Time Series Forecasting with Applications in Macroeconomics and Energy



Siddharth Arora

Somerville College, University of Oxford

Thesis submitted to the University of Oxford

for the degree of Doctor of Philosophy

Hilary Term 2013

To,

The Supreme Consciousness,

The Substratum of Us All.

ACKNOWLEDGEMENTS

I am grateful to my parents for being an enormous source of inspiration and for their monumental support and guidance over the years. I consider myself lucky to have tremendous support from my younger brother Mayank, and his wonderful wife Alankrita. I thank them both for also pestering me to submit this thesis soon, and special thanks to the love of my life, my niece, Samayra. I am blessed to have guidance from my Guru, Swami Atmananda. His teachings of Advait Vedanta have always inspired me to strive to work selflessly, to try and realize the non-dual nature of our true self, and I thank him from the bottom of my heart. The love and encouragement from my family has been monumental in every aspect of my life.

I am most thankful to my supervisor, James W. Taylor, for his excellent guidance, patience and support for my research. His thorough critique has been extremely valuable in improving the quality of this thesis. James has been a brilliant supervisor, and I extend my unreserved sincere gratitude to him.

I am grateful to my assessors, Dolores Romero Morales and Lilian de Menezes, for taking time out from their busy schedule to examine this thesis. I am also thankful to Dolores and Christoph Reisinger for examining previous versions of this thesis, and for their very useful comments. I wish my gratitude to my tutor, Karin Erdmann, for her support and help regarding college matters.

I am thankful to my colleagues at the Oxford centre for industrial and applied mathematics (OCIAM) and at the Saïd business school (SBS), especially Athanasios Tsanas, Jooyoung Jeon Shilan Mistry and Delioma Oramas Dorta, for stimulating discussions on a wide range of topics. I am thankful to my wonderful friends, Amit Mehndiratta, Angeliki Xifara, Arpit Mittal, Astrid Erber (Ally), Chetak Gandhi, Christine Lo, Elisa Vergari, Fahd Baig, Felix Castellanos, Ghia Osseiran, Jo Aldhouse, Matthias Mang, Sabyasachi Baboo, Sara Kalim, Stephen Haben and Vaanathi Sundaresan, for their enormous support. Life in Oxford was beautiful because of them. I extend my sincere gratitude to M.S. Santhanam, Vijay Sheorey and Prasanta K. Panigrahi for their help and guidance.

I am extremely grateful to the trustees of the Felix scholarship for their financial support that made this study possible. I thank Somerville College, University of Oxford, for awarding me the Lord Jenkins scholarship, SBS for awarding me a research assistantship, and International Institute of Forecasters for funding a part of this research.

I am thankful to OCIAM and SBS for providing me access to departmental computational facilities to conduct my research.

Last but not the least; I thank every individual whose hard work has formed the basis of my scientific research today.

Time Series Forecasting with Applications in Macroeconomics and Energy

Siddharth Arora

Somerville College, University of Oxford

Summary of the Thesis Submitted for the Degree of Doctor of Philosophy

Hilary Term 2013

The aim of this study is to develop novel forecasting methodologies. The applications of our proposed models lie in two different areas: macroeconomics and energy. Though we consider two very different applications, the common underlying theme of this thesis is to develop novel methodologies that are not only accurate, but are also parsimonious.

For macroeconomic time series, we focus on generating forecasts for the US Gross National Product (GNP). The contribution of our study on macroeconomic forecasting lies in proposing a novel nonlinear and nonparametric method, called weighted random analogue prediction (WRAP) method. The out-of-sample forecasting ability of WRAP is evaluated by employing a range of different performance scores, which measure its accuracy in generating both point and density forecasts. We show that WRAP outperforms some of the most commonly used models for forecasting the GNP time series.

For energy, we focus on two different applications: (1) Generating accurate short-term forecasts for the total electricity demand (*load*) for Great Britain. (2) Modelling Irish electricity smart meter data (consumption) for both residential consumers and small and medium-sized enterprises (SMEs), using methods based on kernel density (KD) and *conditional* kernel density (CKD) estimation.

To model load, we propose methods based on a commonly used statistical dimension reduction technique, called singular value decomposition (SVD). Specifically, we propose two novel methods, namely, discount weighted (DW) intraday and DW intraweek SVD-based exponential smoothing methods. We show that the proposed methods are competitive with some of the most commonly used models for load forecasting, and also lead to a substantial reduction in the dimension of the model.

The load time series exhibits a prominent intraday, intraweek and intrayear seasonality. However, most existing studies accommodate the ‘double seasonality’ while modelling short-term load, focussing only on the intraday and intraweek seasonal effects. The methods

considered in this study accommodate the ‘triple seasonality’ in load, by capturing not only intraday and intraweek seasonal cycles, but also intrayear seasonality.

For modelling load, we also propose a novel rule-based approach, with emphasis on special days. The load observed on special days, e.g. public holidays, is substantially lower compared to load observed on normal working days. Special day effects have often been ignored during the modelling process, which leads to large forecast errors on special days, and also on normal working days that lie in the vicinity of special days. The contribution of this study lies in adapting some of the most commonly used seasonal methods to model load for both normal and special days in a coherent and unified framework, using a rule-based approach. We show that the post-sample error across special days for the rule-based methods are less than half, compared to their original counterparts that ignore special day effects.

For modelling electricity smart meter data, we investigate a range of different methods based on KD and CKD estimation. Over the coming decade, electricity smart meters are scheduled to replace the conventional electronic meters, in both US and Europe. Future estimates of consumption can help the consumer identify and reduce excess consumption, while such estimates can help the supplier devise innovative tariff strategies. To the best of our knowledge, there are no existing studies which focus on generating density forecasts of electricity consumption from smart meter data. In this study, we evaluate the density, quantile and point forecast accuracy of different methods across one thousand consumption time series, recorded from both residential consumers and SMEs. We show that the KD and CKD methods accommodate the seasonality in consumption, and correctly distinguish weekdays from weekends.

For each application, our comprehensive empirical comparison of the existing and proposed methods was undertaken using multiple performance scores. The results show strong potential for the models proposed in this thesis.

CONTENTS

List of Figures 1

List of Tables 7

Abbreviations 9

Frequently Used Notations and Symbols 12

Chapter 1 **INTRODUCTION** ……. **14**

1.1 Nonlinear and Nonparametric Modelling Approaches for Forecasting the US GNP ……. 15

1.2 Short-term Load Forecasting using Discount Weighted SVD-based Methods ……. 16

1.3 Short-term Forecasting of Anomalous Load using Rule-based Methods ……. 18

1.4 Modelling Electricity Smart Meter Data Based on Kernel Density Estimation ……. 20

Chapter 2 **NONLINEAR AND NONPARAMETRIC MODELLING APPROACHES FOR PROBABILISTIC FORECASTING OF THE US GROSS NATIONAL PRODUCT** ……. **22**

2.1 Introduction ..…… 22

2.1.1 Motivation and Challenges ……. 23

2.1.2 Historical Review ……. 24

2.1.3 Contributions 27

2.2 GNP Characteristics 28

2.3 Parametric Modelling Approaches 30

2.3.1 Autoregressive Models 30

2.3.2 Self-exciting Autoregressive Models 32

2.3.3 Markov-switching Autoregressive Models 37

2.4 Nonlinear and Nonparametric Modelling Approaches 42

2.4.1 Motivation and Rationale for the Proposed Methodology 43

2.4.2 Nearest Neighbour Method 46

2.4.3 Kernel Regression 53

2.4.4 Weighted Random Analogue Prediction Method 56

2.5 Results and Observations 60

2.6 Summary and Concluding Remarks 84

Chapter 3 **SHORT-TERM LOAD FORECASTING USING DISCOUNT WEIGHTED SVD-BASED EXPONENTIAL SMOOTHING 86**

3.1 Introduction 86

3.1.1 Motivation and Challenges 86

3.1.2 Overview of SVD 88

3.1.3 Historical Review 90

3.1.4 Contributions 92

3.2 Load Characteristics 93

3.3 Forecasting Methods 96

3.3.1 HWT Exponential Smoothing Method ..……..…….**.....** 96

3.3.2 Seasonal ARMA …….……..…….... 99

3.3.3 Seasonal ANN Method …...…….…….……... 101

3.4 SVD-based Exponential Smoothing Methods ……..…….…….... 103

3.4.1 Intraday SVD-based Exponential Smoothing …..…….…..... 104

3.4.2 Intraweek SVD-based Exponential Smoothing …..…….…..... 107

3.4.3 Discount Weighted Intraday SVD-based Exponential Smoothing …….…….. 110

3.4.4 Discount Weighted Intraweek SVD-based Exponential Smoothing …….…….. 114

3.5 Results and Observations …..…...…….…….... 115

3.6 Summary and Concluding Remarks …….…… 122

Chapter 4 **SHORT-TERM FORECASTING OF ANOMALOUS LOAD USING RULE-BASED TRIPLE SEASONAL METHODS** ….....**.** **123**

4.1 Introduction ……..…… 123

4.1.1 Motivation and Challenges ……. 123

4.1.2 Historical Review …….……... 126

4.1.3 Contributions **..**…… 129

4.2 Anomalous Load Characteristics ……..... 131

4.3 Forecasting Methods ……..……...……...…… 135

4.3.1 Rule-based Triple Seasonal Modified HWT Exponential Smoothing Method …...…….……..….... 135

4.3.2 Rule-based Triple Seasonal ARMA …….……... 139

4.3.3 Rule-based Triple Seasonal ANN Method 140
4.3.4 Rule-based Triple Seasonal Modified Intraday SVD-based Exponential Smoothing Method 141
4.3.5 Rule-based Triple Seasonal Modified Intraweek SVD-based Exponential Smoothing Method 142
4.4 Formulating Rules 143
4.4.1 Rule 1 144
4.4.2 Rule 2 145
4.4.3 Rule 3 146
4.4.4 Rule 4 148
4.5 Results and Observations 149
4.5.1 Results using load for Great Britain 150
4.5.2 Results using load for France 166
4.6 Summary and Concluding Remarks 170
Chapter 5 **MODELLING ELECTRICITY SMART METER DATA BASED ON KERNEL DENSITY ESTIMATION** **171**
5.1 Introduction 171
5.1.1 Motivation and Challenges 172
5.1.2 Historical Review 174
5.1.3 Contributions 176
5.2 Smart Meter Data Characteristics 178
5.3 Modelling based on the Kernel Density Estimator 186
5.3.1 Unconditional and Conditional Kernel Density Estimation ... 186

5.3.2 Kernel Bandwidth Selection and Boundary Correction …….. 188
5.3.3 Unconditional Kernel Density Estimation (KD-U) …..…….. 190
5.3.4 Separate Kernel Density Estimation for each Period of the Week (KD-W) ……..….... 190
5.3.5 CKD Estimation Conditional on the Period of Week (CKD-W) ……...……..….... 191
5.3.6 CKD Estimation Conditional on the Period of Week and Period of Day (CKD-WD) ….....……..….... 192
5.3.7 KD Estimation based on Type of Intraday Cycle (KD-IC) … 193
5.3.8 CKD Estimation Conditional on Type of Intraday Cycle (CKD-IC) ……...……..….... 194
5.3.9 CKD Estimation Conditional on Lagged Consumption (CKD-Lag) ……..……...……..….... 195
5.4 Empirical Analysis ……..…….……..…….…. 195
5.4.1 Evaluating Density Forecasts ……...…….…….. 198
5.4.2 Evaluating Quantile Forecasts ……. 204
5.4.3 Evaluating Point Forecasts …….……. 209
5.5 Deriving Prediction Intervals for Electricity Cost ……...…. 210
5.6 Summary and Concluding Remarks ……..……..…. 217
Chapter 6 **SUMMARY AND DIRECTIONS FOR FUTURE RESEARCH ... 219**
6.1 Nonlinear and Nonparametric Modelling Approaches for Forecasting the US GNP ……...…..…….……...…… 219
6.2 Short-term Load Forecasting using Discount Weighted SVD-based

Methods ……..…..…….……...……... 222

6.3 Short-term Forecasting of Anomalous Load using Rule-based Methods ……...…..……...……...……. 223

6.4 Modelling Electricity Smart Meter Data Based on Kernel Density Estimation ……...…….……...……. 225

6.5 Directions for Future Research ……...….……...……. 227

References 231

LIST OF FIGURES

2.1 Log returns of the quarterly US Gross National Product (GNP) ……..... 29

2.2 Comparison of point forecast accuracy of the benchmarks – unconditional mean and random walk in forecasting US GNP, quantified using the Root Mean Square Error (RMSE) ……..……..…..……. 65

2.3 Evaluation of point forecast accuracy of parametric and nonparametric models on the out-of-sample US GNP time series for horizons (look ahead time) ranging from one quarter to sixteen quarters (four years), quantified using the Root Mean Square Error (RMSE) ……..……. 66

2.4 Evaluation of point forecast accuracy of parametric and nonparametric models on the out-of-sample US GNP time series for horizons ranging from one quarter to sixteen quarters (four years), quantified using the Mean Absolute Error (MAE)……. 68

2.5 Evaluation of density forecast accuracy of parametric and nonparametric models on out-of-sample US GNP time series for horizons ranging from one quarter to sixteen quarters (four years), quantified using the Continuous Ranked Probability Score (CRPS) …..... 70

2.6 One-step-ahead out-of-sample density forecasts for US GNP generated using the proposed model ……...……..…….. 71

2.7 Evaluation of bias and coverage for the proposed model based on a reliability diagram ……..... 72

2.8 One-step-ahead out-of-sample predictive error density obtained using parametric and nonparametric models ……...……..… 76

2.9 Probability integral transform (PIT) histogram for one-step-ahead out-of-sample forecasts obtained using parametric and nonparametric models …….. 78

2.10 Autocorrelation function (ACF) for PIT values for one-step-ahead out-of-sample forecasts obtained using parametric and nonparametric models ……..……...…. 79

2.11 Evaluation of density forecast accuracy of parametric and nonparametric models on out-of-sample UK GDP time series for horizons ranging from one quarter to sixteen quarters (four years), quantified using the Root Mean Square Error (RMSE) …..….. 82

2.12 Evaluation of density forecast accuracy of parametric and nonparametric models on out-of-sample UK GDP time series for horizons ranging from one quarter to sixteen quarters (four years), quantified using the Mean Absolute Error (MAE) ……... 83

2.13 Evaluation of density forecast accuracy of parametric and nonparametric models on out-of-sample UK GDP time series for horizons ranging from one quarter to sixteen quarters (four years), quantified using the Continuous Ranked Probability Score (CRPS) ……..... 84

3.1 Pictorial representation of the different steps involved in applying SVD to the load data ……..…….……...…… 90

3.2 Half-hourly load for Great Britain stretching from 1 January 2001 to 31 December 2009. The vertical dashed line denotes the time index that divides the time series into non-overlapping estimation and evaluation set …….…… 94

3.3 Load for a winter fortnight (16-29 January, 2005) and a summer fortnight (10-23 July, 2005) …… 95

3.4 Average intraday profile for each day of the week, computed using only the estimation period (2001-2008) ……..…...……...……. 95

3.5 First three intraday feature vectors for the British log load time series ……... 106

3.6 First three interday feature series for the British log load time series …….. 106

3.7 First three intraweek feature vectors for the British log load time series ……. 108

3.8 First three interweek feature series for the British log load time series ……..……. 109

3.9 First intraday feature vector plotted over time …..……... 111

3.10 First intraweek feature vector plotted over time …..……. 111

3.11 Out-of-sample MAPE for the original and discount weighted versions of the intraday and intraweek SVD-based exponential smoothing methods ……..…… 119

3.12 Out-of-sample MAPE for the triple seasonal versions of the HWT exponential smoothing, SARMA and the ANN method, along with the simple benchmarks and combination of the proposed SVD-based exponential smoothing methods ……...… 120

4.1 Half-hourly load for Great Britain stretching from 1 January 2001 to 31 December 2009 …….. 132

4.2 Load profile for a spring bank holiday Monday (26 May 2008), a normal Monday (19 May 2008) from the preceding week, and a normal Monday (2 June 2008) from the following week ……..…… 133

4.3 Load profile for New Year's Day (1 January), Easter Monday (24 March), Spring Bank Holiday Monday (26 May), and Christmas Day (25 December) observed in the year 2008 ……... 133

4.4 Load profile for Christmas Day (25 December), observed across all years within the estimation sample, stretching from 2001 to 2008 …… 134

4.5 Seasonal indices for the intraday, intraweek, and intrayear cycle …… 137

4.6 Out-of-sample MAPE across special days for the original and rule-based triple seasonal HWT exponential smoothing ….....…… 155

4.7 Out-of-sample MAPE across special days for the original and rule-based SARMA ... 156

4.8 Out-of-sample MAPE across special days for the original and rule-based seasonal ANN ……..… 157

4.9 Out-of-sample MAPE across special days for the original and rule-based intraday SVD ……..……..… 158

4.10 Out-of-sample MAPE across special days for the original and rule-based intraweek SVD ……..……..… 159

4.11 Out-of-sample MAPE across special days for the best performing method from Figures 4.6, 4.7, 4.8, 4.9 and 4.10, along with the simple benchmarks ……...… 162

4.12 Out-of-sample MAPE across normal days, for the two best performing rule-based methods from Figure 4.11 (rule-based HWT and SARMA, both used with Rule 3), the corresponding original versions, and their combination ……. 163

4.13 Out-of-sample MAPE across special days, obtained using RB-HWT and RB-SARMA (with Rule 3), plotted against different periods of the day, for horizon equal to: a) six hours, and b) twenty four hours ……...…… 164

4.14 Out-of-sample MAPE across normal days, obtained using RB-HWT and RB-SARMA (with Rule 3), plotted against different periods of the day, for horizon equal to: a) six hours, and b) twenty four hours ……...…… 165

