision-makers in the electricity sector and should be monitored regularly. The inaccurate method of prediction may negatively impact dealing with some challenges, including load shedding, and as a result, the challenge conspires to restrict economic recovery. Long-term electricity demand forecasting is essential to avoid the shortage of electricity that may hamper South African economic growth. South Africa needs reliable forecasts of electricity demand to mitigate the risk of electricity shortages in the future to guarantee sustainability.

Chapter 3

Methods

3.1 Introduction

Chapter 3 outlines the methodology for fitting the monthly, quarterly and daily peak electricity demand using quantile regression models. It presents an overview of the quantile regression, generalised additive quantile regression, and then later explores the history of extreme value theory and extremal mixture models briefly. The methods such as semi-parametric extremal mixture, AQR and NLQR models with their implementations based on extreme quantiles are discussed. The AQR models are becoming increasingly popular in many applications as they are robust and flexible. Chapter 3 has also included a brief explanation of the scoring rules. This chapter shows how to build optimal reliable, and efficient computational tools with linear programming algorithms and give some insights for combining them with some quantile regression techniques. Linear programming also known as linear optimisation is a method to achieve the best outcome for example, maximum profit or lowest cost in a mathematical model whose requirements are represented by linear relationships. Analysing the impact of parametric setting using linear programming on the solution of interested variables help to improve the model. The study will improve and extends the existing QR forecasting methods based on parametric linear programming method to adjust the coefficient parameters of the linear programming problem. According to Portnoy et al. (1997), the more advanced linear programming algorithms with initial phase of preprocessing is designed for reducing the problem dimensions by picking out redundant variables and constraints. However, the normal preprocessing strategies for linear programming problems are not exceptionally well suitable to the applications of QR. The analysis of this study will be done using the "R" statistical package version 4.05 (Team, 2021). The linear programming formulation of the QR model in equation (3.14) will be modelled using the following algorithms: Barrodale and Roberts algorithm for l_1 -regression (Koenker and d'Orey (1987); Koenker and d'Orey (1994) and Koenker (1994)), Frisch-Newton interior point method (Portnoy et al. (1997)), Frisch-Newton approach after preprocessing (Portnoy et al. (1997)) and Frisch-Newton algorithm (Portnoy et al., 1997).

3.2 Quantile regression

3.2.1 Brief history of quantile regression

In the previous decades, many papers on QR modelling have been proposed by (Koenker and Bassett Jr, 1978) as one of the early pioneers. QR model offers an important interpretation as well as a statistical solution to solve the linear optimisation problem (Kaza, 2010). Moreover, it is robust to outliers and more flexible when the distribution of the outcome is not specified as parametric assumptions (Huang et al., 2017). Many statistical software programs use QR commands and procedures. However, the most widely used software for QR models is the R package 'quantreg' (Koenker et al., 2018).

3.2.2 Linear quantile regression model

QR was introduced by Koenker and Bassett Jr (1978) and is described in detail in Koenker (2005) as an extension of classical least-squares estimation of conditional mean models to conditional quantile function. An updated QR theory captures this method as a particular centre of the distribution, minimizing the weighted absolute sum of deviations (Hao and Naiman, 2007). The conditional quantile function is given in equation(3.1).

$$Q_\tau(y_t|\boldsymbol{x}_t) = \boldsymbol{x}_t^T\boldsymbol{\beta}_\tau + \varepsilon_{t,\tau}, \tag{3.1}$$

where $0 < \tau < 1$ and $Q_\tau(y_t|\boldsymbol{x}_t)$ denotes the conditional function for the τ^{th} quantile of the electricity demand. Given a set of training data $G = (y_t, \boldsymbol{x}_t^T)^T$ for $t = 1, ..., G$, the parameter $\boldsymbol{\beta}_\tau$ can be estimated as

$$\widehat{\boldsymbol{\beta}}_\tau = arg\ min \frac{1}{G}\left\{\tau \sum_{t:y_t \geq \boldsymbol{x}_t^T\boldsymbol{\beta}_\tau} |y_t - \boldsymbol{x}_t^T\boldsymbol{\beta}_\tau| + (1-\tau)\sum_{t:y_t < \boldsymbol{x}_t^T\boldsymbol{\beta}_\tau} |y_t - \boldsymbol{x}_t^T\boldsymbol{\beta}_\tau|\right\}. \tag{3.2}$$

The parameter $\widehat{\boldsymbol{\beta}}_\tau$ in equation (3.2) is usually estimated by using linear programming methods called least absolute deviation (LAD) regression. If $\tau = 0.5$, equation (3.2) is now given by

$$\widehat{\boldsymbol{\beta}}_\tau = \frac{1}{G}\sum_{t=1}^{G}|y_t - \boldsymbol{x}_t^T\boldsymbol{\beta}_\tau|. \tag{3.3}$$

Equation (3.2) can also be written as

$$\widehat{\boldsymbol{\beta}}_\tau = \frac{1}{G}\sum_{t=1}^{G}\rho_\tau(y_t - \boldsymbol{x}_t^T\boldsymbol{\beta}_\tau) = \frac{1}{G}\sum_{t=1}^{G}(\tau - \mathbf{1}_{y_t - \boldsymbol{x}_t^T\boldsymbol{\beta}_\tau < 0})(y_t - \boldsymbol{x}_t^T\boldsymbol{\beta}_\tau), \tag{3.4}$$

where ρ_τ is a check function such that $\rho_\tau(b) = \tau_b$ if $b \geq 0$ and $\rho_\tau(b)$=$(\tau - 1)b$ if $b < 0$, $\mathbf{1}_B$ is the indicator function of the event B.

The general l_1 -regression problem given in equation (3.5) is a modification of equation (3.2) and discussed in detail in (Portnoy et al., 1997):

$$\arg\min_{b \in B} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_i^T \boldsymbol{b} \right), \tag{3.5}$$

where $\rho_\tau(s) = s[\tau - \boldsymbol{I}(s < 0)]$ for $\tau \epsilon (0,1)$ and $\boldsymbol{b} \in \mathbb{R}^p$. From equation (3.5), we get:

$$\min \left\{ \tau z^T k + (1-\tau) z^T l | y = Xb + k - l, (k,l) \in \mathbb{R}^2 n_+ \right\} \tag{3.6}$$

and dual formulations are given by equations (3.7) and (3.8), respectively.

$$\max \left\{ y^T p | X^T p = 0, \; p \; \epsilon [\tau - 1, \tau]^n \right\} \tag{3.7}$$

let $w = d + 1 - \tau$,

$$\max \left\{ y^T w | X^T w = (1 - \tau) X^T z, w \epsilon [0,1]^n \right\}. \tag{3.8}$$

Equation (3.7) or (3.8) of the QR problem fits very well into the normal formulations of interior points methods for linear programs with bounded variables. Taieb et al. (2016) propose the boosting AQR procedure for forecasting uncertainty using electricity smart meter data. The authors consider boosting procedure to estimate an AQR model for a set of quantiles of the future distribution. According to Taieb et al. (2016), the k^{th} -step ahead forecast for the conditional function for the τ^{th} -quantile of the electricity demand in equation (3.1) can be estimated using the pinball loss function defined by

$$L(y_t, Q_\tau) = \begin{cases} \tau(y_t - \widehat{Q}_\tau) & if \quad y_t \geq \widehat{Q}_\tau; \\ (1-\tau)(\widehat{Q}_\tau - y_t) & if \quad y_t < \widehat{Q}_\tau, \end{cases} \tag{3.9}$$

where $\widehat{Q}_\tau$ denotes the quantile forecast and y_t represents the actual value of peak electricity demand. QRA is another recent forecast combination technique used to compute the prediction interval. Despite its popularity and simplicity, it is noted that combining forecasts has not been performed widely in the area of long-term peak electricity demand forecasting using quarterly or monthly data. However, it has been performing well in the practice of electricity price forecasting (see, Maciejowska et al. (2016), Nowotarski and Weron (2015), among others). According to Taieb et al. (2016) the QRA model is given by

$$y_t = g_k(\boldsymbol{x}_t) + \varepsilon_t, \tag{3.10}$$

where $\boldsymbol{x}_t = (\boldsymbol{y}_t, \boldsymbol{z}_t)$; $\boldsymbol{z}_t$ is a vector of exogenous variables known at time t; $\boldsymbol{y}_t$ is a vector of past demand occurring prior to time t; ε_t denotes the model error term with $E[\varepsilon_t] = 0$ and $E[\boldsymbol{x}_t \varepsilon_t] = 0$.

3.2.3 Nonlinear quantile regression

NLQR is an extension of linear quantile regression in which the models are nonlinear in their parameters, whereas the linear-in-parameters quantile regression model is given by:

$$q_{Y|X}(\tau) = X^T \beta(\tau). \tag{3.11}$$

The NLQR model (nonlinear in parameters) is given in Equation (3.12) and was discussed in detail in Koenker (2005):

$$q_{Y|X}(\tau) = g(X, \beta_0(\tau)) \tag{3.12}$$

and the corresponding estimator is given by:

$$\hat{\beta}(\tau) = \arg\min_{\boldsymbol{b} \in \boldsymbol{B}} \sum_{i=1}^{n} \rho_\tau \left(y_i - g(x_i, \boldsymbol{b}) \right), \tag{3.13}$$

where $g(x_i, \boldsymbol{b})$ is a function with unknown parameters, $\boldsymbol{B} \in \mathbb{R}^p$ and $\hat{\beta}(\tau)$ is the unknown regression coefficient for the τ^{th} -quantile. The NLQR model is formed by replacing the linear quantile regression model with the quantile curve in Equation (3.12). The study by Koenker and Park (1996) proposed the interior point algorithm approach for computing NLQR estimates. The proposed estimation is used to solve Equation (3.13).

3.2.4 Estimation of parameters

Using linear programming formulation and adopting the algorithm by Wang and Jiang (2012), the QR model can be written as given in equation (3.14):

$$\begin{aligned} \min_{u,v} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\tau u_i^+ + (1-\tau) u_i^- \right) + \sum_{j=1}^{p} w_j^{t-1} v_j \right\} \\ \text{subject to} \\ u_i^+ - u_i^- = \boldsymbol{Y}_i - \boldsymbol{X}_i^T \boldsymbol{\beta}, i = 1, ..., n, \\ u_i^+ \geq 0, u_i^- \geq 0, i = 1, ..., n, \\ v_j \geq \beta_j, v_j \geq -\beta_j, j = 1, ..., n, \end{aligned} \tag{3.14}$$

where u_i^+, u_i^-, v_j are slack variables and w_j^t are weights. The linear programming formulation of the QR model in (3.14) can be solved by the following algorithms:

- **Barrodale and Roberts algorithm:** This algorithm is based on l_1 -regression (Koenker and d'Orey (1987) and Koenker and d'Orey (1994)). The algorithm is known to be effective for relatively large problems which have several thousands of observations. It can be used to calculate the entirely quantile regression process. Moreover it also implements a scheme for the estimated confidence intervals parameters based on inversion of a rank test as described in detail by (Koenker, 1994).

- **Frisch-Newton interior-point algorithm:** The algorithm is more useful for problems than those that can be solved using the Barrodale and Roberts algorithm. The method is well-suitable for large sample size n and small p problems, especially when the parametric dimension of the model is small. A detailed discussion of this algorithm is given in (Portnoy et al., 1997).

- **Sparse Frisch-Newton algorithm:** This algorithm is useful for large problems with large parametric dimensions. It uses the Frisch Newton algorithm but exploits sparse algebra to calculate iterates and is helpful, especially when the factor variables are included.

3.3 Generalised additive quantile regression

3.3.1 Brief history of generalised additive quantile regression

It is argued in Wahba (1980) that the reduced rank penalised smoothing started in the 1980s. This argument is supported by Parker and Rice (1985).

The authors Wahba (1980) and Parker and Rice (1985) carried out important and useful studies using regularisation, regression splines including generalised cross-validation used in solving noisy data with approximation methods. In the context of GAMs, Hastie and Tibshirani (1990) used penalised reduced rank smoothers to represent GAMs. The authors are convinced that in GAMs, the response variable is expressed as a sum of smoothing functions. Detailed discussion on smoothing splines methodology is discussed in Wood (2017); Kim et al. (2021); Morin et al. (2021); Meng et al. (2021); Nanfuka et al. (2021) among others.

3.3.2 Additive quantile regression model

Additive quantile regression (AQR) is a hybrid model that combines quantile regression with generalised additive models. This study uses the quantile generalised additive model (quantGAM) developed by Gaillard et al. (2016) and extended by Fasiolo et al. (2020a). The quantGAM is given in equation (3.15).

$$y_{t,\tau} = \sum_{j=1}^{p} s_{j,\tau}(x_{tj}) + \varepsilon_{t,\tau}; \quad \tau \in (0,1), \tag{3.15}$$

where $s_{j\tau}$ represents the smoothing spline function s at the j^{th} parameter at the τ^{th} quantile and is given in equation (3.16).

$$s_{j\tau}(x) = \sum_{k=1}^{q} \beta_{kj} b_{kj}(x_{tj}), \tag{3.16}$$

where β_{kj} denotes the j^{th} parameter and $b_{kj}(x)$ represents the k^{th} basis function with dimension q. The parameter estimates of Equation (3.15) are obtained by minimising the function given in Equation (3.17).

$$q_{Y|X}(\tau) = \sum_{t=1}^{n} \rho_\tau \left(y_{t,\tau} - \sum_{j=1}^{p} s_{j,\tau}(x_{tj}) \right), \tag{3.17}$$

where $q_{Y|X}(\tau)$ is the extreme conditional quantile function of τ and $\rho_\tau(u) = u[\tau - \boldsymbol{I}(u < 0)]$ is a check function. This study is interested in estimating extreme conditional quantiles for $\tau \epsilon (0, 1)$.

3.3.3 Parameter estimation

Parameters of the quantGAMs are estimated using the R package, "qgam" developed by Fasiolo et al. (2020b). In a Bayesian quantile regression framework, the likelihood function, $\pi(y|\beta)$, which is used to update the prior, $\pi(y|\beta)$ does not exist due to the fact that quantile regression is based on the pinball loss (Bissiri et al. (2016); Fasiolo et al. (2020b)). As a result, a belief-updating framework is used (Bissiri et al. (2016); Fasiolo et al. (2020b)). The posterior distribution of the belief updating of $\pi(y|\beta)$ is given by (Bissiri et al. (2016); Fasiolo et al. (2020b)) and defined as follows:

$$\pi(\beta|y) \propto \psi e^{-\psi \sum_{i=1}^{n} L(y_i, \beta)} \pi(\beta), \tag{3.18}$$

where $\psi > 0$ determines the relative weight of the loss, $L(y_i, \beta)$ and the prior, $\pi(\beta)$. The parameter, ψ, is known as the learning rate (Fasiolo et al., 2020b). If $\psi = \frac{1}{\sigma}$, the posterior distribution in (3.18) is given by the following expression (Fasiolo et al., 2020b):

$$\pi(\beta|y) \propto \Pi_{i=1}^{n} \frac{\tau(1-\tau)}{\sigma} exp\left\{ -\rho_\tau \left(\frac{y_i - \mu}{\sigma} \right) \right\} \pi(\beta) \tag{3.19}$$

The density, $\pi_{AL}(y|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} exp\left\{ -\rho_\tau \left(\frac{y_i-\mu}{\sigma} \right) \right\}$, is the asymmetric Laplace density, where μ, σ and τ are the location, scale and asymmetry parameters, respectively.

3.3.4 Partial linear additive quantile regression

Brief history of partial linear additive quantile regression

The partial linear additive quantile regression (PLAQR) model is a special kind of additive quantile regression model that mitigates the curse of dimensionality while allowing for some covariates to have a non-linear relationship with the response variable (Xingyu, 2016). This study uses the Barrodale and Roberts algorithm for l_1 -regression to estimate the PLAQR models with pairwise interactions and without interactions that follow the structure of (Sherwood and Wang, 2016). The model is chosen as it accommodates non-linearity and the non-linear components are estimated using B-spline functions. Following the work done by Wu (2013), another alternative structure to PLAQR is partially linear additive conditional quantile regression (PLACQR) which avoids the curse of dimensionally by restricting the function form to where only univariate unknown function that has to be estimated. Furthermore, this study is based on the work of PLAQR extended recently by Wang et al. (2020) who proposed an efficient estimation for the parameter in additional partial linear with missing covariates, followed by Liang et al. (2021) who used quantile regression of partially linear single-index model with missing observations; among others. In this study, PLAQR models with nonparametric additive and linear parametric components for peak electricity demand to analyse the South African data are considered.

Partial linear additive quantile regression model

Generalised additive models developed by Hastie and Tibshirani (1990) allow flexibility in modelling linear predictors as a sum of smooth functions. The partially linear additive quantile regression model Hoshino (2014) is given in equation (3.20).

$$y_{t,\tau} = \beta_{0,\tau} + \sum_{j=1}^{p_1} m_{j,\tau}(x_j) + \sum_{j=1}^{p_2} \beta_{j,\tau} z_j + \varepsilon_{t,\tau}, \tau \in (0,1), \tag{3.20}$$

where $y_{t,\tau},\ t = 1, ..., n$ is the response variable, x_j are continuous variables, $m_{j,\tau}$ are smooth functions, z_j are linear variables, $\beta_{j,\tau}$ are parameters and $\varepsilon_{t,\tau}$ is the quantile error term. Equation (3.15) can be written in vector form as:

$$\boldsymbol{Y} = \boldsymbol{m_\tau(X)} + \boldsymbol{Z^T \beta_\tau} + \boldsymbol{\varepsilon_{t,\tau}}, \tau \in (0,1), \tag{3.21}$$

where $\boldsymbol{Y}$ is a vector outcome variable, $\boldsymbol{X}$ is a vector of continuous variables, $\boldsymbol{Z}$ is a vector of linear variables and $\boldsymbol{m}$ is unknown univariate function. The parameter estimates of equation (3.21) are obtained by minimising the following function:

$$Q(\boldsymbol{\beta}, m(.)) = \sum_{i=1}^{n} \rho_\tau \left(Y_i - Z_i^T \beta - m(X_j)\right) + \sum_{i=1}^{n} \lambda_i \int \left(m''(t)\right)^2 dt, \tag{3.22}$$

where $\rho_\tau(t) = \tau t - tI(t < 0)$ is the quantile loss function and λ_i is a non-negative smoothing parameter. The residuals $\varepsilon_{t,\tau}$ are assumed to be auto-correlated. The following procedure is used to reduce the residual autocorrelations. Let $f(y) = \beta_{0,\tau} + \sum_{j=1}^{p_1} m_{j,\tau}(x_j) + \sum_{j=1}^{p_2} \beta_{j,\tau} z_j$, then equation (3.21) reduces to $y_t = f(y) + \varepsilon_{t,\tau}$. Assuming $\varepsilon_{t,\tau}$ follows an AR(1) process the procedure is as follows:

$$\begin{aligned} y_{t,\tau} &= f(y) + \phi_1 \varepsilon_{t,\tau} + a_{t,\tau} \\ y_{t,\tau} - \phi_1 \varepsilon_{t,\tau} &= f(y) + a_{t,\tau}. \end{aligned} \tag{3.23}$$

Let $y^*_{t,\tau} = y_{t,\tau} - \phi_1 \varepsilon_{t,\tau}$, resulting in

$$y^*_{t,\tau} = f(y) + a_{t,\tau}, \tag{3.24}$$

where ϕ_1 is the parameter of the AR(1) model and $a_{t,\tau}$ is white noise with mean zero and variance, σ^2_a. We then check for residual autocorrelation. The general procedure is as follows: Fit an appropriate SARIMA $(p, 0, q) \times (P, 0, Q)_s$ model. Subtract the fitted values of the residuals of the SARIMA $(p, 0, q) \times (P, 0, Q)_s$ from y_t to get y^*_t. Regress y^*_t on the covariates. Check for residual autocorrelation in the new model. If the residuals are still autocorrelated, repeat the process until the desired results are achieved. The general appropriate SARIMA $(p, 0, q) \times (P, 0, Q)_s$ or SARMA $(p, q) \times (P, Q)_s$ model is given by

$$\begin{aligned} \phi(B)\Phi(B)\varepsilon_{t,\tau} = \theta(B)\Theta(B)a_t \\ y_{t,\tau} = f(y) + \varepsilon_{t,\tau} \\ \Longrightarrow \ \varepsilon_{t,\tau} = y_{t,\tau} - f(y) \\ \phi(B)\Phi(B)(y_{t,\tau} - f(y)) = \theta(B)\Theta(B)a_t, \end{aligned} \tag{3.25}$$

where $\phi(B)$ = non-seasonal AR(p), $\Phi(B)$ = seasonal AR(p), $\theta(B)$ = non-seasonal MA(q), $\Theta(B)$ = seasonal MA(q) and s is the number of observations per season.

Estimation of parameters

An adaptation of the Barrodale and Roberts algorithm for ℓ_1- regression is used to estimate the PLAQR model parameters. The simplex approach to solving the general ℓ_1- regression problem (Portnoy et al., 1997) is given as follows:

$$\min_{\beta\in\mathbb{R}^p}\sum_{i=1}^{n}\rho_\tau\left(Y_i - Z_i^T\beta - m(X_j)\right) + \sum_{i=1}^{n}\lambda_i\int\left(m''(t)\right)^2 dt. \tag{3.26}$$

This is reformulated as a linear programming problem. We introduce $2n$ artificial variables $(u_i, v_i\ :\ 1,...,n)$ to represent the negative and positive parts of the vector of residuals. This results in a new problem.

$$\max\left\{y^T d|\boldsymbol{Z}d = 0, d\in[-1,1]^n\right\} \tag{3.27}$$

and setting $a = d + \frac{1}{2}\boldsymbol{e}$ results in

$$\max\left\{y^T a|\boldsymbol{Z}a = \frac{1}{2}Z^T\boldsymbol{e}, a\in[0,1)^n\right\}. \tag{3.28}$$

3.4 Extremal mixture models

This section gives the foundation and original development of extreme value theory. The modelling procedures discussed include historical and contemporary techniques. Contemporary techniques mentioned in this study include work that involves predicting extreme high and low conditional quantiles using semi parametric mixture model. The importance of combining EVT and QR models as well as the strengths and weaknesses of the AQR, extreme mixture and nonlinear quantile regression models are also discussed.

3.4.1 Semi-parametric extremal mixture model

One direction that has recently attracted much research attention is semi-parametric QR models. Semi-parametric QR maintains the flexibility of non-parametric models while maintaining the explanatory power. Semi-parametric mixture models of sample selection have received considerable attention in recent years as researchers explore various schemes to relax the parametric specification employed in the seminal work (Koenker and Hallock, 2001). Let $X_{t_1}, ..., X_{t_n}$ denote DPED, where $t_i, i = 1, ..., n$ is a sequence of times $(0 \leq t_1 \leq ... \leq t_n \leq T_{\text{max}})$. Suppose the random variable X_{t_i} has a distribution function F_{t_i}. We seek to estimate extreme quantiles, i.e $F_t^{-1}(\tau)$ for $0.950 \leq \tau \leq 0.9999$. Now if F_t is in the domain of attraction of the Fréchet distribution, then the excess distribution function given in equation (3.29)

$$F_{t,\tau}(x) = 1 - \frac{1 - F_t(x)}{1 - F_t(\tau)}, \; x \in [\tau, \infty) \tag{3.29}$$

can be estimated by a Pareto distribution in equation (3.30) Durrieu et al. (2018):

$$G_{\tau,\theta}(x) = 1 - \left(\frac{x}{\tau}\right)^{-\frac{1}{\theta}}, \; x \in [\tau, \infty), \tag{3.30}$$

where $\theta > 0$ and $\tau \geq x_0$ (τ is the unknown threshold). Considering a semi-parametric mixture model (bulk model and tail model) given by Durrieu et al. (2015):

$$F_{t,\tau,\boldsymbol{\theta}}(x) = \begin{cases} F_t(x) & \text{if } x \in [x_0, \tau] \\ 1 - (1 - F_t(\tau))(1 - G_{\tau,\theta}(x)) & \text{if } x > \tau, \end{cases} \tag{3.31}$$

where $\tau \geq x_0$ represents the threshold. For any $p \in (0, 1)$, the extreme quantile of X_t is given by:

$$q_p(t,h) = \begin{cases} F^{-1}_{t,h}(p) & \text{if } p < p_\tau \\ \tau \left(\frac{1-p_\tau}{1-p}\right)^{\theta_{t,h,\tau}} & \text{otherwise,} \end{cases} \tag{3.32}$$

where

$$F_{t,h}(x) = \frac{1}{\sum_{i=1}^{n} K\left(\frac{t_i - t}{h}\right)} \times \sum_{i=1}^{n} K\left(\frac{t_i - t}{h}\right) \boldsymbol{I}_{X_{t_i} \leq x}, \tag{3.33}$$

and $K(.)$ is a kernel function, h is the bandwidth and $\boldsymbol{I}$ is an indicator function.

3.4.2 Threshold selection

In peaks-over-threshold models, the threshold is normally estimated first before fitting the desired model to the exceedances (Wu and Qiu, 2018). A sufficiently high threshold is vital to guarantee the stability of the parameters. If the threshold is incorrectly chosen at some value larger than τ, the number of observations on which the distribution is fit becomes smaller, which leads to unstable parameter estimates. However, if the threshold is too high, it produces fewer excesses to estimate the scale and shape parameters, resulting in a higher variance (Verster et al., 2013). Classically, a quantile-based approach to find an appropriate threshold is used. The study provides a threshold selection method given in Equation (3.31) that effectively detects whether a sample follows a certain distribution F. However, it is noted that other threshold selection methods might be equally valid, depending on the circumstances. The thresholds are estimated as the $95^{th}, 99^{th}, 99.9^{th}$ and 99.99^{th} percentiles using quantile regression. Therefore, it is quite important to plot the DPED and also to superimpose the threshold estimate as shown in Figures 6.2, 6.3 and 6.4, respectively. In fitting a distribution, a high enough threshold is chosen. The threshold, τ is determined by fitting a boundary corrected extremal mixture model, discussed in MacDonald et al. (2013), on the positive residuals extracted from the fitted nonlinear detrending model discussed in Section 3.2.3. The estimated threshold is high enough to satisfy the model's assumptions but low enough to capture a reasonable number of observations. The estimated thresholds are satisfactory, while their excesses are modelled as distribution realisations.

3.4.3 Parameter estimation

The extreme conditional quantiles estimator approach is one way of modelling extremes by fitting the generalised extreme value distribution (GEVD) or GPD where the location (μ), shape (ξ) and scale (σ) parameters depend on either parametric or nonparametric covariates. According to Wang and Li (2013), this approach captures the covariate effects at different tails of the response distribution. However, the limitation of extreme conditional quantile estimators is instability with heavy-tailed distributions due to scarce data in the tails of the distributions (Wang and Li, 2013). To predict the extremely high and extremely low conditional quantiles between 0.9999 and 0.95, this study proposes a Pareto distribution model given in equation (3.30) and the domain of attraction of the Fréchet distribution in equation (3.29). Let $\widehat{\gamma}(Y_{n-k+1,n,\ldots,Y_{n,n}})$ be an estimator for the extreme value index based on k, where k is the upper order statistics. Now, if k is substituted by the estimated quantile curves for a fixed point x_0, then $\widehat{\gamma}(x_0) = \widehat{\gamma}(\widehat{Q}(\tau_{p-k}|x_0, \ldots, \widehat{Q}(\tau_p|x_0))$ for $k < p$. The extreme quantile estimator for $Y|X = x_0$ is given as follows

(Velthoen, 2016):

$$\widehat{Q}(\tau_n|x_0) = \left(\frac{1-\tau_{p-k}}{1-\tau_n}\right)^{\widehat{\gamma}(x_0)} \widehat{Q}(\tau_{p-k|x_0}), \tag{3.34}$$

where $X = x_0$ are increasing sequence of p. Table 3.1 shows a summary of the strengths and weaknesses of the additive quantile regression, extremal mixture and nonlinear quantile regression models.

Table 3.1: Comparison of models.

Models	**Strengths**	**Weaknesses**
Model1 (AQR)	1. A hybrid model that combines GAMS with QR. 2. Estimation is distribution free. 3. Robust to outliers in the response variable.	1. Requires a smoothing function of the covariates. 2. Parameters are harder to estimate. 3. Does not give any details about the size of high level of possible exceedances.
Model2 (EM)	1. Semi-parametric extremal mixture model. 2. Based on one covariate, which is $t = 1, ..., n$.	1. Has limitations on accuracy and stability. 2. Very sensitive to numbers and location of the measured points.
Model3 (NLQR)	1. Inference is performed based on large sample approximation. 2. Robust to outliers in the response variable.	1. Requires a smoothing parameter. 2. Outliers only have influence on quantile curves close to them, i.e they affect extreme quantiles

3.5 Variable selection and evaluation metrics

3.5.1 Variable selection

The three variable selection methods for penalised regression techniques are Lasso, CV and elastic net. Unlike other penalised regression techniques, in Lasso some of the coefficients are shrunken all the way to zero. Lasso guarantees both variable selection and shrinkage simultaneously. The Lasso estimates for the OLS regression model is given by

$$\hat{\boldsymbol{\beta}} = arg\ min \sum_{j=1}^{p} (\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}_j)^2 + \lambda_1 \sum_{j=1}^{p} |\boldsymbol{\beta}_j|, \tag{3.35}$$

where, $\boldsymbol{y} = (\boldsymbol{y}_1, ..., \boldsymbol{y}_n)^T$ is the response vector and $\boldsymbol{x}_j = (\boldsymbol{x}_{1j}, ..., \boldsymbol{x}_{nj})^T, j = 1, ..., p$ are the linear independent predictors, λ is the Lasso regularization parameter , $\sum_{j=1}^{p} |\boldsymbol{\beta}_j| \leq t$ is called Lasso penalty and t is the Lasso turning parameter (Zou et al., 2007). If $p > n$, the Lasso selects at most n variables. There is however an adaptive Lasso method and this method modifies the Lasso penalty by applying weights to each parameter that forms the Lasso constraint. Elastic net is a regularization and variable selection method suggested by (Zou and Hastie, 2005). The method removes the limitation on the number of selected variables and also encourages grouping effect that Lasso fails to do. The elastic net estimator is given by

$$\hat{\boldsymbol{\beta}} = arg\ min \sum_{j=1}^{p} (\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}_j)^2 + \lambda_1 \sum_{j=1}^{p} |\boldsymbol{\beta}_j| + \lambda_2 \sum_{j=1}^{p} \boldsymbol{\beta}_j^2, \tag{3.36}$$

where, $\sum_{j=1}^{p} |\boldsymbol{\beta}_j| \leq t_1$ and $\sum_{j=1}^{p} \boldsymbol{\beta}_j^2 < t_2$ are elastic net penalties, t_1 and t_2 are elastic net turning parameters. Cross-validation is another commonly used variable selection method that uses part of the training data set to fit the model and the remaining part estimates the prediction error.

3.5.2 Model selection and accuracy measures for forecast model

This section considers the model selection criteria for choosing quantile regression models. On modelling the long-term daily peak electricity demand using the PLAQR technique, the accuracy measures such as RMSE, MAE and MAPE are considered. To quantify the performance of estimation, this study uses root mean squared error (RMSE). RMSE is a good measure of accuracy, but only to compare forecasting errors of different PLAQR models for a particular variable. It is given by

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{t=1}^{m}(y_t - \hat{y}_t)^2}, \tag{3.37}$$

where y_t is the true quantile of a distribution and $\hat{y}_t$ is estimated quantile. Mean absolute error (MAE) is another quantity used to measure how close the forecasts or predictions are from the outcomes and it is given by

$$\text{MAE} = \frac{1}{m}\sum_{t=1}^{m}|y_t - \hat{y}_t| \tag{3.38}$$

Moreover, mean absolute percentage error (MAPE) is a measure of accuracy used given by

$$\text{MAPE} = \frac{1}{m}\sum_{t=1}^{m}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \tag{3.39}$$

Based on the findings from work done by Chernozhukov (2005), the present study built on EQR procedures to estimate extremal conditional quantiles. This is done using the scoring rules to evaluate the developed models' predictive accuracy. Scoring rules are considered in this study, as they are significantly important in:

- parameter estimation,
- evaluating the predictive performance of extreme conditional quantiles and AQR models; and
- obtaining the probabilities of rare events.

The scoring rules designed for comparative forecasting evaluation, namely; CRPS, logS, DSS, pinball loss (PL) and interval width (IW), are discussed and used to compare the AQR, EM and NLQR models at both extremely high and extremely low conditional quantiles as shown in Tables 6.2 and 6.3, respectively.

3.5.3 Forecast evaluation

The scoring rules evaluated the probability forecasts by assigning a real number to the predictive distribution and observations. The score values were used to compare the models. The scoring rule $S(y, p)$ is defined as a special case of a loss function $L(y_t, q_\tau)$, measuring the negative worth of behaviour when the variable Y turns out to be y_t, where q_τ and y_t denote the quantile forecast and actual value of the DPED (Wei, 2016). Moreover, the proper scoring rules are preferable if the calibration and sharpness are measured simultaneously. This study used the scoring rules designed for the comparative forecasting evaluation, namely: continuous ranked probability score (CRPS), logarithmic score (LogS), Dawid–Sebastiani score (DSS), pinball loss (PL) and interval width (IW).

Continuous Ranked Probability Score

The CRPS has recently attracted much attention in forecasting performance as it quantifies and takes into consideration both sharpness and calibration (Gneiting and Katzfuss, 2014). The CRPS is defined by:

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} \{F(x) - \boldsymbol{I}(y \leq x)\}^2 dx, \tag{3.40}$$

where y is the actual observation, $\boldsymbol{I}(.)$ is an indicator function equal to one for $y \leq x$ and $F(x)$ is the cumulative distribution function (CDF).

Logarithmic Score

The LogS is the only proper local score that assumes the regularity conditions of $f(y, \theta)$ for almost all the actual observations, where θ is a parameter (Lerch et al., 2017). The LogS is given by:

$$\text{LogS}(F, y) = -\text{Log}(f(y)), \tag{3.41}$$

where $f(y)$ is the probability density function (PDF) and F is a strictly proper scoring rule relative to the probability distribution.

Dawid–Sebastiani Score

The DSS depends on the first and second moments of the forecast. The DSS is given by Gneiting and Katzfuss (2014):

$$\text{DSS}(F, y) = \frac{(y - \mu_F)^2}{\sigma_F^2} + 2\text{Log}\left(\sigma_F\right), \tag{3.42}$$

where F is the predictive distribution of y with first and second moments μ_F and σ_F^2, respectively, $(y - \mu_F)^2$ is a squared error score and μ_F and σ_F^2 denote the mean and variance of the predictive distribution F, respectively.

Pinball Loss Function

The pinball loss (PL) function is relatively easy to use and is given as:

$$PL(q_{\tau,t}) = \begin{cases} 2(1-\tau)|y_t - q_{\tau,t}|, & \text{if } y_t < q_{\tau,t}, \\ 2\tau|y_t - q_{\tau,t}|, & \text{if } y_t \geq q_{\tau,t}, \end{cases} \tag{3.43}$$

where $q_{\tau,t}$ is the quantile forecast at a time t with probability τ and y_t is the observed value of the DPED at time t. The interpretation of $PL(q_{\tau,t})$ is made easier by the inclusion of the multiplier number 2 (Hyndman, 2020). When $\tau = 0.5, PL_{0.5,t} = |y_t - q_{\tau,t}|$, which is the same as the absolute error. Hence, $PL(_{\tau,t})$ is generally interpreted as an absolute error.

Estimated Intervals' Widths

The interval width (IW) is the difference between the estimated upper and lower quantile values. It is given in Equation (3.44) as:

$$\text{IW}_t = q_{\tau,t} - q_{(1-\tau,t)}, \tag{3.44}$$

where $q_{\tau,t}$ and $q_{(1-\tau,t)}$ are the upper and lower quantiles, respectively. The coverage probability (CP) is a procedure for constructing random regions to produce an interval covering the true value. The CP is used to evaluate the performance of estimated intervals. It is considered a property of the interval-producing procedure, independent of the particular sample to which such a procedure is applied. It also evaluates the reliability of the estimated interval widths and is given in Equation (3.45) by:

$$\text{CP} = \frac{1}{h}\sum_{t=1}^{h} I_t, \qquad I_t = \begin{cases} 1 & \text{if} \quad y_t \epsilon (L_t, U_t) \\ 0 & \text{otherwise} \end{cases}, \tag{3.45}$$

where L_t and U_t are the lower and upper specification limits, respectively.

3.6 Conclusion

In this chapter, quantile regression-based models have been considered to predict the long-term peak electricity demand in South Africa. The research methodologies based on electricity demand history alone are no longer good enough to predict better. The old historical peak electricity demand data is not supposed to be used if the detailed and more recent data are recommended for the forecasting procedure. Some of the methods mentioned in this chapter apply to traditional electricity demand predictions. The conventional long-term electricity demand forecasting methods use different scenarios to predict future demand. In conclusion, one can suggest that the quantile regression methods proposed in this study are very accurate, simple and more powerful techniques to predict electricity demand in the long term. Examples of this are: fQRA model that fits very well in QPED data as given in Figure 4.5; PLAQR model with pairwise interactions given in Figure 5.4 which shows a good fit to the density of DPED data in both panels; AQR model that also works very well at both extremely high and low quantile levels as shown by the model comparisons in Table 3.1 among others. Different techniques with different objectives for long-term peak electricity demand forecasting must be compared with other forecasts, for example, short-term or medium-term forecasts. Hence, more accurate data and powerful techniques can provide new opportunities to model electricity demand and quantify the uncertainty of future electricity demand for better electricity security management in the country.

Chapter 4

Long-term peak electricity demand forecasting in South Africa using quantile regression averaging approach

4.1 Introduction

The South African government has noted a relatively slow growth of the renewable energy industry. However, the popularity of peak electricity demand forecasting is mainly gained by understanding the various uncertainties associated with the decision-making processes of industries. The study welcomes, as outlined in the South African energy policy by Marquard (2006) that the development of renewable energy policy activities is greatly influenced by factors such as the periodical occurrence of various policy crises as a result of economic instability. To deal with long-term uncertainties within the energy sector, more accurate long-term peak electricity demand forecasting methods are used in decision-making and planning purposes. Inaccurate forecasting may seriously result in poor decision-making and bad policies establishment.

For instance, the South African government, electricity consumers, producers and industries need accurate forecasts for planning future events to assist in long-term development plans for the country to minimise risks. The long-term electricity demand forecasting methods are dependent on the choice of variable selections and, very importantly, assumptions made for the long run. Researchers have dealt with long-term electricity demand forecasting using different models. For example; OFQR, Multiple Simple Linear Regression (MSLR); ANNs; QR and time series analysis models have been used for long-term electricity demand forecasting((Kandil et al., 2000); (Al-Hamadi and Soliman, 2005); (Mokilane, 2018); (Al-Saba and El-Amin, 1999) and (Ringwood et al., 2001) among others).
The electricity demand in South Africa is higher in winter than in summer and is influenced especially by rare events like extreme weather conditions. While electricity demand is still increasing due to population growth, long-term electricity demand forecasting is essential to avoid the shortage of electricity that may hamper South African economic growth. South Africa needs reliable forecasts of electricity demand to mitigate the risk of shortages in the future to guarantee sustainability.

4.2 Empirical results

4.2.1 Monthly and quarterly demand models

The data consist of 28 South African weather stations divided into interior and coastal regions with monthly and quarterly peak electricity demand (MPED and QPED) data, temperature, lagged demand and calendar effects. The MPED and QPED are the dependent variables, while temperature, lagged demand and calendar effects are used as predictor variables. For MPED, the temperature effects for coastal are AMTC, maxTC and minTC, defined as average monthly coastal, average maximum and minimum coastal temperatures, respectively. The effects for the interior include AMTI, maxTI and minTI, defined as average monthly interior, average maximum and minimum interior temperatures, respectively.

Likewise, for QPED, the temperature effects for coastal are AQTC, maxTC and minTC, defined as average quarterly coastal, average maximum and minimum coastal temperatures, respectively. The effects for the quarterly interior include AQTI, maxTI and minTI, defined as average quarterly interior and average maximum and minimum interior temperatures. The temperature effects for both monthly and quarterly peak electricity demand coastal and interior are given by the average minimum of coastal and interior temperatures (AminTCI), an average of average coastal and interior temperatures (AATCI), average maximum of coastal and interior temperatures (AmaxTCI), the difference between the average maximum of coastal and interior temperatures (DmaxTCI), the difference between an average minimum of coastal and interior temperatures (DminTCI) and the difference between the average of average monthly coastal and interior temperatures (DAAMTCI) and the difference between the average of average quarterly coastal and interior temperatures (DAAQTCI). Furthermore, day type (day of the week), a day before a holiday (DBH), day after the holiday (DAH) and day holiday (DH) are defined as calendar effects.

Monthly model

We use long-term peak electricity demand models to forecast electricity demand at inland and coastal regions. The QR models are used for both monthly and quarterly data. This study will explore the inclusion of monthly temperature covariates that will also be included in the peak electricity demand parameters through heating and cooling degree days, which can be calculated as:

$$\mathrm{HDD}_t = \max\left(T_{\text{ref}} - T_t, 0\right)$$

and

$$\mathrm{CDD}_t = \max\left(T_t - T_{\text{ref}}, 0\right).$$

The total monthly heating degree-days (MHDD_t) and the total monthly cooling degree-days (MCDD_t) are calculated as:

$$\mathrm{MHDD}_t = \max\left[\sum_{j=1}^{m} \mathrm{HDD}_{t,j} - \sum_{j=1}^{m} \mathrm{CDD}_{t,j}, 0\right]$$

and

$$\mathrm{MCDD}_t = \max\left[\sum_{j=1}^{m} \mathrm{CDD}_{t,j} - \sum_{j=1}^{m} \mathrm{HDD}_{t,j}, 0\right]$$

which reduces to

$$\mathrm{MHDD}_t = \max\left[\sum_{j=1}^{m} \left(\max\left(T_{\text{ref}} - T_{t,j}, 0\right)\right) - \sum_{j=1}^{m} \left(\max\left(T_{t,j} - T_{\text{ref}}, 0\right)\right), 0\right]$$

and

$$\mathrm{MCDD}_t = \max\left[\sum_{j=1}^{m} \left(\max\left(T_{t,j} - T_{\text{ref}}, 0\right)\right) - \sum_{j=1}^{m} \left(\max\left(T_{\text{ref}} - T_{t,j}, 0\right)\right), 0\right],$$

respectively, where m is the number of days in month t, T_t is average daily temperature on day t and T_{ref} is the reference temperature which separates cold temperatures from hot temperatures. Based on the description of variable selection given in section 3.1, the following MPED models for electricity demand are proposed: Model 1:

Linear quantile regression (LQR) without interactions model is given as

$$\begin{aligned} Q_\tau(MPED|\boldsymbol{x}_t) = \alpha_0 &+ \alpha_1(DH) + \alpha_2(DAH) + \alpha_3(DBH) + \alpha_4(Daytype) \\ &+ \alpha_5(CDDAminTCI) + \alpha_6(CDDAmaxTCI) \\ &+ \alpha_7(HDDAAMTCI) + \alpha_8(HDDAminTCI) \\ &+ \alpha_9(noltrend) + \varepsilon_{t,\tau}, \end{aligned} \tag{4.1}$$

where $\alpha_0, \alpha_1, ..., \alpha_9$ are constants, CDDAminTCI and CDDAmaxTCI are cooling degree days average minimum and maximum coastal and interior temperatures, HDDAAMTCI is monthly heating degree days average of average coastal and interior temperature, HDDAminTCI is heating degree days average minimum coastal and interior temperature, noltrend is a non-linear trend, $\varepsilon_{t,\tau}$ is residual error.

Model 2: LQR with interactions model is given by

$$\begin{aligned} Q_\tau(MPED|\boldsymbol{x}_t) = \alpha_0 &+ \alpha_1(DAH) + \alpha_2(DBH) + \alpha_3(Daytype) \\ &+ \alpha_4(CDDAminTCI * Daytype) + \alpha_5(CDDAmaxTCI) \\ &+ \alpha_6(HDDAAMTCI) + \alpha_7(HDDAminTCI) \\ &+ \alpha_8(noltrend) + \varepsilon_{t,\tau}. \end{aligned} \tag{4.2}$$

Model 3: QRA is given in equation (4.3)

$$MPED = \alpha_0 + \alpha_1(fOLS) + \alpha_2(fGAM) + \alpha_3(fQR) + \varepsilon_{t,\tau}, \tag{4.3}$$

where fOLS, fGAM and fQR are ordinary least squares, generalised additive model and quantile regression forecasts respectively.

Quarterly model

Similarly, we can drive the formulae for calculating the total quarterly heating and cooling degree-days $QHDD_t$ and $QCDD_t$ respectively as in the monthly formulae. The quarterly temperature covariate will be included in the long-term peak electricity demand parameters through heating and cooling degree-days. Based on the description of variable selection given in Section 3.5.1, the following QPED models for electricity demand are proposed:

Model 4: LQR without interactions model is given as

$$\begin{aligned} Q_\tau(QPED|\boldsymbol{x}_t) = {} & \alpha_0 + \alpha_1(DH) + \alpha_2(DAH) + \alpha_3(Daytype) \\ & + \alpha_4(CDDAminTCI) + \alpha_5(CDDAmaxTCI) \\ & + \alpha_6(CDDAAQTCI) + \alpha_7(noltrend) \\ & + \varepsilon_{t,\tau}, \end{aligned} \tag{4.4}$$

where HDDAAQTCI is quarterly heating degree days average of average coastal and interior temperature.

Model 5: LQR with interactions model is given as

$$\begin{aligned} Q_\tau(QPED|\boldsymbol{x}_t) = {} & \alpha_0 + \alpha_1(DH) + \alpha_2(DAH) + \alpha_3(Daytype) \\ & + \alpha_4(CDDAminTCI) + \alpha_5(CDDAmaxTCI) \\ & + \alpha_6(HDDAAQTCI * Dytype) \\ & + \alpha_7(noltrend) + \varepsilon_{t,\tau}. \end{aligned} \tag{4.5}$$

Model 6: QRA is given in equation (4.6)

$$QPED = \alpha_0 + \alpha_1(fOLS) + \alpha_2(fGAM) + \alpha_3(fQR) + \varepsilon_{t,\tau}. \tag{4.6}$$

4.2.2 Parameter estimation

The QR model defined in equation (3.1) will be helpful in this analysis. The parameter for OLS over $\boldsymbol{\beta}$ is estimated by minimizing:

$$\sum_{t=1}^{n} r(y_t - \boldsymbol{x}_t^T\boldsymbol{\beta}) = \sum_{t=1}^{n}(y_t - \boldsymbol{x}_t^T\boldsymbol{\beta})^2, \tag{4.7}$$

where r is the quadratic loss function and the parameter for MLE over $\boldsymbol{\beta}$ based on the sample $\{\boldsymbol{x}_t, y_t\}_{t=1}^{n}$ is calculated by maximizing

$$L(\beta) \propto exp\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n}(y_t - \boldsymbol{x}_t^T\boldsymbol{\beta})^2\}. \tag{4.8}$$

4.2.3 Benchmark models

Stochastic gradient boosting

Stochastic gradient boosting (SGB) is a modification of Gradient boosting (GB) which is a machine learning technique. It fits an additive model in a stage-wise way. The additive model can take the form given in equation (4.9) (Hastie et al. (2005)).

$$f(x) = \sum_{m=1}^{M}\beta_m b(x;\gamma_m), \tag{4.9}$$

where $b(x;\gamma_m) \in \mathbb{R}$ are functions of x which are characterised by the expansion parameters γ_m, β_m. The parameters β_m and γ_m are fitted in a stage-wise way, a process which slows down over-fitting (Hastie et al. (2005)). With SGB, a random sample of the training data set is taken without replacement. A more detailed discussion of this is found in Friedman (2002).

Support vector regression

Support vector regression (SVR) which is based on support vector machines (SVM) uses different kernel functions which map low dimensional data to high dimensional space. SVR is introduced by Vapnik (2013) and estimates a regression function of the form given in equation (4.10).

$$f(x) = x^T w + b, \tag{4.10}$$

where b is a scalar and ω is a vector of weights. Equation (4.10) can be reformulated as a quadratic programming problem (QPP) by introducing an ε-insensitive loss function as given in equation (4.11) (Vapnik (2013)).

$$\begin{aligned} \min_{(\omega,b,\xi_1,\xi_2)} \ & \frac{1}{2}\omega^T\omega + C(e^T\xi_1 + e^T\xi_2) \\ \text{s.t.} \quad & y_i - x_i^T\omega - b \leq \varepsilon + \xi_1 \\ & x_i^T\omega + b - y_i \leq \varepsilon + \xi_2 \\ \text{and} \quad & \xi_{1i}, \xi_{2i} \geq 0, i = 1, ..., m, \end{aligned} \tag{4.11}$$

where $C > 0, \varepsilon > 0$ are input parameters, $\xi_1 = (\xi_{11}, ..., \xi_{1m})^T$ and $\xi_2 = (\xi_{21}, ..., \xi_{2m})$ are slack variables.

4.2.4 Exploratory data analysis

The study analyses three different benchmark models, which are GAMs, SVR and SGB. The use of at least two benchmark models is consistent with current trends in forecasting. Current trends in forecasting use a variety of statistical models, including comparative analysis with machine learning models. In this study, we use one statistical learning benchmark model, the generalized additive model, and two machine learning models, stochastic gradient boosting and support vector regression. Using at least two benchmark models helps us see how good our model is against well-known methods that produce accurate forecasts. See for instance, (Makridakis et al., 2020). GAMs are considered a simple and powerful technique because they provide more flexible predictor functions that uncover hidden patterns and regularize the predictor functions. Moreover, this is one technique that is mostly indispensable in analysing long-term peak electricity demand forecasting. This machine learning technique is considered as one of the benchmark models in MPED and QPED analysis.

4.2.5 Results

Figure 4.1 and 4.2 display monthly and quarterly peak electricity demand plots for the period 2000 to 2014, respectively. Hence, Figure 4.3 shows monthly and quarterly average temperature plots. It also shows strong seasonal effects. The smallest RMSE, MAE and MAPE values in the OLS forecast model (M4) for both MPED and QPED are 79.95 MW, 72.45 MW and 0.22 MW, respectively. Likewise, the smallest RMSE, MAE and MAPE in the QRA model (M6) are 76.5 MW, 51.61 MW and 0.16 MW, respectively.

Table 4.4 shows the error levels (RMSE=398.35 MW, MAE=325.73 MW, MAPE=0.98 MW) of SVR to be lower than the error level (RMSE=728.19 MW, MAE=558.98 MW, MAPE=166 MW) of SGB model. Figure 4.6 and 4.7 show the SVR and SGB benchmark model forecasts for MPED. We can see that SVR forecasts in Figure 4.6 fit very well compared to SGB forecasts in Figure 4.7. Figure 4.4 and 4.5 show the actuals of both MPED and QPED with quantile regression forecast (fQR), generalised additive model forecast (fGAM) and quantile regression averaging forecast (fQRA), respectively. Figure 4.5 shows fQRA fits very well in QPED data.

4.2.6 Discussion of results

Figure 4.1 and 4.2 show MPED and QPED plots for the period 2000 to 2014 with density, q-q and box plots, respectively. Both MPED and QPED plots show upward trends with strong seasonality. The overall upward drifts indicate the presence of increasing trends in the time series. The seasonal variations produce the regular fluctuations from year to year but appear similar across months and quarters.

Table 4.1 shows the summary statistics for both monthly and quarterly electricity demand data ranging from 1 January 2000 to 31 March 2014. The minimum and maximum values are compared and also used to assess the spreading behaviour of both quarterly and monthly electricity demand. The minimum values of MPED and QPED from January - March 2000 and on Monday, January 2000 are 24083 and 24760 MW, respectively. Moreover, the maximum value of both MPED and QPED from April-June 2007 and on Thursday, 24 May 2007 is 37158 MW. The kurtosis values of MPED and QPED are -0.31 and -0.21, respectively. The values indicate that both MPED and QPED data do not perfectly follow the normal distribution. Furthermore, negative kurtosis shows that the distribution has lighter tails and flatter peaks than the normal distribution. Table 4.1 also reveals that both monthly and quarterly electricity demand data are left-skewed.

In Table 4.2, M1 and M2 represent OLS forecast without and with interactions, respectively, for the MPED model, while M4 and M5 represent OLS forecast without and with interactions, respectively, for the QPED model. Likewise, in Table 4.3, M1 and M2 represent linear QR forecast without and with interactions, respectively, for MPED model. In contrast, M4 and M5 represent linear QR forecasts without and with interactions, respectively. For the QPED model, M3 and M6 are QRA forecasts for the MPED and QPED model, respectively, and M7 and M8 are GAMs forecasts for both MPED and QPED models. The values of RMSE=79.95 MW, MAE=72.45 MW and MAPE=0.22 MW for M4 in Table 4.2 are less than those of M1, M2 and M5 models. Likewise, the values of RMSE=76.55 MW, MAE=51.61 MW and MAPE=0.16 MW for M6 in Table 4.3 are less than that of M1, M2, M3, M4, M5, M7 and M8 models. Moreover, RMSE, MAE and MAPE values for M6 (QRA for QPED) are less than that of M4 (OLS for MPED). The RMSE, MAE, and MAPE results for M6 (QRA for QPED) also confirm that the model is doing well compared to that of M3 (QRA for MPED). In addition, the values of RMSE, MAE and MAPE for benchmark (SVR and SGB) models in Table 4.4 are larger than that of the M6 model.

Figures 4.8 and 4.10 show the results of cross-validation estimates of the mean squared prediction error for the ridge, Lasso and elastic net models in dependence on the regularization parameter (log of lambda). The numbers on top of each figure show the non-zero regression coefficients. The mean squared prediction errors suggest the smaller values of the parameter that shrinks the coefficients to optimal. Figures 4.9 and 4.11 show the coefficient paths for ridge, Lasso and elastic net model independence on log of lambda for α= 0, 1 and 0.5 respectively. The analysis was also done repeatedly using different values of α.

The comparison of MPED cross-validation estimates of the mean squared prediction error for the ridge, Lasso and elastic net in Figure 4.8 shows that the variable selection method called Lasso is superior compared to other methods. There are also noticeable differences in cross-validation estimates for all the methods Lasso gives the best estimates. Using Lasso and the elastic net regression model for model selection, only one non-zero coefficient shows that the function has chosen the second vertical line on the cross-validation. This indicates that the model might be doing a good job. In Figure 4.9, all coefficients of the ridge regression model are essentially zero when the log of lambda is 15. However, as we relax lambda, the coefficients move away from zero in a nice smooth way. In fitting Lasso, many r squared are explained as quite heavily shrunk coefficients, but coefficients increase (grow very large) towards the end. This may suggest that the model might be overfitted at the end of the path.

Furthermore, the elastic net regression model looks like Lasso with slight differences and this confirms that it is an extension of Lasso proposed by (Vapnik, 2013). However, the elastic net does not perform satisfactorily because there are two shrinkage procedures such as ridge and Lasso. Double shrinkage might introduce unnecessary bias. Likewise, Figure 4.10 shows the QPED model selection of ridge, Lasso and elastic net where the dotted red lines show the cross-validation curve. The two dashed lines in Figure 4.8 also show the log of lambda values that gives the minimum mean square errors for cross-validation (left dashed line) and that is within one standard error(right dashed line). Figure 4.11 shows all coefficients of the ridge regression model to be zeros when the value of lambda is 14 or more. It can be seen from Figure 4.11 that Lasso regression shrinks the regression coefficients to zero as lambda increases. For instance, when the log of lambda = 2, there are six non-zero coefficients and when the log of lambda = 0 all coefficients are zero.

Table 4.1: Summary statistics for monthly and quarterly electricity demand data (1 January 200-31 March 2014).

	Mean	**Median**	**Min**	**Max**	**Kurtosis**	**Skewness**	**St.Dev**
Monthly data	31560	31822	24083	37158	-0.31	-0.42	30.44
Quarterly data	32452	32675	24760	37158	-0.21	-0.54	30.71

Table 4.2: OLS forecast model comparisons.

	M1	**M2**	**M4**	**M5**
RMSE	327.91	338.12	79.95	172.17
MAE	252.06	262.83	72.45	128.41
MAPE	0.77	0.80	0.22	0.37

Table 4.3: QR, GAM and QRA forecast models comparison.

	M1	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**
RMSE	324.89	345.38	315.46	87.92	87.97	76.55	315.96	165.94
MAE	249.31	273.41	231.70	73.55	73.64	51.61	235.01	123.57
MAPE	0.76	0.83	0.71	0.22	0.22	0.16	0.72	0.36

Table 4.4: Benchmark model comparisons: Monthly peak electricity demand.

	SVR	SGB
RMSE	398.35	728.19
MAE	325.73	558.98
MAPE	0.98	1.66

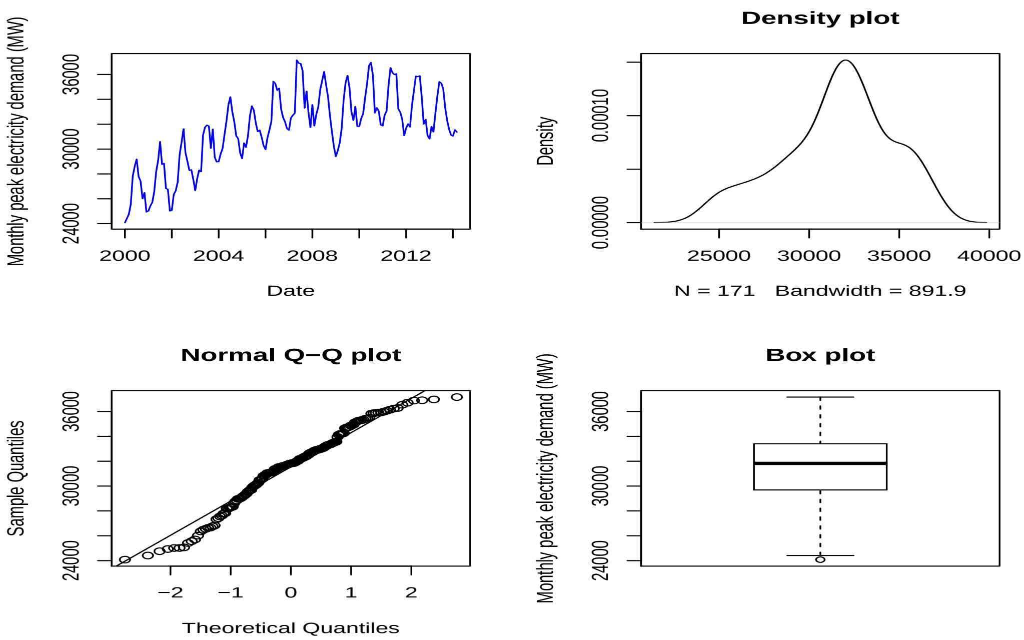

Figure 4.1: Monthly peak electricity demand(top left), Density plot (top right), Q-Q plot (bottom left) and Box plot (bottom right).

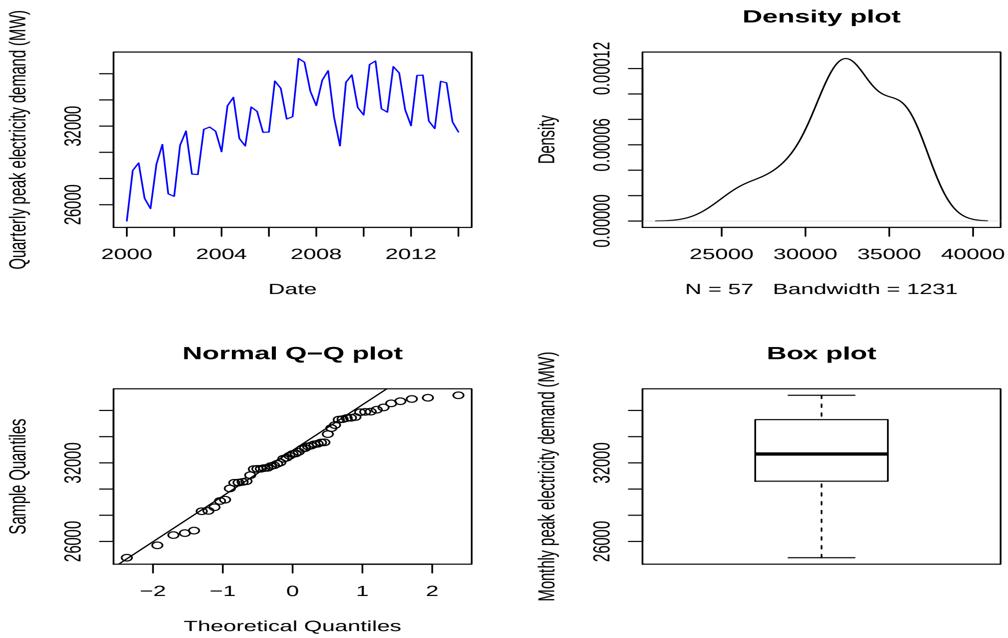

Figure 4.2: Quarterly peak electricity demand (top left), Density plot (top right), Q-Q plot (bottom left) and Box plot (bottom right).

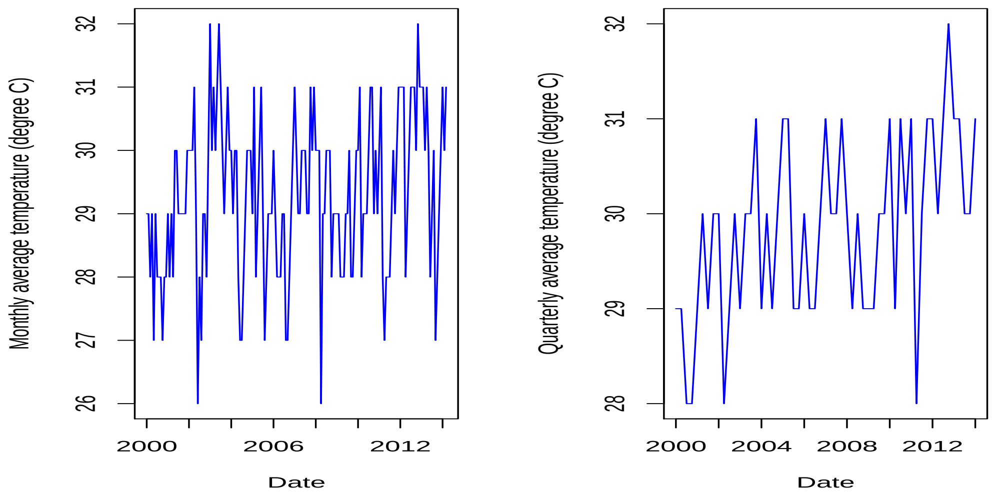

Figure 4.3: Monthly average temperature plot (left side) and Quarterly average temperature plot (on the right).