4.15 Out-of-sample MAPE across special days, obtained using RB-HWT, RB-SARMA, and RB-ANN (all with Rule 5), and their corresponding original versions, using the

French load time series ……...…….. 169

5.1 Consumption time series for a residential consumer …… 178

5.2 Consumption time series for an SME ……...…… 179

5.3 Average intraday profile for each day of the week for: (a) a residential consumer, and (b) an SME ……..…… 180

5.4 SME consumption across different periods of the week using twenty four weeks of data……...…….……...……..… 192

5.5 Consumption across different periods of the week for an SME. Each period includes observations belonging to that period, plus observations belonging to periods with the same intraday cycle…….…….…… 194

5.6 Discounting weights for different values of the discount factor ……..…… 197

5.7 Density forecasts and corresponding actual observations for: a) a residential consumer, and b) an SME …… 199

5.8 Mean CRPS for density forecasting of the 800 residential consumers …… 200

5.9 Mean CRPS for density forecasting of the 200 SMEs …….…… 201

5.10 Density forecasts and corresponding actual observations for an SME, for forecast origin belonging to a Sunday mid-night ……...…… 203

5.11 CRPS for density forecasts of the 200 SMEs generated using the KD-IC Method for moving window of lengths: a) two months, b) four months, and c) six months …….. 204

5.12 Mean unconditional coverage for quantile forecasting across the 800 residential consumers for horizons corresponding to: a) 6 hours, b) 12 hours, and c) 24 hours … 205

5.13 Mean unconditional coverage for quantile forecasting across the 200 SMEs for horizons corresponding to: a) 6 hours, b) 12 hours, and c) 24 hours. …….. 206

5.14 Mean Hit % for across the 800 residential consumers for forecasting of the: a) 5% quantile, b) 50% quantile (median), and c) 95% quantile …… 207

5.15 Mean hit % for across the 200 SMEs for forecasting of the: a) 5% quantile, b) 50% quantile (median), and c) 95% quantile …… 208

5.16 Mean MAE for point forecasting of the 800 residential consumers ……..…… 209

5.17 Mean MAE for point forecasting of the 200 SMEs ……..……..…… 210

5.18 Prediction intervals and corresponding actual cost for a residential consumer …..... 211

5.19 Difference between the total actual cost associated with switching tariffs and the total actual cost associated with using a single allocated tariff, across 580 residential consumers, for the post-sample period, using seven different criteria ……. 216

LIST OF TABLES

2.1 Parameter estimates of an autoregressive model for US GNP……..... 31

2.2 Parameter estimates of the self-exciting threshold autoregressive model for US GNP 37

2.3 Parameter estimates of the Markov-switching autoregressive model for US GNP….. 41

2.4 Parameter estimates for the nearest neighbour, kernel regression, and proposed weighted random analogue prediction method for US GNP……. 60

2.5 Kolmogorov-Smirnov test statistic calculated for the different parametric and nonparametric models ...……... 80

3.1 Parameter estimates for the original and discount weighted versions of the intraday and intraweek SVD-based exponential smoothing methods …… 117

4.1 Parameter estimates for the original and rule-based modified HWT exponential smoothing used in conjunction with four different rules …….…… 151

4.2 Parameter estimates for the original and rule-based modified intraday SVD-based exponential smoothing method used in conjunction with four different rules ……..… 153

4.3 Parameter estimates for the original and rule-based modified intraweek SVD-based exponential smoothing method used in conjunction with four different rules ……..… 154

4.4 List of special days in France observed in the year 2009, and the corresponding historical special day referred to via a selection of different cases using Rule 5 ….… 168

5.1 Time-of-use tariffs for residential consumers ……..…..… 183

5.2 Number of residential consumers allocated as per different tariffs and stimuli ……... 184

5.3 Time-of-use tariffs for SMEs …….……..…..… 185

5.4 Number of SMEs allocated as per different stimuli ……..…… 185

5.5 Model parameters (bandwidths and discount factor) for different kernel-based methods averaged across residential consumers and SMEs ……....… 196

5.6 Number of times different tariffs were selected for the residential consumers, over the whole post-sample data ……..…..……... 214

5.7 Number of consumers who achieved lower costs by: a) switching tariffs, and, b) using a single allocated tariff, and who found no cost difference between the two tariff schemes …… 217

ABBREVIATIONS

ACF	Autocorrelation Function
AIC	Akaike Information Criterion
ANN	Artificial Neural Networks
ARMA	Autoregressive Moving Average
CDF	Cumulative Density Function
CKD	Conditional Kernel Density
CKD-IC	CKD Estimation Conditional on Type of Intraday Cycle
CKD-Lag	CKD Estimation Conditional on Lagged Consumption
CKD-W	CKD Estimation Conditional on the Period of Week
CKD-WD	CKD Estimation Conditional on the Period of Week and Period of Day
CRPS	Continuous Ranked Probability Score
DGP	Data Generating Process
DWR	Discount Weighted Regression
DW-SVD	Discount Weighted Singular Value Decomposition
ECDF	Empirical Cumulative Density Function
EWR	Exponential Weighted Regression
EWPCA	Exponential Weighted Principal Component Analysis
GNP	Gross National Product
GDP	Gross Domestic Product
HWT ES	Holt-Winters-Taylor Exponential Smoothing

IC	Intraday Cycle
ICA	Independent Component Analysis
ID	Intraday
IW	Intraweek
KD	Kernel Density
KD-IC	KD Estimation based on Type of Intraday Cycle
KD-U	Unconditional Kernel Density Estimation
KD-W	Separate Kernel Density Estimation for each Period of the Week
KS	Kolmogorov-Smirnov
LL	Log-likelihood
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Square Error
MS-AR	Markov-Switching Autoregressive
MW	Mega Watt
NID	Normally and Independently Distributed
NN	Nearest Neighbour
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PDF	Probability Density Function
PIT	Probability Integral Transform
RAP	Random Analogue Prediction
RB	Rule-based

RMSE	Root Mean Square Error
RMSPE	Root Mean Square Percentage Error
SARMA	Seasonal Autoregressive Moving Average
SBC	Schwarz Bayesian Criterion
SETAR	Self-Exciting Threshold Autoregressive
SMA	Seasonal Moving Average
SME	Small and Medium-sized Enterprises
SRW	Seasonal Random Walk
SSE	Sum of Squared Errors
SVD	Singular Value Decomposition
TAR	Threshold Autoregressive
WRAP	Weighted Random Analogue Prediction

FREQUENTLY USED NOTATIONS AND SYMBOLS

The description of the notations and symbols is provided in the chapter they have been used. A description of frequently used notations and symbols is also provided below.

E	Expectation
σ	Standard deviation
μ	Mean
N	Length of time series
$\boldsymbol{\theta}$	Parameter vector
$\boldsymbol{x}$	Vector
$\boldsymbol{X}$	Matrix
$\lfloor \rfloor$	Floor operator
t	Time index
$x_t, \hat{x}_t$	Actual observation at time index *t* and its corresponding forecast
$\boldsymbol{\epsilon}_t$	Normally and independently distributed process
N_r	Number of distinct regimes
P	Probability
L	Lag operator
$\boldsymbol{d}$	Distance vector
h	Forecast horizon: number of time steps in future
m_1	Total number of half-hourly periods in the intraday seasonal cycle
m_2	Total number of half-hourly periods in the intraweek seasonal cycle

$m_3(t)$	Total number of half-hourly periods in the intrayear seasonal cycle for period t
H_0	Null hypothesis

Chapter 1

INTRODUCTION

The need for forecasts stems from the challenges of efficient planning and control. We require forecasts for well-informed policy and decision-making, be it in the field of economics, energy or biomedicine. Given the potential impact of these policies and decisions on a system and our societies, it is crucial that forecasts be accurate. We aim to develop novel forecasting methodologies in this thesis, which are both accurate and parsimonious. For macroeconomics, we focus on modelling the US gross national product (GNP) using nonlinear and nonparametric models, as discussed in Chapter 2. For energy, we focus on two different applications: (1) Generating accurate short-term forecasts for the total electricity demand (*load*) for Great Britain. (2) Modelling Irish electricity smart meter data (consumption) for both residential consumers and small and medium-sized enterprises (SMEs), using methods based on kernel density (KD) and *conditional* kernel density (CKD) estimation. We discuss short-term load forecasting in Chapters 3 and 4, while Chapter 5 describes modelling electricity smart meter data. A summary and conclusion of the thesis, along with a brief discussion on potential interesting lines of future work is provided in Chapter 6.

This chapter provides a brief introduction of the different application areas considered in this thesis, and discusses the main contributions of this study. Specifically, Sections 1.1, 1.2, 1.3 and 1.4 provide an introduction for Chapters 2, 3, 4, and 5, respectively.

1.1 Nonlinear and Nonparametric Modelling Approaches for Forecasting the US GNP

Over the years, nonlinear models have attracted a great deal of attention, as most time series arising from the real-world exhibit nonlinear behaviour (Kantz and Schreiber, 2004; Hastie *et al.*, 2009). The aim of our study is to examine nonlinear and nonparametric modelling approaches for generating accurate forecasts for the Gross National Product (GNP) time series. Nonparametric models, unlike parametric models, do not make strong assumptions about the form of the function, and rather attempt to learn the underlying functional form from the data itself. We chose this application due to its crucial role in policy-making, and the fact that macroeconomic time series exhibit regime shifts, i.e. periods of growth and recession, which makes the task of generating accurate GNP forecasts quite challenging. Note that GNP is a commonly used macroeconomic indicator.

In Chapter 2, the contribution of our study lies in proposing a novel, nonlinear and nonparametric method, called the *weighted random analogue prediction* (WRAP) method, for generating forecasts of the US GNP. The rationale of the proposed method is to identify a set of similar observed patterns in the data, and quantify the behaviour of these patterns to generate accurate forecasts. The proposed method is appealing for three reasons: (1) WRAP can provide a transition between different regimes in the GNP time series. (2) WRAP does not require strong prior assumptions about the form of the data generating process and provides sufficient flexibility to generate forecast distributions with properties such as *skewness* (asymmetric), *fat tails* (long-tailed distribution) and *multimodality* (multiple modes). (3) WRAP is based on only two parameters, which allows us to estimate model parameters based on the in-sample density forecast performance without incurring huge computational costs. The out-of-sample forecasting

ability of WRAP is evaluated by employing a range of different performance scores, which measure its accuracy in generating both point and density forecasts.

Although we demonstrate the applicability of WRAP using the GNP time series, this method can be easily adapted to different applications. In addition to using US GNP, we investigated the efficacy of WRAP in modelling the UK Gross Domestic Product (GDP) time series. Furthermore, we adapt WRAP to generate decadal forecasts (for lead times varying from one year up to ten years ahead) using the global surface air temperature (anomaly) time series. We found WRAP to be competitive with the decadal climate prediction system (DePreSys), a global climate model (GCM) based on the Hadley centre coupled model, on both point and density forecast accuracy.

1.2 Short-term Load Forecasting using Discount Weighted SVD-based Methods

We discuss the challenges in modelling load and propose novel methodologies for short-term load forecasting in Chapters 3 and 4. Short-term load forecasts correspond to lead times varying from one half-hour up to one day ahead. Accurate load forecasts are crucial for safe and efficient operations of a power system. Also, it has been shown by Bunn (2000) that inaccurate short-term load forecasts can have substantial financial implications for energy markets. The load time series exhibits a prominent intraday, intraweek and intrayear seasonality. However, most existing studies accommodate the 'double seasonality' while modelling short-term load, focussing only on the intraday and intraweek seasonal effects (Kim *et al*., 2000; Fidalgo and Lopes, 2005). It has been shown recently by Taylor (2010a) that accommodating the intrayear seasonality in load, along with the intraday and intraweek seasonal cycles, leads to an improvement in short-term

load forecast accuracy. The models proposed in Chapters 3 and 4 are aimed at accommodating the 'triple seasonality' in load time series.

In Chapter 3, the contribution of our study lies in proposing two novel methods for load forecasting, based on a commonly used statistical *dimension reduction* technique called singular value decomposition (SVD). Specifically, we propose discount weighted (DW) intraday and DW intraweek SVD-based exponential smoothing methods. Some of the existing methodologies for load forecasting rely on building separate models for different periods of the seasonal cycle (Ramanathan *et al*., 1997; Cottet and Smith, 2003; Cancelo *et al*., 2008). This results in a high-dimensional model, which is inefficient and less parsimonious. To overcome this limitation, SVD-based methods have been proposed for forecasting the intraday time series by Shen and Huang (2005, 2008a, 2008b), Taylor *et al*. (2006), and Taylor (2010b, 2012). The advantage of using SVD is that it transforms the data to an orthogonal space, and allows the focus to be on forecasting only the components that capture a major proportion of variance in the data, thereby reducing the dimensionality of the model to be considered. A limitation with the existing SVD-based methods is that they do not employ discounting within the SVD framework, and ignore the intrayear seasonal effects while modelling load. The incorporation of discounting within the modelling framework allows more emphasis to be placed on more recent observations.

We extend the SVD-based methods used by Taylor (2010b, 2012), and propose two DW SVD-based methods, such that the new methods: (1) Achieve dimension reduction using a procedure in which more recent observations are assigned more weights through the use of discount factors. (2) Accommodate triple seasonality in the existing model formulation of intraday and intraweek SVD-based exponential smoothing methods for the load data. (3) Adapt with the inclusion of new observations through sequential updating.

We show that the proposed methods provide a reduction in the dimensionality of the model, and are competitive with some of most commonly used models for short-term load forecasting.

1.3 Short-term Forecasting of Anomalous Load using Rule-based Methods

In Chapter 4, the contribution of our study lies in proposing a novel rule-based approach for forecasting short-term load, with emphasis on special days. The load observed on special days, e.g. public holidays and long weekends, is referred to as anomalous load. Special days are associated with a substantially lower load compared to normal working days. Although the forecast accuracy of conventional statistical methods have been shown to be competitive overall, their performance under anomalous load conditions has been less than satisfactory. Forecasting of anomalous load has often been ignored during the modelling process (Smith, 2000; Hippert *et al*., 2005), and has been left to the expert judgement of the central controller of the electricity grid.

Most studies on short-term forecasting of normal and anomalous load have employed regression-based methods with dummy variables for special days (see Cancelo *et al*., 2008; Soares and Medeiros 2008; and Dordonnat *et al*., 2008). Specifically, these regression-based models incorporate dummy variables for identifying the special days, and different periods of the seasonal cycle. To include a correction coefficient with each dummy variable results in the final model being highly parameterized and difficult to interpret. Due to this reason, different special days have often been classified as being the same in regression-based modelling approaches. We note that this is a very restrictive assumption, as inspection of the data reveals that different special days exhibit different load profiles, which may change over the years.

To model load for special days, we propose a rule-based approach. The rationale for using a rule-based approach is that it allows the incorporation of prior expert knowledge into the statistical modelling framework.

The contribution of this study lies in proposing a novel rule-based approach for modelling normal and special days in a coherent and unified framework, whereby we: (1) Formulate the rules subjectively by inferring them directly from the data, such that each special day is allowed to have a unique profile. (2) Adapt the conventional univariate methods, namely: triple seasonal Holt-Winters-Taylor (HWT) exponential smoothing, triple seasonal autoregressive moving average (ARMA), artificial neural networks (ANNs), and triple seasonal variants of the intraday and intraweek SVD-based exponential smoothing methods, to model load for normal and special days, by using them in conjunction with the formulated rules. (3) Generate forecasts using the rule-based methods, and provide a comparison of the proposed methods with a range of different benchmarks, across both normal and special days.

We emphasize that the formulated rules can easily be adapted for different load time series. In this study, we adapt the rules, developed using load for Great Britain, to model normal and anomalous load for France. We report that the mean absolute percentage error (MAPE) results for the rule-based methods (for the best performing rules, across special days) were about half the size of those for the corresponding original models that do not accommodate special day effects, for both Great Britain and French load time series.

Although we demonstrate the applicability of the rule-based methods for load forecasting, the proposed methodology can easily be adapted for other applications. Some examples where this approach could potentially be useful for forecasting includes call centre arrivals, hospital

admissions, and transportation counts, as the corresponding time series exhibit seasonality and anomalous conditions pose significant modelling challenges.

1.4 Modelling Electricity Smart Meter Data Based on Kernel Density Estimation

In Chapter 5, the contribution of our study lies in generating accurate density forecasts for electricity smart meter data. This chapter investigates the efficacy of different methods based on KD and CKD estimation for modelling electricity consumption, recorded from smart meters. A smart meter is an electronic device that measures electricity consumption at the installed facility, and transmits this information to the consumer and the energy supplier/operator on a near real-time basis. We choose this application as future estimates of consumption can help the consumer identify and reduce excess consumption, while such estimates can also help the supplier devise innovative time-of-use (TOU) tariffs aimed at their target consumers.

Most existing studies on smart meter data focus on estimating the impact of TOU pricing strategies on electricity consumption. To the best of our knowledge, there are no studies which focus on generating density estimates of electricity consumption from smart meter data.

The contribution of this study lies in accurate modelling of electricity smart meter data, whereby we: (1) Adapt different KD and CKD methods to generate forecasts of electricity consumption, for both residential consumers and SMEs. (2) Use discount factors within the density estimation framework, to allow more emphasis to be placed on more recent observations. (3) Evaluate density, quantile and point forecasts of electricity consumption using a range of different performance scores. (4) Derive prediction intervals for electricity cost based on density

forecasts of electricity consumption, for different available tariffs, and identify the tariff that would incur the lowest cost at the consumer's end. (5) Investigate if it is more cost effective for the consumer to: a) switch between different available tariffs, or b) use a single pre-allocated tariff at all times.