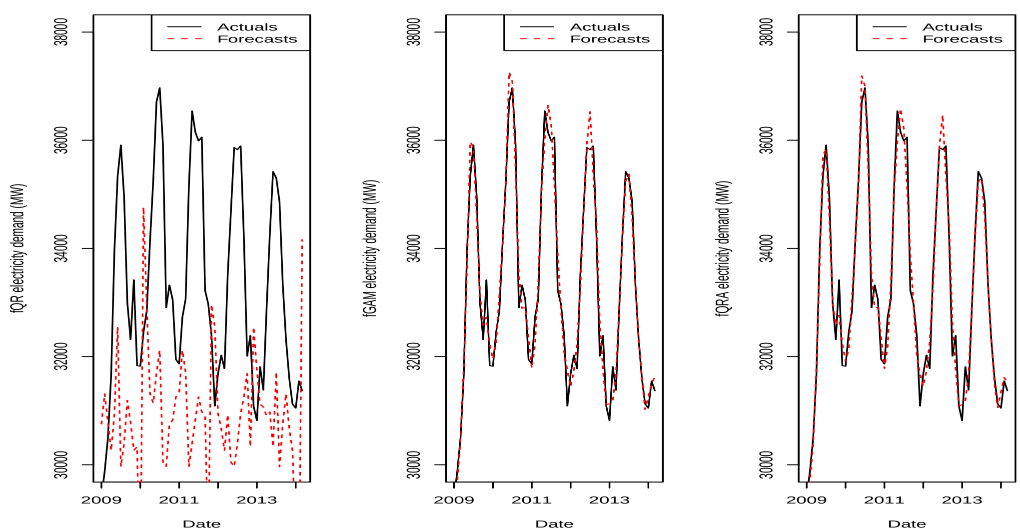

Figure 4.4: Actuals and forecasts for MPED.

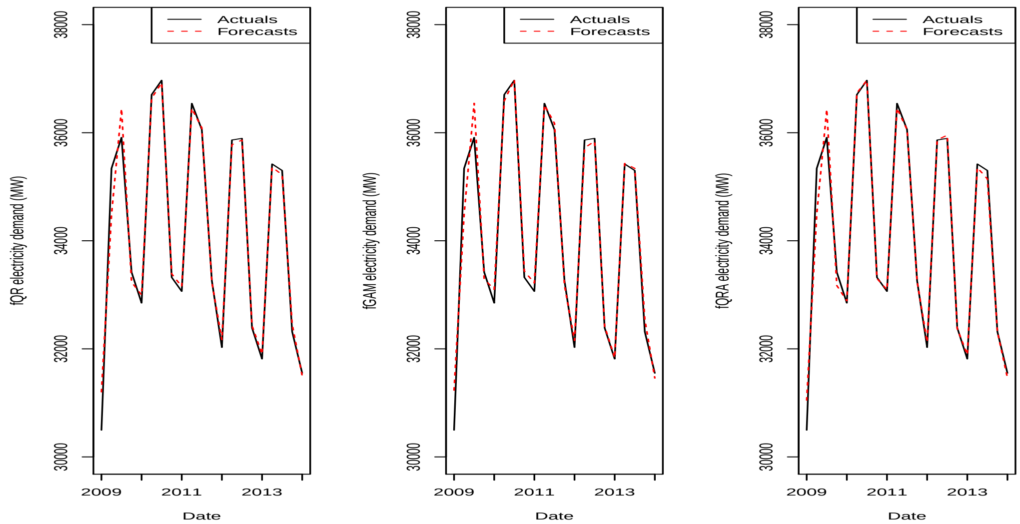

Figure 4.5: Actuals and forecasts for QPED.

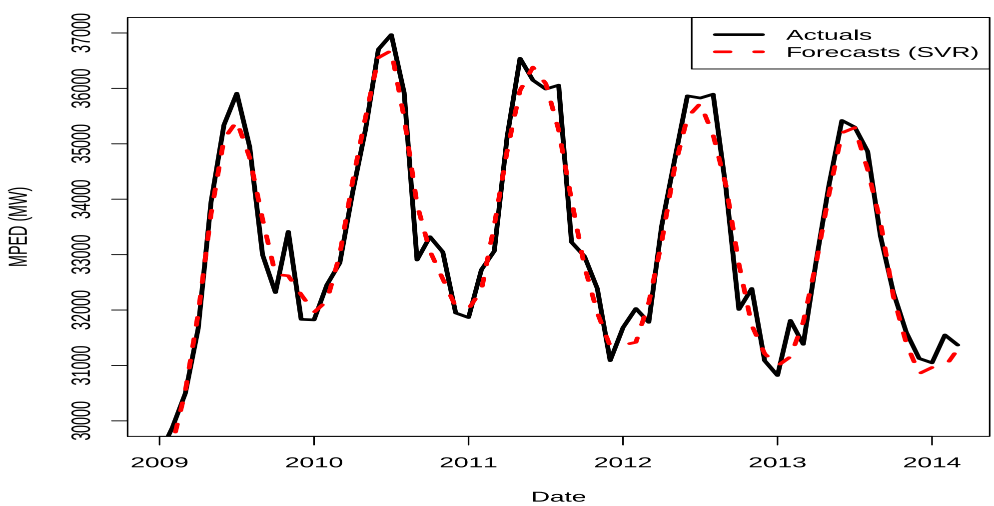

Figure 4.6: SVR forecasts for MPED.

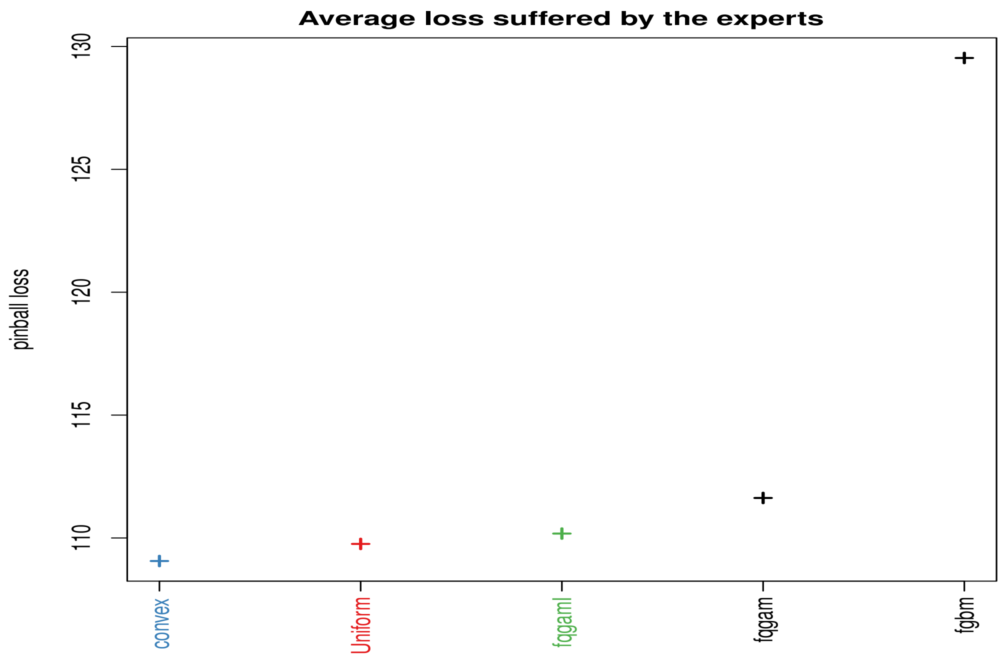

Figure 4.7: SGB forecasts for MPED.

4.3 Conclusion

In this study, the long-term monthly and quarterly peaks electricity demand forecasting using South African data have been presented. Therefore, the policies to control the electricity supply are required to make sure that the electricity demand is met to support the South African community. The more precise long-term forecasting models are developed by using monthly and quarterly data and then applying QR models. The QR models are becoming more attractive to long-term electricity demand modelling these days as it enables us to construct new quantile functions easily. This study will contribute to all research application areas, especially in short, medium, or long-term peak electricity demand methodology. QRA model in QPED has the smallest RMSE, MAE and MAPE compared to other models. The three variable selection methods are investigated to overcome the variable selection problem for long-term electricity demand forecasting. The Lasso model gives the best estimates compared to other variable selection methods. The empirical results of the proposed models are suitable for long-term electricity demand forecasting and for testing policies geared towards the expansion of the electricity infrastructure that should be intensified in South Africa to cope with the increasing demand exercised by the country's economic growth rapid industrialisation programme. QRA is a new technique and performed well in QPED data for long-term electricity demand and it did not perform badly in MPED data. However, it is not advisable to rely only on a single approach for determining the electricity demand in such a fast-growing and changing system of the South African economy.

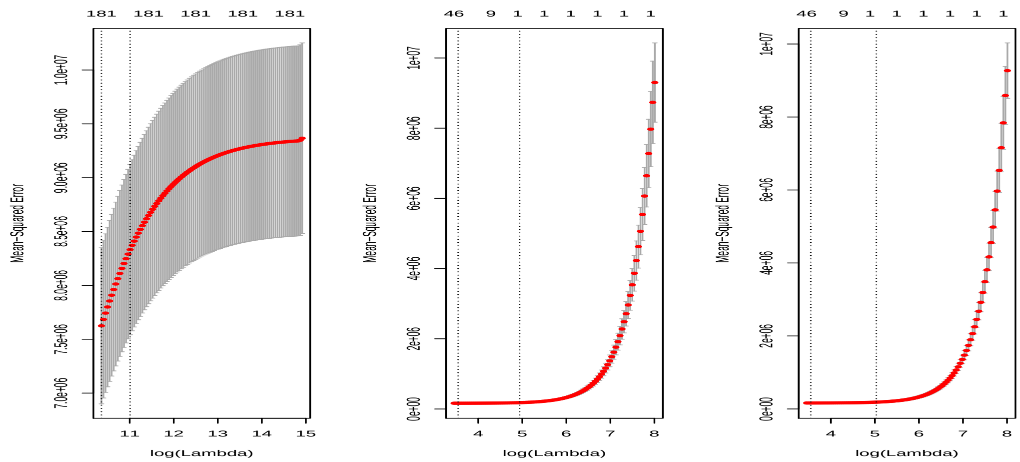

Figure 4.8: MPED cross validation estimates of the mean squared prediction error for ridge (left side), Lasso (center) and elastic net (on the right) as the function of log λ.

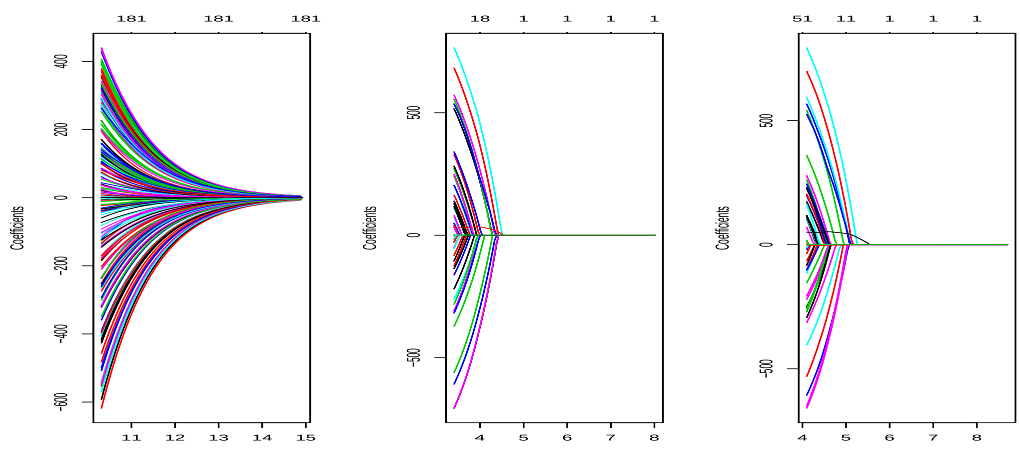

Figure 4.9: Coefficient estimates for ridge(left side), Lasso(centre) and elastic net (right side) for MPED data plotted versus log λ.

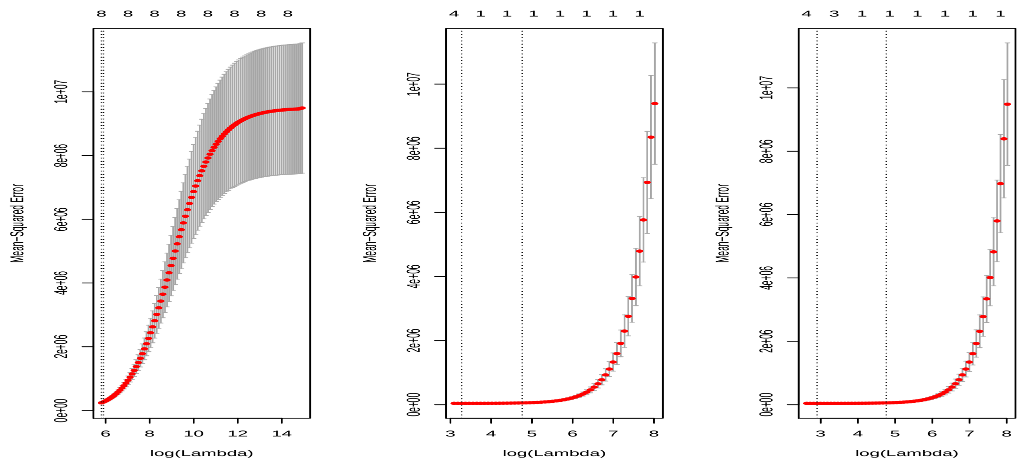

Figure 4.10: QPED cross validation estimates of the mean squared prediction error for ridge (left side), Lasso (center) and elastic net (on the right) as the function of log λ.

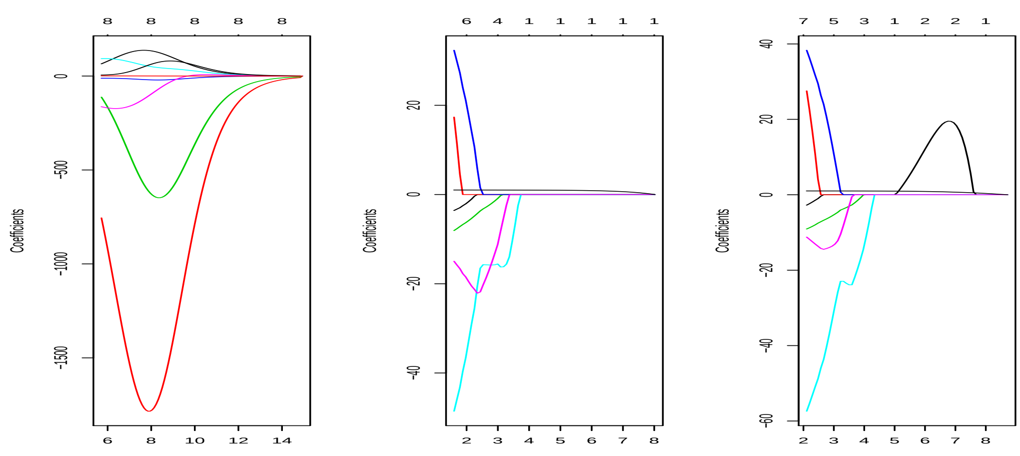

Figure 4.11: Coefficient estimates for ridge(left side), Lasso(centre) and elastic net (right side) for QPED data plotted versus log λ.

Chapter 5

Peak electricity demand forecasting using partially linear additive quantile regression models

5.1 Introduction

Planning for capacity expansion, including medium-term risk assessment, requires accurate predictions of peak electricity demand. Although many studies have been done on short, medium, or long-term peak electricity demand forecasting, the application and investigation of PLAQR models in forecasting peak electricity demand has not been carried out extensively in the literature. QR was introduced by Koenker and Bassett Jr (1978) as a mechanism for estimating models for the conditional median and the full range of other conditional quantiles, thereby giving a more complete statistical analysis of relationships among random variables than the conditional mean functions. It is an extension of the univariate quantile estimation. For an authoritative overview of the quantile regression methodology, see a monograph by Koenker (2005). GAMs, on the other hand, were developed by Hastie and Tibshirani (1990). PLAQR models are a combination of GAMs developed by Hastie and Tibshirani (1990) and QR models where the conditional quantile function comprises a linear parametric component and a nonparametric additive component Hoshino (2014). Among the first to introduce partially linear models are Engle et al. (1986) where they analysed the relationship between electricity usage and temperature.

In the literature, regression methods (including QR) have been used in modelling and forecasting time series data. Hoshino (2014) developed a two-step approach for estimating a PLAQR model. The usefulness of the developed PLAQR model is seen through an application to a real data set. Double-penalized quantile regression partially linear additive models are discussed by (Jiang, 2015). The study compares the methodology using a simulation study and an application to a real data set. Bayesian partially linear additive quantile regression models are used in simulation studies by (Hu et al., 2015). The models are then applied to two real data sets. For modelling time series data with multiple seasonalities, Bien et al. (2013) developed Lasso for hierarchical pairwise interactions in regression-based models. In another study, a method which satisfies the strong hierarchy for learning linear interaction models is presented by (Lim and Hastie, 2015). Results show the developed method to be comparable with past methods. However, the technique caters to both continuous and categorical variables.

Modelling and forecasting of peak electricity demand in South Africa has been studied to some extent in the literature (Rasuba et al. (2010); Sigauke and Chikobvu (2012); Sigauke et al. (2013); among others). Forecasts for electricity demand in South Africa for the period 2010-2035 are presented by Rasuba et al. (2010). Sigauke and Chikobvu (2012) developed an additive model for forecasting daily winter peak electricity demand in South Africa. The study shows that electricity demand in South Africa is highly sensitive to cold temperatures compared to hot temperatures. The modelling of extreme daily increases in peak electricity demand using South African data is studied by Sigauke et al. (2013) in which the authors focus on tail quantiles of the distribution of daily peak electricity demand. In this study, PLAQR models are applied to the modelling and forecasting of DPED on South African data.

5.2 Empirical results

5.2.1 Exploratory data analysis

The South African DPED for the national system from January 2007 to December 2013 is used in this study. The data for January 2007 to December 2010 is used for training and the remaining data for testing. The variables considered in this study are DPED as the response variable; the predictors are lagged demand denoted as DPEDLag1, DPEDLag2; calendar variables represented by "month", holiday denoted as DH, day of the week represented by Daytype; minimum electricity demand denoted by minED, average daily electricity demand denoted by AED and nonlinear trend represented by noltrend. Temperature data is from 28 South African weather stations. In this study, South Africa is split into two main thermal regions, i.e., coastal and inland. The temperature variables are average daily coastal temperature (ADTC), average maximum and minimum coastal temperature (maxTC and minTC), average minimum, average maximum, average daily interior temperature (minTI, maxTI and ADTI), respectively, average minimum of coastal and interior temperatures (AminTCI), average of average daily coastal and interior temperatures (AADTCI), average maximum of coastal and interior temperatures (AmaxTCI), the difference between average minimum of coastal and interior temperatures (DminTCI), the difference between average maximum of coastal and interior temperatures (DmaxTCI), the difference between the average of average daily coastal and interior temperatures (DADTCI). Averaging and finding the differences between temperature values of the two thermal regions are in line with work done by Fan and Hyndman (2012) in which they had two temperature sites. Out of 19 variables, ten are selected using Lasso via hierarchical interactions. The selected variables are: Daytype, noltrend, month, minED, AED, minTI, DminTCI, DADTCI, DPEDLag1 and DPEDLag2 and the pairwise interacting variables are (Daytype, DPEDLag1) and (noltrend, DPEDLag2). The final model is:

$$\begin{aligned}\log(\text{DPED}) = \text{DH} + \text{month} + s(\text{Daytype}) + s(\text{minED}) + s(\text{AED})\\ +s(\text{minTI}) + s(\text{DminTCI}) + s(\text{DADTCI}) + s(\text{noltrend})\\ +s(\text{DPEDLag1}) + s(\text{DPEDLag2}) + s(\text{Daytype,DPEDLag1})\\ +s(\text{noltrend,DPEDLag2}) + \varepsilon_{t,\tau}, \qquad (5.1)\end{aligned}$$

where s represents a B-spline.

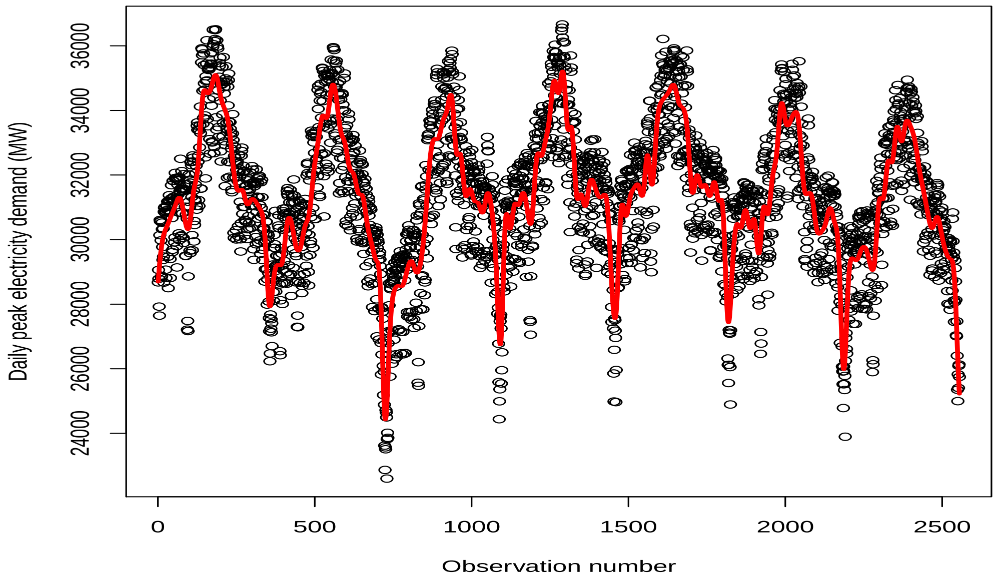

Figure 5.1: Daily peak electricity demand with a superimposed nonlinear trend.

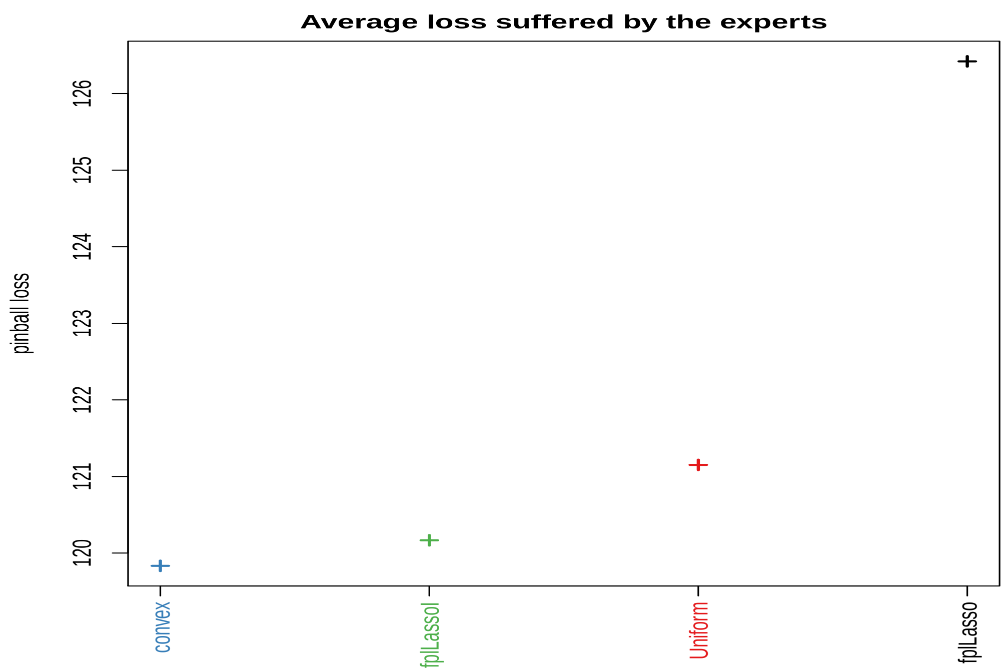

Figure 5.2: Average loss suffered by the models.

Table 5.1: Model comparisons.

	M1	M2	Combined
RMSE	326.34	306.44	306.08
MAE	252.84	240.33	239.67
MAPE	0.814	0.773	0.771

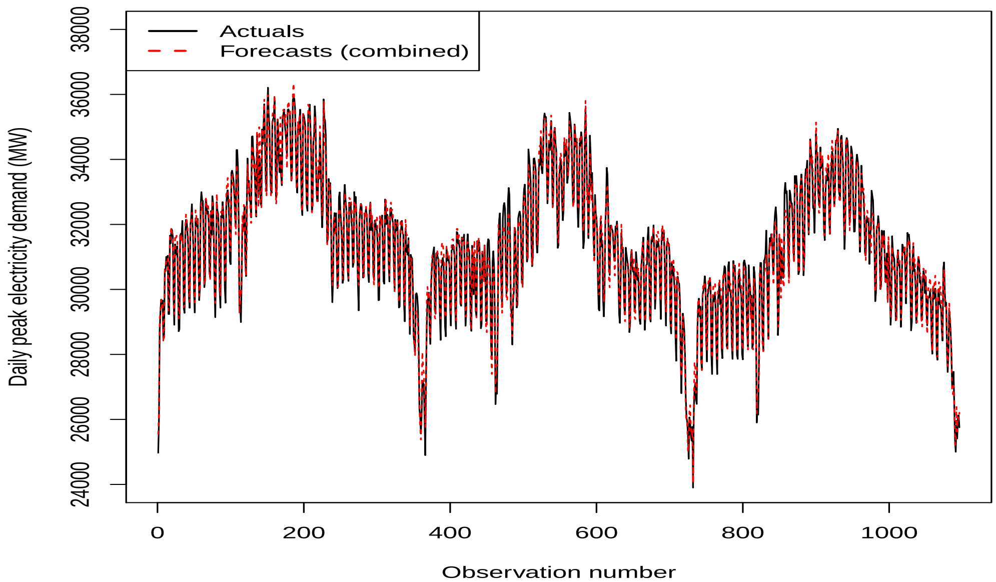

Figure 5.3: Actual demand superimposed with combined forecasts (January 2011 - December 2013.

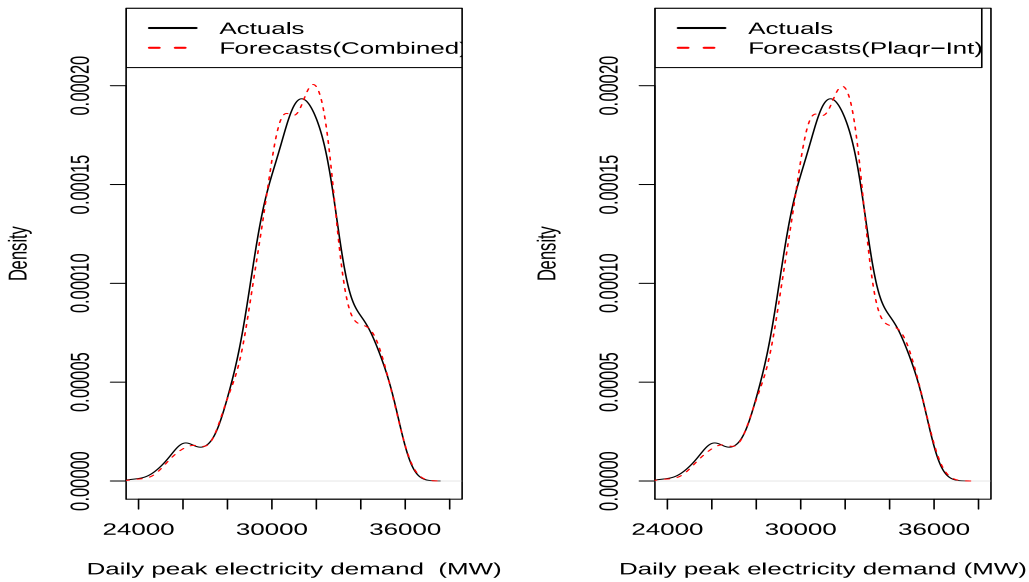

Figure 5.4: Density plots of actual demand superimposed with densities from combined forecasts and PLAQR model with interactions.

5.2.2 Results

Based on the RMSE, MAE and MAPE, model M2 is better than model M1. There is a slight improvement in the accuracy measures after combing the forecasts, as shown in Table 5.1.

5.2.3 Discussion of results

Figure 5.1 shows a plot of DPED with a superimposed nonlinear trend. Based on the pinball loss function, the average loss suffered by the PLAQR model with pairwise interactions (fplLassoI, i.e., model M2) is smaller than that of the PLAQR model without interactions (fplLasso, i.e., model M1) as seen in

Figure 5.2. The weights assigned to the forecasts from these two models are 0.14 for model M1 and 0.86 for the M2 model. Table 5.1 gives a summary of the accuracy measures for the models M1 and M2, respectively, together with the combined forecasts. A plot of DPED (solid black line) and forecasts (dashed red line) given in Figure 5.3 shows that the forecasts follow the actual DPED very well. The densities of the forecasted demand from the combined forecasts and the PLAQR model with pairwise interactions given in Figure 5.4 show a good fit for the DPED data density in both panels.

5.3 Conclusion

In this paper, we studied the application of PLAQR models to modelling and forecasting peak electricity demand. The conditional quantile function in a PLAQR consists of a nonparametric additive component and a linear parametric component. Empirical results have shown the usefulness of PLAQR models. Variable selection is initially done using Lasso via hierarchical interactions. Two models are considered, one without interactions and one with interactions. The model with pairwise interactions produced more accurate forecasts than the model without interactions. The forecasts from the two models are combined using a forecast combination algorithm in which the average loss suffered by the models is based on the pinball loss function.

Chapter 6

Prediction of extreme conditional quantiles of electricity demand

6.1 Introduction

South Africa needs to identify the key goals for the nation and how the electricity sector fits among its priorities. It makes sense to invest in a country that ensures the sustainability of renewable energy and energy efficiency. It might also be argued that inaccurate forecasting may seriously result in poor decision-making and bad policies. However, predicting electricity demand, especially for extreme conditions, is important for policymaking in utilising renewable energies. Hence, the quantile regression method is useful for such a purpose. To diversify the energy mix, the South African government has developed a renewable energy independent power producer procurement program (REIPPPP). The REIPPPP has proven to be a very successful program, especially in bringing renewable energy projects to commercial operations over the past five years Larmuth and Cuellar (2019). The country also came up with a new plan, namely the national development plan (NDP). The NDP is a plan for infrastructure development from 2013 to 2030. Moreover, the NDP is also an excellent guiding economic plan that sets the GDP growth target per annum for the country to be able to meet its economic, social and political objectives (Stands, 2015).

In identifying the national goals relevant to establishing renewable energy policy objectives, South Africa needs to identify the key goals for the nation and how the electricity sector fits in among its priorities. The country also needs to take urgent actions to ensure the sustainability of renewable energy and energy efficiency by 2030. According to Do (2015), renewable and energy efficiency have a positive impact on electricity demand during peak hours.

It is important to predict extremely high quantiles of electricity demand accurately. Uncertainties related to electricity demand have to be considered when predicting electricity demand. The superiority of extreme conditional quantile models over the least squares (conditional mean) model in the extremes of the conditional distribution is well established in the literature. Hence, accurate prediction of electricity demand at extreme levels could produce more useful information to decision-makers on the sustainability of renewable energy, energy generation and energy purchases. Furthermore, the accurate prediction of extreme electricity demand distributions would significantly mitigate load shedding and overloading and allow energy-efficient storage.

The main contribution of this paper is the use of AQR, EM and NLQR models in estimating the extremely high and extremely low quantiles of electricity demand using South African data. Such a study has not been carried out elsewhere to the best of our knowledge.

The highlights of the study are summarised as follows:

- The study carried out a comparative analysis of EM, AQR and NLQR models in predicting extremely high and low daily peak electricity demand;
- The identification of how electricity demand will change in the distribution networks in five to fifteen years in the future;
- The prediction of extremely high quantiles of DPED could help system operators know the possible largest demand that will enable them to supply adequate electricity to consumers;
- Knowing the possible largest electricity demand at a given point in time can help system operators shift demand to off-peak periods.