The methods considered in this study capture the seasonality in consumption, and enable a nonlinear and nonparametric estimation of its density. We show that the KD and CKD methods accommodate the seasonality in consumption, and correctly distinguish weekdays from weekends.

Although we demonstrate the applicability of the different KD and CKD methods for modelling consumption, the data-driven methods proposed in this study for smart electricity meter data can potentially be used for modelling data from smart gas or water meters.

Chapter 2

NONLINEAR AND NONPARAMETRIC MODELLING APPROACHES FOR PROBABILISTIC FORECASTING OF THE US GROSS NATIONAL PRODUCT

2.1 Introduction

The Gross National Product (GNP) of a country reflects the total value of all goods and services produced by its nationals. It is a macroeconomic measure that provides an overall estimate of the economic health of a country. A range of parametric models have been employed for modelling the GNP. The aim of this study is to investigate the efficacy of simple nonlinear and nonparametric methods for probabilistic forecasting of US GNP.

Classical parametric modelling approaches model the dependent variable using a set of independent variables (predictors or explanatory variables) by making certain assumptions about the true functional form f of the Data Generating Process (DGP). Nonparametric models, unlike parametric models, do not make strong assumptions about the form of the function, and rather attempt to learn the function from the data itself. It is, however, assumed that f is a smooth, continuous function. Nonlinearity is defined as deviations from the classical Gauss-Markov assumptions (linearity, homogeneity and independence). In this section, we discuss the motivation and challenges in modelling macroeconomic time series, the rationale for using a nonparametric modelling approach for GNP, and provide the contributions of this study.

2.1.1 Motivation and Challenges

Accurate estimates of future trends in GNP are important for informed decision-making and government policy-making. An accurate GNP forecast facilitates the smooth functioning of an economy through timely changes in policy, providing improved decision-making and leading to greater stability of financial markets. The dynamics underlying GNP time series is complex, and we only have a few numbers of quarterly post-war observations, which makes the task of generating reliable economic forecasts extremely challenging. Numerous time series models are available for forecasting economic output. Autoregressive models were initially applied to US GNP, and have been extended to nonlinear structures, such as the Self-exciting Threshold Autoregressive (SETAR) and Markov-switching Autoregressive (MS-AR) models (Tong and Lim, 1980; Hamilton, 1989; Terasvirta and Anderson, 1992; Tiao and Tsay, 1994; Potter, 1995). We show that while these nonlinear models fit the in-sample data better than linear models, the out-of-sample forecast performance of extremely simple methods, such as the unconditional mean is competitive compared with both previously published linear and nonlinear models for GNP time series. Motivated by Occam's razor and the forecasting competitiveness of the unconditional mean, we seek parsimonious methods. It is noteworthy that Occam's razor is a principle of parsimony, which states that: "Among competing hypotheses, the hypothesis with the fewest assumptions should be selected." In the context of this work, the principle of parsimony motivates the use of models which are based on simple assumptions, incorporate few parameters, and can generate accurate forecasts. We propose a nonlinear and nonparametric method based on nearest neighbour and kernel regression. The nonparametric methods employed in this study incorporate only a few parameters, and the model estimation framework relies on optimizing the performance of in-sample density forecasts. Model estimation framework based

on the in-sample density forecast performance has also been recently employed by Jeon and Taylor (2012).

One of the major limitations in modelling macroeconomic time series is the presence of regimes in the DGP. Note that some systems are also known to exhibit regime shifts, whereby the DGP switches from one plausible regime to the other with time. For example, the economy is characterized by the presence of two regimes, namely growth and recession, while financial variables are known to exhibit regimes of high and low volatility. This phenomenon of shifts in regime requires that the DGP is modelled using two or more sub-models, perhaps having the same model structure but a different set of model parameters (see, for example, Tong, 1978; Tong and Lim, 1980; Hamilton, 1989; Clements *et al*., 2004). To overcome these challenges, we propose a nonparametric modelling approach that can provide a transition between regimes of growth and contraction, without making prior assumptions about the presence or number of regimes.

2.1.2 Historical Review

A diverse range of time series models are available for generating GNP forecasts. Nonlinear regime switching models have become extremely popular in contemporary econometrics (Terasvirta and Anderson, 1992; Potter, 1995; Krolzig, 2001; Clements *et al*., 2003; Clements *et al*., 2004), providing alternatives to Box and Jenkins (1970) class of linear models. Regime switching models like the SETAR models (Tiao and Tsay, 1994; Potter, 1995), and MS-AR models (Hamilton, 1989) have been employed for modelling US GNP. These models are structured to be able to characterize regimes of growth and recession in a business cycle. There remains no clear consensus on whether nonlinear models are better than their linear counterparts

from the perspective of out-of-sample forecast performance, even when linearity is rejected statistically (Diebold and Nason, 1990; De Gooijer and Kumar, 1992). Similar results were reported by Swanson and White (1995, 1997) who compared the out-of-sample point forecast accuracy of multivariate adaptive linear and nonlinear models (using a range of different macroeconomic variables), and reported that linear models often outperformed nonlinear models in simulated real time. Note a univariate model employs only historical GNP observations for generating forecasts, while a multivariate model uses different macroeconomic variables in addition to the historical GNP observations. Following Hamilton (1989), Potter (1995), Clements and Krolzig (1998), we employ univariate models in this study. Stock and Watson (1999) found linear models to be more accurate than nonlinear models for forecasting a range of different macroeconomic time series, while Dueker and Neely (2007) reported that a simple benchmark generated more accurate point forecasts compared to a nonlinear regime-switching model.

The lack of concrete evidence in support of nonlinear regime-switching models over linear models as superior forecasting tools has led to considerable amounts of research into the structural conditions and performance scores under which one class of models may fare better than the other. In this context, nonlinearity is defined based on the presence of different regimes in the DGP. Note that GNP follows an asymmetric cyclical process, giving rise to different regimes (growth and recession), with growth periods being much longer than recessionary periods, see Terasvirta and Anderson (1992). As defined in Granger and Terasvirta (1993), a business cycle is asymmetric if the average distance from peaks and troughs in the time series, is not equal to the average distance from troughs to peaks. Terasvirta (1995) found evidence of nonlinearity in the long annual GNP, and used the Exponential Smooth Transition Autoregressive (ESTAR) model to characterize business cycles in the time series. Terasvirta and

Anderson (1992) argued that regime-switching models may perform better than the linear models, if there are a sufficient number of observations in each separate regime of the time series. This implies that regime-switching models, like SETAR and MS-AR, are in general better able to characterize the regime switching events underlying the time series. As also pointed out by Clements and Krolzig (1998), the tendency of an economy to return quickly to a mean level from a negative shock/recession, is a prominent nonlinear switching behaviour. Hence, a time series characterized by a substantial number of observations in the recessionary regime would presumably be better modelled by nonlinear regime-switching models, compared to the classical linear models. To translate this property of regime switching models into enhanced forecast accuracy requires that the out-of-sample and in-sample data exhibit some minimum degree of switching nonlinearity. Hansen (1992) and Hansen (1996b) found weak evidence of rejecting linear models in favour of nonlinear models for quarterly GNP time series. Potter (1995) compared the error variance and forecast accuracy of nonlinear and linear models and reported nonlinearity in US GNP time series, while Brock and Sayers (1988) found little evidence of nonlinearity in real US GNP. Using GNP time series, Clements *et al.* (2003) focused on measuring the degree of nonlinearity (based on a significance level at which test of null of linearity was rejected for heteroscedasticity-robust and standard Hansen tests) under which nonlinear models may be superior to linear models. Panagiotidis (2002) employed various tests for nonlinearity for the US and Canadian macroeconomic time series and reported nonlinearity in the Canadian time series, while the results for US were mixed based on the type of test employed. Marcellino *et al.* (2003) compared forecasting accuracy of univariate and multivariate models using macroeconomic data, and reported that pooled univariate AR models were competitive with multivariate models. Ocal (2000) employed threshold autoregressive models

for UK macroeconomic time series and reported that they were not considerably superior to their linear counterparts. Morley and Piger (2004) argued that reproducing the features of a business cycle requires a nonlinear approach. Engel *et al.* (2005) demonstrated that regime-switching models explain the shape and variability of a business cycle better than linear models, but do so at the expense of a very strong rebound effect, which makes expansions continue for longer durations than they do in reality. Though parametric regime switching models (SETAR and MS-AR) have shown to be useful for reproducing business cycles and analysing macroeconomic time series, we aim to investigate if these models can generate more accurate point and density GNP forecasts on the out-of-sample data compared to nonparametric modelling approaches that incorporate fewer parameters. Using data for inflation and GDP, Kapetanios *et al*. (2007) showed that point forecasts from MS-AR and SETAR are slightly more accurate than AR. In recent studies, Buchmann and Hubrich (2010) compare the out-of-sample density performance of AR, vector AR, and MS-AR using the GDP time series, for horizons ranging from one to four quarters ahead. They evaluate density forecasts, and find that MS-AR offers improvement in forecast accuracy over linear modelling approaches. Using GNP time series, we find that density forecasts from MS-AR are competitive with both AR and SETAR, however, none of these models outperform an unconditional forecast.

2.1.3 Contributions

The contributions reported in this thesis relate to accurate forecasts for US GNP for horizons ranging from one to sixteen quarters (four years), based on a novel nonlinear and nonparametric model. We show that our model outperforms previously published linear and nonlinear models proposed for forecasting US GNP, including simple benchmarks. We provide a systematic

comparison between classical linear and nonlinear parametric models using a range of performance scores.

Specifically, the contributions of this study are:

1. Systematic comparison of out-of-sample point and density forecast performance of classical linear and nonlinear models using a range of performance scores,

2. Proposing a parsimonious nonlinear and nonparametric model for generating accurate forecasts that can be easily generalized to different forecasting applications, and,

3. Generating forecasts for the US GNP using the proposed model, which convincingly outperform some of the most commonly used models for forecasting this time series. The proposed model incorporates less than half of the total parameters used in classical models.

2.2 GNP Characteristics

The GNP data comprises seasonally adjusted quarterly prices of real US GNP (in billions of dollars). As analysed in Potter (1995), the shorter time series measures US GNP at 1982 dollar prices dating from 1947Q1 to 1990Q4[1]. The longer, more recent time series contains US GNP recordings from 1947Q1 to 2008Q3 at 2000 dollar prices[2]. The first GNP series is downloaded from the online data archive of the Journal of Applied Econometrics, while the second time series was obtained from the Federal Reserve Economic Data II (FRED II) affiliated with the Federal Reserve Bank of St. Louis. The two time series are different in terms of the number of quarterly GNP observations and the value of dollar prices at which they have been measured.

[1] http://econ.queensu.ca/jae/1995-v10.2/potter/potter.data
[2] http://research.stlouisfed.org/fred2/series/GNPC96

Our results were found to be consistent for both the GNP time series; however, for the purpose of demonstration, we present results corresponding to the longer time series in this thesis.

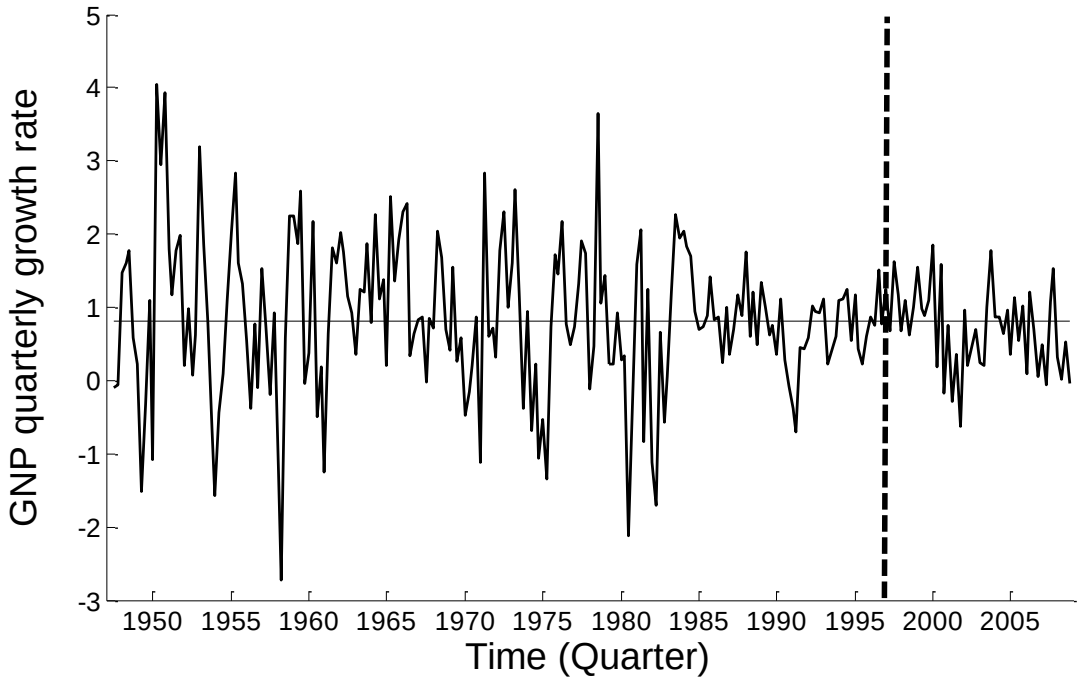

Figure 2.1— Plot of percentage log returns of the quarterly US GNP from period 1947Q2-2008Q3. The horizontal line corresponds to the unconditional mean of the whole time series, while the vertical dashed line denotes the partition between the training and testing set.

We estimate models on the log of quarterly growth rate (first differences) of the GNP (as shown in Figure 2.1), i.e. $x_t = 100 \log(y_t/y_{t-1})$, where y_t is the actual GNP value at time instant t. This transformation is based on the assumption that raw GNP quarterly observations are generated from a non-stationary process. Following Clements and Krolzig (1998), we multiply the quarterly returns by 100, in order to obtain percentage returns.

2.3 Parametric Modelling Approaches

This section describes commonly used linear and nonlinear parametric models for modelling GNP time series, namely the Auto-Regressive (AR), Self-Exciting Auto-Regressive (SETAR) and Markov-Switching Auto-Regressive (MS-AR) models, and discusses the process of parameter estimation for these models.

2.3.1 Autoregressive Models

The AR model is a linear model that belongs to the Box and Jenkins (1970) class of models. This model characterizes the time series assuming a linear relationship between observations. An AR model of order p is denoted as AR(p) and can be represented using the following form:

$$x_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \, x_{t-i} + \varepsilon_t, \qquad where \;\; \varepsilon_t \sim \text{NID}(0, \sigma^2) \tag{2.1}$$

where x_t is an observation in the time series at time t, α_0 is a constant, α_i denotes AR model parameters for $i = 1, 2, \ldots, p$, and ε_t is an Normally and Independently Distributed (NID) process with mean zero and variance σ^2. Note that p corresponds to the number of lagged variables in this model. The model coefficients can be estimated using ordinary least squares (OLS). We use Akaike Information Criterion (AIC) proposed by Akaike (1973) to select the model order. The AIC is represented as:

$$AIC(p, N) = \frac{1}{N_1} RSS + 2 \, \hat{\sigma}^2 \frac{p}{N_1}, \tag{2.2}$$

where residual sum of squares, $RSS = \sum_{i=1}^{N_1} e_i^2$, such that e_i (residual) is the difference between the actual response and the model response, $\hat{\sigma}^2$ is the variance of the residuals, p is the model order, N is the total number of observations used for training, while $N_1 = N - p$. The first term

on the right hand side of equation 2.2 measures the goodness of model fit, while the second term imposes a penalty on the number of free parameters employed in achieving that fit. If the model is over-parameterized, the RSS would be low, but the penalty term would be high, and vice-versa for an under-parameterized model. The model with the smallest AIC is selected as the best model from a set of candidate models for a given dataset. Note that the absolute value of AIC does not convey any meaningful information; and only relative values are meaningful when comparing competing models. Another popular criterion is the Schwarz Bayesian Criterion (SBC) proposed by Schwarz (1978), which is similar in essence to the AIC, with a slight difference that SBC imposes a stricter penalty on the inclusion of free parameters. Empirical forms of IC (EIC) provide a data-driven method for model selection. It has been shown by Billah *et al*. (2006) and Taylor (2008) that model selection based on EIC can lead to more accurate forecasts, compared to the case when models are selected using either AIC or SBC. The estimates of AR model parameters (obtained using OLS) for the in-sample period 1947Q2-1996Q4 are presented in Table 2.1. The value of AR model order was chosen as four using AIC.

TABLE 2.1

PARAMETER ESTIMATES FOR THE AR(4) MODEL AND CORRESPONDING STANDARD ERRORS (IN PARENTHESIS)

Parameter	**Estimate (Standard Error)**
α_0	0.6499 (0.1132)
α_1	0.3120 (0.0719)
α_2	0.1284 (0.0751)
α_3	-0.0832 (0.0751)
α_4	-0.1259 (0.0716)
σ	0.9624
Total Parameters	5

2.3.2 Self-exciting Threshold Autoregressive (SETAR) Models

The presence of structural shifts may lead to distinct regimes in the time series. The rationale for using regime switching models, like the SETAR model, is to try and model each regime separately. Hence, in regime switching models, each distinct regime is characterized by a unique parameter vector. Note that classical linear models do not assume the presence of multiple regimes in the DGP.

The challenges of using regime-switching models are:

a) Finding the evidence in favour of using regime-switching models over linear models,
b) Estimating the number of regimes in the time series,
c) Estimating the model parameters for each separate regime, and,
d) Identifying the times at which different regimes apply, and being able to switch between regimes at the right times.

In this section, we focus on understanding the assumptions and applications of a commonly used regime switching model, namely the SETAR model.