6.2 Empirical results

6.2.1 Exploratory data analysis

Table 6.1 presents the summary statistics of the DPED. During the sampling period, the minimum and maximum DPED values are 17 605MW and 37 158MW, respectively. The skewness value is -0.232, showing that the distribution of the DPED is skewed to the left. The density (Panel (c)) and box (Panel (d)) plot given in Figure 6.1 confirm that the distribution of the DPED is negatively skewed.

Table 6.1: Descriptive statistics of DPED.

Var	**Min**	**Q1**	**Mean**	**Median**	**Q3**	**Max**	**Skew**	**Kurt**
DPED	17605	25706	28688	29149	31596	37158	-0.232	2.287

Figure 6.1 (a) shows a plot of the DPED. The DPED data are not normally distributed as shown by Panel (b) of Figure 6.1. A plot of the DPED superimposed with a nonlinear trend variable is given in Figure 6.2. A penalised cubic regression spline is used to obtain the nonlinear trend, used as a covariate for all the models, except for the extremal mixture model, where a linear trend is used as a covariate.

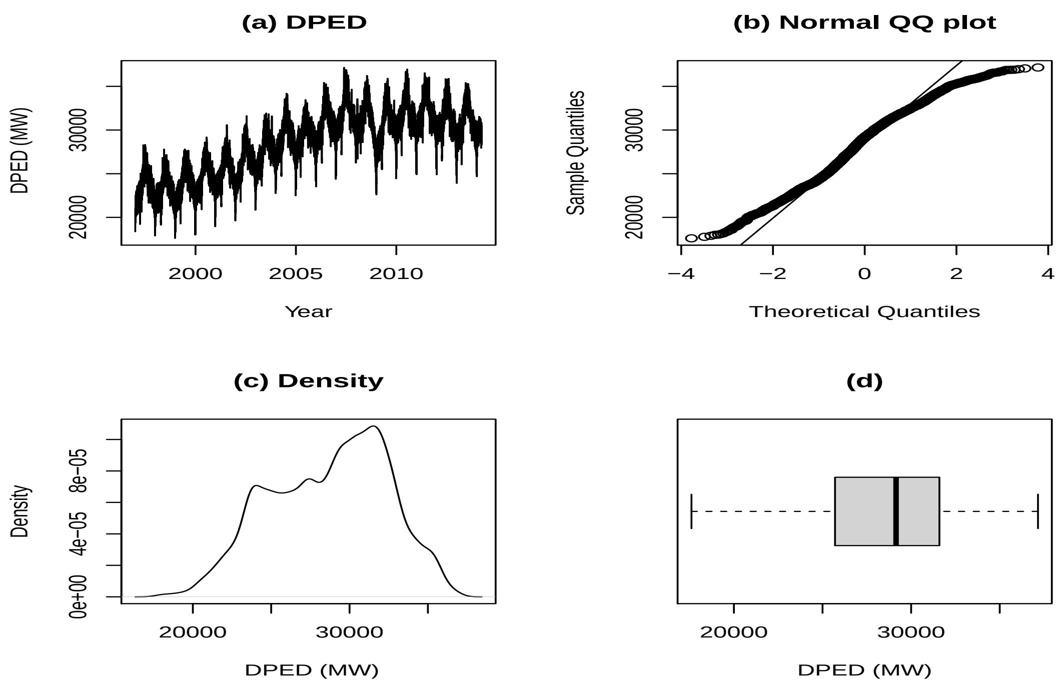

Figure 6.1: DPED plot (**a**), Q-Q plot (**b**), density plot (**c**) and box plot (**d**).

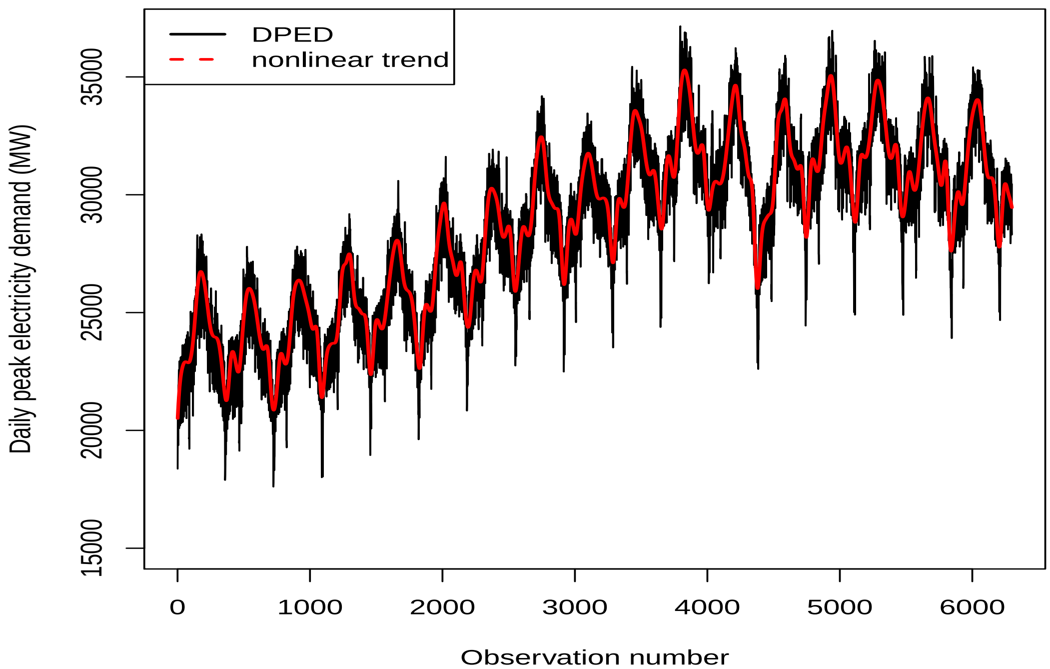

Figure 6.2: Daily peak electricity demand superimposed with a nonlinear trend.

6.2.2 Results

The probabilistic accuracy measures (scoring rules) are calculated as follows: we fit a parametric distribution to the forecasts and then estimated its parameters. The study considered a comparative analysis of three models, i.e., the AQR (M1), EM (M2) and NLQR (M3) models at 0.95, 0.99 and 0.9999 quantile forecasts. To evaluate the forecast accuracy measures of the AQR, EM and NLQR models, we used the CRPS, LogS, DSS, interval width and the pinball losses. The lower the values of the scoring rules, the better the prediction performance. The scoring rules are then computed based on the fitted parametric distribution. A comparison of the models at both extremely high and extremely low quantiles are given in Tables 6.2 and 6.3, respectively. At the 0.95 quantiles, the EM model has the lowest CRPS, LogS and DSS, making it the best fitting model at this quantile level. The AQR model is the best fit model at the 0.99, 0.999 and 0.9999 quantiles based on the CRPS and DSS evaluation metrics. Based on the prediction interval width coverage probability of 0.98 as shown in Table 6.2, all the models had valid coverage probabilities since they are all greater than 0.98. However, the EM model provided the largest coverage probability, 0.9886, and has the least number of observations below the 0.01 and 0.99 quantiles, respectively. As for the extremely low quantiles, the NLQR model has the smallest CRPS, LogS and DSS at the 0.05 quantile. At the 0.01 quantile, the AQR model has the smallest LogS and DSS values. The NLQR model is the best fit at the 0.001 quantiles based on the CRPS, LogS and DSS evaluation metrics.

Table 6.2: Model comparisons (extremely high quantiles).

95.0 th percentiles (0.95 quantile)				
Models	**CRPS**	**LogS**	**DSS**	**PL**
M1 (AQR)	2144.546	9.6088	17.4418	165.9795
M2 (EM)	2069.789	9.5560	17.3701	209.2725
M3 (NLQR)	2155.875	9.6116	17.4495	161.9122
M4 (Median)	2155.875	9.6116	17.4495	165.5629
99.0th percentiles (0.99 quantile)				
Models	**CRPS**	**LogS**	**DSS**	**PL**
M1 (AQR)	**2125.457**	**9.5947**	**17.4256**	43.2765
M2 (EM)	2131.232	inf	17.4315	56.8979
M3 (NLQR)	2163.629	inf	17.4598	42.6107
M4 (Median)	2163.629	inf	174598	43.2509
99.9th percentiles (0.999 quantile)				
Models	**CRPS**	**LogS**	**DSS**	**PL**
M1 (AQR)	2172.784	inf	17.4759	5.529
M2 (EM)	2426.084	inf	17.7509	7.9267
M3 (NLQR)	2190.201	inf	17.4958	5.3599
M4 (Median)	2190.201	inf	17.4958	5.4869
99.99th percentiles (0.9999 quantile)				
Models	**CRPS**	**LogS**	**DSS**	**PL**
M1 (AQR)	2168.997	inf	17.4747	0.6116
M2 (EM)	2945.86	inf	18.3882	1.0272
M3 (NLQR)	2202.221	inf	17.5149	0.6053
M4 (Median)	2202.221	inf	17.5149	0.6274
Interval widths for the 0.01 and 0.99 quantiles (CP = 0.98)				
Models	**Ave IW**	**Cov Prob**	**Below 0.01 quantile**	**Above 0.99 quantile**
M1 (AQR)	5364	0.9833	59	46
M2 (EM)	6807	0.9886	52	20
M3 (NLQR)	5312	0.9806	65	57
M4 (Median)	5385	0.9843	56	43

Table 6.3: Model comparisons (extremely low quantiles).

5.0th percentiles (0.05 quantile)				
Models	**CRPS**	**LogS**	**DSS**	**PL**
M1 (AQR)	3826.125	10.3962	19.1919	226.0744
M2 (EM)	4225.651	10.5635	19.6090	287.7971
M3 (NLQR)	3789.174	10.3804	19.1481	220.9624
M4 (Median)	3789.174	10.3804	19.1481	224.6259
1.0th percentiles (0.01 quantile)				
Models	**CRPS**	**LogS**	**DSS**	**PL**
M1 (AQR)	4529.206	10.6899	20.0098	64.0034
M2 (EM)	5089.073	10.9302	20.6903	79.2499
M3 (NLQR)	4523.984	10.6905	20.0264	63.6298
M4 (Median)	4523.984	10.6905	20.0264	64.45815
0.1th percentiles (0.001 quantile)				
Models	**CRPS**	**LogS**	**DSS**	**PL**
M1 (AQR)	5050.386	10.9086	20.6212	7.793396
M2 (EM)	5257.614	11.0022	20.9055	8.342102
M3 (NLQR)	5033.644	10.9037	20.6112	7.756549
M4 (Median)	5033.644	10.9037	20.6112	7.781235

Figures 6.3–6.5, respectively, show the DPED superimposed with the 0.95, 0.99 and 0.01 quantiles from the M1, M2 and M3 models, respectively. All figures reflect the peak identification in the electricity demand based on the 0.95, 0.99 and 0.999 quantiles of the DPED data. Furthermore, Figures 6.3–6.5 show the solid black line depicting the DPED from January 1997 to May 2014, the red dotted line the M1 model distribution, the blue dotted line the M2 model and the green dotted line the M3 model distribution, respectively.

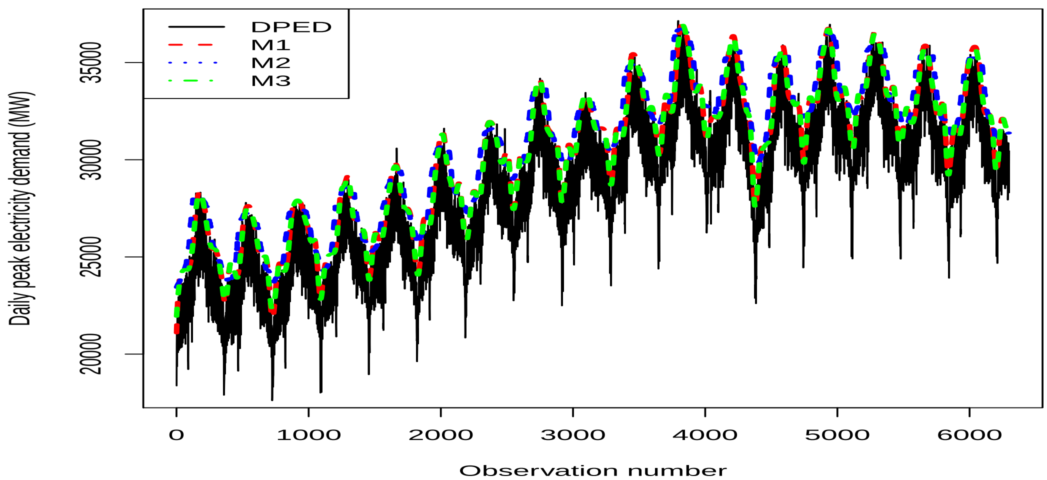

Figure 6.3: DPED superimposed with the 0.95 quantiles from the M1, M2 and M3 models.

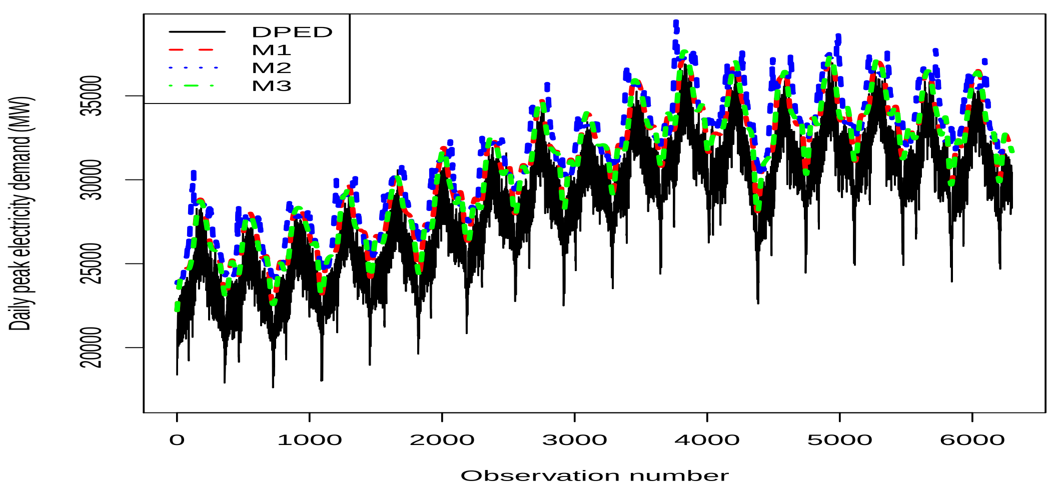

Figure 6.4: DPED superimposed with the 0.99 quantiles from the M1, M2 and M3 models.

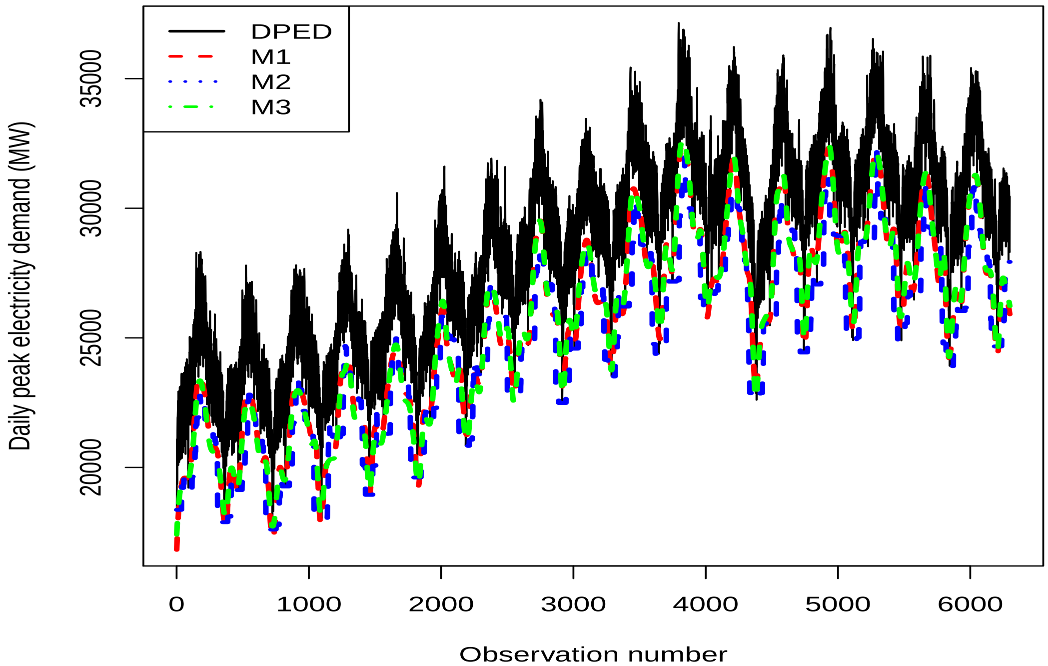

Figure 6.5: DPED superimposed with the 0.01 quantiles from the M1, M2 and M3 models.

Figures 6.6–6.8, respectively, show three box plots of the 0.99, 0.999 and 0.9999 quantiles with the AQR, EM and NLQR models. The box plots for the EM model have shorter left tails and longer right tails than the other two models, AQR and NLQR, respectively. The medians of the AQR and NLQR models are all at the same level; however, their box plots show very different distributions for the DPED data.

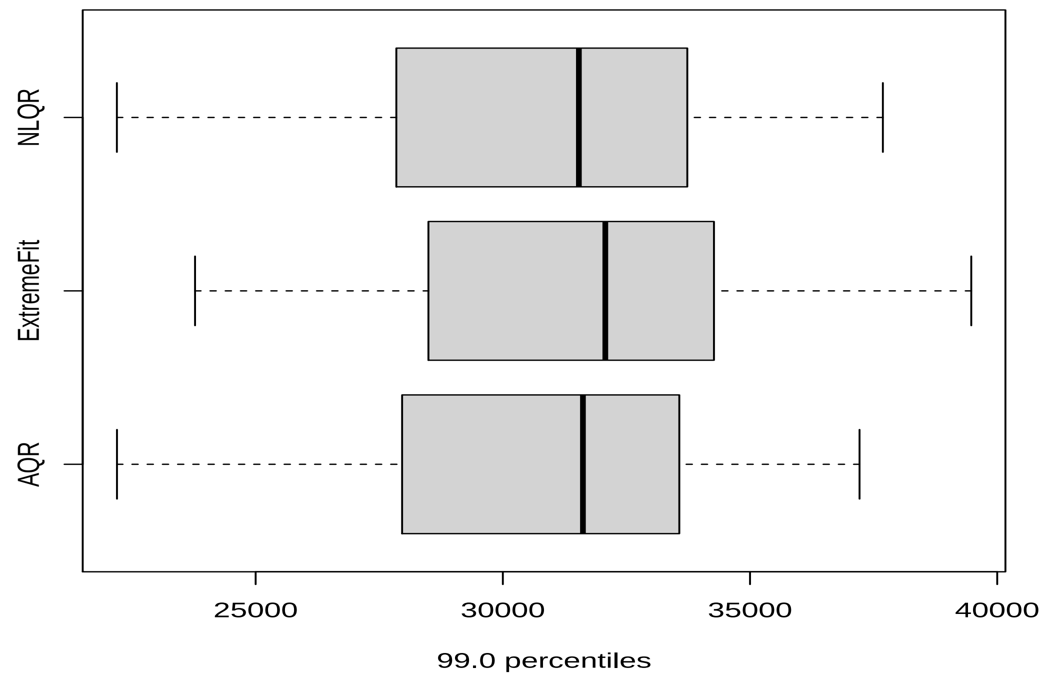

Figure 6.6: Box plots of the 0.99 quantile.

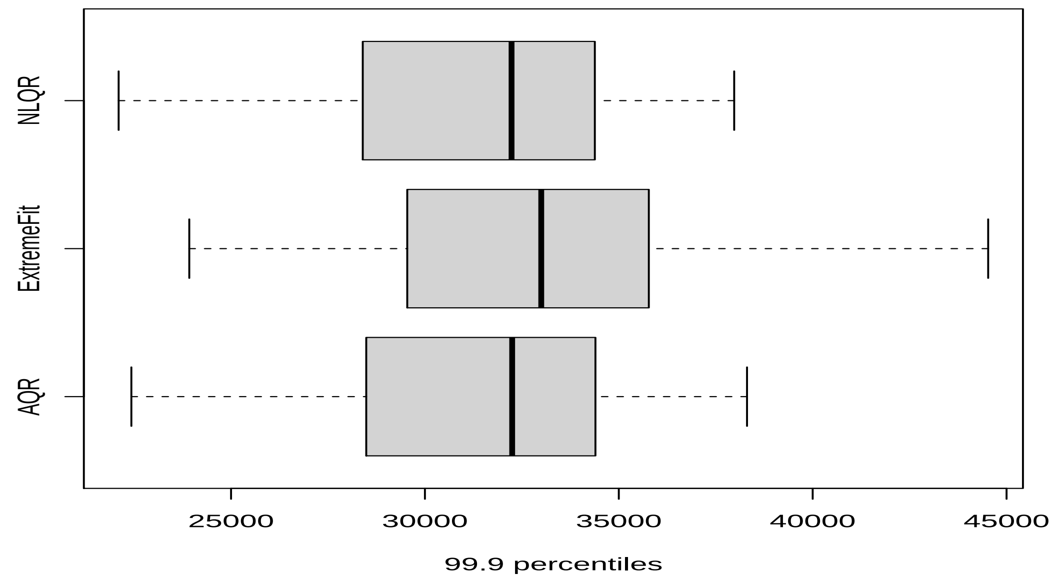

Figure 6.7: Box plots of the 0.999 quantile.

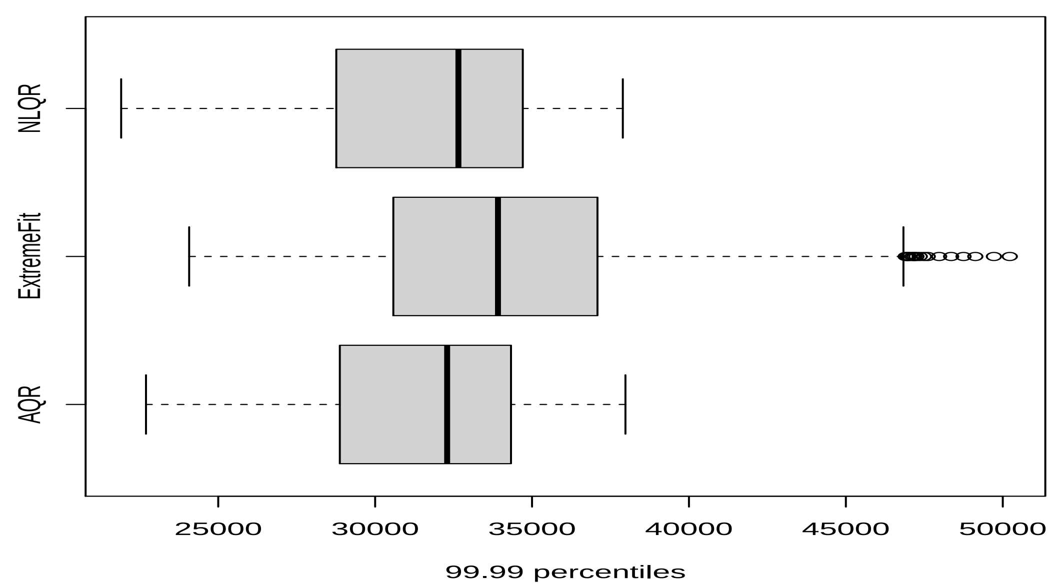

Figure 6.8: Box plots of the 0.9999 quantile.

Figure 6.9 shows the forecast distributions for the DPED data from 1997 to 2021. It also summarises the forecast distributions drawn at each specific date interval. For every five years, the histogram represents a sample from the model's forecast distribution.

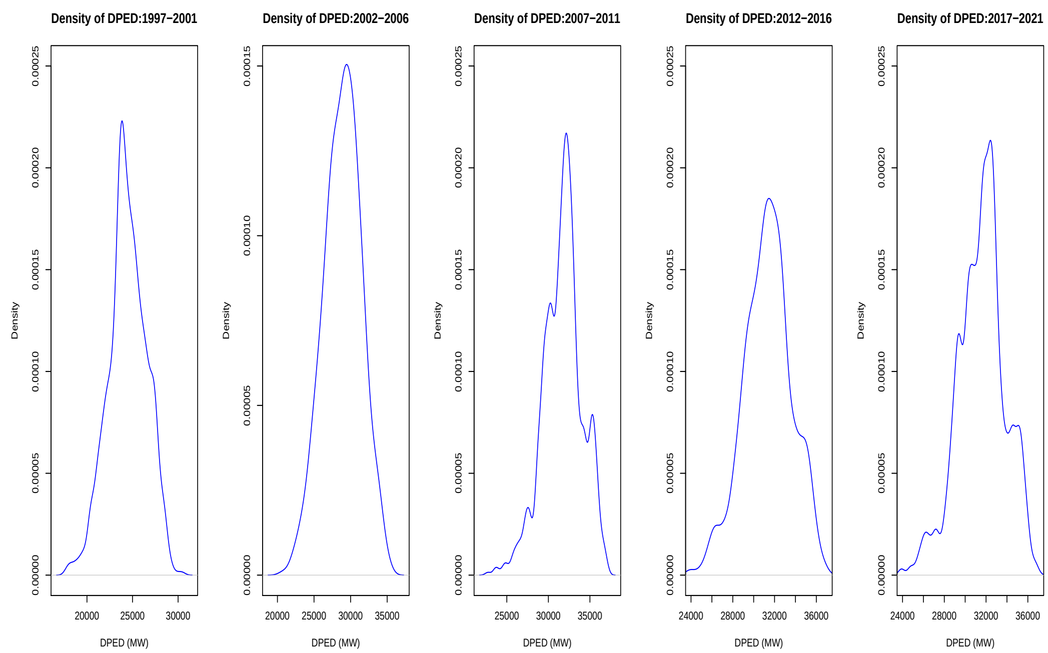

Figure 6.9: Distributions for the DPED for every five years.

The Murphy diagrams in Figures 6.10–6.12 show empirical scores and differences in scores for the M1 and M2, M1 and M3 and M2 and M3 models, respectively. A negative difference means that the regime-switching forecast is preferable. It must be noted that Murphy diagrams (Figures 6.10–6.12) might lead to inconclusive situations in which neither of the three forecast methods dominates the other. As a result, it would be unhelpful in decision-making. The p-values of the three models for three different forecasting horizons are analysed. The forecast accuracies of the 0.95, 0.99, 0.999 and 0.9999 quantiles are summarised. Since all p-values were less than 0.05, we, therefore, rejected the null hypothesis that there is no significant difference in the predictive abilities amongst the models. The forecasting performance of all the models is significant. Hence, the forecasting accuracy of M1 is better than that of the M2 and M3 models in extremely high quantiles, and the M3 model is better than the M1 and M2 models in extremely low quantiles.

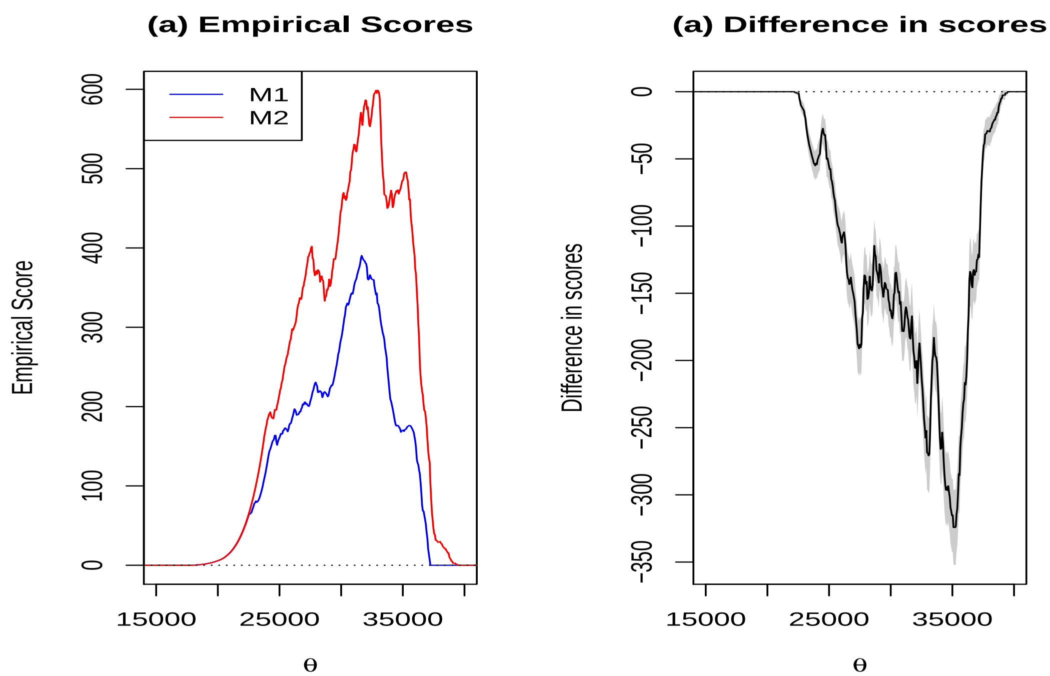

Figure 6.10: Murphy diagrams for the comparison of the 0.99 quantiles of AQR99 and Extremal99 (M1 and M2).

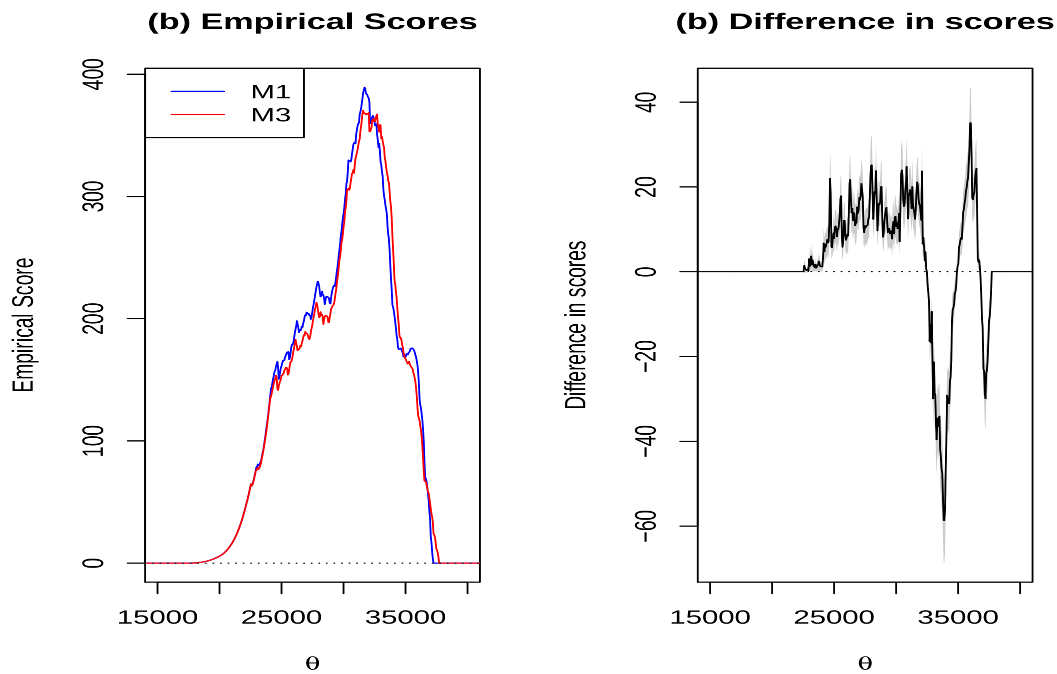

Figure 6.11: Murphy diagrams for the comparison of the 0.99 quantiles of AQR99 and NLQR99 (M1 and M3).