SETAR models are a subclass of the Threshold Autoregressive (TAR) models. The TAR models were proposed by Tong (1978) and further discussed by Tong and Lim (1980). The TAR model assigns each observation to a particular regime, and builds a linear model for the regime. The choice of a regime is based on the value of a threshold variable, whereby the role of the threshold variable is to force the model to switch between different regime models. This makes TAR a *piecewise linear AR model*. When the threshold variable is an observation in the time series, the TAR model is described as 'self-exciting', hence the name SETAR.

A SETAR model composed of N_r regimes is denoted as SETAR $(N_r; p_1, p_2, \dots, p_{N_r})$, and is represented as:

$$x_t = \alpha_{0j} + \sum_{i=1}^{p_j} \alpha_{ij}\, x_{t-i} + \varepsilon_{tj}\,, \qquad where \;\; \varepsilon_{tj} \sim \text{NID}(0, \sigma_j^2) \tag{2.3}$$

where α_{ij} is the ith autoregressive coefficient for a given regime index j that obeys $r_{j-1} \leq x_{t-d} < r_j$, d is the delay order, while r_j for $j = 1, 2, \dots, N_r - 1$ are the *thresholds* that divide the time series into N_r different regimes, p_j is the AR model order for regime j, and ε_{tj} is a NID process with mean zero and variance σ_j^2. The values of thresholds are chosen from observations in the time series, i.e. thresholds are restricted to be *endogenous*, and the choice of AR model coefficients depends on the thresholds r_j and the *delay order* d.

The above general representation of the SETAR model for N_r separate regimes specializes to the two-regime model, given as:

$$x_t = \begin{cases} \alpha_{01} + \sum_{i=1}^{p_1} \alpha_{i1}\, x_{t-i} + \varepsilon_{t1}\,, & if \;\; x_{t-d} \leq r \qquad where \;\; \varepsilon_{t1} \sim \text{NID}(0, \sigma_1^2) \\ \alpha_{02} + \sum_{i=1}^{p_2} \alpha_{i2}\, x_{t-i} + \varepsilon_{t2}\,, & if \;\; x_{t-d} > r \qquad where \;\; \varepsilon_{t2} \sim \text{NID}(0, \sigma_2^2) \end{cases} \tag{2.4}$$

Each regime is governed by an AR model, and the models are separated by the lagged observation value x_{t-d} and the value of the threshold r. The magnitude of delay is given by the delay order d. From the perspective of economic modelling, the condition $x_{t-d} \leq r$ may correspond to a regime of recession, while $x_{t-d} > r$ may correspond to the regime of growth. For an epidemiological time series, the same conditions may represent inhibition and outbreak of an epidemic (Watier and Richardson, 1995).

As evident from equation 2.4, the SETAR model requires the estimation of thresholds r_j, delay order d, the autoregressive model parameters α_{ij}, and the model orders p_j. The problem of estimating r_j and d has been addressed in Tong (1983), who suggested a grid search method. The idea is to vary r_j and d over a set of possible values, and select a particular set of parameter values that minimizes the overall RSS. The thresholds r_j are varied so as to take lagged values of the observed time series, while d is varied from one to some maximum delay order d_{max}. Hansen (1996b) estimated a SETAR model by allowing the delay order to take values of one, two, and five, while the threshold was varied in the range of the 15th and 85th percentiles of distribution of the time series. Given the delay order and threshold, the model order p_j is varied across regimes, and a corresponding AR model for each regime estimated. As pointed out by Potter (1995), the estimation of AR model parameters can be undertaken by running a least squares regression on each regime, for a given value of r_j and d. Assuming a bi-regime SETAR model, the estimated parameter set is chosen from the in-sample time series that minimizes the following term (used for SETAR model by Clements and Krolzig, 1998):

$$N_L \ln\hat{\sigma}_L^2 + N_U \ln \hat{\sigma}_U^2 + 2(p_L + 1) + 2(p_U + 1), \tag{2.5}$$

where N_L and N_U are number of observations in the lower and upper regime, $\hat{\sigma}_L^2$ and $\hat{\sigma}_U^2$ are the corresponding variances of the residuals, while p_L and p_U are the model orders within each regime. The above term can be viewed as an IC for regime switching models, whereby the first two terms give an estimate of the goodness of model fit, while the last two terms impose a penalty on the model complexity in each regime. Restricting the range of threshold values to lie within the 15th and 85th percentile of the time series distribution ensures that each regime always

has a certain minimum number of observations during the process of parameter estimation. The problem of estimating the number of regimes N_r has been discussed in detail by Hansen (1999).

Using only the in-sample time series for model estimation, we find that for the shorter GNP time series a SETAR(2;2,2) model is suitable, while a SETAR(2;1,1) model is appropriate for the longer, more recent GNP time series. We estimate a SETAR model using the value of delay length $d = 2$, and restrict the model order to be same for both regimes, as assumed earlier for US GNP by Tiao and Tsay (1994) and Clements and Krolzig (1998). From the evaluation of a SETAR model for the two GNP time series, we find that the majority of observations (78%) fall in the upper regime (growth regime) for the shorter time series, while about 74% of the observations belong to the upper regime for the second case. The optimum estimate of threshold model r for shorter time series is -0.0580 using SETAR(2;2,2) (as previously reported by Clements and Krolzig, 1998), while for the longer time series we obtain $r = 0.2808$. In this study, we estimated the model parameters by restricting the range of the threshold to lie within 15th and 85th percentile of the in-sample data distribution, as used earlier by Hansen (1996b) and Clements and Krolzig (1998). The results for parameter estimates are presented in Table 2.2.

SETAR models have been employed for modelling of US GNP (see Tiao and Tsay, 1994; Potter, 1995), epidemiological time series (Watier and Richardson, 1995), foreign exchanges (Boero and Marrocu, 2002), and Canadian Gross Domestic Product (GDP), see Feng and Liu (2002). Potter (1995) validated the rationale for employing SETAR over AR, by comparing the goodness of in-sample fit of SETAR with AR using the US GNP, and reported that the error variance of SETAR was lower than that of AR. To ensure that the reduction in error variance was not due to SETAR over-fitting the data, Potter (1995) also compared the out-of-sample forecasts from the two models, and found SETAR to generate more accurate forecasts than the

AR model. Tiao and Tsay (1994) compared the forecast accuracy of SETAR and AR models using US GNP time series and reported that the performance of SETAR compared to AR improves if forecasts are evaluated separately when conditioned on the regime. Similar results were reported later by Boero and Marrocu (2002) for Euro effective exchange rate. Note that if the majority of observations lie in one particular regime, then the model evaluation is dominated by the goodness of fit in that regime.

Evaluating the model conditioned on the regimes ensures that each regime is treated equally, irrespective of the number of observations lying in that regime. If most of the observations in a time series belong to a single regime, then fitting a linear AR model across the whole time series is similar to fitting a SETAR model across regimes. Hence, typically there would be no substantial improvement in the forecast accuracy when SETAR is preferred over AR for a time series mostly comprising observations belonging to a particular regime. For this reason, both in-sample and out-of-sample data should consist of a substantial number of observations belonging to different regimes to allow SETAR to produce more accurate forecasts compared to a global AR model.

Clements and Smith (1999) pointed out that the results of Tiao and Tsay (1994) were misleading because they used information from the whole time series, e.g. they did not report out-of-sample results. The major limitations of SETAR model are that the forecast accuracy is quite sensitive to the parameter estimation error, and the goodness of in-sample fit does not necessarily translate into the superior out-of-sample forecast, as pointed by Clements and Krolzig (1998).

TABLE 2.2

PARAMETER ESTIMATES FOR SETAR(2;1,1) MODEL AND CORRESPONDING STANDARD ERRORS (IN PARENTHESIS)

Parameter	Estimate (Standard Error)
Lower Regime	
$\alpha_{0\mathrm{L}}$	0.4507 (0.1803)
$\alpha_{1\mathrm{L}}$	0.2806 (0.1506)
σ_{L}	1.2359
N_{L}	52
Upper Regime	
$\alpha_{0\mathrm{U}}$	0.6166 (0.1081)
$\alpha_{1\mathrm{U}}$	0.3399 (0.0769)
σ_{U}	0.8952
N_{U}	145
r	0.2808
d	2
Total Parameters	8

2.3.3 Markov-switching Autoregressive (MS-AR) Models

There are two main approaches to modelling regime shifts in a time series. One approach, as adopted in SETAR, is to assume that the underlying law governing a regime shift can be captured through an endogenous threshold variable, along with past realizations of the time series. The second approach, as adopted in MS-AR models, is to model the cycles of growth and recession based on a regime variable that is latent (not observable), and that the evolution of the regime variable is governed by a Markov chain. The choice of model coefficients at any given instant of time depends on the regime variable. Note that both SETAR and MS-AR are nonlinear

models, i.e. they deviate from the classical Gauss-Markov assumptions, and can be viewed as nonlinear extensions of the linear AR model.

Some of the early contributions include the work of Goldfeld and Quandt (1973) on Markov-switching regression, which was later extended by Hamilton (1989) to model the returns (first-order differences) of quarterly US GNP time series. The major accomplishment of the model proposed by Hamilton (1989) was that it captured the periods of growth and recession in the business cycle as established by the National Bureau of Economic Research (NBER) with great accuracy.

Hamilton (1989) proposed a bi-regime, MS-AR model for US GNP. The bi-regime Markov-switching model with AR order *p* can be denoted as MS(2)-AR(p), and is defined as:

$$x_t = \mu(s_t) + \sum_{i=1}^{p} \alpha_i (x_{t-i} - \mu(s_{t-i})) + \varepsilon_t, \qquad where \quad \varepsilon_t \sim \text{NID}(0, \sigma^2) \tag{2.6}$$

where x_t is an observation in the time series with time index t, α_i are the AR coefficients for $i = 1, 2, \dots, p$, and s_t is a regime variable of the system such that, $s_t = 1$ corresponds to growth and $s_t = 2$ corresponds to recession. The mean $\mu(s_t)$ switches between the two regimes, such that $\mu(s_t)$ is positive if $s_t = 1$, and negative if $s_t = 2$. The transition between the two regimes depends on the variable s_t, governed by a first order Markov chain with transition probabilities *P* given as:

$$P_{rj} = P(s_{t+1} = j \mid s_t = r), \tag{2.7}$$

for all $r, j \in \{1, 2\}$ and $P_{r1} + P_{r2} = 1$. From the above representation of the model (equation 2.6), it can be inferred that MS(2)-AR(p) assumes that the switching between the two distinct regimes is governed by the evolution of s_t.

For the Hamilton's (1989) bi-regime MS(2)-AR(p) model, the parameter vector that needs to be estimated is $\boldsymbol{\theta} = (\mu(1), \mu(2), \alpha_1, \alpha_2, \dots, \alpha_p, \sigma, P_{11}, P_{22})$. The parameter vector $\boldsymbol{\theta}$ is chosen so as to maximize the likelihood of the observations. The estimation of this parameter vector has been discussed in detail by Hamilton (1990) and Krolzig (1997), based on maximizing the likelihood using expectation maximization. Assuming a normal distribution for residuals, the conditional probability density function for the time series x_t given state variables s_t and s_{t-1} and the previous observation vector $\boldsymbol{x}_{t-1} = (x_1, x_2, \dots, x_{t-1})$, is defined as:

$$f(x_t|\, s_t, s_{t-1}, \boldsymbol{x}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ - \frac{(x_t - \mu(s_t) - \sum_{i=1}^{p} \alpha_i (x_{t-i} - \mu(s_{t-i})))^2}{2\sigma^2} \right\}. \tag{2.8}$$

Note from equation 2.6 that $\varepsilon_t = (x_t - \mu(s_t) - \sum_{i=1}^{p} \alpha_i (x_{t-i} - \mu(s_{t-i}))) \sim \text{NID}(0, \sigma^2)$. We need to find the parameter vector $\boldsymbol{\theta}$ that maximizes the likelihood $L(\theta)$ given by:

$$L(\theta) = \sum_{t=2}^{T} f(x_t|\boldsymbol{x}_{t-1}), \tag{2.9}$$

where T is the forecast origin. The conditional density of the observation x_t given the observation vector $\boldsymbol{x}_{t-1}$, can be written in terms of the state transition probabilities as,

$$f(x_t|\boldsymbol{x}_{t-1}) = \sum_{s_t = 1}^{2} \sum_{s_{t-1} = 1}^{2} f(x_t|s_t, s_{t-1}, \boldsymbol{x}_{t-1})\, p(s_t, s_{t-1}|\boldsymbol{x}_{t-1}). \tag{2.10}$$

Using the chain rule for conditional probabilities, $P(A \cap B|C) = P(A|B \cap C)P(B|C)$, equation 2.10 can be written as:

$$f(x_t|\boldsymbol{x}_{t-1}) = \sum_{s_t=1}^{2}\sum_{s_{t-1}=1}^{2} f(x_t|s_t, s_{t-1}, \boldsymbol{x}_{t-1})\, p(s_t|s_{t-1}, \boldsymbol{x}_{t-1}) p(s_{t-1}|\boldsymbol{x}_{t-1}),$$

$$= \sum_{s_t=1}^{2}\sum_{s_{t-1}=1}^{2} f(x_t|s_t, s_{t-1}, \boldsymbol{x}_{t-1})\, p(s_t|s_{t-1}) p(s_{t-1}|\boldsymbol{x}_{t-1}), \quad (2.11)$$

as $p(s_t|s_{t-1}, \boldsymbol{x}_{t-1}) = p(s_t|s_{t-1})$, follows from the property of Markov chain. The basic idea is to make an inference about state variables at time t, based on information available up to time t-1. Once the observation x_t becomes available, we need to update the observation vector $\boldsymbol{x}_t = \{x_t, \boldsymbol{x}_{t-1}\}$, and the conditional probability $p(s_t, s_{t-1}|\boldsymbol{x}_t)$. Note that the term $f(x_t|s_t, s_{t-1}, \boldsymbol{x}_{t-1})$ (as used in equation 2.11) can be obtained from equation 2.8. Hence, the problem of computing the likelihood reduces to estimating the transitional probabilities $p(s_t|s_{t-1})$ and multiplying them with the conditional probability $p(s_{t-1}|\boldsymbol{x}_{t-1})$ of the state variable. At the first iteration, we initialize $P(s_1 = 1|\, \boldsymbol{x}_1) = P(s_1 = 1) =$ 0.5. The conditional probability is computed and updated at each time step with the inclusion of an extra observation in order to iteratively maximize the log-likelihood function. The final procedure of updating the conditional joint probability for state variables can be formulated as follows:

$$p(s_t, s_{t-1}|\boldsymbol{x}_t) = p(s_t, s_{t-1}|x_t, \boldsymbol{x}_{t-1}), \quad (2.12)$$

$$= \frac{p(x_t, s_t, s_{t-1}|\boldsymbol{x}_{t-1})}{f(x_t|\boldsymbol{x}_{t-1})},$$

$$= \frac{p(x_t|s_t, s_{t-1}, \boldsymbol{x}_{t-1}) p(s_t, s_{t-1}|\boldsymbol{x}_{t-1})}{f(x_t|\boldsymbol{x}_{t-1})}.$$

We find the estimates for Hamilton's MS(2)-AR(p) model for the GNP time series, as shown in Table 2.3. Based on AIC, we find that the optimum AR model order p for GNP time series (using only in-sample observations) is five. The parameter estimates for the shorter GNP time

series were found to be in accordance with the MS-AR parameters reported earlier for GNP by Clements and Krolzig (1998).

For simplicity, we assume that the residuals are normally distributed in this case, though one can obtain parameter estimates for the MS-AR models using a range of different distributions using the above procedure.

TABLE 2.3

PARAMETER ESTIMATES FOR MS(2)-AR(5) MODEL AND CORRESPONDING STANDARD ERRORS (IN PARENTHESIS)

Parameter	**Estimate (Standard Error)**
Mean	
μ_L	-1.3208 (0.3888)
μ_H	0.9729 (0.1449)
AR(5) Coefficients	
α_1	0.3899 (0.0732)
α_2	0.1998 (0.0742)
α_3	-0.1753 (0.0687)
α_4	-0.1550 (0.0685)
α_5	-0.0231 (0.0677)
σ	0.81226
Transitional Probability	
P_{LL}	0.2050
P_{HH}	0.9547
Total Parameters	10

It should be noted that though MS-AR model performs well in identifying the presence of separate regimes in the time series and has found applications in business cycle dating (Hamilton, 1989), the superiority of MS-AR over AR remains doubtful when out-of-sample

forecast performance is assessed (Clements and Krolzig, 1998). Compared to SETAR, MS-AR is superior in terms of forecasting US GNP at shorter horizons (one to five quarters ahead) when evaluated using the RMSE (for the out-of-sample period 1991Q1-1996Q2), as shown by Clements and Krolzig (1998). Based on standard likelihood ratio tests, Hansen (1992) and Hansen (1996a) reported no substantial evidence to reject an AR model over MS-AR. Dueker and Neely (2007) evaluated the out-of-sample forecast performance of the MS-AR model for exchange rate returns, and found that MS-AR does not perform better than a naïve forecast based on a simple moving average. Similar results were reported earlier by Goodwin (1993), who applied Hamilton's model to eight different economies and found that the MS-AR model only provides a slight improvement in the forecast ability compared to the linear models.

Bessec and Bouabdallah (2005) showed that a major limitation of MS-AR models lies in its tendency to misclassify future regimes. Misclassification of future regimes is a major limitation of the model, as it would lead to the selection of a wrong set of model coefficients for future regimes, leading to inaccurate out-of-sample forecasts. Markov switching models have been discussed in detail by Kim and Nelson (1999). A review of economic and financial time series forecasting using nonlinear models has been provided by Clements *et al*. (2004).