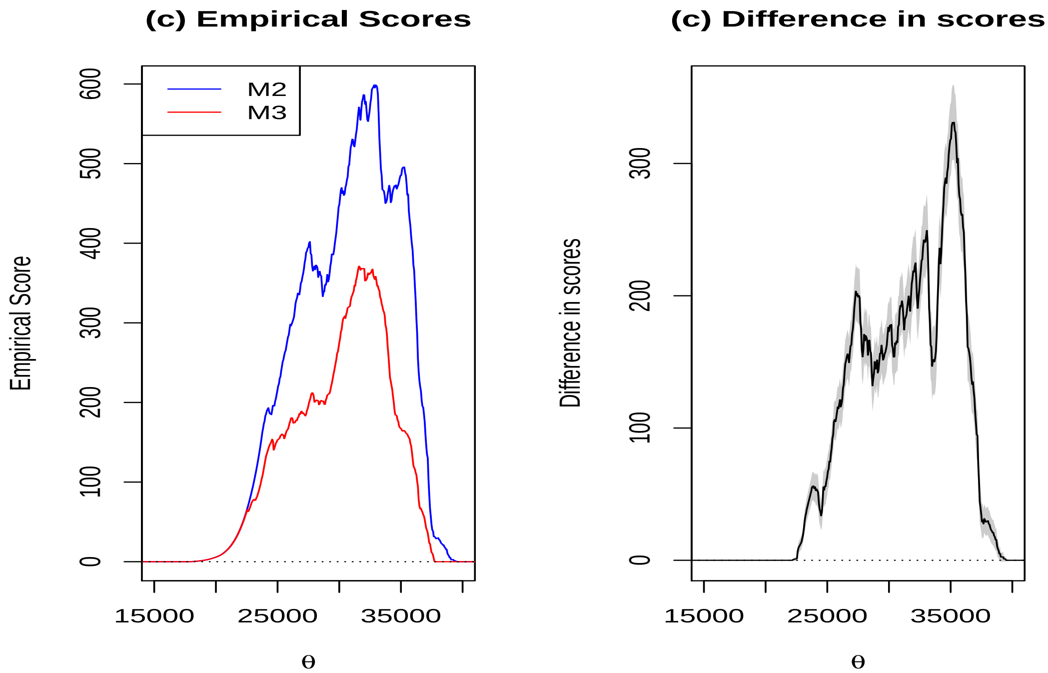

Figure 6.12: Murphy diagrams for the comparison of the 0.99 quantiles of Extremal99 and NLQR99 (M2 and M3).

6.2.3 Discussion of the results

The current study is motivated by the need to accurately predict the greatest possible demand for electricity at any given point in time. This can help system operators shift demand to off-peak periods, including generating units during peak periods. The study carried out a comparative analysis of the AQR, EM, and NLQR models in predicting extremely high and extremely low daily peak electricity demand quantiles. Other approaches, for example Beirlant et al. (2004), Gajowniczek and Zabkowski (2017) and Gardes and Girard (2010) among others, underestimated conditional quantiles or failed to estimate extreme conditional quantiles from their proposed models.

The additive quantile regression models are becoming increasingly popular in many applications as they are flexible and robust. The AQR model does not require a predetermined function fit; however, it determines the best fit from the DPED data under extremely high quantiles. Table 6.2 shows the AQR model as a powerful method as suggested by the low evaluation metrics. Table 6.3 shows NLQR to be the best fitting model for extremely low quantiles. The predictive performance of the models is evaluated based on three evaluation metrics, the CRPS, LogS and DS, respectively. The models are compared based on these scoring rules. The comparative forecast evaluation generally applauds the critical features of out-of-sample forecasts in modelling comparison. The AQR model shows the smallest values of the scoring rules in all three extremely high quantiles, except for one quantile. All sets of scores' (CRPS, LogS and DSS) values in Table 6.2 suggest that the M1 model was the best, while Table 6.3 indicates that M3 is the best fitting model. Hence, the model M1 provided the highest predictive accuracy at the 0.95 and 0.99 quantiles, respectively, as given in both Figures 6.3 and 6.4. In this study, accurate predictions in extreme conditional quantiles of electricity demand are necessary for planning power systems and assessing investment projects in South Africa. Based on the Murphy diagrams, M1 (AQR model) has the highest predictive ability compared to the extremal mixture and nonlinear quantile regression models. The question of model comparisons in both extremely high and low quantiles in this study arises from the need to:

- address the uncertainties that seem to be ignored in practice (for example, the uncertainty in the process that is generating the occurrence of the extreme events);

- quantify the uncertainties in the estimated parameters of the distribution;
- predict extremely high quantiles of daily peak electricity demand. This helps system operators know the possible largest demand, which will enable them to supply adequate electricity to consumers and shift load to off-peak periods.

6.3 Conclusion

The paper illustrates the prediction of extreme conditional quantiles of electricity demand using South African data. It contributes to the solutions to the challenges of extreme quantile forecast evaluations. In particular, the probabilistic accuracy measures such as CRPS, Logs, DSS and PL are introduced to evaluate the forecast performances. An extremal mixture model in which a kernel density fit the bulk model and a Pareto distribution fit to the tail model is compared with an additive quantile regression and a nonlinear quantile regression model. The empirical results show that the AQR model works well at extremely high and low quantile levels as it produces the most accurate predictions. Furthermore, the additive quantile regression and nonlinear quantile regression models fit well based on the Murphy diagrams for comparing the 0.99 quantiles compared to the other models. The results from this study could be useful to decision-makers in power utility companies in predicting extremely high and low electricity demand, thereby assisting them in managing the risk of overprediction and underprediction.

Chapter 7

Reliable predictions of peak electricity demand and reliability of power system management

7.1 Introduction

It is important to have accurate and reliable DPED forecasts as they are useful in reliability assessments using indices such as generation reserve margin (GRM), loss of load probability (LOLP), loss of load expectation (LOLE), forced outage rates (FORs) including availability factors (AFs) among others (Čepin, 2011). Power system reliability is the probability that an electric power system can perform a required function under given conditions for a given time interval (Čepin, 2011). Some of the important measures in reliability modelling include reliability indexes such as the LOLE, which is derived using daily peak load Čepin (2011), including, among others, the extreme peak load frequency (EPLF), which is the average number of peak loads above a sufficiently high threshold.

7.2 Empirical results

7.2.1 Exploratory data analysis

Data from January 2008 to December 2011 is used for training, while the remaining data is from January 2012 to December 2013 for testing. Operational forecasts will be from 1 January 2014 to 31 December 2016. The additive quantile regression models are without interactions and with pairwise interactions. Figure 7.1 presents a summary of the descriptive statistics of DPED. As shown by the skewness and kurtosis values in Figure 7.2, DPED data does not follow a normal distribution. This is confirmed by the Jarque-Bera test at the five percent significance level. We present two of some of the reliability indices, GRM and LOLE, calculated using the following formulae:

$$GRM = \frac{capacity\ in\ service - peak\ load}{capacity\ in\ service} \times 100 \tag{7.1}$$

and

$$LOLE_p = \sum_{i=1}^{n} p_i(c_i - DPED_i)\ days/period, \tag{7.2}$$

where, $LOLE_p$ is the value of LOLE in period p in n days, c_i is the available capacity on day i, $DPED_i$ is the forecasted peak load on day i and p_i is the probability of loss of load on day i (Čepin, 2011).

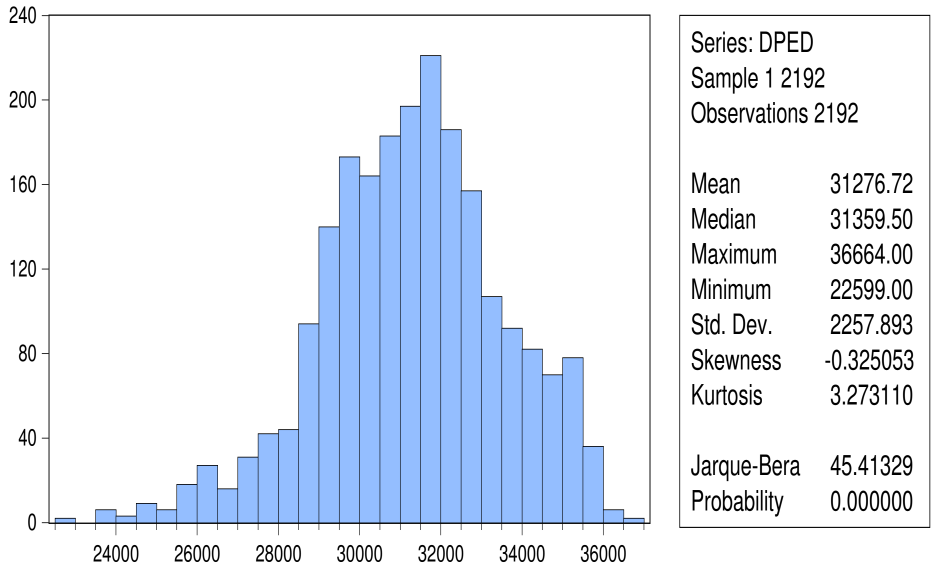

Figure 7.1: Summary Statistics of DPED (1 January 2008 to 31 December 2013).

Figure 7.2 shows a time series plot of DPED at the top left panel. The density, normal quantile to quantile (QQ) and box plots show that DPED data is non-normal. Likewise, Figure 7.9 presents operational forecasts of DPED with density, quantile to quantile and box plots. The diagnostic plots for the point process model fitted to exceedances above the threshold ($\hat{\tau} = 1260.616MW$) are given in Figure 7.11.

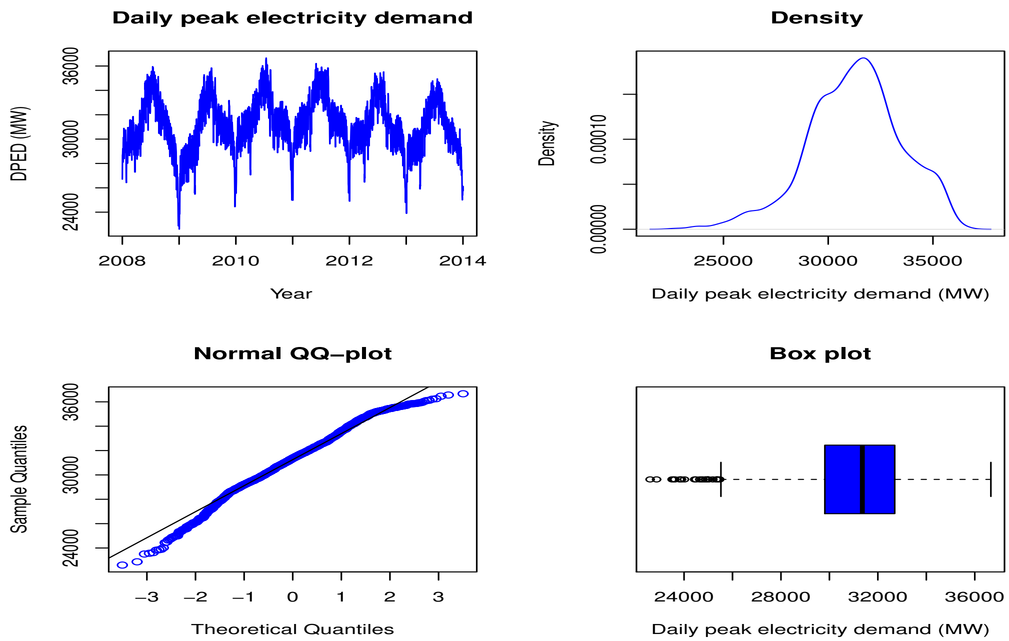

Figure 7.2: Daily peak electricity demand from 1 January 2008 to 31 December 2013.

7.2.2 Results

Based on the pinball loss, model M5 (fqgamOS) has the smallest value of 457.0219, as shown in Table 7.2 confirming that it is the best fitting model. M5 is the AQR model without interactions. We shall use the GRM of 40036MW as shown in Table 7.3. In 2012 the peak demand was 35706MW and it was on Monday 20 July. In 2013 it was 35681MW and it was on Monday 1 July and in 2014, it was 35480MW and it was on Tuesday 29 July. Although there is a general increase in the reserve margins for the forecasted peak DPED for the five years 2012-2016 as shown in Table 7.3, all the predicted reserve margins are below 15%. The extreme peaks plot of DPED from 1 January 2012 to 31 December 2016 is given in Figure 7.12. Figure 7.13 shows that most daily peak loads occur between hours 18:00 and 20:00.

7.2.3 Discussion of results

A plot of DPED overlaid with a non-linear trend (solid curve) for the period 2008 to 2014 is given in Figure 7.3.

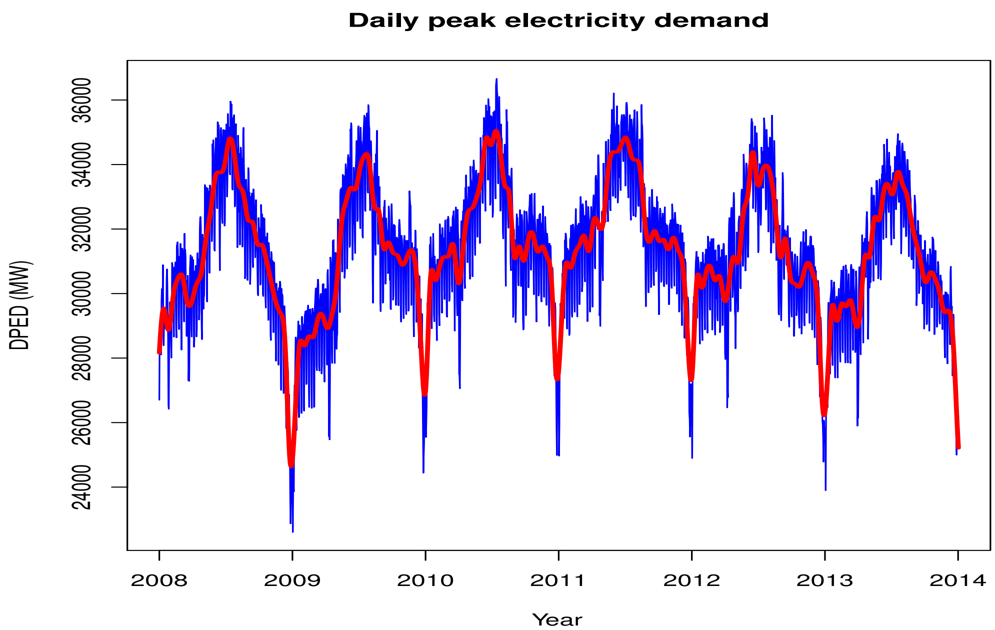

Figure 7.3: Plot of DPED overlaid a non-linear trend (solid curve).

The models developed are M1 (fqgam), M2 (fqgamI) and M3 (gbm). The opera R package developed by Gaillard et al. (2016) is used in combining the forecasts from the three models. The weights assigned to the forecasts from the three methods are 0.0917, 0.717 and 0.191 for M1, M2 and M3 models, respectively. Table 7.1 and Figure 7.4 both shows the estimated pinball losses. Before combining the forecasts, model M2 is the best model since it has the lowest pinball loss value. After combining the forecasts, the pinball loss of the convex model (M4) is the lowest.

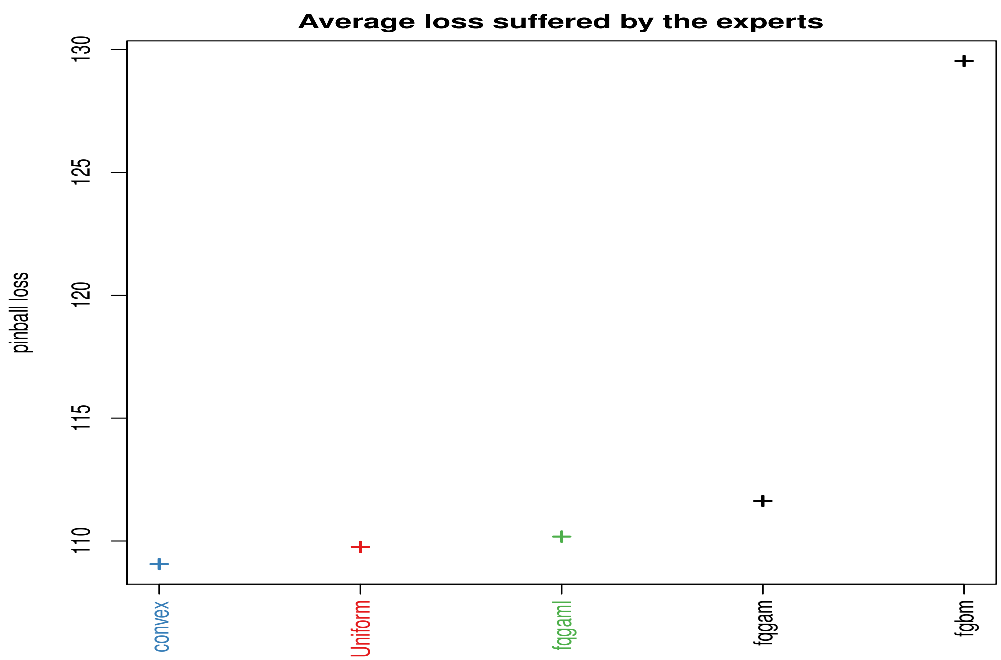

Figure 7.4: Plot of the experts (models) against the pinball loss.

Table 7.1: Error measures (forecast given covariates: 1 Jan 2012 to 31 Dec 2013).

	M1(fqgam)	**M2(fqgamI)**	**M3(fgbm)**	**M4(Combined)**
Pinball loss	111.6279	110.1792	129.5315	108.5632
CRPS	1152.902	1152.313	1156.65	1151.936

Table 7.2: Error measures (operational forecasts: 1 Jan 2012 to 31 Dec 2013).

	M5(fqgamOS)	**M6(fqgamOSI)**	**M7(fgbmOS)**
Pinball loss	457.0219	458.1457	458.3564

Table 7.3: Estimated reserve margins.

Year	**Capacity(MW)**	**Peak demand(MW)**	**Reserve margin(%)**
2012	40036	35706	10.82
2013	40036	35681	10.88
2014	40036	35480	11.38
2015	40036	35012	12.55
2016	40036	35000	12.58

Table 7.4: Transition matrix of the states.

	Extreme	**increase**	**Decrease**
Extreme	0.40707965	0.5309735	0.0619469
increase	0.03344482	0.6948161	0.2717391
Decrease	0.03061224	0.3469388	0.6224490

A plot of actual DPED overlaid with combined forecasts for the testing period, 1 January 2012 to 31 December 2013 is given in Figure 7.5. The forecast values follow remarkably well with the actual demand.

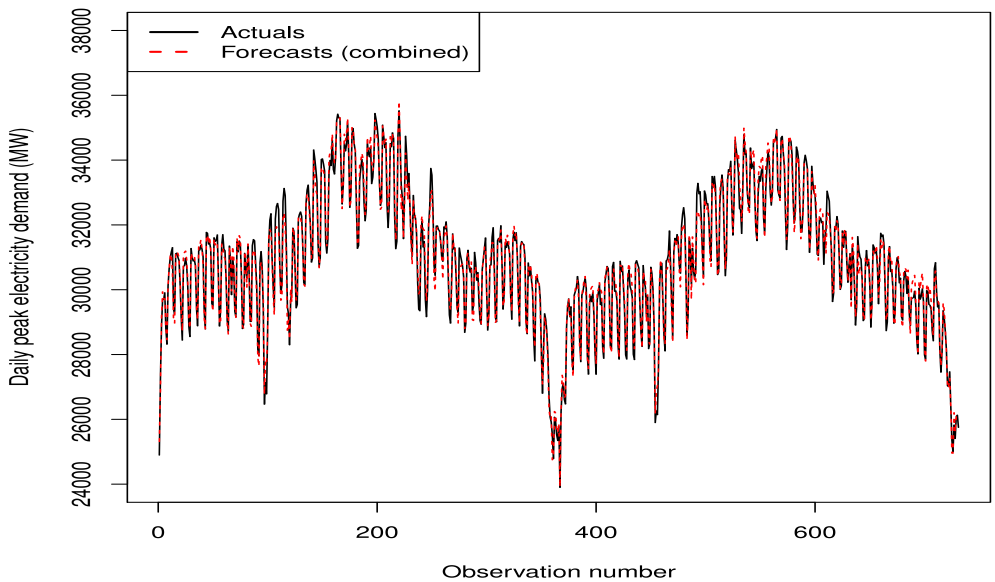

Figure 7.5: Actual demand overlaid with combined forecasts (1 January 2012 to 31 December 2013).

Actual demand density plot overlaid with those from the combined and AQR models with interactions (fqgamI) are given in Figure 7.6.

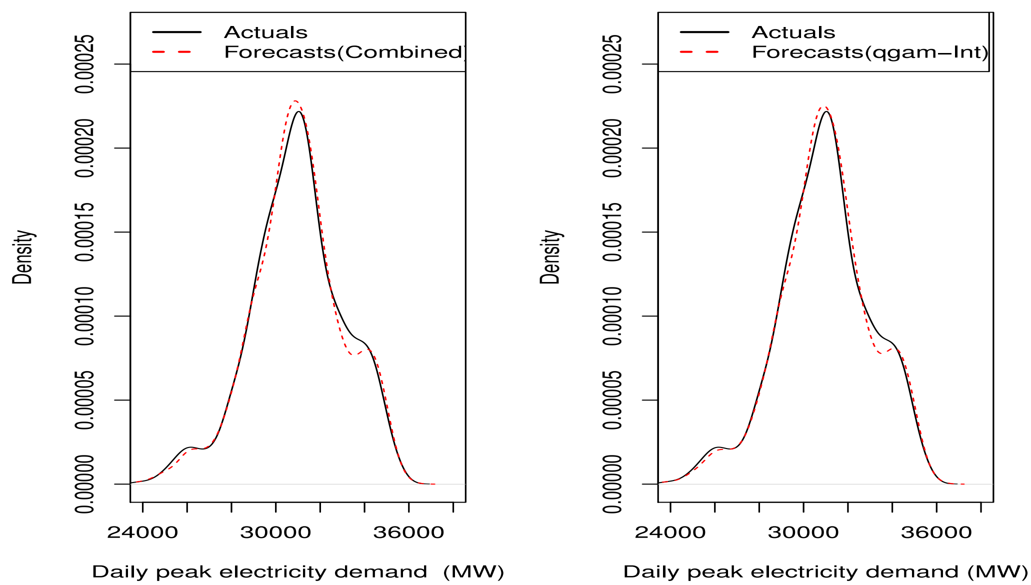

Figure 7.6: Actual demand overlaid with densities from combined and quantGAM with interaction forecasts (1 January 2012 to 31 December 2013).

The DPED density for the forecasted demand (1 January 2012 to 31 December 2013) together with densities for quantiles $q_0.01$ and $q_0.99$ are shown in Figure 7.7. The $q_0.99$ shows that its highly unlikely that DPED demand will exceed 38 000MW and the $q_0.01$ also shows that it is highly unlikely for DPED to be below 23 500MW. This is important for the system operator who has to plan the scheduling and dispatching of electricity to customers and strategic managers for capacity expansion planning. Figure 7.8 shows a summary statistics for operational forecasts of DPED with skewness and kurtosis values. Figure 7.10 shows the estimated threshold ($\hat{\tau} = 1260.616MW$) which is a vertical line separating the bulk model (kernel density) from the tail model (GPD). A three-state discrete Markov Chain problem is proposed in modelling frequency of occurrences of DPED. The three states are extreme ($\tau > 1260.616$), increase ($0 \leq \tau \leq 1260.616$) and decrease ($\tau < 0$). The transition matrix and transition diagram are given in Table 7.4 and Figure 7.14 respectively.

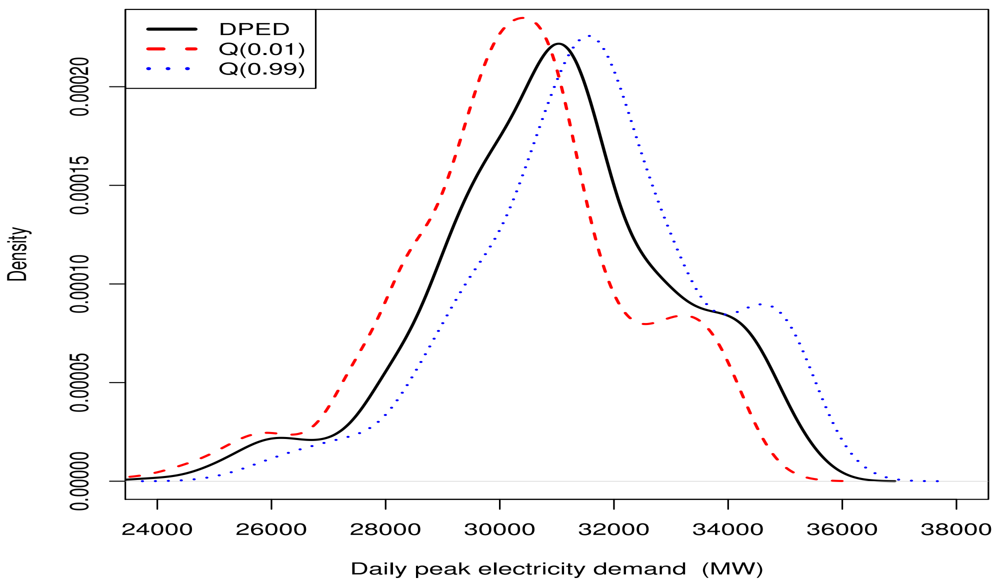

Figure 7.7: DPED density for the forecasted demand (1 January 2012 to 31 December 2013) with densities for quantiles $q_0.01$ and $q_0.99$.

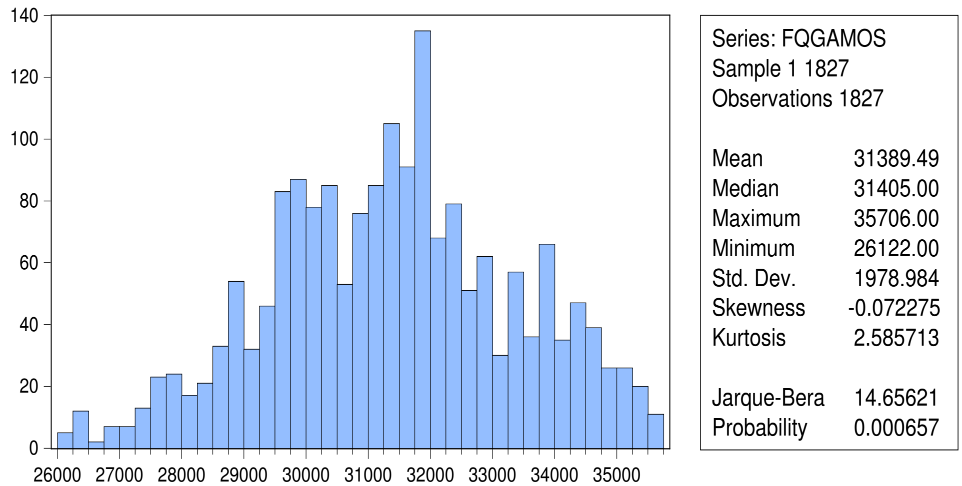

Figure 7.8: Summary statistics for operational forecasts of DPED (1 January 2012 to 31 December 2016).

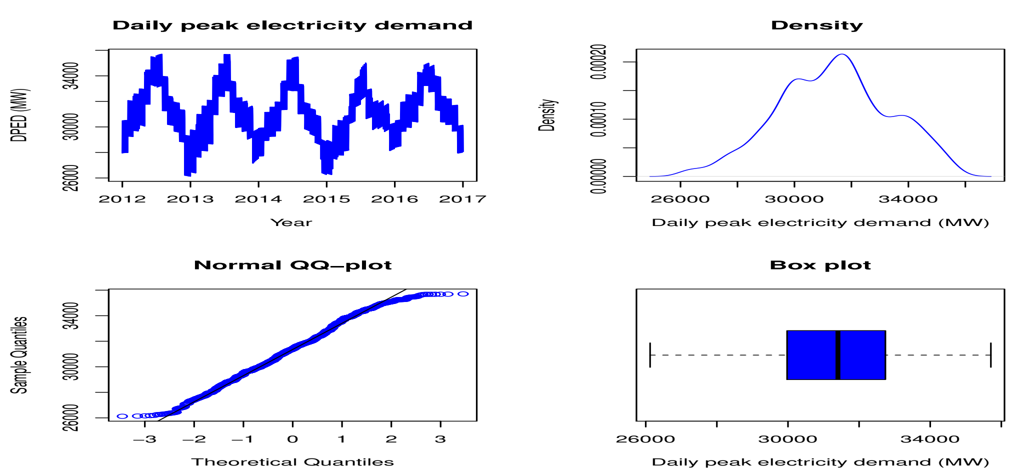

Figure 7.9: Operational forecasts of DPED (1 January 2012 to 31 December 2016).

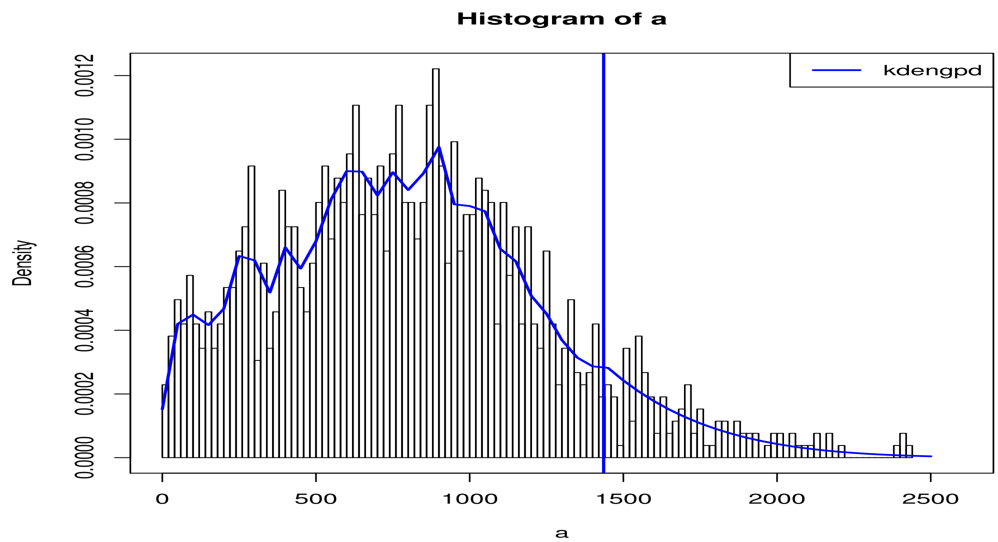

Figure 7.10: Threshold estimation ($\hat{\tau} = 1260.616MW$).

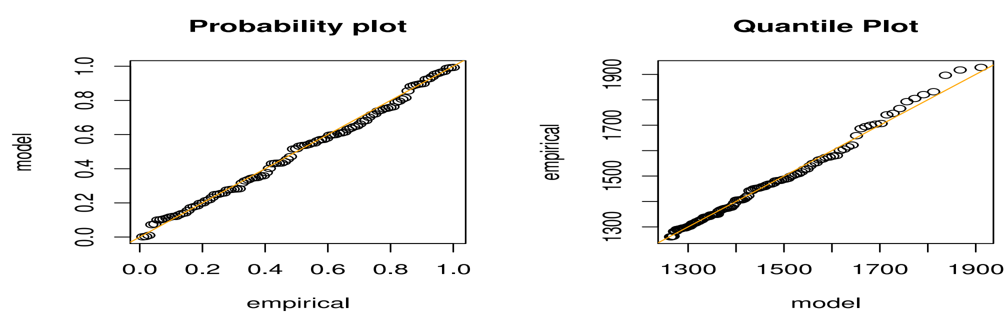

Figure 7.11: Diagnostic plots for the point process model fitted to exceedances above the threshold.