2.4 Nonlinear and Nonparametric Modelling Approaches

This section describes three nonlinear and nonparametric models that can effectively capture any temporal dependence structure in a time series, as measured by, for example, the similarity of previous *states* to the current state. As defined in the literature on nonlinear dynamics (see Kantz and Schreiber, 2004), a state is a point in the vector space (also called a state space or phase space), whereby the state space is a collection of all such states. A state is basically a

vector of time-delayed observations. We discuss the *nearest neighbour* method, *kernel regression* and propose a novel method, *weighted random analogue prediction* (WRAP). We provide motivation and rationale for the proposed methodology, describe the model design, and explain the process of parameter estimation for these models based on the in-sample density forecast performance.

2.4.1 Motivation and Rationale for the Proposed Methodology

The motivation for investigating a nonparametric modelling approach based on nearest neighbour and kernel regression is that it can cope with nonlinearities in the DGP, requires the estimation of a few parameters, and can provide a transition between regimes. The proposed nonlinear model, like the classical nonlinear regime-switching models, deviates from the classical G-M assumptions of linearity, however, the condition of homogeneity (which can be tested, for example, using a heteroscedasticity-robust test) may or may not hold true for our model.

The WRAP model can be viewed as a hybrid between the nearest neighbour method and kernel regression, and is an extension of the Random Analogue Prediction (RAP) model proposed by Paparella *et al.* (1997), as described in section 2.4.3. An *analogue* is defined as a past state that is similar to the current state, whereby the similarity between states is quantified using a distance metric. The rationale of the proposed model lies in utilizing similar patterns in GNP to generate forecasts. Conceptually, the proposed model attempts to generate a forecast for a given state, based on the trajectories of similar past states. Though parametric models, like AR, SETAR and MS-AR, have been employed for analysing the characteristics of GNP based on the interpretation of model parameters (Clements and Krolzig, 1998; Terasvirta and Anderson, 1992), their forecast performance over simpler benchmarks has been less than satisfactory

overall. Given that we have few GNP training observations, a plausible reason for the poor out-of-sample forecasting performance of parametric models is due to the models over-fitting the in-sample data. Note that MS-AR incorporates ten parameters. This motivates us to investigate if parsimonious models, which require the estimation of only a few parameters, could be employed over parametric models for out-of-sample forecasting. Furthermore, the regime switching models make strong prior assumptions about the functional form of the DGP, whereby the model presumably identifies the exact time period of regime shifts and adequately characterizes the behaviour of GNP in each separate regime. However, it has been shown by Bessec and Bouabdallah (2005) that regime-switching models have a tendency to misclassify future regimes on the out-of-sample data. Also, using regime-switching models, the form of error distribution needs to be assumed beforehand. In order to avoid making potentially incorrect assumptions about the DGP, we attempt to learn the functional form from the data itself using a nonparametric approach. This approach does not require assumptions about the form of error distribution, as opposed to previous nonlinear modelling approaches, see Terasvirta and Anderson (1992). We investigate nonlinear models in order to tackle nonlinearity in the DGP. Note that if the recent GNP observations belong to a period of recession/contraction, then in order to generate forecasts, this method would find similar historical states in the past that includes observations belonging to a recessionary period, and would utilize their future behaviour to issue forecast for the current period. This allows the proposed model to provide a transition between regimes, without making prior assumptions about the presence and number of regimes in the time series. Also, since WRAP is not based on assumptions regarding the presence of distinct regimes, a limited number of observations in a particular regime do not upset the forecasting performance of the model, as opposed to the classical regime switching models.

However, if the in-sample and out-of-sample data come from different DGP, then the forecasting performance of analogue based prediction models would be poor. For details on analogue based forecasting models, see Lorenz (1969), Woodcock (1979), Paparella *et al*. (1997), Obled *et al.* (2002), and, Bliefernicht and Bardossy (2007).

The modelling framework employed for nonparametric models considered in this study relies on parameter estimates obtained by optimizing the performance of in-sample density forecasts, quantified using the Continuous Ranked Probability Score (CRPS). The proposed model is simple in that it requires estimation of only a few parameters. Note that it is computationally very expensive for one to estimate parameters based on CRPS for regime switching models that comprise a relatively large number of parameters. Estimating model parameters based on density forecast performance ensures that the estimation framework takes into account the divergence between the true and fitted distribution, and not just a point estimate (even if it maximizes in-sample likelihood). Also, the proposed model does not require the forecast evaluation to be conditioned on separate regimes of GNP time series to convincingly and consistently outperform classical linear and nonlinear models (AR, SETAR and MS-AR) across varying forecast horizons, as contrasted with the forecast evaluation scheme employed by Tiao and Tsay (1994). To condition the model evaluation on regimes requires assessing the model's performance separately on regimes of growth and recession.

The idea behind the analogue forecasting model is to identify a subset of closest analogues to the current state, and issue the *image* (observation corresponding to next instance in time) of the analogues to generate forecast. Given a discrete time series, x_t, for $t = 1, 2, \dots, N$, a m-dimensional delay vector can be constructed as $\boldsymbol{x}_t = (x_{t-m+1}, x_{t-m}, \dots, x_t)$. For this study we estimate m based on the minimization of in-sample density forecast error. As proposed by

Lorenz (1969), one forecasting strategy for nonlinear systems is to search for the previous state (analogue) that is similar to the current state $\boldsymbol{x}_t$, and issue the image of the analogue as a forecast for the current state. This forecasting scheme of issuing the image of the closest analogue was later modified by Paparella *et al*. (1997), who proposed the RAP model which took into account the behaviour of a collection of analogues (discussed in detail in Sub-section 2.4.4). Bliefernicht and Bardossy (2007) adopted an analogue prediction model for generating probabilistic forecast for precipitation, focusing on extreme events. In the context of macroeconomic forecasting, this approach is based on the assumption that the dynamics underlying the economic time series follows a pattern, such that the trajectory of the economy from the current state would be quite similar to the trajectories of similar past states, given that a similar economic situation had been witnessed. The focus of our study lies in generating more accurate point and density forecasts compared to AR, SETAR and MS-AR using the proposed model, especially for time series that exhibits regime shifts.

2.4.2 Nearest Neighbour (NN) Method

The nearest neighbour method is a nonparametric method, having a range of applications in regression, classification, and data mining tasks, see Hastie *et al.* (2009). This method relies on the assumption that sequences of past states are similar to sequences following the current state. Hence, based on our knowledge of the outcome of similar past states, a plausible estimate of the outcome of future states can be inferred, see Kantz and Schreiber (2004). Nearest neighbour method has been successfully applied for forecasting a range of time series. Mulhern and Caprara (1994) employed nearest neighbour models for forecasting market response. Fernández-Rodríguez *et al*. (1999) used NN-methods for forecasting the exchange rates of nine currencies, and reported that an NN-model was able to slightly outperform an Autoregressive Integrated

Moving Average (ARIMA) model and a *random walk benchmark*. Forecasts using a random walk benchmark are generated by issuing the current time series observation as a forecast for the next time step. Similar results were reported by Barkoulas *et al.* (2003) who adopted NN-method for forecasting US interest rates. They reported that the NN-method outperforms Box and Jenkins (1970) models in terms of out-of-sample forecast accuracy. One major drawback of the NN-method is that the forecast accuracy relies heavily on the choice of a single parameter used to denote the size of neighbourhood. Furthermore, the performance of NN-methods is also dependent on the choice distance function, which is used to quantify the similarity between the state vectors. The distance function can either be global (same distance function used throughout the state space), or local (different distance functions used for different parts of the state space). Furthermore, the local distance functions can be classified as being query-based or point-based, for details, see Atkenson *et al*. (1997).

The rationale behind the NN-method lies in identifying an optimum neighbourhood size (for a given loss function), and utilizing the information from *neighbouring states* (analogues or states lying within the optimum neighbourhood) to issue a forecast. The size of the neighbourhood is commonly defined either in terms of the number of states (k-NN), radii (distance) from the current state (r-NN) or fraction of total number of states (f-NN). If S is the total number of available states and $f \in [0, 1]$ is the optimum neighbourhood fraction size, the total number of states within the neighbourhood would be given by $f \times S$. The plausible effects of defining the neighbourhood size using these metrics (number, radii, and fraction) on forecast accuracy and the procedure for estimating optimum neighbourhood size are discussed later in this section. The steps involved in using the NN-method for forecasting are,

a) Choosing a metric to denote the size of the neighbourhood (e.g. number, radii, or fraction),
b) Choosing a distance metric to quantify the similarity between different states (e.g. Euclidian distance, Mahalanobis distance, or Chebyshev distance),
c) Estimating the optimum size of the neighbourhood, and,
d) Quantifying the collective behaviour of neighbouring states, and issuing an out-of-sample forecast.

Given an optimum neighbourhood size, one can quantify the collective behaviour of neighbouring states using, for example, the mean, or build a local linear model for neighbouring states, and use the estimated local model for forecasting (see Cleveland *et al.*, 1988; Barkoulas *et al*., 2003; Regonda *et al.*, 2005).

The size of the neighbourhood subspace can be defined in terms of radii, i.e. distance of the farthest neighbouring state from the current state (r-NN). The disadvantage of this method, however, is that the size of the neighbourhood does not adapt to the change in density of states around the current state. Specifically, let $r(\boldsymbol{x}_t)$ denote the radii of the optimum neighbourhood centred at the current state $\boldsymbol{x}_t$, and $d(\boldsymbol{x}_t)$ denote the density of states lying within the neighbourhood. The density of states $d(\boldsymbol{x}_t)$ is directly proportional to the number of neighbouring states and inversely proportional to the volume of the neighbourhood centred at $\boldsymbol{x}_t$. If we consider a different state, $\boldsymbol{x}_{t+1}$, such that $\boldsymbol{x}_{t+1} \neq \boldsymbol{x}_t$, the density of states in the vicinity of this new state would most likely change (if states are not uniformly spaced) with respect to the last state, i.e. $d(\boldsymbol{x}_{t+1}) \neq d(\boldsymbol{x}_t)$. Note that a state is a point in the state space (or vector space). Hence, the location of the state $\boldsymbol{x}_{t+1}$ is different than the location of $\boldsymbol{x}_t$ in the state space for $\boldsymbol{x}_{t+1} \neq \boldsymbol{x}_t$. If the neighbourhood size is defined in terms of radii, we have $r(\boldsymbol{x}_{t+1}) = r(\boldsymbol{x}_t)$,

i.e. the size of neighbourhood remains same, despite the change in location and density of states in the state space. This is a major shortcoming of r-NN, as the neighbourhood would tend to incorporate fewer states around the low density states, possibly introducing high variability in the model. On the other hand, high density states would encompass a large number of states in the neighbourhood, due to which the model would suffer from high bias. A simple solution to this problem is to determine the neighbourhood size in terms of the number of states (a strategy we call k-NN). This approach ensures that the radii of the optimum neighbourhood is not fixed, and adapts to the change in data density in order to include k nearest states. One of the shortcomings of the k-NN approach, however, is that the neighbourhood size does not adapt to the change in the size of the available data. Note that the size of data available for generating forecasts varies when a rolling forecast scheme is employed (details provided in Section 2.5). This shortcoming can be overcome by defining the neighbourhood size in terms of a fraction of the total available states, f-NN. Note that though f-NN adapts to the change in data density and size, the optimum value of f is estimated using only the in-sample data, and is held fixed while generating out-of-sample forecasts in this study.

This method requires the estimation of a parameter, f, which denotes the optimum fraction of available states (i.e. nearest neighbours). Estimating the size of the neighbourhood in terms of a fraction has the advantage that the radius at each local neighbourhood adapts with the density of data in the state space. Using this approach, the radius of the neighbourhood $r(\boldsymbol{x}_t)$ can vary with the location (defined by the current state vector $\boldsymbol{x}_t$) in the state space. If the density of state vectors in the neighbourhood centred at $\boldsymbol{x}_t$ is low, the radius of the neighborhood will be relatively large in order to include a sufficient number of state vectors. In contrast, if the density is high, the effective radius will be relatively small. This flexibility allows the model to find the

balance between bias and variance. Note that the concept of denoting the neighbourhood size in terms of an optimum fraction of available states has been used earlier by Barkoulas *et al.* (2003) for forecasting US interest rates.

To compute a forecast for x_{t+1}, we use the current state $\boldsymbol{x}_t = (x_{t-m+1}, x_{t-m}, \dots, x_t)$, and calculate the distance vector $\boldsymbol{D}_{t-m} = (D_m, D_{m+1}, \dots, D_{t-m})$, where $D_i = ||\boldsymbol{x}_t - \boldsymbol{x}_i||_2$, and $t > m$. A set of nearest states is identified by defining an integer mapping ϑ_i: $\mathbb{Z} \rightarrow \mathbb{Z}$, such that:

$$D_{\vartheta_1} \leq D_{\vartheta_2} \leq \dots \leq D_{\vartheta_k}, \tag{2.13}$$

where D_{ϑ_k} corresponds to the distance of the k^{th}nearest state from the current state under consideration. The local neighbourhood is defined by the radius $r(\boldsymbol{x}_t) = D_{\vartheta_k}$, where $k = \lfloor f \times N_t \rfloor$, such that $N_t = t - 2m + 1$. We use Euclidean distance in this study. Note that for out-of-sample forecasting, $t \in [T, N)$, where T is the forecast origin, and N is the length of the time series. Probabilistic forecasts are generated by simulating an ensemble of Q scenarios by sampling randomly from the discrete set of k neighbouring state vectors centred at the current state, as explained later in this section. The point forecast for a particular horizon (h) is obtained by computing the mean of an ensemble of Q scenarios. This specific approach used to sample from the neighbourhood takes into account the behaviour of similar past states, allowing us to generate probabilistic forecasts for the current state.

The optimum neighbourhood fraction f and the dimension of the state-space m are estimated using *cross-validation*. Cross-validation is a statistical method commonly used for obtaining reliable estimates of parameters by taking into account the goodness of fit and generalizability of the model, see Hastie *et al.* (2009). The advantage of this approach is that it tends to ensure that the estimated model does not over-fit the data. This is done by splitting the data into different

training (learning) and testing (validation) parts. The parameters for f-NN are estimated by minimizing the one-step ahead (next step) in-sample density forecast error quantified using the Continuous Ranked Probability Score (CRPS). We denote h-step ahead forecast for x_t by $\hat{x}_{t+h}$, whereby $h = 1$ corresponds to one-step ahead forecast, while $h > 1$ corresponds to multi-step ahead forecasts. In order to estimate the optimum value of f and m, the in-sample time series is divided into two unequal parts. Note that we use only the in-sample observations for model estimation, while the out-of-sample observations are reserved completely for model evaluation. Moreover, observations corresponding to the last twenty five percent of the in-sample dataset are used as the cross-validation hold out sample. Specifically, the in-sample time series denoted by $\boldsymbol{y}_t$, where $\boldsymbol{y}_t = (x_1, x_2, \dots, x_T)$, T being the forecast origin, is divided into two unequal parts. Note that for this study, we divide the in-sample set in the ratio of 3:1, and one can estimate parameters for different ratios as well. The first part of the in-sample set ($\boldsymbol{y}_J$) comprises observations ranging from x_1 to x_J, while the second half ($\boldsymbol{y'}_T$) includes observations from x_{J+1} up to x_T. Hence, the in-sample data has been divided into a training part ($\boldsymbol{y}_J$), and a validation part ($\boldsymbol{y'}_T$). Note that since the in-sample data is divided in the ratio of 3:1, we have $J = 3T/4$. Using the observations from the training part, probabilistic forecasts are generated and evaluated for the validation part. In order to generate probabilistic forecasts for the observations in validation set (say observation x_{J+1}), we compute a distance vector (based on the Euclidean distance) between the current state vector $\boldsymbol{x}_J$ and previous state vectors ($\boldsymbol{x}_m, \boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_{J-m}$). The distance vector is sorted based on the integer mapping ϑ_i (equation 2.13). Using the sorted distance vector, we identify a set of closest states to the current state $\boldsymbol{x}_J$, which lie within the neighbourhood defined by f. Note that we include previous states only up to time index $J - m$ in this case, so that none of the observations in the current state vector $\boldsymbol{x}_J$ match with observations

in the previous states vectors. Once the neighbouring states (centred at $\boldsymbol{x}_J$) have been identified based on the sorted distance vector, images (observation corresponding to the next instance in time) of these neighbouring states are randomly sampled and issued as a probabilistic forecast for x_{J+1}. Specifically, if say $\boldsymbol{x}_s$ is one of the closest states to the current state $\boldsymbol{x}_J$, i.e. $\boldsymbol{x}_s$ lies within the neighbourhood (defined by f) centred at $\boldsymbol{x}_J$, the image (next instance) of $\boldsymbol{x}_s$, i.e. x_{s+1}would be issued as a forecast for x_{J+1}. Random sampling from the images of neighbouring states is employed for generating density forecasts. We then role forward one step, generate forecast for x_{J+2}, and keep rolling forward till we obtain forecast for x_T. The values of m and f that minimize CRPS on the validation set are used for generating forecasts on the out-of-sample data.

The general idea behind this method is that, at each forecast horizon (lead time into future), previous states (to the current state) are ranked based on their distance to the current state, and only the top (100×f)% of the closest historical states are used in forecast generation. Note that for $k = 1$ ($f = 1/N_t$), the f-NN method corresponds to the *local analogue method*, in which the future trajectory of the most similar state is issued as a forecast, see Lorenz (1969). The advantage of generating an ensemble of Q scenarios for nonlinear models is that it is based on few assumptions regarding the forecast distribution at each horizon. Different forecast scenarios for a current state are generated by random sampling of the past states. This nonlinear and nonparametric approach provides sufficient flexibility to generate forecast distributions with properties such as *skewness* (asymmetry), *fat tails* (long tailed distribution) and *multimodality* (multiple modes). This flexibility allows the model to generate accurate density forecasts, and generalize to different datasets.

For generating out-of-sample multi-step forecasts (horizon $h > 1$), we iterate the procedure for each individual forecast scenario. The trajectory for a given scenario is obtained by employing the forecast of the previous time steps to include an additional state vector. For example, for $h = 2$, when estimating $\hat{\boldsymbol{x}}_{t+2}$, we would use the forecast $\hat{x}_{t+1}$ to update the state vector $\hat{\boldsymbol{x}}_{t+2} = (x_{t-m+2}, x_{t-m+1}, \dots, \hat{x}_{t+1})$. In this way, the scenarios for horizon $h > 1$ depend on all the previous forecasts. Note that as we generate multi-step ahead forecasts, the total number of previous states (to the current state) increases. Further details on generating multi-step ahead forecast using nonlinear models can be found in Granger and Terasvirta (1993). The advantage of using f-NN to generate multi-step ahead forecasts is that the neighbourhood size takes into account the additional available state vector. This means that $r(\boldsymbol{x}_{t+1})$ (radii of a neighbourhood centred at any given state $\boldsymbol{x}_{t+1}$) computed using f-NN, adapts not only to the density of data in the state space, but also to the number of available states at each horizon, unlike the case when the neighbourhood size is defined based on the number of states or a fixed radii.

2.4.3 Kernel Regression

Kernel regression can be viewed as an extension of the nearest neighbour method that relies on an additional assumption that the closer the past state is to the current state (as quantified by distance metric), the more likely it becomes for their trajectories to match. Note that in f-NN, the states are weighted equally, irrespective of their different distances from the current state. Sensitivity to similarity is crucial for analogue prediction methods because if a previous state is relatively far from the current state, then it is less likely to provide an accurate representation of the current state. Kernel regression overcomes this limitation by assigning weights to the previous states based on their degree of similarity to the current state.

Kernel regression assigns relatively higher weights to the states that are closer to the current state. The kernel is a smooth function centred at each sampled state, and the width of the kernel is controlled by a single bandwidth parameter. The kernel is basically a weighting function, whereby the rate of weight decay depends on the bandwidth of the kernel. A larger kernel bandwidth corresponds to a slower rate of weight decay (flat weighting function), while a small kernel bandwidth leads to fast decay of weights. The problem of choosing the appropriate kernel and estimating the optimum bandwidth are discussed later in this section. Kernel regression methods have been successfully employed for forecasting wide range of time series recently. Binner *et al.* (2009) employed kernel regression for forecasting US inflation, while Sun and Chen (2008) applied kernel regression for traffic forecasting. Asber *et al.* (2007) applied kernel regression models for short-term forecasting of electric load time series. The steps involved in employing kernel regression for a time series are as follows:

a) Selecting a kernel function (e.g. Gaussian kernel, Epanechnikov kernel, and tricube kernel),

b) Selecting a distance function (e.g. Euclidian distance, Mahalanobis distance, or Chebyshev distance),

c) Estimating the bandwidth parameter, and,

d) Issuing forecasts, using the optimum bandwidth estimated from the in-sample data.

As pointed out by Mays *et al.* (2001), the estimation of the kernel bandwidth is crucial, as a large bandwidth introduces high bias (low variance), while a small kernel bandwidth introduces high variance (low bias). Estimating an optimum bandwidth based on the trade-off between bias and variance is crucial, because low variance leads to a poor generalization of the model on the

out-of-sample data, whereas low bias points towards the model under-fitting the data. The estimation of the kernel bandwidth using cross-validation has been discussed previously by Fan and Yim (2004). The advantage of estimating the bandwidth using cross-validation is that it guards against the possibility of model over-fitting the data. This ensures that the model can generalize to the out-of-sample data, thereby leading to an improved out-of-sample forecast performance. A literature review on bandwidth estimation in kernels can be found in Atkeson (1997). Regarding the choice of kernel functions, some of the common choices include the uniform, Epanechnikov, Gaussian, triangular, Nadaraya-Watson (1964), and tri-cubic kernel. For this study we employ the Gaussian kernel:

$$k_{t\tau} = e^{-\beta d(\boldsymbol{x}_t, \boldsymbol{x}_\tau)^2}, \tag{2.14}$$

where β is the kernel bandwidth and $d(\boldsymbol{x}_t, \boldsymbol{x}_\tau)$ is the Euclidian distance between the two states $\boldsymbol{x}_t$ and $\boldsymbol{x}_\tau$. The kernel function $k(\cdot)$ is assumed to be real valued function satisfying $\int |k(x)| dx < \infty$, and, $\int k(x) dx = 1$, see Hastie *et al*. (2009).

We further extend the notion of similarity between two state vectors, quantified using the Euclidean distance, to also take into account their time of occurrence. Specifically, the similarity between two states, say $\boldsymbol{x}_t$ and $\boldsymbol{x}_\tau$ depends on their Euclidean distance, and the difference between their time of occurrence $\Delta t'_{t\tau}$, where $\Delta t'_{t\tau} = t - \tau$, for $t > \tau$. The advantage of this method is that it not only depends on how close (in state space) a previous state is to the current state, but also takes into account how recent or old it is with respect to the current state. This approach is based on the assumption that somewhere in the past, there are states that would provide an estimate about the future state, but the farther back we look; the less relevant such states becomes for forecasting. This method incorporates Gaussian kernel:

$$k_{t\tau} = e^{-\beta d(\boldsymbol{x}_t, \boldsymbol{x}_\tau)^2} \times e^{-\alpha(\Delta t'_{t\tau})^2}, \quad (2.15)$$

where $\Delta t'_{t\tau}$ is the time difference between state vectors $\boldsymbol{x}_t$ and $\boldsymbol{x}_\tau$, for $t > \tau$. This method of ensures that the recent state vectors have higher weight, along with the state vectors that are relatively similar to the current state vector as quantified by the distance metric. This method is based on the idea of using a double kernel estimator as proposed by Yu and Jones (1998). Recently, Harvey and Oryshchenko (2011) have proposed a modification of the kernel density estimator to capture changes in different quantiles over time.

The bandwidth parameters $\{\alpha, \beta\}$ and the optimum dimension m for kernel regression are estimated based on minimization of the in-sample density error (quantified by CRPS) using cross-validation. To estimate the parameters, the in-sample data is divided into two unequal parts in the ratio 3:1, namely the training and validation part. Density forecasts are generated for the validation part using past states from the training part. Note that the next instances of past states are sampled based on their allotted weights, as controlled by the bandwidth parameters. The set of parameters that minimizes one-step ahead in-sample density forecast error are chosen as the optimum parameter set and are employed for out-of-sample forecasting.

2.4.4 Weighted Random Analogue Prediction (WRAP) Method

The WRAP model can be viewed as a hybrid between the f-NN method and kernel regression. WRAP estimates an optimum neighbourhood, and assigns weights to the state vectors within the neighbouring (defined by f) based on their degree of similarity to the current state vector. This model requires the estimation of dimension m, parameter f that gives an estimate of the size of the optimum neighbourhood, and the parameters $\{\alpha, \beta\}$, that controls the bandwidth of the kernel employed for weighting state vectors within the neighbourhood. The WRAP model is an

extension of the Random Analogue Prediction (RAP) model, proposed by Paparella *et al.* (1997). The general idea behind the RAP model is to estimate an optimum neighbourhood size, and sample next instances of the neighbouring states (also called analogue) randomly from the neighbourhood to issue forecasts. Specifically, RAP defines neighbourhood size in terms of number of states (as considered in *k*-NN), whereby the probability of sampling an analogue is inversely proportional to the distance between the corresponding neighbouring state and the current state. As pointed out by Paparella *et al*. (1997), the advantage of RAP model lies in its parsimony and its ability to generate accurate predictions for nonlinear time series. The WRAP model extends RAP by: (a) incorporating a double kernel for appropriately weighting the analogues, (b) defining the neighbourhood size in terms of fraction of total states, so that the model adapts with an increase in the size of data while generating multi-step ahead forecasts using a rolling forecast origin, and, (c) using an estimation framework based on the in-sample density forecast performance, quantified using the CRPS. Given that RAP is a data-based modelling approach; the major disadvantage of this approach lies in its sensitivity to the uncertainty in observations, and availability of an inadequate number of observations. It is noteworthy that the use of analogues for prediction was proposed by Lorenz (1969). Woodcock (1979) showed the usefulness of analogues to improve regression forecasts. Obled *et al*. (2002) employed analogue method for forecasting precipitation, while Bliefernicht and Bardossy (2007) employed analogues for generating probabilistic forecasts for precipitation focusing on the extreme events.

The WRAP model, in addition to estimating an optimum neighbourhood size, as undertaken in RAP, allots weights to the discrete set of *k* state vectors lying within the neighbourhood based on the kernel employed. The weights are defined by incorporating Gaussian kernels, whereby the

similarity of past states to the current state is quantified by the Euclidean distance and their time difference. For a given m, the estimation of the parameter set $\{f, \alpha, \beta\}$ is undertaken using cross-validation, by varying the parameters over a grid, and choosing the parameter set that minimizes the one-step ahead, in-sample density forecast error. Specifically, the in-sample data is divided into two parts in the ratio of 3:1, as explained previously in Sub-section 2.4.2. The value of f is varied so as to include only the closest state, and incremented up to one so as to include all states within the training set. Once the states lying within the neighbourhood have been identified for a given f, weights are allotted to the neighbouring states so that similar states get higher weight, whereby the rate of decay of weights is controlled by the bandwidth parameters $\{\alpha, \beta\}$, using the double kernel $k_{t\tau} = e^{-\beta d(\boldsymbol{x}_t, \boldsymbol{x}_\tau)^2} \times e^{-\alpha(\Delta t_{t\tau})^2}$ for $t > \tau$. Note that the double kernel has a kernel with one sided decaying weights, whereby the rate of weight decay is controlled by the parameter α. The in-sample density forecasts are generated using weighted random sampling. Note that the parameter estimation framework for WRAP is similar to that of f-NN, with the additional difference that the states within the neighbourhood are weighted based on a kernel (as explained in Section 2.4.3). The WRAP model is suitable for characterizing time series exhibiting regime shifts, as it depends on exploiting similarity to past sequence of states, specifically the location and trajectory of the past states, and also their time of occurrence. Due to the small number of parameters used in WRAP compared to regime-switching models, it is relatively easier to estimate parameters based on the density forecast performance, and also update the parameters with inclusion on new data. Using the WRAP method, we are basically defining a local probability sub-space (using weights obtained from a double kernel) for the discrete set of k state vectors in the neighbourhood (defined by parameter f), and sampling from this neighbourhood (using a weighted random sampling) to generate forecasts. The forecast

performance of parametric (AR, SETAR and MS-AR) and nonparametric models (f-NN, Kernel Regression and WRAP) is evaluated in the next section. Since f-NN method is an extension of the RAP method, we only include f-NN as a sophisticated benchmark in this study. It is noteworthy that both f-NN and RAP rely on generating forecasts by sampling next instances of similar states lying within the neighbourhood. As opposed to RAP, which defines neighbourhood size in terms of the number of states, f-NN defines neighbourhood size in terms of a fraction of total available states. This allows f-NN to adapt with an increase in the size of the available dataset, as we generate multi-step ahead forecasts by rolling the forecast origin forward through each quarter in the post-sample data. Moreover, RAP uses a linear weighting scheme, while f-NN allots equal weights to all states within the neighbourhood.

For the f-NN model, we estimated f = 0.13 and m = 5. The optimum kernel bandwidth β for the GNP time series was found to be 0.70 (for kernel regression). Note that using a double kernel with a kernel that has one sided weights (for the time of occurrence), did not improve the forecast accuracy of the model for GNP time series, compared to the case when the single kernel was used. Hence, we present results for the single kernel (that quantified similarity based on Euclidean distance) in this thesis. The parameter estimation procedure for nonparametric models used cross-validation. In this thesis, we use a Gaussian kernel and select Euclidean distance for quantifying the similarity between state vectors. The estimates of the parameter set $\{f, \beta\}$ for the WRAP model were found to be $\{0.13, 0.12\}$. The parameter f was varied such that the corresponding number of neighbouring states was within the range $[1, (3T/4)-m]$, T being the forecast origin and m being the dimension, while β (kernel bandwidth) was varied in the range [0.01, 1] in increments of 0.01. The estimates for nonparametric models, applied to the longer GNP time series are presented in Table 2.4.

TABLE 2.4

PARAMETER ESTIMATES FOR THE *f*-NN, KERNEL REGRESSION AND WEIGHTED ANALOGUE PREDICTION (WRAP) MODEL BASED ON MINIMIZATION OF IN-SAMPLE DENSITY FORECAST ERROR

Parameter	f **– NN**	**Kernel Regression**	**WRAP**
f	0.13	-	0.13
β	-	0.70	0.12
m	5	5	5
Total Parameters	2	2	3

The aforementioned reasons make the proposed model appealing conceptually, which is reflected in its ability to generate more accurate point and density forecasts than parametric models for the GNP time series, as shown in the next section. It is noteworthy that the success of nonparametric models considered in this study relies on the appropriate selection of the distance metric and distribution function, in addition to the proper estimation of model parameters. The parameters for f-NN, kernel regression, and WRAP were estimated based on the in-sample density forecast performance of different models, as quantified using the CRPS.

2.5 Results and Observations

In this section, we evaluate the out-of-sample point and density forecast performance of the parametric and nonparametric models. The point forecast performance for a particular forecast horizon h is evaluated using the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), given by:

$$RMSE_h = \sqrt{\frac{1}{N-T-h+1}\sum_{i=T+h}^{N}(x_i - \hat{x}_i)^2}, \tag{2.16}$$

$$MAE_h = \frac{1}{N-T-h+1}\sum_{i=T+h}^{N}|x_i - \hat{x}_i|, \tag{2.17}$$

where $RMSE_h$ is the RMSE at horizon h, x_i is the actual observation, $\hat{x}_i$ is the forecast, T is the forecast origin and N is the length of the time series. In order to evaluate the density forecast performance, we use the empirical form of CRPS. The advantage of using CRPS for quantifying density forecast performance is that it takes into account both *sharpness* and *calibration*. The first property, sharpness, rewards the model if the forecast density has a small spread and is highly peaked around the actual observation. The second property, calibration, rewards the model if the agreement between the forecast density and actual data density is good, for further details on CRPS, see Gneiting *et al*. (2007). Note that CRPS can be viewed as the distributional analogue of the mean absolute error, and is defined in terms of the predictive cumulative distribution function (CDF) denoted by F as follows:

$$CRPS = \int_{-\infty}^{\infty}\{F(y) - \mathbf{1}(y \geq x)\}^2 dy, \tag{2.18}$$

where x is the actual observation, and $\mathbf{1}$ is an indicator function that equals one for $y \geq x$. The empirical form of CRPS, as given by Gneiting and Raftery (2007), is represented as:

$$CRPS = E_F|X - x| - \frac{1}{2}E_F|X - X'|, \tag{2.19}$$

where X and X' are independent samples drawn from the forecasts density function, each having the same distribution F, E_F is the expectation with respect to the distribution F, while x is the actual observation.

We use equation 2.19 to estimate CRPS as it is relatively convenient to compute compared to equation 2.18. Note that for $X = X'$, i.e. when F reduces to a point forecast, CRPS is same as the MAE. CRPS has been employed recently for assessing density economic forecasts. Gneiting and Ranjan (2011a) propose a weighted version of the CRPS for economic applications, whereby the weighting allows more emphasis on the quantiles of interest. Gneiting and Ranjan (2011b) provide details of density forecasting studies in economics based on utilizing prediction pools. Gneiting and Thorarinsdottir (2010) discuss the issues regarding the choice of benchmarks in density forecasting for economic time series, and show that forecasts from a simple no-change model are competitive with expert forecasts when evaluated using CRPS. Using CRPS, we find that an unconditional benchmark is more accurate than parametric models for forecasting US GNP.

A series of forecasts for horizons ranging from one to sixteen quarters ahead is generated using a rolling forecast scheme, as previously used for GNP time series by Clements and Krolzig (1998). In this study, the GNP time series dating from 1947Q2-2008Q3, is divided into 1947Q2-1996Q4 (199 observations) for training and from 1997Q1-2008Q3 (47 observations) for testing. Specifically, we first generate a sequence of 47 one-step ahead forecasts from 1997Q1 up to 2008Q3 by rolling the forecast origin through the out-of-sample data. We then include the one-step-ahead forecast for 1997Q1 in the model, and generate a two-step forecast for 1997Q2. We continue rolling forward so as to obtain a sequence of 46 two-step ahead forecasts. The above process is repeated for increasing horizons, until we obtain 31 sixteen-step ahead forecasts. Note

that we do not estimate horizon-specific models, but rather estimate a single model by minimizing one-step ahead error using the in-sample data. Having estimated the model, we then iterate the forecasts forward across multiple time periods. Multi-step ahead forecasts for the nonlinear models were generated using Monte Carlo simulations, by constructing an ensemble of Q scenarios. In this study, we chose Q equal to ten thousand. Note that the nonparametric models employed in this study do not rely on assumptions regarding the forecast distribution at any horizon, and different forecast scenarios for the current state are generated by random sampling of the next instances of similar past states. For kernel regression and WRAP, we employ weighted random sampling to generate density forecasts, whereby the weights are derived based on a kernel (as explained in the previous section). The model parameters for all models were estimated using only the in-sample data and then held fixed, and each model had the same amount of information for making out-of-sample forecasts. The advantage of this scheme is that it includes previous step forecasts in the model for generating forecasts for a given horizon. Using the same forecasting scheme, Clements and Krolzig (1998) reported that AR is a more robust out-of-sample GNP predictor compared to both SETAR and MS-AR. A comparison of iterated and direct forecasting schemes has been provided by Marcellino *et al*. (2006). Iterated forecasts are generated using a model estimated based on one-step ahead error, whereby multi-step ahead forecasts are constructed by iterating the forecasts forward for multiple periods. On the other hand, a direct forecast is generated using a horizon-specific model. Using a range of different macroeconomic time series, Marcellino *et al*. (2006) show that iterated forecasts are more accurate than direct forecasts, whereby the performance of iterated forecasts improves with longer horizons.