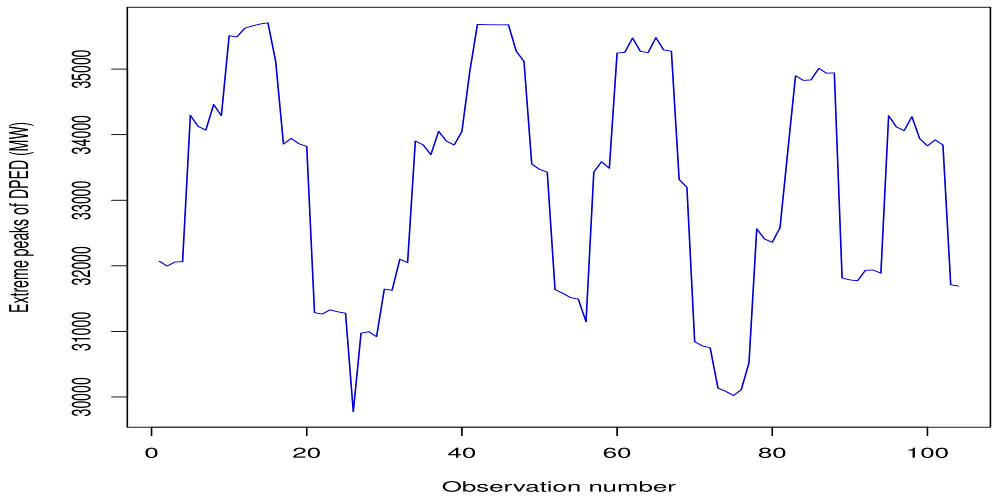

Figure 7.12: Extreme peaks of DPED for the period 1 January 2012 to 31 December 2016.

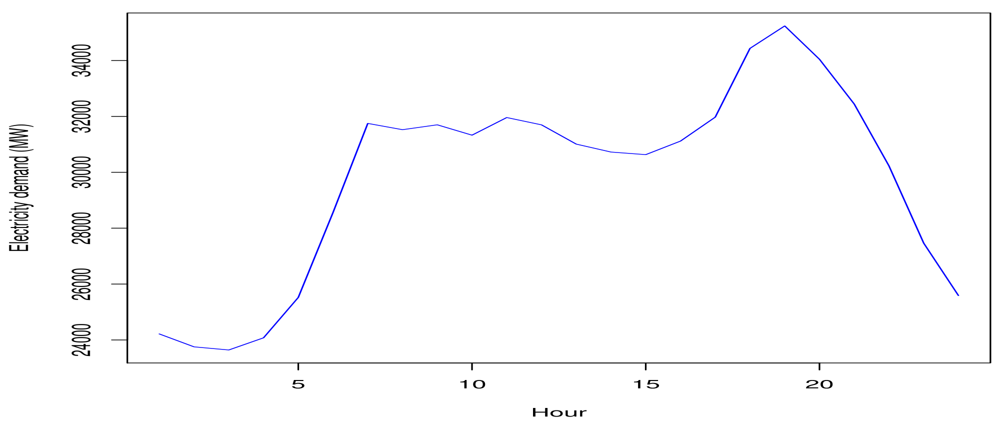

Figure 7.13: Typical daily load profile for South Africa (Thursday 13 June 2013).

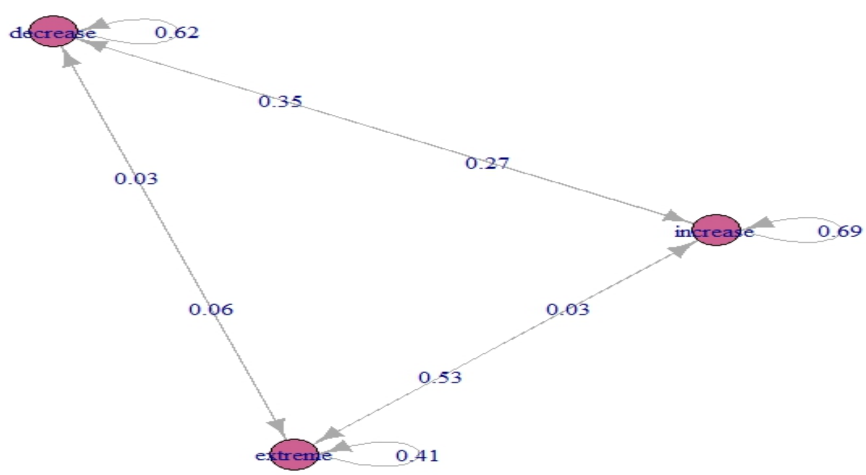

Figure 7.14: Transition diagram of the three states.

7.3 Conclusion

Power system reliability requires a probabilistic characterization of extreme peak loads, which results in severe stress to the system and causes problems to the grid. Therefore, accurate predictions of long-term electricity demand are very important as such forecasts can be used in the timing and rate of occurrence of such extreme peak loads. This is particularly important in reliability modelling. Severe stress on the grid system occurs when the peak electricity demands amplitude exceeds a predetermined sufficiently high threshold which is determined using extreme value theory methods. The findings of this study provide appropriate best-fitting models for assessing the system performance and identifying the system's weak points. An analysis of power systems reliability in the presence of random extreme peak loads is of great importance because peak loads affect the technical planning of long-term electricity demand. Hence, modelling the upper tail of the distribution of peak loads is important in reliability modelling.

Chapter 8

Conclusion

8.1 Introduction

Energy security has become an important issue recently, not just in South Africa but also worldwide. Therefore, the interventions are necessary to improve forecasting techniques' performance and efficiency. In this study, quantile regression approaches are utilised comparatively as the comprehensive forecasting solution for long-term peak electricity demand using South African data. This chapter summarises the main conclusions about modelling and forecasting long-term peak electricity. It is organised as follows: Sections 8.2 and 8.3 discuss the overall modelling summary and key findings. Section 8.4 gives the study's concluding remarks, while Sections 8.5 and 8.6 present the study's limitations and future research directions.

8.2 Summary of modelling by chapters

In Chapter 4, the study starts by overviewing how long-term electricity demand forecasting is important in the South African economy. It also highlights the challenges of using inaccurate forecasting methods in decision-making. The chapter identifies the linear relationship between the response and the predictor variables using GAMs, OLS and QRA models. Chapter 4 further explored the inclusion of monthly and quarterly temperature covariates in the peak electricity demand parameters through heating and cooling degree days. The variable selection in this chapter is done comparatively using ridge, Lasso, cross-validation and elastic net. Chapter 5 presents the application of peak electricity demand forecasting using PLAQR models. This chapter applies the proposed model to model and forecast the DPED using Lasso via hierarchical pairwise interactions. Chapter 6 discusses the comparative analysis of the AQR model with EM and NLQR models at extremely high and extremely low conditional quantiles of DPED. The AQR model shows the best results at extremely high conditional quantiles compared to other models. Chapter 7 gives the exploratory analysis of system reliability and applicability of the AQR model for determining DPED. The AQR model is considered in the models with interactions and without interactions. Several error measures for probability forecasting, including PL and CRPS, are used in this study to assess a forecast's predictive performance.

8.3 Modelling discussion and summary of key findings

This section aims to compare QR models' performances and summarises the key findings of the study:

- QRA model is compared with QR and GAM based on accuracy measures. The findings show that the proposed model produces the most accurate forecasts compared to QR and GAM. It performed very well in QPED data than in MPED data. The ordinary least square forecast (fOLS), fQR, fQRA and fGAMs for both MPED and QPED models based on RMSE, MAE and MAPE accuracy measures in Tables 4.2 and 4.3 are considered. The forecast plot in Figure 4.5 shows that fQRA fits very well in QPED. Hence, it is clear that QRA gives accurate results in QPED data for long-term peak electricity demand than in MPED.

- In the application and investigation of PLAQR models, fplLasso (M1) and fplLassoI (M2) with combined forecasts in Tables 5.1 and 7.1 are compared. The results show that the combined forecasts are more accurate than other models.

- In predicting extreme conditional quantiles of electricity demand, the AQR model is compared with EM and NLQR models as shown in Tables 6.2 and 6.3, respectively. The comparison is based on probabilistic accuracy measures such as CRPS, LogS, DSS, interval width and the pinball loss. The findings show that the AQR produces the most accurate predictions at high and low conditional quantile levels.

- In the section on empirical results, the AQR (M5) model is compared with fgamOSI (M6) and fgbmOS (M7). The findings show that the AQR model provides the best fitting and smallest value, as shown in Table 7.2.

8.4 Concluding remarks

Energy security has become an important issue in recent years. As a result, long-term peak electricity demand forecasting techniques are necessary to quantify the uncertainty of future electricity demand for better electricity security management in South Africa. The long-term electricity demand forecasting solution is discussed, and more importantly, the best forecasting model for future peak electricity demand is suggested. The question of modelling the extreme high and low quantiles of long-term peak electricity demand is important as there is a need to know the largest demand to supply adequate electricity to consumers in the country. The increased electricity demand has severely impacted the prices and living costs that make South African electricity more expensive. Moreover, modelling the upper tail distribution of peak electricity demand is significant as there are times when high load values exceed the maximum generated electricity, which could cause problems to the system, including grid instability. Considering the various methods proposed and key findings obtained, this study provides answers to the objectives stated in Chapter 1 (Section 1.3) as follows: First, Lasso proposed linear quantile regression (LQR) with and LQR without interactions for both MPED and QPED models. The forecasting accuracy of MPED and QPED models are compared using OLS, QR and QRA models. The three benchmark models, GAMs, SVB and SGB, are also considered for long-term peak electricity demand forecasting. To compare coefficient estimates for the ridge, Lasso and elastic net, the MPED and QPED data are utilised.

Secondly, the application of peak electricity demand forecasting using PLAQR models is presented with the inclusion of a nonlinear trend covariate, which is determined using a penalised cubic smoothing spline. The proposed model is applied to forecast the daily peak electricity demand using Lasso via hierarchical pairwise interactions.

Thirdly, the study conducted a comparative analysis of EM, AQR and NLQR models to predict extremely high and low daily peak electricity demand. The comparative analysis is done at different quantile levels. The NLQR model is nonlinear in the parameters and is formed by replacing the LQR model with a quantile curve. The three models are considered based on their strengths and weaknesses. However, the comparison is based on the scoring rules: CRPS, LogS, DSS, PL, and IW. The comparison for both extremely high and low quantiles addresses the uncertainties that might be ignored in practice, quantifies the uncertainties in the estimated distribution parameters, and predicts the high and low quantiles of daily peak electricity demand.

Fourthly, to look at the application of the AQR model, the study uses the temperature variables derived by splitting South African data into inland and coastal regions. The temperature variables include average daily temperature, average minimum, and average maximum temperature, among others. The study discusses the forecasting of the covariates needed to be used to forecast the response variable.

8.5 Limitations of the thesis

The summary of limitations that impacted the findings of the thesis is as follows:

- A limitation of this study is that the sample data is from 1997-2014. It would have been good to cover the period of the COVID-19 pandemic (that is, from March 2020 to date) in South Africa. In July 2020, South Africa was among the worst-affected countries globally in COVID-19 cases as the number of infections increased. However, in this study, the data covering the pandemic period was unavailable to the researchers. This would have helped see the impact of the pandemic on electricity demand. When data is made available, future research should include this period.

- Another limitation is that other DPED data sources of long-term electricity demand may produce different results contrary to the one we have used.

- This study used R package software to analyse the QR models; other packages available may produce different results. Likewise, other methods other than QR can also be applied and give different results.

- The AQR and NLQR models fitted well on comparing 99 percentiles to other models based on Murphy diagrams. However, the AQR model needs to give details about the size of the high level of possible exceedances, and parameters are hard to estimate. In the NLQR model, outliers only influence the quantile curve close to them (it affects extreme quantiles).

- Although the study presented superior empirical results with a comparative analysis of QR models with other traditional approaches, it does not give a clinical conclusion in comparing Murphy diagrams as the figures might lead to an inconclusive situation. Neither of the AQR, EM and NLQR forecast models dominates the other, and as a result, it would be unhelpful in decision making.

- Many pieces of literature have extensively dealt with EQR, which consists of one-stage and two-stage approaches. The limitation of this study on using QR in estimating extreme conditional quantiles is that a one-step extreme conditional quantile procedure based on QR underestimates these conditional quantiles. The two-step procedure was not useful for practical purposes and hence will not be considered for future purposes (Beirlant et al., 2004).

8.6 Future research studies

In light of the study findings, the possible future research directions are as follows:

- Introducing a three-step extreme conditional quantile procedure might be another way to overcome the limitations of one-step and two-step procedures. The procedure is useful for practical and reliable inference estimating extreme conditional quantiles.

- In this study, we applied extreme quantile prediction methods to South African electricity demand data. Future research areas should carry out similar studies using Bayesian semiparametric additive quantile regression for the impact of covariates on the complete conditional distribution of a response variable. Applying the Bayesian semiparametric additive quantile regression approach is a sensible alternative to the classical QR method and may help estimate underlying quantile regression functions.

- The future long-term electricity demand distribution uncertainties have to be considered when evaluating the accuracy of forecasting quantiles at extreme levels. Hence, the various time series forecasting techniques or other soft computing techniques such as genetic algorithms, fuzzy logic, ANN, quantile regression neural network (QRNN) and generalised lambda distribution (GLAD) model, among others, for prediction and uncertainty estimation related to extreme conditional quantile analyses can be proposed and lead to different results. Long-term electricity demand forecasting needs to be studied to avoid the shortage of electricity that prevents economic growth in the coming time. As a developing country, South Africa needs to know their electricity demand to avoid any challenges in the future. Therefore, sustainability will be controlled, and policies may be established. He and Zhang (2020) followed by Cannon (2011) proposed the probability density forecasting of wind power based on multi-core parallel quantile regression neural network (MPQRNN) and QRNN models, respectively. The forecasting techniques are useful in giving the flexibility of nonlinear and nonparametric models in the probability regression context of prediction.

- Other possible future research areas would be the use of sample data of DPED that cover the period of the COVID-19 pandemic in South Africa. Recently, the electricity demand in South Africa has shown a significant decline due to the introduction of measures adopted by the government to contain the COVID-19 pandemic. The impacts of the COVID-19 pandemic are currently not threatening the energy sector only in South Africa but also influencing all industries around the world (Buechler et al., 2020).

- Other future research directions consider the inclusion of more covariates such as meteorological, economic, and calendar variables to capture their effects on predicting extremely high and extremely low electricity demand.

References

M. Agyei-Sakyi, Y. Shao, O. Amos, and A. Marymargaret. Determinants of electricity consumption and volatility-driven innovative roadmaps to one hundred percent renewables for top consuming nations in africa. *Sustainability*, 13(11):6239, 2021.

H.M. Al-Hamadi and S.A. Soliman. Long-term/mid-term electric load forecasting based on short-term correlation and annual growth. *Electric power systems research*, 74(3):353–361, 2005.

T. Al-Saba and I. El-Amin. Artificial neural networks as applied to long-term demand forecasting. *Artificial Intelligence in Engineering*, 13(2):189–197, 1999.

H.M. Alhajeri, A. Almutairi, A. Alenezi, and F. Alshammari. Energy demand in the state of kuwait during the covid-19 pandemic: technical, economic, and environmental perspectives. *Energies*, 13(17):4370, 2020.

M. Alkhraijah, M. Alowaifeer, M. Alsaleh, A. Alfaris, and D.K. Molzahn. The effects of social distancing on electricity demand considering temperature dependency. *Energies*, 14(2):473, 2021.

D.E Allen, A.K. Singh, R.J. Powell, M. McAleer, J. Taylor, and L. Thomas. The volatility-return relationship: insights from linear and non-linear quantile regressions. *KIER Discussion Paper*, 831, 2012.

O.T. Altinoz and E. Mengusoglu. Cloud-based long term electricity demand forecasting using artificial neuro-fuzzy and neural networks. In *2015 9th International Conference on Electrical and Electronics Engineering (ELECO)*, pages 977–981. IEEE, 2015.

A.A. Bazmia, M. Davoodya, and G. Zahedia. Electricity demand estimation using an adaptive neuro-fuzzy network: a case study from the state of johor, malaysia. *International Journal*, 3(4), 2012.

C.N. Behrens, H.F. Lopes, and D. Gamerman. Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling*, 4(3):227–244, 2004.

J. Beirlant, T.D. Wet, and Y. Goegebeur. Nonparametric estimation of extreme conditional quantiles. *Journal of statistical computation and simulation*, 74(8):567–580, 2004.

T.L. Berning. *Improved estimation procedures for a positive extreme value index*. PhD thesis, Stellenbosch: University of Stellenbosch, 2010.

J. Bien, J. Taylor, and R. Tibshirani. A lasso for hierarchical interactions. *Annals of Statistics*, 41(3):1111–1141, 2013.

P.G. Bissiri, C.C. Holmes, and S.G. Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 78(5):1103, 2016.

L. Blázquez, N. Boogen, and M. Filippini. Residential electricity demand in spain: new empirical evidence using aggregate data. *Energy economics*, 36:648–657, 2013.

J.B. Bremnes. A comparison of a few statistical models for making quantile wind power forecasts. *Wind Energy*, 9(1-2):3–11, 2006.

E. Buechler, S. Powell, T. Sun, C. Zanocco, N. Astier, J. Bolorinos, J. Flora, H. Boudet, and R. Rajagopal. Power and the pandemic: exploring global changes in electricity demand during covid-19. *Available online: https://doi.org/10.48550/arXiv.2008.06988, Accessed on 17-11-2022*, 2020.

A.J. Cannon. Quantile regression neural networks: Implementation in r and application to precipitation downscaling. *Computers & geosciences*, 37(9): 1277–1284, 2011.

J. Carreau and Y. Bengio. A hybrid pareto model for asymmetric fat-tailed data: the univariate case. *Extremes*, 12(1):53–76, 2009.

M. Čepin. *Assessment of power system reliability: methods and applications*. Springer Science & Business Media, 2011.

V. Chernozhukov. Extremal quantile regression. *The Annals of Statistics*, 33 (2):806–839, 2005.

S. Coles. Classical extreme value theory and models. In *An introduction to statistical modeling of extreme values*, pages 45–73. Springer, 2001.

A. Daouia, L. Gardes, and S. Girard. On kernel smoothing for extremal quantile regression. *Bernoulli*, 19(5B):2557–2589, 2013.

C. Davino, M. Furno, and D. Vistocco. *Quantile regression: theory and applications*. John Wiley & Sons, 2013.

B.V. de Melo Mendes and H.F. Lopes. Data driven estimates for mixtures. *Computational statistics & data analysis*, 47(3):583–598, 2004.

M. Devaine, P. Gaillard, Y. Goude, and G. Stoltz. Forecasting electricity consumption by aggregating specialized experts. *Machine Learning*, 90(2): 231–260, 2013.

X. d'Haultfoeuille, A. Maurel, and Y. Zhang. Extremal quantile regressions for selection models and the black-white wage gap. Technical report, National Bureau of Economic Research, 2014.

T.A. Diriba, L.K. Debusho, and J. Botai. Modeling extreme daily temperature using generalized pareto distribution at port elizabeth, south africa. In *Annual Proceedings of the South African Statistical Association Conference*, volume 2015, pages 41–48. South African Statistical Association (SASA), 2015.

L.P.C. Do. Using quantile regression for modeling of electricity price and demand. Master's thesis, NTNU, 2015.

F. F. do Nascimento, D. Gamerman, and H.F. Lopes. A semiparametric bayesian approach to extreme value estimation. *Statistics and Computing*, 22(2):661–675, 2012.

G. Durrieu, I. Grama, Q.K. Pham, and J.M. Tricot. Nonparametric adaptive estimation of conditional probabilities of rare events and extreme quantiles. *Extremes*, 18(3):437–478, 2015.

G. Durrieu, I. Grama, K. Jaunatre, Q.K. Pham, and J.M. Tricot. extremefit: A package for extreme quantiles. *Journal of Statistical Software*, 87(1): 1–20, 2018.

N. Elamin. Quantile regression model for peak load demand forecasting with approximation by triangular distribution to avoid blackouts. 2018.

P. Embrechts, S.I. Resnick, and G. Samorodnitsky. Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3(2):30–41, 1999.

R.F Engle, C.W.J. Granger, J. Rice, and A. Weiss. Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American statistical Association*, 81(394):310–320, 1986.

S. Fan and R.J. Hyndman. Forecasting electricity demand in australian national electricity market. In *2012 IEEE Power and Energy Society General Meeting*, pages 1–4. IEEE, 2012.

M. Fasiolo, S.N Wood, M. Zaffran, R. Nedellec, and Y. Goude. Fast calibrated additive quantile regression. *Journal of the American Statistical Association*, pages 1–11, 2020a.

M. Fasiolo, S.N. Wood, M. Zaffran, R. Nedellec, and Y. Goude. qgam: Bayesian non-parametric quantile regression modelling in r. *arXiv preprint arXiv:2007.03303*, 2020b.

R.A. Fisher and L.H.C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society*, volume 24, pages 180–190. Cambridge University Press, 1928.

J. Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85, 2005.

J.H. Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.

A. Frigessi, O. Haug, and H. Rue. A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5(3):219–235, 2002.

F.U.P. Futures. Quantile forecasting with ensembles and combinations. *Business Forecasting: The Emerging Role of Artificial Intelligence and Machine Learning*, page 371, 2021.

P. Gaillard, Y. Goude, and R. Nedellec. Additive models and robust aggregation for gefcom2014 probabilistic electric load and electricity price forecasting. *International Journal of forecasting*, 32(3):1038–1050, 2016.

K. Gajowniczek and T. Zabkowski. Two-stage electricity demand modeling using machine learning algorithms. *Energies*, 10(10):1547, 2017.

L. Gardes and S. Girard. Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels. *Extremes*, 13(2):177–204, 2010.

D.H. Gebremeskel, E.O. Ahlgren, and G.B. Beyene. Long-term evolution of energy and electricity demand forecasting: The case of ethiopia. *Energy Strategy Reviews*, 36:100671, 2021.

C. Gibbons and A. Faruqui. Quantile regression for peak demand forecasting. *Available at SSRN 2485657*, 2014.

T. Gneiting and M. Katzfuss. Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1:125–151, 2014.

T. Gonen. *Electric power distribution engineering*. CRC press, 2015.

E.J. Gumbel. *Statistics of extremes*. Columbia university press, 1958.

N. Guo, W. Chen, M. Wang, Z. Tian, and H. Jin. Appling an improved method based on arima model to predict the short-term electricity consumption transmitted by the internet of things (iot). *Wireless Communications and Mobile Computing*, 2021, 2021.

L. Hao and D.Q. Naiman. *Quantile regression*. Number 149. Sage, 2007.

T. Hastie, R. Tibshirani, et al. Bayesian backfitting (with comments and a rejoinder by the authors. *Statistical Science*, 15(3):196–223, 2000.

T.J Hastie and R.J Tibshirani. Generalized additive models, volume 43 of monographs on statistics and applied probability. Chapman and Hall, 1990.

Trevor Hastie, Robert Tibshirani, Jerome Friedman, and James Franklin. Reviews-the elements of statistical learning: data mining, inference and prediction. *Mathematical Intelligencer*, 27(2):83–84, 2005.

Y. He and W. Zhang. Probability density forecasting of wind power based on multi-core parallel quantile regression neural network. *Knowledge-Based Systems*, 209:106431, 2020.

S. Hedden and S. Hedden. Gridlocked: A long-term look at south africa's electricity sector. *Institute for Security Studies Papers*, 2015(15):24, 2015.

T. Hong and S. Fan. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 32(3):914–938, 2016.

T. Hong, J. Wilson, and J. Xie. Long term probabilistic load forecasting and normalization with hourly information. *IEEE Transactions on Smart Grid*, 5(1):456–462, 2013.

C. Hor, S. Watson, D. Infield, and S. Majithia. Assessing load forecast uncertainty using extreme value theory. *16th PSCC, Glasgow*, 2008.

T. Hoshino. Quantile regression estimation of partially linear additive models. *Journal of Nonparametric Statistics*, 26(3):509–536, 2014.

Y. Hu. Extreme value mixture modelling with simulation study and applications in finance and insurance. 2013.

Y. Hu, K. Zhao, and H. Lian. Bayesian quantile regression for partially linear additive models. *Statistics and Computing*, 25(3):651–668, 2015.

Q. Huang, H. Zhang, J. Chen, and M.J.J.B.B. He. Quantile regression models and their applications: a review. *J Biom Biostat*, 8(10.4172):2155–6180, 2017.

R.J. Hyndman. Quantile forecasting with ensembles and combinations. *Available online:https://robjhyndman.com/publications/quantile-ensembles/, Accessed on 11-01-2021*, 2020.

R.J. Hyndman and S. Fan. Density forecasting for long-term peak electricity demand. *Power Systems, IEEE Transactions on*, 25(2):1142–1153, 2010.

R. Inglesi. Aggregate electricity demand in south africa: Conditional forecasts to 2030. *Applied energy*, 87(1):197–204, 2010.

R. Inglesi and A. Pouris. Forecasting electricity demand in south africa: A critique of eskom's projections. *South African Journal of Science*, 106(1): 50–53, 2010.

Y. Jang, E. Byon, E. Jahani, and K. Cetin. On the long-term density prediction of peak electricity load with demand side management in buildings. *Energy and Buildings*, 228:110450, 2020.

Y. Jiang. Double-penalized quantile regression in partially linear models. *Open Journal of Statistics*, 5(02):158–164, 2015.

M.S. Kandil, S.M. El-Debeiky, and N.E. Hasanien. Long-term forecasting for fast developing utility using a knowledge-based expert system. *IEEE Transaction on Power System*, 17(2):491–496, 2000.

M.S. Kandil, S.M. El-Debeiky, and N.E. Hasanien. Long-term load forecasting for fast developing utility using a knowledge-based expert system. *IEEE transactions on Power Systems*, 17(2):491–496, 2002.

N. Kaza. Understanding the spectrum of residential energy consumption: a quantile regression approach. *Energy policy*, 38(11):6574–6585, 2010.

J.D.J. Ketchen and D.D. Bergh. *Research methodology in strategy and management*. Emerald Group Publishing, 2006.

K.R. Kim, I.L. Dryden, H. Le, and K.E. Severn. Smoothing splines on riemannian manifolds, with applications to 3d shape space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83(1):108–132, 2021.

R. Koen and J. Holloway. Application of multiple regression analysis to forecasting south africa's electricity demand. *Journal of Energy in Southern Africa*, 25(4):48–58, 2014.

Renée Koen, Thandulwazi Magadla, and Paul Mokilane. Developing long-term scenario forecasts to support electricity generation investment decisions. In *43rd Annual Conference of the Operations Research Society of South Africa*, page 9, 2014.

R. Koenker. Confidence intervals for regression quantiles. In *Asymptotic statistics*, pages 349–359. Springer, 1994.

R. Koenker. *Quantile regression*. Number 38. Cambridge University Press, 2005.

R. Koenker and G. Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

R. Koenker and V. d'Orey. A remark on algorithm as 229: Computing dual regression quantiles and regression rank scores. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43(2):410–414, 1994.

R. Koenker and K. Hallock. Quantile regression: An introduction. *Journal of Economic Perspectives*, 15(4):43–56, 2001.

R. Koenker and B.J. Park. An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, 71(1-2):265–283, 1996.

R. Koenker, S. Portnoy, P.T. Ng, A. Zeileis, P. Grosjean, and B.D. Ripley. Package quantreg. *Cran R-project. org*, 2018.

R.W. Koenker and V. d'Orey. Algorithm as 229: Computing regression quantiles. *Applied statistics*, pages 383–393, 1987.

T. Küçükdeniz. Long term electricity demand forcesting: An alternative approach with support vector machines. *Istanbul university of engineering sciences*, 1:45–53, 2010.

C.S. Lai, G. Locatelli, A. Pimm, X. Wu, and L.L. Lai. A review on long-term electrical power system modeling with energy storage. *Journal of Cleaner Production*, 280:124298, 2021.

T. Lancaster and S. Jae Jun. Bayesian quantile regression methods. *Journal of Applied Econometrics*, 25(2):287–307, 2010.

J. Larmuth and A. Cuellar. An updated review of south african csp projects under the renewable energy independent power producer procurement programme (reipppp). In *AIP Conference Proceedings*, volume 2126, page 040001. AIP Publishing LLC, 2019.

K. Larsen. Gam: The predictive modeling silver bullet. *Multithreaded. Stitch Fix*, 30, 2016.

M.E. Lebotsa, C. Sigauke, A. Bere, R. Fildes, and J.E. Boylan. Short term electricity demand forecasting using partially linear additive quantile regression with an application to the unit commitment problem. *Applied Energy*, 222:104–118, 2018.

D. Lee, W.K. Li, and T.S.T. Wong. Modeling insurance claims via a mixture exponential model combined with peaks-over-threshold approach. *Insurance: Mathematics and Economics*, 51(3):538–550, 2012.

S. Lerch, T.L. Thorarinsdottir, . Ravazzolo, and T. Gneiting. Forecaster's dilemma: extreme events and forecast evaluation. *Statistical Science*, pages 106–127, 2017.

Y. Li and B. Jones. The use of extreme value theory for forecasting long-term substation maximum electricity demand. *IEEE Transactions on Power Systems*, 35(1):128–139, 2019.

H.Y. Liang, B.H. Wang, and Y. Shen. Quantile regression of partially linear single-index model with missing observations. *Statistics*, 55(1):1–17, 2021.

M. Lim and T. Hastie. Learning interactions via hierarchical group-lasso regularization. *Journal of Computational and Graphical Statistics*, 24(3): 627–654, 2015.

H. Lu, X. Ma, and M. Ma. A hybrid multi-objective optimizer-based model for daily electricity demand prediction considering covid-19. *Energy*, 219: 119568, 2021.

M. Ma and Z. Wang. Prediction of the energy consumption variation trend in south africa based on arima, ngm and ngm-arima models. *Energies*, 13 (1):10, 2020.

A. MacDonald, C.J. Scarrott, D. Lee, B. Darlow, M. Reale, and G. Russell. A flexible extreme value mixture model. *Computational Statistics & Data Analysis*, 55(6):2137–2157, 2011.

A. MacDonald, C.J. Scarrott, and D.S. Lee. Boundary correction, consistency and robustness of kernel densities using extreme value theory. *Unpublished manuscript*, 2013.

K. Maciejowska, J. Nowotarski, and R. Weron. Probabilistic forecasting of electricity spot prices using factor quantile regression averaging. *International Journal of Forecasting*, 32(3):957–965, 2016.

K. Makatjane. Bootstrapping uncertainty intervals for return period of extreme monthly electricity consumption in south africa.

S. Makridakis, E. Spiliotis, and V. Assimakopoulos. Predicting/hypothesizing the findings of the m4 competition. *International Journal of Forecasting*, 36(1):29–36, 2020.

S. Manikandan. Measures of central tendency: Median and mode. *Journal of pharmacology and pharmacotherapeutics*, 2(3):214, 2011.

D. Maposa. *Statistics of extremes with applications to extreme flood heights in the Lower Limpopo River Basin of Mozambique*. PhD thesis, 2016.

A. Marquard. *The origins and development of South African energy policy.* PhD thesis, University of Cape Town, 2006.

N. Maswanganyi, C. Sigauke, and E. Ranganai. Peak electricity demand forecasting using partially linear additive quantile regression models. In *Annual Proceedings of the South African Statistical Association Conference*, volume 2017, pages 25–32. South African Statistical Association (SASA), 2017.