In this study, we use the unconditional mean forecast as a benchmark. For point forecasts, we compute the unconditional mean of the in-sample data, and issue it as a forecast for the out-of-sample data. Note that, in order to make h-step ahead forecasts, original data only up to time index $T-h+1$ is used (T being the forecast origin). The actual number of data points employed to compute RMSE/MAE varies with the forecast horizon. The whole out-of-sample data (starting from $T+1$) is employed to compute one-step ahead error, while the two-step ahead error is computed starting from $T+2$ until the last available quarter. The process is repeated until we obtain 16-quarter ahead forecast error. We find that the unconditional forecast performs well for all horizons compared to the classical models, across both the time series. This result points towards the efficacy and robustness of the unconditional method as a benchmark for GNP time series. We compare the random walk against the unconditional mean benchmark in terms of forecast accuracy.

In order to issue h-step ahead forecast using the random walk benchmark (also known as persistence), the current observation is issued as a forecast for the h-step ahead observation. Hence, for $h=1$, we basically issue the current observation as a forecast for the next step. Regarding the choice of benchmarks, most of the classical and recent work on GNP and GDP forecasting use the random walk model as a benchmark, but none compare their forecasts with the unconditional forecast (see Potter, 1995; Clements and Krolzig, 1998; Ang *et al.*, 2006, Zheng and Rossiter, 2006; Barhoumi *et al.*, 2008).

Forecast comparison between the unconditional and random walk benchmark, as quantified using RMSE, is provided in Figure 2.2. The results are consistent when the evaluation of benchmarks is based on MAE. Note that the *f*-NN method defaults to the unconditional density forecast as *f* approaches one.

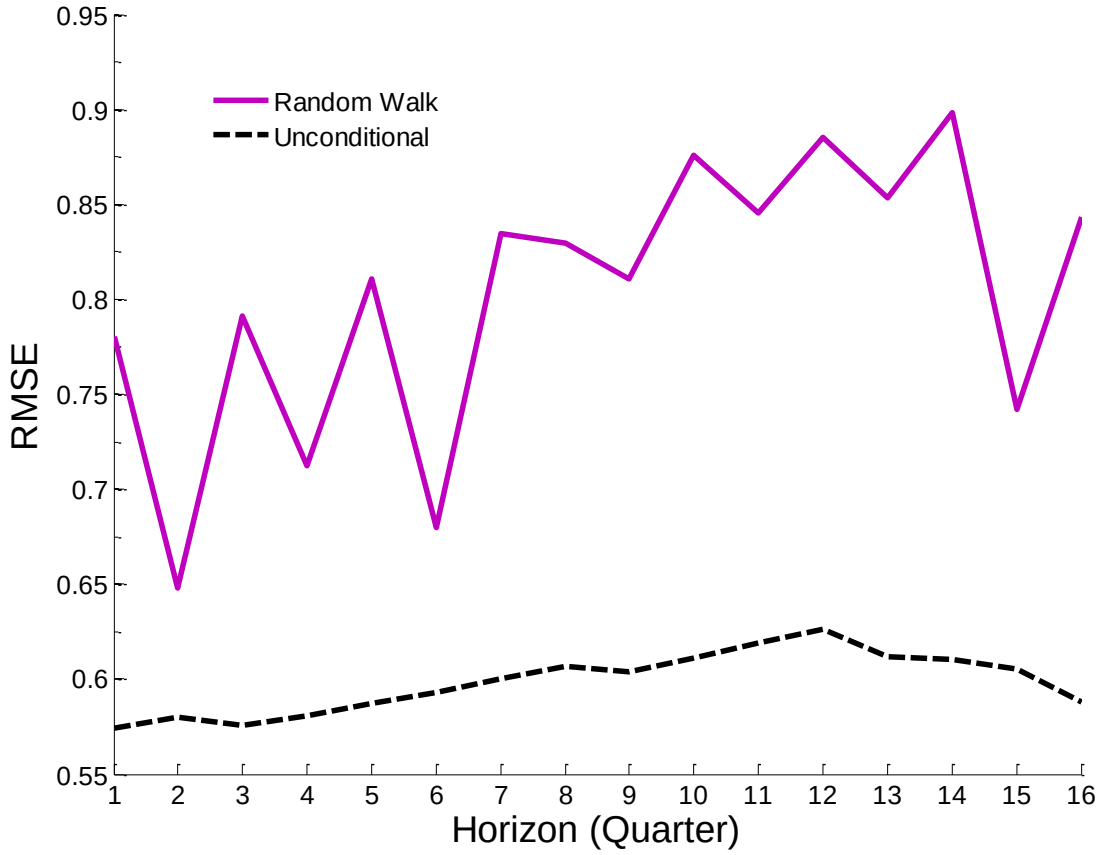

Figure 2.2— Comparison of random walk (persistence) and unconditional benchmark in terms of out-of-sample point forecast accuracy for US GNP, whereby point forecast performance is quantified using the Root Mean Squared Error (RMSE). Model evaluation period: 1997Q1 – 2008Q3.

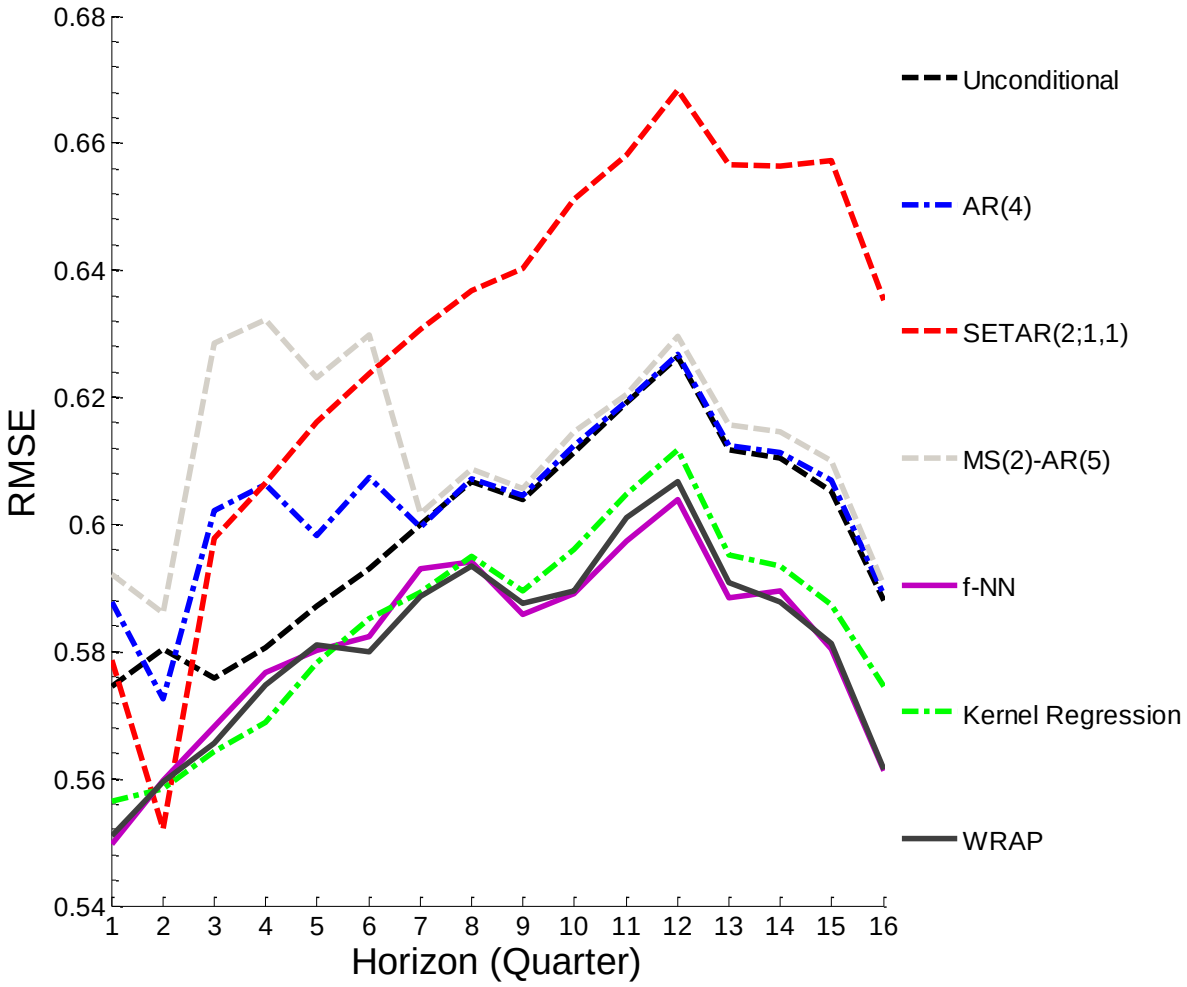

Figure 2.3— Out-of-sample point forecast performance of the parametric models (AR, SETAR and MS-AR), unconditional method and nonparametric models (*f*-NN, Kernel Regression and WRAP) for the longer US GNP time series (evaluation period: 1997Q1-2008Q3). Model evaluation uses the root mean square error (RMSE).

The out-of-sample point forecast performance of different models is presented in Figure 2.3 and Figure 2.4. When point forecasts are evaluated using RMSE (Figure 2.3), we find that *f*-NN, kernel regression and WRAP generate more accurate forecasts compared to AR, SETAR and MS-AR, except for horizon $h = 2$ when SETAR is slightly better. Forecasts from the SETAR model are found to be more accurate for the longer time series compared to the shorter time

series (1947Q2-1990Q4) at short horizons. The improved performance of SETAR may be attributed to the relatively larger percentage of observations belonging to the lower regime in the longer GNP time series.

It can be seen from Figure 2.3 that *f*-NN performs best on longer horizons. Given the limited number of samples available for computing RMSE for longer horizons, however, *f*-NN was not the best model overall. SETAR(2;1,1) performs well only for shorter horizons, whereas the forecast performance of AR(4) and MS(2)-AR(5) is not distinguishably superior to the unconditional forecast. In the case of MAE (Figure 2.4), the nonparametric models (*f*-NN, kernel regression and WRAP) outperform all other models on all horizons.

From the CRPS results (Figure 2.5), it is evident that the nonparametric models (*f*-NN, kernel regression and WRAP) consistently outperform the parametric models (AR, SETAR and MS-AR) on all horizons. The superiority of nonparametric models is enhanced when quantifying their density forecasts using CRPS. The superior density forecasts of nonparametric models could be related to their parameter estimation process which optimizes CRPS. Also, note that it is not straightforward to estimate more than, say, four parameter values by optimizing CRPS, since it is computationally expensive to generate and evaluate density forecasts for different parameter values, making the parameter estimation for AR, SETAR and MS-AR models extremely difficult.

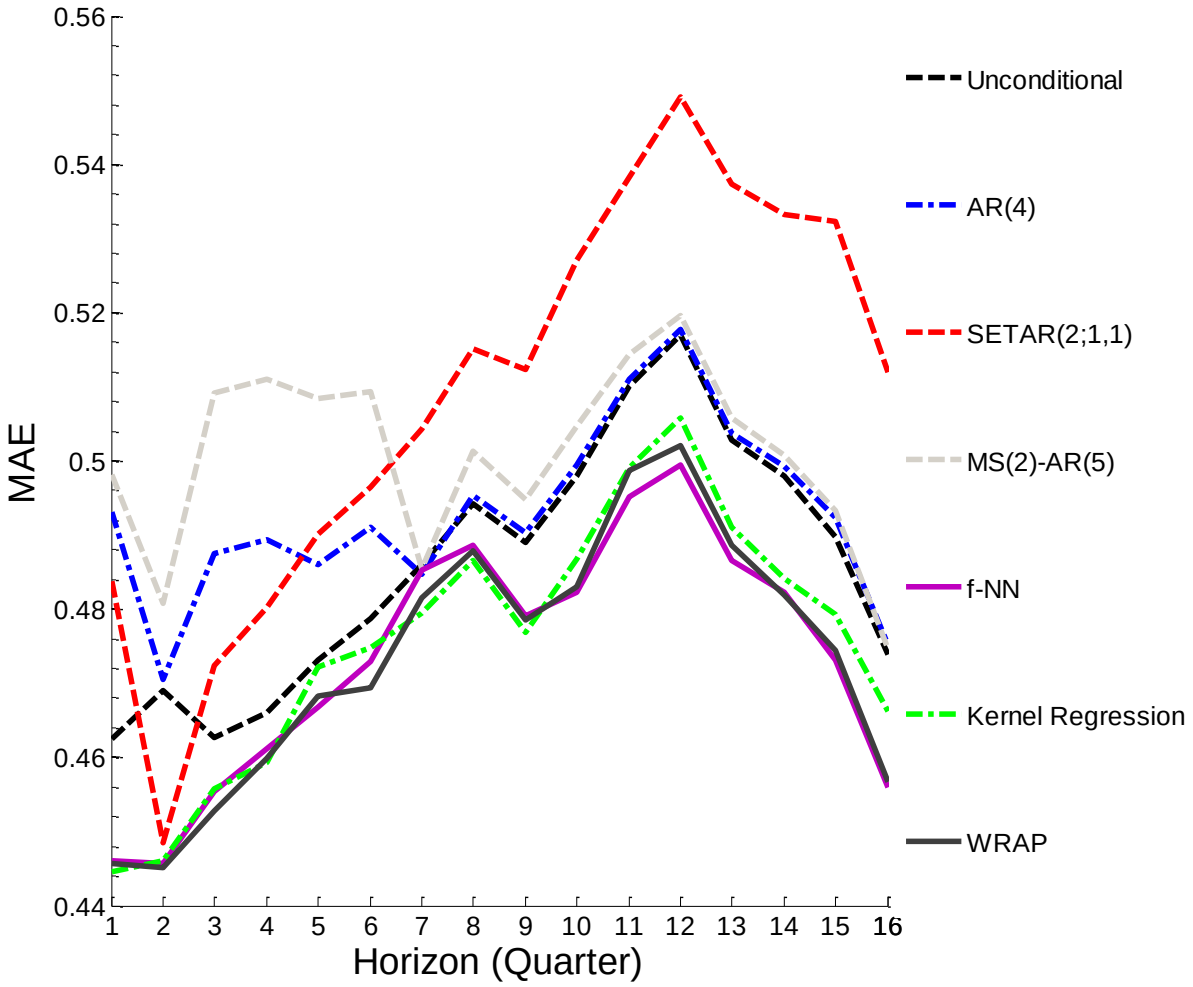

Figure 2.4— Out-of-sample point forecast performance of the parametric models (AR, SETAR and MS-AR), unconditional method and nonparametric models (*f*-NN, Kernel Regression and WRAP) for the longer US GNP time series (evaluation period: 1997Q1-2008Q3). Model evaluation uses the mean absolute error (MAE).

Interestingly, Clements and Krolzig (1998), and Galbraith (2003) have reported similar results based on the comparison of out-of-sample point macroeconomic forecasts obtained from different parametric models. Clements and Krolzig (1998) showed that if the DGP is nonlinear, SETAR and MS-AR are competitive with AR only for one and two-quarter ahead forecasts. They report that for horizons beyond two-quarters, AR is the most accurate model, while MS-AR is worse than AR, but more accurate than SETAR. Galbraith (2003) defines content horizon as

the forecast horizon beyond which forecasts obtained using the unconditional mean are no worse than the forecasts obtained from the model under consideration, and report that an AR model is not superior to the unconditional mean beyond two quarters for forecasting US GDP. Among nonparametric models, we find that both *f*-NN and WRAP are more accurate than kernel regression, and all three models are superior to the unconditional benchmark. Note that these results are in accordance with the out-of-sample point forecast results based on RMSE and MAE reported in Figure 2.3 and 2.4, whereby the nonparametric model forecasts are more accurate than the unconditional forecast.

We also evaluated the point forecasting performance of the nonparametric models when the parameter estimation process minimized in-sample RMSE. This was achieved by generating point forecasts using nonparametric models, and choosing the parameters which minimized one-step ahead RMSE on the in-sample validation set. We found that the nonparametric models outperformed classical models on most forecast horizons when quantified using both RMSE and MAE. The superior forecast performance of the proposed model (WRAP) compared to the unconditional forecast across different horizons, demonstrates the ability of the proposed model at generating both short (one quarter) and long-term (four years) forecasts. The model parameters for all models were estimated using only the training sample and then held fixed, and each model had exactly the same amount of information for making out-of-sample forecasts.

It is noteworthy that the number of observations available for forecast evaluation decreases with an increase in the forecast horizon, as we use a rolling forecast origin through the out-of-sample data. For example, we have 47 observations to evaluate one-step ahead forecasts, while for evaluating sixteen-step ahead forecasts, we only have 31 observations. Hence, for a given

model, it is not informative to compare RMSE/MAE values across different horizons using this forecasting scheme.