N. Maswanganyi, E. Ranganai, and C. Sigauke. Long-term peak electricity demand forecasting in south africa: A quantile regression averaging approach. *AIMS Energy*, 7(6):857–882, 2019.

P.E. McSharry, S. Bouwman, and G. Bloemhof. Probabilistic forecasts of the magnitude and timing of peak electricity demand. *Power Systems, IEEE Transactions on*, 20(2):1166–1172, 2005.

C. Meng, J. Yu, Y. Chen, W. Zhong, and P. Ma. Smoothing splines approximation using hilbert curve basis selection. *arXiv preprint arXiv:2109.11727*, 2021.

A.A. Mir, M. Alghassab, K. Ullah, Z.A. Khan, Y. Lu, and M. Imran. A review of electricity demand forecasting in low and middle income countries: The demand determinants and horizons. *Sustainability*, 12(15):5931, 2020.

P. Mokilane. Density forecasting for long-term electricity demand in south africa using quantile regression. *South African Journal of Economic and Management Sciences*, 21(1):1–14, 2018.

P. Mokilane, P. Debba, V.S.S. Yadavalli, and C. Sigauke. Bayesian structural time-series approach to a long-term electricity demand forecasting. 2019.

P.M Mokilane, P. Debba, V.S.S. Yadavalli, and C. Sigauke. Long-term electricity demand forecasting using a generalised additive mixed quantile averaging (gammqv). 2018.

L. Morin, R. Brenner, K. Derrien, and K. Dorhmi. Periodic smoothing splines for fft-based solvers. *Computer Methods in Applied Mechanics and Engineering*, 373:113549, 2021.

P.P.N. Moroke. *Extreme Value Theory as a financial risk measure of the South African stock market*. PhD thesis, North-West University, 2019.

L. Moutinho and G.D. Hutcheson. *The SAGE dictionary of quantitative management research*. Sage, 2011.

M. Nanfuka, F. Berntsson, and G. Kakuba. Solving a cauchy problem for the heat equation using cubic smoothing splines. *Applicable Analysis*, pages 1–16, 2021.

N. Norouzi, G.Z. de Rubens, S. Choupanpiesheh, and P. Enevoldsen. When pandemics impact economies and climate change: exploring the impacts of covid-19 on oil and electricity demand in china. *Energy Research & Social Science*, 68:101654, 2020.

J. Nowotarski and R. Weron. Computing electricity spot price prediction intervals using quantile regression and forecast averaging. *Computational Statistics*, 30(3):791–803, 2015.

R.L. Parker and J.A. Rice. Discussion of some aspects of the spline smoothing approach to nonparametric curve fitting by bw silverman. *Journal of the Royal Statistical Society, Series B*, 47:40–42, 1985.

U. Perwez, A. Sohail, S. F. Hassan, and U. Zia. The long-term forecast of pakistan's electricity supply and demand: An application of long range energy alternatives planning. *Energy*, 93:2423–2435, 2015.

S. Portnoy, R. Koenker, et al. The gaussian hare and the laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statistical Science*, 12(4):279–300, 1997.

E. Ranganai. Quality of fit measurement in regression quantiles: an elemental set method approach. *Statistics & Probability Letters*, 111:18–25, 2016.

M. Rasuba, S. Khuluse, and C.D. Elphinstone. Forecasts for electricity demand in south africa (2010–2035) using the csir sectoral regression model. 2010.

M. Rasuba, S. Khuluse, and C.D. Elphinstone. Forecasts for electricity demand in south africa (2010–2035) using the csir sectoral regression model, 2017.

J.V. Ringwood, D. Bofelli, and F.T. Murray. Forecasting electricity demand on short, medium and long time scales using neural networks. *Journal of Intelligent and Robotic Systems*, 31(1-3):129–147, 2001.

A.I. Saba and A.H. Elsheikh. Forecasting the prevalence of covid-19 outbreak in egypt using nonlinear autoregressive artificial neural networks. *Process safety and environmental protection*, 141:1–8, 2020.

S. Shenoy and D. Gorinevsky. Stochastic optimization of power market forecast using non-parametric regression models. In *2015 IEEE Power & Energy Society General Meeting*, pages 1–5. IEEE, 2015.

B. Sherwood and L. Wang. Partially linear additive quantile regression in ultra-high dimension. *The Annals of Statistics*, 44(1):288–317, 2016.

C. Sigauke and D. Chikobvu. Short-term peak electricity demand in south africa. *African Journal of Business Management*, 6(32):9243, 2012.

C. Sigauke and M.M. Nemukula. Modelling extreme peak electricity demand during a heatwave period: A case study. *Energy Systems*, 11(1):139–161, 2020.

C. Sigauke, A. Verster, D. Chikobvu, et al. Tail quantile estimation of heteroskedastic intraday increases in peak electricity demand. *Open Journal of Statistics*, 2(4):435–442, 2012.

C. Sigauke, A. Verster, and D. Chikobvu. Extreme daily increases in peak electricity demand: Tail-quantile estimation. *Energy policy*, 53:90–96, 2013.

R.L. Smith. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, pages 367–377, 1989.

I.N. Soyiri, D.D. Reidpath, and C. Sarran. Forecasting peak asthma admissions in london: an application of quantile regression models. *International journal of biometeorology*, 57(4):569–578, 2013.

S.R. Stands. *Utility-scale renewable energy job creation: an investigation of the South African renewable energy independent power producer procurement programme (REIPPPP)*. PhD thesis, Stellenbosch: Stellenbosch University, 2015.

N. Szarka, M. Eichhorn, R. Kittler, A. Bezama, and D. Thrän. Interpreting long-term energy scenarios and the role of bioenergy in germany. *Renewable and Sustainable Energy Reviews*, 68:1222–1233, 2017.

S.B. Taieb, R. Huser, R.J Hyndman, and M.G Genton. Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. *IEEE Transactions on Smart Grid*, 7(5):2448–2455, 2016.

R Core Team. R: A language and environment for statistical computing (version 4.0. 5)[programming language]. *R Foundation for Statistical Computing*, 2021.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

V. Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

J.J. Velthoen. Non-parametric extreme quantile estimation for the common shaped tail model: Forecasting extreme precipitation by post-processing precipitation from a numerical weather prediction model. Master's thesis, , Delf, the Netherland, 2016.

A Verster, D De Waal, and S van der Merwe. Selecting an optimum threshold with the kullback-leibler deviance measure, 2013.

G. Wahba. *Spline bases, regularization, and generalized cross validation for solving approximation problems with large quantities of noisy data*. University of WISCONSIN, 1980.

H.J. Wang and D. Li. Estimation of extreme conditional quantiles through power transformation. *Journal of the American Statistical Association*, 108(503):1062–1074, 2013.

X. Wang, Y. Song, and S. Zhang. An efficient estimation for the parameter in additive partially linear models with missing covariates. *Journal of the Korean Statistical Society*, 49(3):779–801, 2020.

Ying-Ming Wang and Peng Jiang. Alternative mixed integer linear programming models for identifying the most efficient decision making unit in data envelopment analysis. *Computers & Industrial Engineering*, 62(2):546–553, 2012.

W. Wei. *Calibration tests to evaluate probabilistic forecasts and temporal modelling to monitor infectious diseases*. PhD thesis, University of Zurich, 2016.

S.N. Wood. *Generalized additive models: an introduction with R*. CRC press, 2017.

S.N. Wood and N.H. Augustin. Gams with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological modelling*, 157(2-3):157–177, 2002.

J.G. Wright, T. Bischof-Niemz, J.R. Calitz, C. Mushwana, and R. Van Heerden. Long-term electricity sector expansion planning: A unique opportunity for a least cost energy transition in south africa. *Renewable Energy Focus*, 30:21–45, 2019.

C. Wu. *Essays on High-dimensional Nonparametric Smoothing and Its Applications to Asset Pricing*. PhD thesis, University of Cincinnati, 2013.

G. Wu and W. Qiu. Threshold selection for pot framework in the extreme vehicle loads analysis based on multiple criteria. *Shock and Vibration*, 2018, 2018.

T. Xingyu. Boosting for partially linear additive models. 2016.

K. Yu, Z. Lu, and J. Stander. Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3):331–350, 2003.

X. Zhao, C. Scarrott, L. Oxley, and M. Reale. Extreme value modelling for forecasting market crisis impacts. *Applied Financial Economics*, 20(1-2): 63–72, 2010.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

H. Zou, T. Hastie, R. Tibshirani, et al. On the degrees of freedom of the lasso. *The Annals of Statistics*, 35(5):2173–2192, 2007.

Publications

Research articles (title pages and abstracts only) from this thesis are given in the next five pages.

AIMS Energy

http://www.aimspress.com/journal/energy

AIMS Energy, 7(6): 857–882.
DOI: 10.3934/energy.2019.6.857
Received: 25 June 2019
Accepted: 30 September 2019
Published: 06 December 2019

Research article

Long-term peak electricity demand forecasting in South Africa: A quantile regression averaging approach

Norman Maswanganyi[1,*], Edmore Ranganai[2] and Caston Sigauke[3]

[1] Department of Statistics and Operations Research, University of Limpopo, Private Bag X1106, Sovenga, 0727, South Africa

[2] Department of Statistics, University of South Africa, Private Bag X6, Florida 1710, South Africa

[3] Department of Statistics, University of Venda, Private Bag X5050, Thohoyandou, 0950, Limpopo, South Africa

* **Correspondence:** Email: nmaswanganyi72@gmail.com; Tel: +27152683680.

Abstract: Forecasting electricity demand in South Africa remains an increasingly national challenge as the government does not sufficiently take into account the impact of the electricity prices in their electricity demand forecast. Effective measures to rapidly reduce the demand of electricity are urgently needed to deal with future electricity prices and government policies uncertainties within the energy industry. Moreover, long-term peak electricity demand forecasting methods are needed to quantify the uncertainty of future electricity demand for better electricity security management. The prediction of long-term electricity demand assists decision makers in the electricity sector in planning for capacity generation. This paper presents an application of quantile regression averaging (QRA) approach using South African monthly and quarterly data ranging from January 2007 to December 2014. Variable selection is done in a comparative manner using ridge, least absolute shrinkage and selection operator (Lasso), cross validation (CV) and elastic net. We compare the forecasting accuracy of monthly peak electricity demand (MPED) and quarterly peak electricity demand (QPED) forecasting models using generalised additive models (GAMs) and QRA. The coefficient estimates for ridge, Lasso and elastic net are estimated and compared using MPED and QPED data.

Keywords: additive quantile regression; cubic smoothing splines; long term peak demand forecasting; penalised regression methods; quantile regression averaging model; temperature

PEAK ELECTRICITY DEMAND FORECASTING USING PARTIALLY LINEAR ADDITIVE QUANTILE REGRESSION MODELS

Norman Maswanganyi [1]
University of Limpopo
e-mail: *nmaswanganyi72@gmail.com*
Caston Sigauke
University of Venda
Edmore Ranganai
University of South Africa

Key words: Additive models, Forecast combination, Lasso, Peak demand forecasting, Quantile regression.

Abstract: The paper presents an application of partially linear additive quantile regression models in forecasting peak electricity demand using South African data from January 2007 to December 2013. Variable selection is done using the least absolute shrinkage and selection operator (Lasso) via hierarchical pairwise interactions in which the main effects (lower order variables) must be in the model if higher order interactions are included. One of the contributions of this paper is the inclusion of a nonlinear trend as one of the covariates which is determined using a penalized cubic smoothing spline.

1. Introduction

Planning for capacity expansion including medium term risk assessment in the electricity sector requires accurate predictions of peak electricity demand. Although many studies have been done on short, medium or long-term peak electricity demand forecasting, application and investigation of partially linear additive quantile regression (PLAQR) models in forecasting peak electricity demand has not been carried out extensively in the literature. Quantile regression (QR), an extension of univariate quantile estimation, was introduced by Koenker and Bassett Jr. (1978) as a mechanism for estimating models for the conditional median and the full range of other conditional quantiles, thereby giving a more complete statistical analysis of relationships among random variables than the conditional mean functions. For an authoritative overview of the quantile regression methodology, refer to the monograph by Koenker (2005). Generalised additive models (GAMs) on the other hand were developed by Hastie and Tibshirani (1990). Partially linear additive quantile regression (PLAQR) models are a combination of generalised additive models (GAMs) developed by Hastie and Tibshirani (1990) and QR models where the conditional quantile function comprises a linear

[1] Corresponding author

energies

MDPI

Article

Prediction of Extreme Conditional Quantiles of Electricity Demand: An Application Using South African Data

Norman Maswanganyi [1,†], Caston Sigauke [2,*,†] and Edmore Ranganai [3,†]

[1] Department of Statistics and Operations Research, University of Limpopo Private Bag X1106, Sovenga 0727, South Africa; nmaswanganyi72@gmail.com
[2] Department of Statistics, University of Venda Private Bag X5050, Thohoyandou 0950, South Africa
[3] Department of Statistics, University of South Africa, Private Bag X6, Florida 1710, South Africa; rangae@unisa.ac.za
* Correspondence: caston.sigauke@univen.ac.za; Tel.: +27-15-962-8135
† These authors contributed equally to this work.

Abstract: It is important to predict extreme electricity demand in power utilities as the uncertainties in the future of electricity demand distribution have to be taken into consideration to achieve the desired goals. The study focused on the prediction of extremely high conditional quantiles (between 0.95 and 0.9999) and extremely low quantiles (between 0.001 and 0.05) of electricity demand using South African data. The paper discusses a comparative analysis of the additive quantile regression model with an extremal mixture model and a nonlinear quantile regression model. The estimated quantiles at each level were then combined using the median approach. The comparisons were carried out using daily peak electricity demand data ranging from January 1997 to May 2014. Proper scoring rules were used to compare the three models, and the model with the smallest score was preferred. The results could be useful to system operators including decision-makers in power utility companies by giving insights and guidance for future electricity demand patterns. The prediction of extremely high quantiles of daily peak electricity demand could help system operators know the possible largest demand that will enable them to supply adequate electricity to consumers and shift demand to off-peak periods. The prediction of extreme conditional quantiles of daily peak electricity demand in the context of South Africa using additive quantile regression, nonlinear quantile regression, and extremal mixture models has not been performed previously to the best of our knowledge.

Keywords: additive quantile regression; extremal mixture model; extreme conditional quantiles; nonlinear quantile regression; scoring rules

Citation: Maswanganyi, N.; Sigauke, C.; Ranganai, E. Prediction of Extreme Conditional Quantiles of Electricity Demand: An Application Using South African Data. *Energies* **2021**, *1*, 0. https://doi.org/

Academic Editor: Surender Reddy Salkuti, Zita Vale and Tiago Pinto
Received: 12 May 2021
Accepted: 04 August 2021
Published:

1. Introduction

In an attempt to diversify the energy mix, the South African government has developed a Renewable Energy Independent Power Producer Procurement Program (REIPPPP). The REIPPPP has proven to be a very successful program especially in bringing renewable energy projects to commercial operations over the past five years [1]. The country also came up with a new plan, namely the National Development Plan (NDP). The NDP is a plan for infrastructure development from 2013 to 2030. Moreover, the NDP is also an excellent guiding economic plan that set the GDP growth target per annum for the country to be able to meet its economic, social and political objectives [2].

In identifying the national goals relevant to establishing renewable energy policy objectives, South Africa needs to identify the key goals for the nation and how the electricity sector fits in among its priorities. The country also needs to take urgent actions to ensure the sustainability of renewable energy and energy efficiency by 2030. According to [3], renewable and energy efficiency have a positive impact on electricity demand during peak hours.

10

Reliable Predictions of Peak Electricity Demand and Reliability of Power System Management

Caston Sigauke
University of Venda

Santosh Kumar
University of Melbourne

Norman Maswanganyi
University of Limpopo

Edmore Ranganai
University of South Africa

CONTENTS

10.1 Introduction

Electricity demand forecasting is very important to system operators who plan for hour-to-hour electricity production to meet demand, at the same time ensuring grid stability, and to strategic managers who plan for medium-term risk assessment and capacity expansion. The chapter presents a point process characterization of peak electricity demand. An analysis of power systems reliability in the presence of random peak loads is also discussed in the chapter. Modeling the upper tail distribution of peak electricity demand is important as there are times when high load values exceed the maximum generated electricity, which could cause problems to the system, including grid instability. A severe stress on the grid system occurs when the peak electricity demands' amplitude exceeds a predetermined sufficiently high threshold that is determined using extreme value theory methods (Chidodo and Lauria, 2012). This high load tests the reliability of the electrical power systems components.

Power system reliability is defined as "the probability that an electric power system can perform a required function under given conditions for a given time interval" (Čepin, 2011). Some of the important measures in reliability modeling include reliability indexes such as the loss of load expectation (LOLE), which is derived using daily peak load (Čepin, 2011), including among others the extreme peak load frequency (EPLF), which is the average number of peak loads above a sufficiently high threshold (Chidodo and Lauria, 2012). Chidodo and Lauria (2012) discuss the probabilistic characterization of peak loads. It is assumed that peak load follows a Poisson process. The authors argue that it is essential for system operators to have information about the occurrences of extreme peak loads as these form part of the power components reliability modeling.

Prediction of probabilistic forecasts of future peak electricity demand is discussed in McSharry et al. (2005). The authors argue that the risk management and assessment of uncertainty in the forecasts can be improved. Several methods are discussed in the literature on short-term to long-term electricity demand forecasting. For recent reviews on these techniques, see Hong and Fan (2016) and Hong et al. (2016), among others. Methodologies

Appendix

R codes

```
attach(analyticdata)
head(analyticdata)
tail(analyticdata)
win.graph()
#library(glmnet)#package for shrinkage methods
library(forecast)
library(glmnet)
library(lars)
library(forecast)
library(glinternet)
library(earth)

# EARTH determining reference temperatures
a <- earth(DPED˜AmaxTCI, minspan=-1)
plotmo(a)
summary(a, digits = 1, style = ”pmax”)#minspan=912,

###########################################################################
#attach(data)
#head(data)
#tail(data)

win.graph()

x=ts(DPED, start=2000, freq=365)
plot(x, type=”p”, ylab=”Daily peak electricity demand (MW)”,
col=”blue”, xlab=”Year”)
(smooth.spline(time(x), x))
lines(smooth.spline(time(x), x, spar=0.1136387), col=”red”, lwd=3)
dpdfits = fitted((smooth.spline(time(x), x, spar=0.1136387)))

write.table(dpdfits, ”˜/fittedvalues.txt”, sep=”\t”)
```

```
##################################################################

#CROSS VALIDATION MODEL SELECTION
attach(dataCV)
head(dataCV)
library(forecast)
y=ts(DPED, start=2000, freq=365)
regress = tslm(y~CDDAADTCI+CDDAmaxTCI+CDDAminTCI+DAH+
Daytype+DBH+DH+DAH+HDDAADTCI+HDDAmaxTCI+
HDDAminTCI+noltrend)
summary(regress)
CV(regress)

################################################################

# FORECASTING

attach(dataDPED)
head(dataDPED)
tail(dataDPED)
win.graph()
plot(DPED, xlab="Days", ylab="DPED")
n <- length(DPED)
n

attach(data)
head(data)
tail(data)

# Save the length of the information regarding
forecasting horizon
h <- nrow(data)-n
h
# Fit an additive regression model OLS
fit1 <- lm(DPED~CDDAADTCI+CDDAmaxTCI+CDDAminTCI
+DAH+Daytype+DBH+DH+HDDAADTCI+HDDAmaxTCI+
HDDAminTCI+noltrend, data[1:n,])# insample
# Return the summary of the model
```

```
summary(fit1)
# Add a new line (the fit of the model) on the
 existing graph
lines(fit1$fit , col="blue")

# Calculate the out-of-sample forecasts , based
on the available information
#on temperature
fcs <- predict(fit1 , data[(n+1):(n+731),])
fcs
plot(fcs)
write.table(fcs,"~/forecastsOLS.txt",sep="\t")

#lines(fcs , col=" red")

############
##QUANTREG
###########

library(quantreg)
fit2 <- rq(DPED~CDDAADTCI+CDDAmaxTCI+CDDAminTCI+
DAH+Daytype+DBH+DH+
HDDAADTCI+HDDAmaxTCI+HDDAminTCI+noltrend ,tau=0.5,
data[1:n,])# insample
summary(fit2)
lines(fit2$fit , col="blue")

fcs2 <- predict(fit2 , data[(n+1):(n+731),])
fcs2
plot(fcs2)
write.table(fcs2,"~/forecastsQR.txt",sep="\t")


library(forecast)
attach(forecasts)
head(forecasts)
win.graph()
accuracy(fols ,dped)
```

```
accuracy(fqr,dped)

y=ts(dped)
plot(y,ylab="Electricity demand(MW)",col="blue",
xlab="Observation number")
lines(fqr,col="red")
legend("topleft",lty=1,col=c("black","red"),
       legend=c("Actuals","Forecasts"))

#####################################################
attach(analytimdata)
head(analytimdata)
tail(analytimdata)
win.graph()
#library(glmnet)#package for shrinkage methods
library(forecast)
library(glmnet)
library(lars)
library(forecast)
library(glinternet)
library(earth)

# EARTH determining reference temperatures
a <- earth(MPED~AmaxTCI,minspan=-1)
plotmo(a)
summary(a, digits = 1,style = "pmax")#minspan=171,

######################################################
#attach(temdata)
#head(temdata)
#tail(temdata)

win.graph()

x=ts(MPED, start=2000, freq=12)
plot(x,ylab="Monthly average temperature (degree C)",
col="blue",xlab="Year")
(smooth.spline(time(x), x))
```

```
lines(smooth.spline(time(x), x, spar=0.1136387),
col="red",lwd=3)
mpdfits = fitted((smooth.spline(time(x), x,
 spar=0.1136387)))

write.table(mpdfits,"~/fittedvalues.txt",sep="\t")


##########################################################

#CROSS VALIDATION MODEL SELECTION
attach(temcvdata)
head(temcvdata)
library(forecast)
y=ts(MPED,start=2000,freq=12)
regress = tslm(x~CDDAADTCI+CDDAmaxTCI+CDDAminTCI+
Daytype+DBH+DH+DAH+
HDDAADTCI+HDDAmaxTCI+HDDAminTCI+noltrend)
summary(regress)
CV(regress)

###########################################################

# FORECASTING

attach(MPEDdata)
head(MPEDdata)
tail(MPEDdata)
win.graph()
x=ts(MPED, start=2000, freq=12)
plot(x,ylab="Monthly peak electricity demand (MW)",
col="blue",xlab="Year")
n <- length(MPED)
n

attach(temdata)
head(temdata)
tail(temdata)
```

```
# Save the length of the information regarding
 forecasting horizon
h <- nrow(temdata)-n
h
# Fit an additive regression model OLS
fit1 <- lm(MPED~CDDAADTCI+CDDAmaxTCI+CDDAminTCI+
DAH+Daytype+DBH+DH+
 HDDAADTCI+HDDAmaxTCI+HDDAminTCI+noltrend,
 temdata[1:n,])# insample
# Return the summary of the model
summary(fit1)
# Add a new line (the fit of the model) on
 the existing graph
lines(fit1$fit, col="blue")

# Calculate the out-of-sample forecasts, based
 on the available information
#on temperature
fcs <- predict(fit1, temdata[(n+1):(n+27),])
fcs
plot(fcs)
write.table(fcs,"~/forecastsOLS.txt",sep="\t")

#lines(fcs, col=" red")

###########
##QUANTREG
##########

library(quantreg)
fit2 <- rq(MPED~CDDAADTCI+CDDAmaxTCI+CDDAminTCI+
DAH+Daytype+DBH+DH+
HDDAADTCI+HDDAmaxTCI+HDDAminTCI+noltrend,tau=0.5,
temdata[1:n,])# insample
summary(fit2)
lines(fit2$fit, col="blue")
```

```
fcs2 <- predict(fit2, temdata[(n+1):(n+27),])
fcs2
plot(fcs2)
write.table(fcs2,"~/forecastsQR.txt",sep="\t")


library(Mforecast)
attach(Mforecast)
head(Mforecast)
win.graph()
accuracy(fols,ped)
accuracy(fqr,ped)

y=ts(ped)
plot(y, ylab="Monthly electricity demand (MW)",
col="blue",xlab="Observation number")
lines(fqr,col="red")
legend("topleft",lty=1,col=c("black","red"),
       legend=c("Actuals","Forecasts"))

#plot of residuals against fitted values
win.graph()
plot(fcs(analytimdata),residuals(analytimdata),
xlab="fitted values",ylab="residuals")
###############################################
attach(anlyteqdata)
head(anlyteqdata)
tail(anlyteqdata)
win.graph()
#library(glmnet)#package for shrinkage methods
library(forecast)
library(glmnet)
library(lars)
library(forecast)
library(glinternet)
library(earth)

# EARTH determining reference temperatures
```

```
a <- earth(Qpedd~DmaxTCI,minspan=-1)
plotmo(a)
summary(a, digits = 1,style = "pmax")#minspan=912,

#####################################################
#attach(tqdata)
#head(tqdata)
#tail(tqdata)

win.graph()
x<-ts(Qpedd[2], start=2000, freq=4)
plot(x, ylab="Quarterly peak electricity demand (MW)",
col="blue",xlab="Observations")
(smooth.spline(time(x), x))
lines(smooth.spline(time(x), x, spar=0.1136387),
col="red",lwd=3)
Qedfits = fitted((smooth.spline(time(x), x,
spar=0.1136387)))

write.table(Qedfits,"~/fittedvalues.txt",sep="\t")


#####################################################

#CROSS VALIDATION MODEL SELECTION OF QPED####
attach(tqcvdata)
head(tqcvdata)
library(forecast)
y=ts(Qpedd,start=2000,freq=4)
regress = tslm(x~CDDAADTCI+CDDAmaxTCI+CDDAminTCI+DAH+
Daytype+DBH+DH+DAH+
HDDAADTCI+HDDAmaxTCI+HDDAminTCI+noltrend)
plot(regress)
summary(regress)
CV(regress)

#########################################################
```

FORECASTING

```
attach(Qpeddata)
head(Qpeddata)
tail(Qpeddata)
win.graph()
x=ts(Qpedd[2], start=2000, freq=4)
plot(x,ylab="Quarterly peak electricity demand (MW)",
col="blue",
      xlab="Days")
n <- length(Qpd)
n

attach(tqdata)
head(tqdata)
tail(tqdata)

# Save the length of the information regarding
forecasting horizon
h <- nrow(tqdata)-n
h
# Fit an additive regression model OLS
fit1 <- lm(Qpd~CDDAADTCI+CDDAmaxTCI+CDDAminTCI+DAH+
Daytype+DBH+DH+
 HDDAADTCI+HDDAmaxTCI+HDDAminTCI+noltrend,
 data=tqdata[1:n,])# insample
plot(fit1)
# Return the summary of the model
summary(fit1)
# Add a new line (the fit of the model) on the
existing graph
lines(fit1$fit, col="blue")

# Calculate the out-of-sample forecasts, based on
 the available information
#on temperature
fcs <- predict(fit1, tqdata[(n+1):(n+13),])
fcs
```

```
plot(fcs)
write.table(fcs,"~/forecastsOLS.txt",sep="\t")

#lines(fcs, col=" red")

############
##QUANTREG
###########

library(quantreg)
fit2 <- rq(Qpd~CDDAADTCI+CDDAmaxTCI+CDDAminTCI+
DAH+Daytype+DBH+DH+
HDDAADTCI+HDDAmaxTCI+HDDAminTCI+noltrend,tau=0.5,
data=tqdata[1:n,])# insample
summary(fit2)
lines(fit2$fit, col="blue")

fcs2 <- predict(fit2, tqdata[(n+1):(n+13),])
fcs2
plot(fcs2)
write.table(fcs2,"~/forecastsQR.txt",sep="\t")


library(forecast)
attach(Qdforecast)
head(Qdforecast)
win.graph()
accuracy(fols,Qed)
accuracy(fqr,Qed)

y=ts(Qed)
plot(y,ylab=" Quarterly electricity demand(MW)",
col="blue",xlab="Observation number")
lines(fqr,col="red")
legend("topleft",lty=1,col=c("black","red"),
       legend=c("Actuals","Forecasts"))
##################################################
#CROSS VALIDATION MODEL SELECTION OF MPED
```

```
library(glmnet)
attach(mcvdata)
head(mcvdata)
tail(mcvdata)
names(mcvdata)
dim(mcvdata)

mcvdata.pred.names = names(mcvdata)[1 : length
(names(mcvdata)) -1]
mcvdata.pred = mcvdata[, mcvdata.pred.names]
set.seed(1)
train=sample(1:nrow(mcvdata.pred),
 nrow(mcvdata.pred)/2)
test=(-train)

x=model.matrix(~., mcvdata.pred[train, ])[, -1]
y = mcvdata$Mped[train]
x.test = model.matrix(~., mcvdata.pred[test, ])
[, -1]
y.test = mcvdata$Mped[test]

lm.mod = lm(y~., data = data.frame(x))
summary(lm.mod)$r.squared
library(car)
vif(lm.mod)


lm.mod1 = lm(Mpd~CDDAADTCI+CDDAmaxTCI+CDDAminTCI+
Daytype+DBH+DH+DAH+
HDDAADTCI+HDDAminTCI+noltrend,data=mdata[1:n,])
summary(lm.mod1)
##############################
lambda.grid = 10^seq(5, -2, length=100)
ridge.mod1 = glmnet(x, y, alpha=0, lambda=lambda.grid)
plot(ridge.mod1, xvar="lambda")

lasso.mod1 = glmnet(x, y, alpha=1, lambda=lambda.grid)
plot(lasso.mod1, xvar="lambda")
```

```
elNet.mod1 = glmnet(x, y, alpha=0.5,
lambda=lambda.grid)
plot(elNet.mod1, xvar="lambda")


##################MPED CV estimate plot######
par(mfrow = c(1,3))
plot(ridge.mod,xvar="lambda",label=TRUE)
plot(lasso.mod,xvar="lambda",label=TRUE)
plot(elNet.mod,xvar="lambda",label=TRUE)
####################################################

set.seed(1)
ridge.cv.out = cv.glmnet(x, y, alpha=0)
set.seed(1)
lasso.cv.out = cv.glmnet(x, y, alpha=1)
set.seed(1)
elNet.cv.out = cv.glmnet(x, y, alpha=0.5)

win.graph()
par(mfrow = c(1,3))
plot(ridge.cv.out)
plot(lasso.cv.out)
plot(elNet.cv.out)

ridge.bestlam = ridge.cv.out$lambda.min
ridge.lam1se = ridge.cv.out$lambda.1se
lasso.bestlam = lasso.cv.out$lambda.min
lasso.lam1se = lasso.cv.out$lambda.1se
elNet.bestlam = elNet.cv.out$lambda.min
elNet.lam1se = elNet.cv.out$lambda.1se

ridge.mod.best = glmnet(x, y, alpha=0,
 lambda=ridge.bestlam)
coef(ridge.mod.best)
ridge.mod.1se = glmnet(x, y, alpha=0,
lambda=ridge.lam1se)
```

```
coef(ridge.mod.1se)

lasso.mod.best = glmnet(x, y, alpha=1,
lambda=lasso.bestlam)
coef(lasso.mod.best)
lasso.mod.1se = glmnet(x, y, alpha=1,
lambda=lasso.lam1se)
coef(lasso.mod.1se)

elNet.mod.best = glmnet(x, y, alpha=1,
 lambda=elNet.bestlam)
coef(elNet.mod.best)
elNet.mod.1se = glmnet(x, y, alpha=1,
lambda=elNet.lam1se)
coef(elNet.mod.1se)

y.lm.train = predict(lm.mod, newdata =
data.frame(x))
sum((y.lm.train - y)^2)
y.lm.test = predict(lm.mod, newdata =
data.frame(x.test))
sum((y.lm.test - y.test)^2)

y.ridge.best.train = predict(ridge.mod.best,
 newx=x)
sum((y.ridge.best.train - y)^2)
y.ridge.best.test = predict(ridge.mod.best,
 newx=x.test)
sum((y.ridge.best.test - y.test)^2)

y.ridge.1se.train = predict(ridge.mod.1se,
 newx=x)
sum((y.ridge.1se.train - y)^2)
y.ridge.1se.test = predict(ridge.mod.1se,
 newx=x.test)
sum((y.ridge.1se.test - y.test)^2)

y.lasso.best.train = predict(lasso.mod.best,
```

```
newx=x)
sum((y.lasso.best.train - y)^2)
y.lasso.best.test = predict(lasso.mod.best,
newx=x.test)
sum((y.lasso.best.test - y.test)^2)

y.lasso.1se.train = predict(lasso.mod.1se,
newx=x)
sum((y.lasso.1se.train - y)^2)
y.lasso.1se.test = predict(lasso.mod.1se,
newx=x.test)
sum((y.lasso.1se.test - y.test)^2)

y.elNet.best.train = predict(elNet.mod.best,
newx=x)
sum((y.elNet.best.train - y)^2)
y.elNet.best.test = predict(elNet.mod.best,
newx=x.test)
sum((y.elNet.best.test - y.test)^2)

y.elNet.1se.train = predict(elNet.mod.1se,
newx=x)
sum((y.elNet.1se.train - y)^2)
y.elNet.1se.test = predict(elNet.mod.1se,
newx=x.test)
sum((y.elNet.1se.test - y.test)^2)
##################################################
attach(mcvdata)
head(mcvdata)
tail(mcvdata)
library(glmnet)
x=model.matrix(Mped~.-1,data=mcvdata)
y=mcvdata$Mped

modd.ridge=glmnet(x,y,alpha=0)
plot(modd.ridge,xvar="lambda",label=TRUE)
modd.lasso=glmnet(x,y,alpha=1)
plot(modd.lasso,xvar="lambda",label=TRUE)
```

```
plot(modd.lasso,xvar="dev",label=TRUE)
modd.elNet=glmnet(x,y,alpha=0.5)
plot(modd.elNet,xvar="lambda",label=TRUE)
########Coefficient estimate plots for MPED####
win.graph()
par(mfrow = c(1,3))
plot(modd.ridge,xvar="lambda",label=TRUE)
plot(modd.lasso,xvar="lambda",label=TRUE)
plot(modd.elNet,xvar="lambda",label=TRUE)

cv.ridge=cv.glmnet(x,y,alpha=0)
plot(cv.ridge)
cv.lasso=cv.glmnet(x,y,alpha=1)
plot(cv.lasso)
coef(cv.lasso)
cv.elnet=cv.glmnet(x,y,alpha=1)
plot(cv.elnet)

win.graph()
par(mfrow = c(1,3))
plot(cv.ridge)
plot(cv.lasso)
plot(cv.elnet)

lasso.tr=glmnet(x[train,],y[train])
lasso.tr
pred=predict(lasso.tr,x[-train,])
dim(pred)
rmse= sqrt(apply((y[-train]-pred)^2,2,mean))
plot(log(lasso.tr$lambda),rmse,type="b",
xlab="Log(lambda)")
lam.best=lasso.tr$lambda[order(rmse)[1]]
lam.best
coef(lasso.tr,s=lam.best)
##################################
attach(tqcvdata)
head(tqcvdata)
tail(tqcvdata)
```

```
library(glmnet)
x=model.matrix(Qpedd~.-1,data=tqcvdata)
y=tqcvdata$Qpedd

modd1.ridge=glmnet(x,y,alpha=0)
plot(modd1.ridge,xvar="lambda",label=TRUE)
modd1.lasso=glmnet(x,y,alpha=1)
plot(modd1.lasso,xvar="lambda",label=TRUE)
plot(modd1.lasso,xvar="dev",label=TRUE)
modd1.elNet=glmnet(x,y,alpha=0.5)
plot(modd1.elNet,xvar="lambda",label=TRUE)

win.graph()
par(mfrow = c(1,3))
plot(modd1.ridge,xvar="lambda",label=TRUE)
plot(modd1.lasso,xvar="lambda",label=TRUE)
plot(modd1.elNet,xvar="lambda",label=TRUE)

cv.ridge=cv.glmnet(x,y,alpha=0)
plot(cv.ridge)
cv.lasso=cv.glmnet(x,y,alpha=1)
plot(cv.lasso)
coef(cv.lasso)
cv.elnet=cv.glmnet(x,y,alpha=1)
plot(cv.elnet)

win.graph()
par(mfrow = c(1,3))
plot(cv.ridge)
plot(cv.lasso)
plot(cv.elnet)
########################################
##QUANTILE REGRESSION##
########################

attach(anlytemdat)
head(anlytemdat)
tail(anlytemdat)
```

```
win.graph()
#library(glmnet) # package for shrinkage methods
library(forecast)
#library(glinternet)
library(earth)
library(mgcv)
library(itsadug)

# EARTH determining reference temperatures
a<- earth(Mped~AmaxTCI, minspan=-1)
plotmo(a)
summary(a, digits = 1, style = "pmax")#minspan=912,

##########################################################
Monthly average temp plot
##########################################################
#attach(mdata)
#head(mdata)
#tail(mdata)

win.graph()

x=ts(Mped[2], start=2000, freq=12)
plot(x, type="p", ylab="Monthly peak electricity
demand (MW)",col="blue",
      xlab="Year")
(smooth.spline(time(x), x))
lines(smooth.spline(time(x), x, spar=0.1136387),
col="red",lwd=3)
mpedfits = fitted((smooth.spline(time(x), x,
spar=0.1136387)))

write.table(mpedfits,"~/fittedvalues.txt",sep="\t")


###########################################################
###########################################################
```

```
#CROSS VALIDATION MODEL SELECTION
attach(mcvdat)
head(mcvdat)
library(forecast)
y=ts(Mped, start=2000, freq=12)
regress = tslm(x~CDDAAMTCI+CDDAmaxTCI+CDDAminTCI+
DAH+Daytype+DBH+DH+
                 HDDAAMTCI+HDDAminTCI+noltrend)
summary(regress)
CV(regress)


##################################################################
##################################################################

# FORECASTING

attach(Mddat)
head(Mddat)
tail(Mddat)
win.graph()
plot(Mpd, xlab="Observations", ylab="MPED")
n <- length(Mpd)
n

attach(mdat)
head(mdat)
tail(mdat)

# Save the length of the information regarding
forecasting horizon
h <- nrow(mdat)-n
h
# Fit an additive regression model OLS
fit1 <- lm(Mpd~CDDAAMTCI+CDDAmaxTCI+CDDAminTCI+
DAH+Daytype+DBH+DH+
             HDDAAMTCI+HDDAminTCI+noltrend,
             data=mdat[1:n,]) # insample
```

```
# Return the summary of the model
summary(fit1)
# Add a new line (the fit of the model) on the
 existing graph
 lines(fit1$fit , col="blue")


# Calculate the out-of-sample forecasts , based
 on the available information
#on temperature
 fcs <- predict(fit1 , mdat[(n+1):(n+63),])
 fcs
 plot(fcs)
 write.table(fcs ,"~/forecastsOLS.txt",sep="\t")

#lines(fcs , col=" red")

############
##QUANTREG
###########

 library(quantreg)
 fit2<- rq(formula=Mpd~CDDAAMTCI+CDDAmaxTCI+
CDDAminTCI+DAH+Daytype+DBH+DH+
                HDDAAMTCI+HDDAminTCI+noltrend ,
                 tau = 0.5,data=mdat[1:n,])# insample
 fit2<- rq(Mpd~CDDAminTCI+CDDAminTCI*Daytype ,
 tau = 0.5,data=mdat[1:n,])# insample
summary(fit2)
 lines(fit2$fit , col="blue")

 fcs2 <- predict(fit2 , mdat[(n+1):(n+63),])
 fcs2
 plot(fcs2)
 write.table(fcs2 ,"~/forecastsQR.txt",sep="\t")
```

```
library(forecast)
attach(M2forecast)
head(M2forecast)
win.graph()
accuracy(fols,Med)
accuracy(fqr,Med)

y=ts(Med)
plot(y,xlab="year", ylab="Electricity demand (MW)",
ylim=c(30000,38000))
lines(fqr,lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs2, col="blue")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
######################
# Fit an additive regression model GAM
library (lattice)#For multi panel graphs
library(mgcv)#For GAM
library(gstat)#For checking spatial patterns
library(gamlss)#For GAM

fit3 <- gam(Mpd~CDDAminTCI+CDDAAMTCI+CDDAmaxTCI+
DH+s(Daytype,bs="re")+HDDAminTCI+
s(noltrend),method="REML",data=mdat[1:n,])
summary(fit3)
lines(fit3$fit, col="blue")

fcs3 <- predict(fit3, mdat[(n+1):(n+63),])
fcs3
plot(fcs3)
write.table(fcs3,"~/forecastsgam.txt",sep="\t")

library(forecast)
attach(MOforecast)
head(MOforecast)
tail(MOforecast)
win.graph()
```

```
accuracy(fols,Med)
accuracy(fqr,Med)
accuracy(fgam,Med)

y=ts(Med)
plot(y,xlab="year", ylab="Electricity demand (MW)",
 ylim=c(30000,38000))
lines(fgam,lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs3, col="blue")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
########Actual and forecast plots################
win.graph()
par(mfrow=c(1,2))
accuracy(fols,Med)
accuracy(fqr,Med)
accuracy(fgam,Med)

y=ts(Med)
plot(y,xlab="year", ylab="fQR electricity demand (MW)",
 ylim=c(30000,38000))
lines(fqr,lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs2, col="blue")
legend("topright",lty=c("solid","dashed"),col=c("black",
"red"),
       c("Actuals","Forecasts"))

y=ts(Med)
plot(y,xlab="year", ylab="fGAM electricity demand (MW)",
 ylim=c(30000,38000))
lines(fgam,lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs3, col="blue")
legend("topright",lty=c("solid","dashed"),col=c("black",
"red"),
       c("Actuals","Forecasts"))

####################
```

```
# Fit QRA for MPED###

library (lattice)#For multi panel graphs
library(mgcv)#For GAM
library(gstat)#For checking spatial patterns
library(gamlss)#For GAM

#####
attach(Moforecast)
head(Moforecast)
win.graph()
z=Med
y <- ts(z)

plot(y, xlab="Observation number", ylab="Monthly load",
ylim=c(30000,38000),xlim=c(0,14))

###########
fit4 <-rq(Med~(fols)+(fqr)+(fgam),tau=0.5,data =
Moforecast[1:n,])# insample
summary(fit4)
lines(fit4$fit , col="blue")
fcs4 <- predict(fit4 , mdat[(n+1):(n+63),])
fcs4
fqra=fitted(fit4)
plot(fqra)
Write.table(fqra,"~/forecastsqra.txt",sep="\t")

library(forecast)
attach(M1forecast)
head(M1forecast)
tail(M1forecast)
win.graph()
accuracy(fols,Med)
accuracy(fqr,Med)
accuracy(fgam,Med)
accuracy(fqra,Med)
```

```
y=ts(Med, start=c(2009,1), freq=12,)
plot(y,xlab="year", ylab="Electricity demand (MW)",
 ylim=c(30000,38000))
lines(fqra, lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs4, col="blue")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
########Actual and forecast plots################
win.graph()
par(mfrow=c(1,2))
accuracy(fols,Med)
accuracy(fqr,Med)
accuracy(fqra,Med)

y=ts(Med, start=c(2009,1), freq=12)
plot(y,xlab="Observation number", ylab=
"fQR electricity demand (MW)",ylim=c(30000,38000))
lines(fqr, lty= "dashed",col="red")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
y=ts(Med, start=c(2009,1), freq=12,)
plot(y,xlab="year", ylab="fQRA electricity demand (MW)",
 ylim=c(30000,38000))
lines(fqra, lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs4, col="blue")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
####################Fig4 QR, GAM and QRA plot####
win.graph()
par(mfrow=c(1,3))
accuracy(fols,Med)
accuracy(fqr,Med)
accuracy(fgam,Med)
accuracy(fqra,Med)
```

```
y=ts(Med, start=c(2009,1), freq=12)
plot(y,xlab="Date", ylab="fQR electricity demand (MW)",
ylim=c(30000,38000))
a=ts(fqr, start=c(2009,1), freq=12)
lines(a,lty= "dashed",col="red")
legend("topright",lty=c("solid","dashed"),col=c("black",
"red"),
       c("Actuals","Forecasts"))
y=ts(Med, start=c(2009,1), freq=12,)
plot(y,xlab="Date", ylab="fGAM electricity demand (MW)",
 ylim=c(30000,38000))
b=ts(fgam, start=c(2009,1), freq=12)
lines(b,lty= "dashed",col="red")
legend("topright",lty=c("solid","dashed"),col=c("black",
"red"),
       c("Actuals","Forecasts"))
y=ts(Med, start=c(2009,1), freq=12)
plot(y,xlab="Date", ylab="fQRA electricity demand (MW)",
 ylim=c(30000,38000))
c=ts(fqra, start=c(2009,1), freq=12)
lines(c,lty= "dashed",col="red")
legend("topright",lty=c("solid","dashed"),col=c("black",
"red"),
       c("Actuals","Forecasts"))
###################################################################
attach(anlyteqdat)
head(anlyteqdat)
tail(anlyteqdat)
win.graph()

#library(glmnet) # package for shrinkage methods
library(forecast)
#library(glinternet)
library(earth)
library(mgcv)
library(itsadug)

# EARTH determining reference temperatures
```

```
a <- earth(Qpedd~DmaxTCI,minspan=-1)
plotmo(a)
summary(a, digits = 1,style = "pmax")#minspan=912,

################Quarterly average temp plot##
#attach(tqdata)
#head(tqdata)
#tail(tqdata)

win.graph()
x<-ts(Qpedd[2], start=2000, freq=4)
plot(x, ylab="Quarterly peak electricity demand (MW)",
col="blue",xlab="Observations")
(smooth.spline(time(x), x))
lines(smooth.spline(time(x), x, spar=0.1136387),
col="red",lwd=3)
Qedfits = fitted((smooth.spline(time(x), x,
spar=0.1136387)))

write.table(Qedfits,"~/fittedvalues.txt",sep="\t")


#####################################################

#CROSS VALIDATION MODEL SELECTION
attach(tqcvdat)
head(tqcvdat)
library(forecast)
y=ts(Qpedd,start=2000,freq=4)
regress = tslm(x~CDDAAQTCI+CDDAmaxTCI+CDDAminTCI+
DAH+Daytype+DH+noltrend)
summary(regress)
CV(regress)

#####################################################

# FORECASTING
```

```
attach(Qeddat)
head(Qeddat)
tail(Qeddat)
win.graph()
x=ts(Qpedd[2], start=2000, freq=4)
plot(x,ylab="Quarterly peak electricity demand (MW)",
col="blue",
      xlab="Days")
n <- length(Qpd)
n

attach(tqdat)
head(tqdat)
tail(tqdat)

# Save the length of the information regarding
forecasting horizon
h <- nrow(tqdat)-n
h
# Fit an additive regression model OLS####
fit4 <- lm(Qpd~CDDAAQTCI+CDDAmaxTCI+CDDAminTCI+
DAH+Daytype+DH
+noltrend,data=tqdat[1:n,])# insample
# Return the summary of the model
summary(fit4)
# Add a new line (the fit of the model) on the
  existing graph
lines(fit4$fit, col="blue")

# Calculate the out-of-sample forecasts, based
on the available information
#on temperature
fcs1 <- predict(fit4, tqdat[(n+1):(n+21),])
fcs1
plot(fcs1)
write.table(fcs1,"~/forecastsOLS.txt",sep="\t")

#lines(fcs, col=" red")
```

```
############
##QUANTREG
###########

library(quantreg)
fit5 <- rq(Qpd~CDDAAQTCI+CDDAmaxTCI+CDDAminTCI+
DAH+Daytype+DH
+noltrend,tau=0.5,data=tqdat[1:n,])# insample
fit5<- rq(Qpd~noltrend,tau=0.5,
data=tqdat[1:n,])# insample
summary(fit5)
lines(fit5$fit, col="blue")

fcs4 <- predict(fit5, tqdat[(n+1):(n+21),])
fcs4
plot(fcs4)
write.table(fcs4,"~/forecastsQR.txt",sep="\t")


library(forecast)
attach(Qaforecast)
head(Qaforecast)
win.graph()
accuracy(fols,Qed)
accuracy(fqr,Qed)

y=ts(Qed)
plot(y,xlab="Date", ylab="Electricity demand (MW)",
ylim=c(30000,38000))
lines(fqr,lty= "dashed",col="red")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
######################
# Fit an additive regression model GAM
library (lattice)#For multi panel graphs
library(mgcv)#For GAM
```

```
library(gstat)#For checking spatial patterns
library(gamlss)#For GAM

fit6 <- gam(Qpd~CDDAminTCI+CDDAAQTCI*Daytype+
DH+s(Daytype,bs="re")+DAH+
s(noltrend),method="REML",data=tqdat[1:n,])
summary(fit6)
lines(fit6$fit, col="blue")

fcs5 <- predict(fit6, tqdat[(n+1):(n+21),])
fcs5
plot(fcs5)
write.table(fcs5,"~/forecastsgam.txt",sep="\t")

library(forecast)
attach(Qoforecast)
head(Qoforecast)
tail(Qoforecast)
win.graph()
accuracy(fols,Qed)
accuracy(fqr,Qed)
accuracy(fgam,Qed)

y=ts(Qed)
plot(y,xlab="Date", ylab="Electricity demand (MW)",
 ylim=c(30000,38000))
lines(fgam,lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs5, col="blue")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
########Actual and forecast plots###########
win.graph()
par(mfrow=c(1,2))
accuracy(fols,Qed)
accuracy(fqr,Qed)

y=ts(Qed)
```

```
plot(y,xlab="Date", ylab="fQR electricity demand (MW)",
ylim=c(30000,38000))
lines(fqr,lty= "dashed",col="red")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
accuracy(fols,Qed)
accuracy(fqr,Qed)
accuracy(fgam,Qed)

y=ts(Qed)
plot(y,xlab="Date", ylab="fGAM electricity demand (MW)",
 ylim=c(30000,38000))
lines(fgam,lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs5 , col="blue")
legend("topright",lty=c("solid","dashed"),col=c("black","red"),
       c("Actuals","Forecasts"))
###################################################
######################
# Fit QRA

library (lattice)#For multi panel graphs
library(mgcv)#For GAM
library(gstat)#For checking spatial patterns
library(gamlss)#For GAM


#####
attach(Qoforecast)
head(Qoforecast)
win.graph()
z=Qed
y <- ts(z)

plot(y, xlab="Date", ylab="Quaterly load",
ylim=c(30000,38000)

###########
```

```
fit7 <-rq(Qed~(fols)+(fqr)+(fgam),tau=0.5,data =
  Qoforecast[1:n,])# insample
summary(fit7)
lines(fit7$fit, col="blue")
fqra=fitted(fit7)
plot(fqra)
write.table(fqra,"~/forecastsqra.txt",sep="\t")

library(forecast)
attach(Qforecast)
head(Qforecast)
win.graph()
accuracy(fols,Qed)
accuracy(fqr,Qed)
accuracy(fgam,Qed)
accuracy(fqra,Qed)

y=ts(Qed)
plot(y,xlab="Date", ylab="Electricity demand (MW)",
ylim=c(30000,38000))
lines(fqra,lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs5, col="blue")
legend("topright",lty=c("solid","dashed"),col=c("black","red"),
       c("Actuals","Forecasts"))
########Actual and forecast plots for QPED##########
win.graph()
par(mfrow=c(1,2))
accuracy(fols,Qed)
accuracy(fqr,Qed)

y=ts(Qed,start=c(2009,1), freq=4)
plot(y,xlab="Date", ylab="fQR electricity demand (MW)",
ylim=c(30000,38000))
lines(fqr,lty= "dashed",col="red")
legend("topright",lty=c("solid","dashed"),col=c("black","red"),
       c("Actuals","Forecasts"))
accuracy(fols,Qed)
accuracy(fqr,Qed)
```

```
accuracy(fqra,Qed)

y=ts(Qed,start=c(2009,1), freq=4 )
plot(y,xlab="Date", ylab="fQRA electricity demand (MW)",
 ylim=c(30000,38000))
lines(fqra,lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs5, col="blue")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
################# QR, GAM and QRA plots######
win.graph()
par(mfrow=c(1,3))
accuracy(fols,Qed)
accuracy(fqr,Qed)
accuracy(fgam,Qed)
accuracy(fqra,Qed)
y=ts(Qed,start=c(2009,1), freq=4)
plot(y,xlab="Date", ylab="fQR electricity demand (MW)",
ylim=c(30000,38000))
m=ts(fqr,start=c(2009,1), freq=4)
lines(m,lty= "dashed",col="red")
legend("topright",lty=c("solid","dashed"),col=c("black","red"),
       c("Actuals","Forecasts"))
y=ts(Qed,start=c(2009,1), freq=4)
plot(y,xlab="Date", ylab="fGAM electricity demand (MW)",
 ylim=c(30000,38000))
l=ts(fgam,start=c(2009,1), freq=4)
lines(l,lty= "dashed",col="red")
legend("topright",lty=c("solid","dashed"),col=c("black","red"),
       c("Actuals","Forecasts"))
y=ts(Qed,start=c(2009,1), freq=4)
plot(y,xlab="Date", ylab="fQRA electricity demand (MW)",
 ylim=c(30000,38000))
k=ts(fqra,start=c(2009,1), freq=4)
lines(k,lty= "dashed",col="red")
legend("topright",lty=c("solid","dashed"),col=c("black","red"),
       c("Actuals","Forecasts"))
```

```
##########
######SVR and SGB forecasts for MPED#######
win.graph()

z <- ts(data_test&Mped, start=c(2009,1), freq=12)
f <- ts(predYtest_L, start=c(2009,1), freq=12)
plot(z,xlab="Date",ylim=c(30000,37000),lwd=3,ylab="MPED (MW)")
lines(f,col="red", lty=2,lwd=3)
legend("topright",col=c("black","red"), lty=1:2,lwd=3,
       legend=c("Actuals", "Forecasts (SVR)"))
################################################
##QUANTILE REGRESSION#

attach(anlytemdat)
head(anlytemdat)
tail(anlytemdat)
win.graph()
#library(glmnet) # package for shrinkage methods
library(forecast)
#library(glinternet)
library(earth)
library(mgcv)
library(itsadug)

# EARTH determining reference temperatures
a<- earth(Mped~AmaxTCI,minspan=-1)
plotmo(a)
summary(a, digits = 1,style = "pmax")#minspan=912,

########################################################
########################################################
#attach(mdata)
#head(mdata)
#tail(mdata)

win.graph()

x=ts(Mped[2], start=2000, freq=12)
```

```
plot(x, type="p", ylab="Monthly peak electricity demand (MW)",
col="blue",
     xlab="Year")
(smooth.spline(time(x), x))
lines(smooth.spline(time(x), x, spar=0.1136387),col="red",
lwd=3)
mpedfits = fitted((smooth.spline(time(x), x, spar=0.1136387)))

write.table(mpedfits,"~/fittedvalues.txt",sep="\t")


################################################################
################################################################

#CROSS VALIDATION MODEL SELECTION
attach(mcvdat)
head(mcvdat)
library(forecast)
y=ts(Mped,start=2000,freq=12)
regress = tslm(x~CDDAAMTCI+CDDAmaxTCI+CDDAminTCI+DAH
+Daytype+DBH+DH+
HDDAAMTCI+HDDAminTCI+noltrend)
summary(regress)
CV(regress)


############################################################
############################################################

# FORECASTING

attach(Mpddat)
head(Mpddat)
tail(Mpddat)
win.graph()
plot(Mpd, xlab="Observations", ylab="MPED")
n <- length(Mpd)
n
```

```
attach(mdat)
head(mdat)
tail(mdat)

# Save the length of the information regarding
forecasting horizon
h <- nrow(mdat)-n
h
# Fit an additive regression model OLS
fit1 <- lm(Mpd~CDDAAMTCI+CDDAmaxTCI+CDDAminTCI+
DAH+Daytype+DBH+DH+
              HDDAAMTCI+HDDAminTCI+noltrend,
              data=mdat[1:n,])# insample

# Return the summary of the model
summary(fit1)
# Add a new line (the fit of the model) on the
  existing graph
lines(fit1$fit, col="blue")


# Calculate the out-of-sample forecasts, based on
  the available information
#on temperature
fcs <- predict(fit1, mdat[(n+1):(n+27),])
fcs
plot(fcs)
write.table(fcs,"~/forecastsOLS.txt",sep="\t")

#lines(fcs, col=" red")

###########
##QUANTREG
##########

library(quantreg)
fit2 <- rq(formula=Mpd~CDDAAMTCI+CDDAmaxTCI+
```

```
CDDAminTCI+DAH+Daytype+DBH+DH+
            HDDAAMTCI+HDDAminTCI+noltrend ,
tau = 0.5,data=mdat[1:n,])# insample
fit2<- rq(Mpd~CDDAminTCI+CDDAminTCI*Daytype,
tau = 0.5,data=mdat[1:n,])# insample
summary(fit2)
lines(fit2$fit , col="blue")

fcs2 <- predict(fit2 , mdat[(n+1):(n+27),])
fcs2
plot(fcs2)
write.table(fcs2,"~/forecastsQR.txt",sep="\t")


library(forecast)
attach(M2forecast)
head(M2forecast)
win.graph()
accuracy(fols ,Med)
accuracy(fqr ,Med)

y=ts(Med)
plot(y,xlab="year", ylab="Electricity demand (MW)",
 ylim=c(31000,38000))
lines(fqr,lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs2, col="blue")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
#####################
# Fit an additive regression model GAM
library (lattice)#For multi panel graphs
library(mgcv)#For GAM
library(gstat)#For checking spatial patterns
library(gamlss)#For GAM

fit3 <- gam(Mpd~CDDAminTCI+CDDAAMTCI+CDDAmaxTCI
+DH+s(Daytype,bs="re")+HDDAminTCI+
```

```
s(noltrend),method="REML",data=mdat[1:n,])
summary(fit3)
lines(fit3$fit, col="blue")

fcs3 <- predict(fit3, mdat[(n+1):(n+27),])
fcs3
plot(fcs3)
write.table(fcs3,"~/forecastsgam.txt",sep="\t")

library(forecast)
attach(Mgforecast)
head(Mgforecast)
win.graph()
accuracy(fols,Med)
accuracy(fqr,Med)
accuracy(fgam,Med)

y=ts(Med)
plot(y,xlab="year", ylab="Electricity demand (MW)",
ylim=c(31000,38000))
lines(fgam,lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs3, col="blue")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
        c("Actuals","Forecasts"))
########Actual and forecast plots#########
win.graph()
par(mfrow=c(1,2))
accuracy(fols,Med)
accuracy(fqr,Med)
accuracy(fgam,Med)

y=ts(Med)
plot(y,xlab="year", ylab="fQR electricity demand (MW)",
ylim=c(31000,38000))
lines(fqr,lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs2, col="blue")
legend("topright",lty=c("solid","dashed"),
```

```
col=c("black","red"),
       c("Actuals","Forecasts"))

y=ts(Med)
plot(y,xlab="year", ylab="fGAM electricity demand (MW)",
ylim=c(31000,38000))
lines(fgam,lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs3 , col="blue")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
###################

#####################
# Fit QRA

library (lattice)#For multi panel graphs
library(mgcv)#For GAM
library(gstat)#For checking spatial patterns
library(gamlss)#For GAM


#####
attach(Mgforecast)
head(Mgforecast)
win.graph()
z=Med
y <- ts(z)

plot(y, xlab="Observation number", ylab="Monthly load",
ylim=c(31000,38000),xlim=c(0,14))

###########
fit4 <-rq(Med~(fols)+(fqr)+(fgam),tau=0.5,data =
  Mgforecast[1:n,])# insample
summary(fit4)
lines(fit4$fit , col="blue")
fqra=fitted(fit4)
```

```
plot(fqra)
write.table(fqra,"~/forecastsqra.txt",sep="\t")

library(forecast)
attach(Mvforecast)
head(Mvforecast)
win.graph()
accuracy(fols,Med)
accuracy(fqr,Med)
accuracy(fgam,Med)
accuracy(fqra,Med)

y=ts(Med)
plot(y,xlab="year", ylab="Electricity demand (MW)",
 ylim=c(31000,38000))
lines(fqra,lty= "dashed",col="red")
lines(x=(n+1):(n+h), fcs3, col="blue")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
########Actual and forecast plots############
win.graph()
par(mfrow=c(1,2))
accuracy(fols,Med)
accuracy(fqr,Med)
accuracy(fqra,Med)

y=ts(Med)
plot(y,xlab="Observation number", ylab=
"fQR electricity demand (MW)",ylim=c(31000,38000))
lines(fqr,lty= "dashed",col="red")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
y=ts(Med)
plot(y,xlab="year", ylab="fQRA electricity demand (MW)",
ylim=c(31000,38000))
lines(fqra,lty= "dashed",col="red")
```

```
lines(x=(n+1):(n+h), fcs4, col="blue")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
####################Fig4 QR, GAM and QRA plot###########
win.graph()
par(mfrow=c(1,3))
accuracy(fols,Med)
accuracy(fqr,Med)
accuracy(fgam,Med)
accuracy(fqra,Med)

y=ts(Med)
plot(y,xlab="Observation number", ylab=
"fQR electricity demand (MW)",ylim=c(31000,38000))
lines(fqr,lty= "dashed",col="red")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
y=ts(Med)
plot(y,xlab="year", ylab="fGAM electricity demand (MW)",
 ylim=c(31000,38000))
lines(fgam,lty= "dashed",col="red")
legend("topright",lty=c("solid","dashed"),
col=c("black","red"),
       c("Actuals","Forecasts"))
y=ts(Med)
plot(y,xlab="year", ylab="fQRA electricity demand (MW)",
ylim=c(31000,38000))
lines(fqra,lty= "dashed",col="red")
legend("topright",lty=c("solid","dashed"),col=c("black","red"),
       c("Actuals","Forecasts"))
################################################################
Write Table 3.1:Comparison of models#############
begin{table}[!h!]
setlength{tabcolsep}{0.02mm}
caption{Comparison of models.}
small    begin{tabular}{cll}
```

```
hline
textbf{Models}  & {\bf Strengths} &  {\bf Weaknesses}\\
hline
Model1(AQR)&1.A hybrid model that combines GAMS &1.Requires
a smoothing
\\
& with QR.  &function of the covariates.\\
&2. Estimation is distribution free. &2. Parameters are harder \\
&3. Robust to outliers in the response &to estimate. \\
&variable. &3. Does not give any       \\
&&details about the size\\
&&of high level of possible\\
&& exceedances.\\ \hline
Model2(EM)&1.Semi-parametric extremal mixture model.&1.Has limitations
\\
&2. Based on one covariate, which is &accuracy and stability.\\
&   t=1,...,n. &2. Very sensitive to \\
&&numbers and location\\
&&of the measured points.\\ \hline
Model3(NLQR)&1.Inference is performed based on &1.Requires a smoothing
\\
&large sample approximation.   &parameter.  \\
&2. Robust to outliers in the  &2. Outliers only have\\
&response variable. &influence on quantile \\
&&curves close to them, i.e\\
& &they affect extreme quantiles\\
hline
end{tabular}
label{t1}
end{table}
```