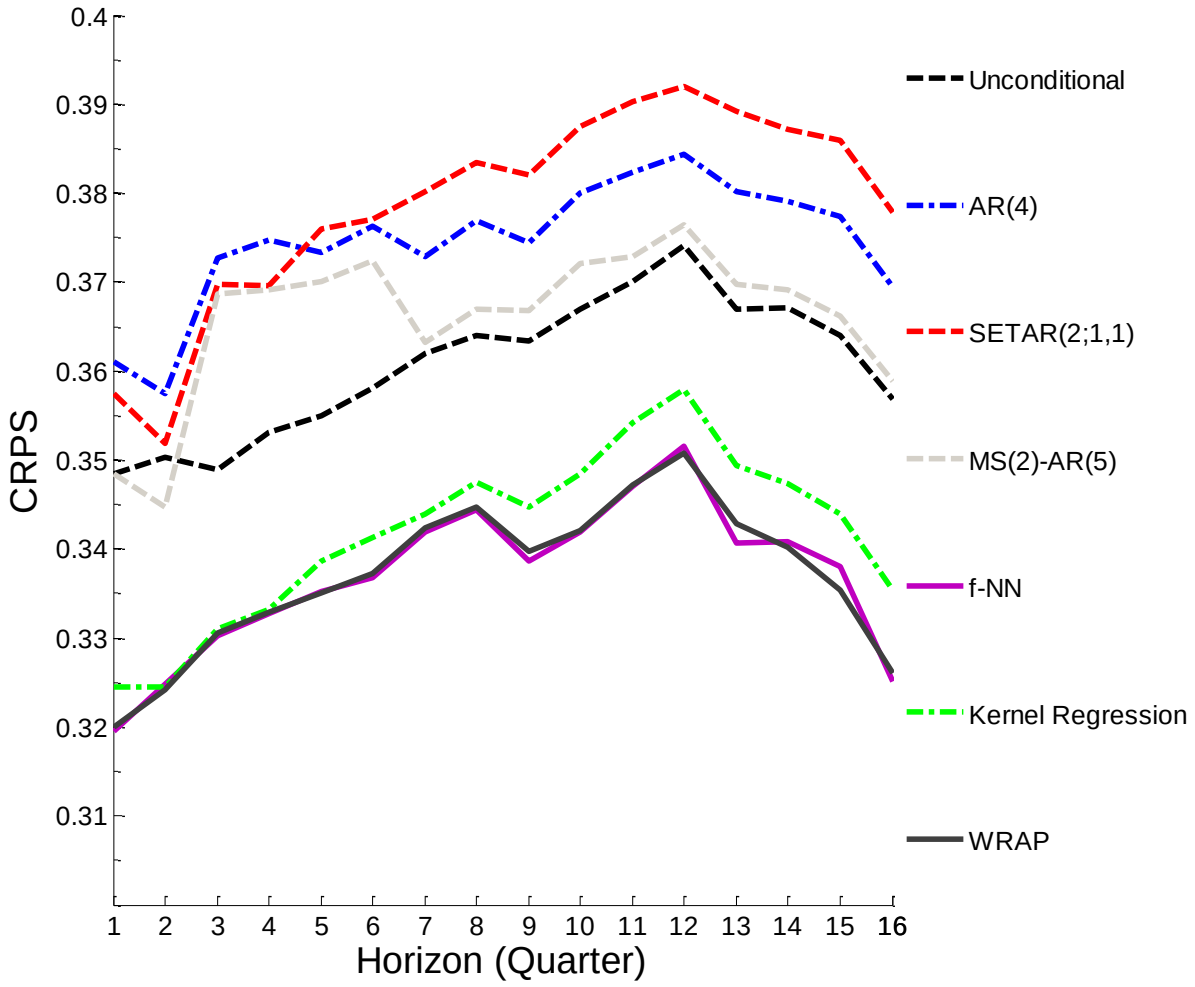

Figure 2.5— Out-of-sample density forecast performance of the parametric models (AR, SETAR and MS-AR), unconditional method and nonparametric models (*f*-NN, Kernel Regression and WRAP) for the longer US GNP time series (evaluation period: 1997Q1-2008Q3). Model evaluation uses the continuous ranked probability score (CRPS).

Moreover, we find that MS(2)-AR(5), which was selected over the fourth order Markov switching model MS(2)-AR(4), (as proposed by Hamilton (1989) for GNP) as the best model according to AIC, performed worse than the MS(2)-AR(4) in terms of forecasting ability. Hence, from the perspective of forecasting, it remains doubtful whether AIC is the right choice for

selecting the model order for the nonlinear models, as also pointed out previously by Clements and Krolzig (1998).

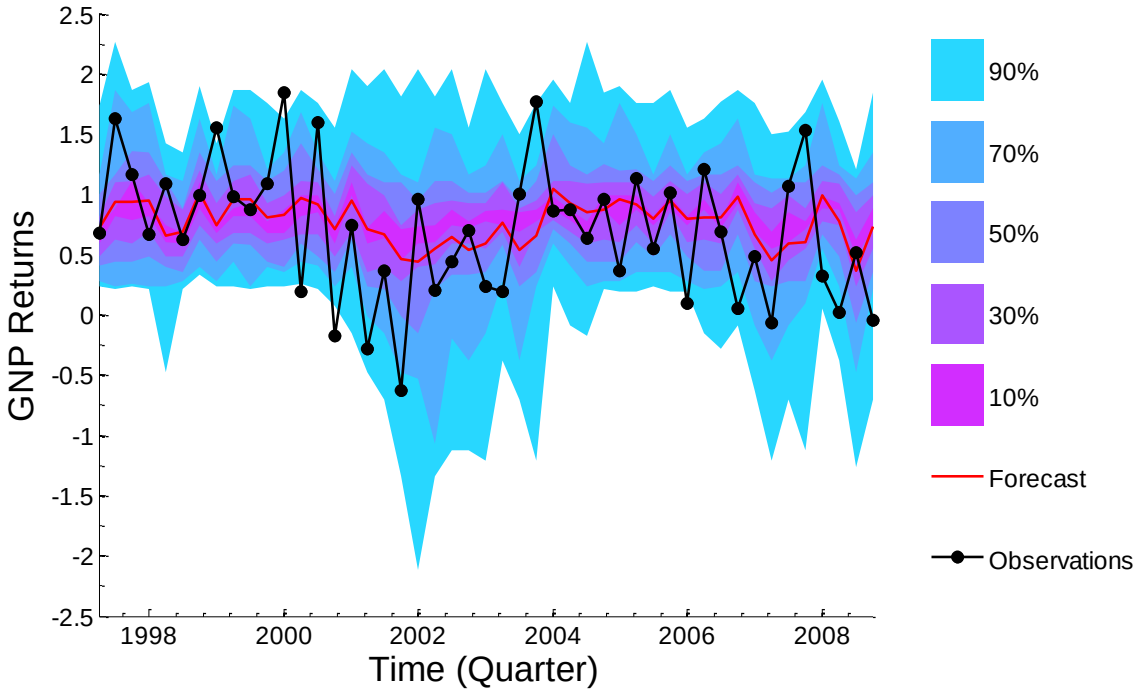

Figure 2.6— One-step ahead out-of-sample density forecasts of GNP quarterly returns from 1997Q1 up to 2008Q3 using the weighted random analogue prediction (WRAP) model. The shaded regions correspond to the different quantiles of the density forecasts centred on the median. The red line corresponds to the point forecasts provided by the mean of the forecast density and the black line indicates the quarterly GNP returns.

To further investigate the suitability of the proposed model in capturing the data density, we generate a sequence of one-step-ahead out-of-sample density forecasts from 1997Q1 up to 2008Q3 using the WRAP model and analyse the calibration of the model. In Figure 2.6, the different shaded regions correspond to the different quantile ranges (centred around the median) of the forecast distribution. Note that there are several instances in time (around 2002, 2007 and 2008) when the GNP recorded a high change with respect to the last quarter. The point forecast corresponding to the particular time of observation fails to capture these sharp changes in the GNP, but it is during these times that there is a marked increase in the uncertainty associated

with the forecasts. It is evident from Figure 2.6, that the forecast density estimates give more informed and valid inference about the changes in GNP, compared to point estimates. This underscores the need for generating density forecasts along with point forecasts, in order to communicate forecast uncertainty to policy-makers.

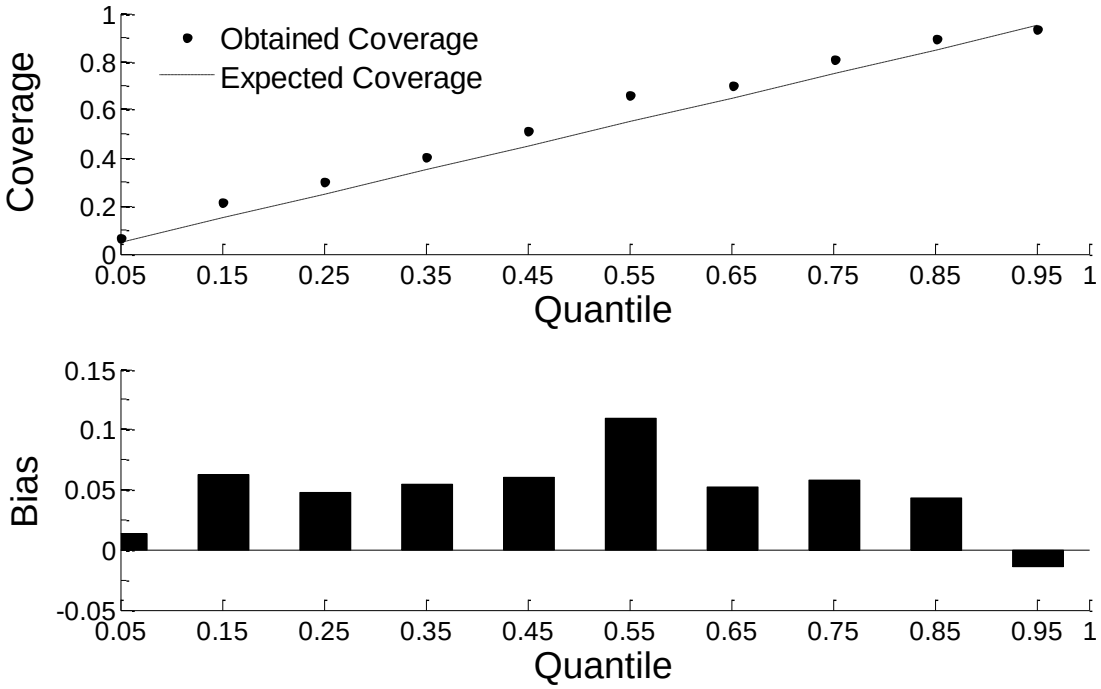

Figure 2.7— Reliability diagram (coverage and bias) for one-step-ahead out-of-sample forecasts from 1997Q1 up to 2008Q3 using the weighted random analogue prediction (WRAP) model. Coverage reflects the degree to which actual observations lie within different quantile ranges of forecast distribution. For a perfect density forecast, the coverage would lie along the diagonal (dashed line), and hence the bias would be zero across all quantiles.

The quality of the density forecasts generated using the WRAP model is further evaluated using a *reliability diagram*. A reliability diagram measures the degree to which the actual observations lie within the different quantile ranges of the density forecast, also known as *coverage*. Note that each quantile range of the density forecast is expected to encompass a specific proportion of the actual observations, as quantified by coverage. The difference between

the expected and obtained coverage is known as *bias*. We find that the bias for the WRAP model across different quantiles is small, as shown in Figure 2.7. The plausible reasons underlying the efficacy of the proposed WRAP model in generating more accurate GNP forecasts on the out-of-sample period, compared to the classical models, can be listed as follows:

1. The WRAP model provides a transition between different finer levels of growth and recession by weighted random sampling of analogues. Note that SETAR and MS-AR models assume two separate regimes in GNP, and hence a "great depression" and "small plunge" in the state of an economy are not distinguished on finer scales. The same holds true for the growth period. Hence, both SETAR and MS-AR do not distinguish between the different regimes of the economy on a finer scale, and broadly classify economic regimes as either "growth" or "recession". This is a major drawback, as one would assume economic transitions to comprise transitional regimes of different magnitudes.

2. The performance of the proposed model is not dependent upon assumptions regarding the presence of regimes in the DGP. Hence, the model does not require a substantial number of observations to lie in distinct regimes (both in and out-of-sample) in order to outperform linear models. Also, the model is parsimonious and nonparametric, and can be easily generalized to different time series.

3. Parameter estimation framework for WRAP is based on density forecast performance. Note that since WRAP is simple, i.e. it is based on the estimation of only three parameters (f, β, and m); it is relatively easy to estimate WRAP parameters using CRPS, compared to AR, SETAR and MS-AR.

For the GNP time series, it is noteworthy that the performance of WRAP is not superior compared to f-NN. Also, it is worth mentioning that though we employed a double kernel for WRAP in this study, the model is easily scalable, i.e. more kernels can be easily added in the modelling framework. Note that the forecasting ability of different models can change under the conditions of a structural break. The presence of structural breaks can adversely impact the out-of-sample forecasting accuracy of time series models even if they have a good in-sample fit, Clark and McCracken (2002). A structural break can be defined as a permanent change in the parameter vector of a model, which leads to instability in parameter estimates. The impact of a structural break on forecast accuracy depends on the type of break and also the form of the model under consideration, Clements and Hendry (1998). It has been argued by Hansen (2001) that ignoring structural breaks can lead to economic relationships go off course, leading to inaccurate forecasts. Clements and Hendry (1998) discuss the use of updating, intercept corrections and differencing to alleviate the impact of a structural change. It has been discussed by Kapetanios and Tzavalis (2005) that a major limitation of regime switching models, like SETAR and MS-AR, is that these models assume the structural parameter changes to be of fixed magnitude, which can lead to poor forecasts. They also mention that models with time varying coefficient assume the structural breaks to be continuous. To overcome these limitations, Kapetanios and Tzavalis (2005) propose a model that allows for structural breaks to be stochastic in nature, both in magnitude and in timing. The nonlinear regime switching models are better suited for forecasting under the presence of structural breaks compared to their linear counterparts. The nonlinear and nonparametric models considered in this study have the flexibility to allow for structural breaks to be of different magnitudes, which makes them more appealing for out-of-sample forecasting compared to regime switching models. Note that in order

to generate forecasts under the conditions of a structural break, the nonparametric models would quantify the behaviour of the past economic periods, which had witnessed a similar structural break as the current period. Also, since the proposed models adapt with the density of states in the state space, the models would tend to include an adequate number of analogous past states (exhibiting similar structural breaks) in the neighbourhood, and utilize their future trajectories to generate forecasts for the current period. This modelling approach allows the nonparametric models to adapt differently to the conditions of structural breaks based on their respective magnitudes. The precondition for the nonparametric models to perform well under the conditions of a structural break for a given period, however, is that a similar break must have already been witnessed before, somewhere in the past.

We show the plot of one-step-ahead out-of-sample predictive error density obtained using different models in Figure 2.8. This figure shows that the predictive error density obtained from the nonparametric models are narrower and centred close to zero compared to the parametric models. We use the two-sided Kolmogorov-Smirnov (KS test) test to show that the difference in predictive error density obtained using parametric and nonparametric models is statistically significant.

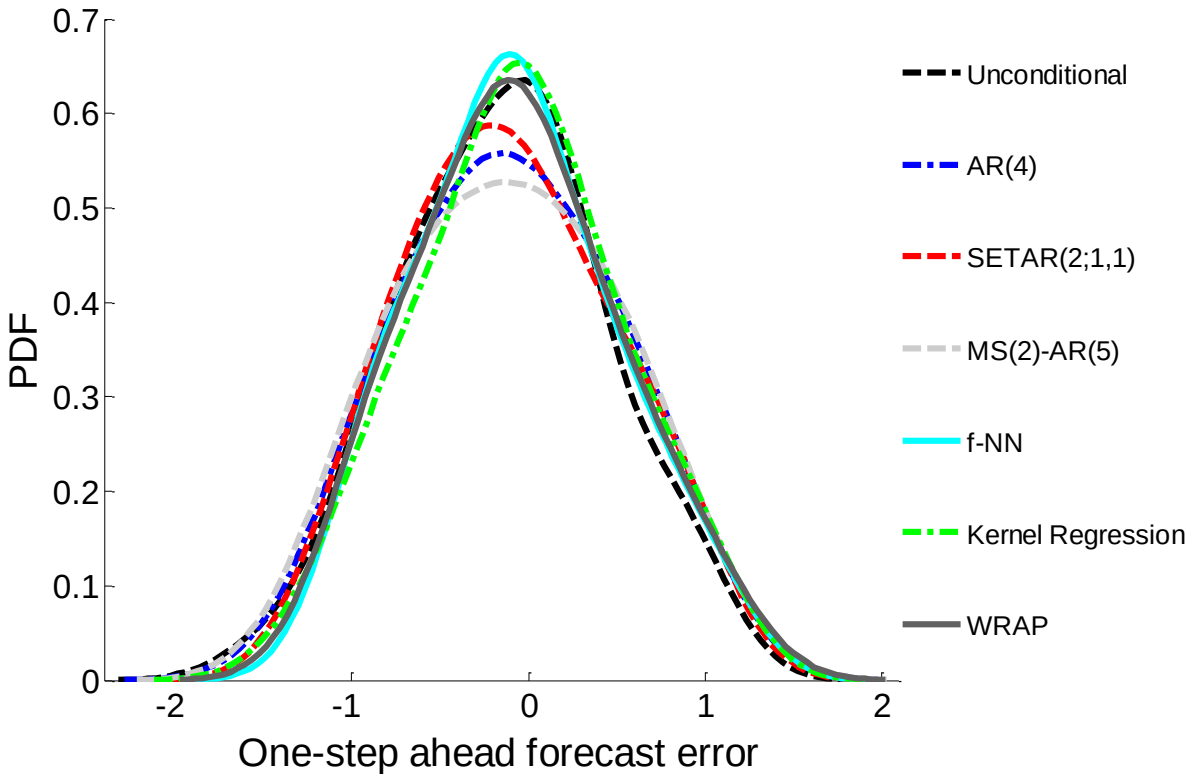

Figure 2.8— One-step ahead out-of-sample predictive error density obtained using (a) AR, (b) SETAR, (c) MS-AR, (d) *f*-NN, (e) kernel regression, and (f) WRAP.

To evaluate the out-of-sample density forecasts from different models, specifically the forecast calibration, we use the Probability Integral Transform (PIT), as also employed by Gneiting *et al.* (2007), Little *et al*. (2009) and Harvey and Oryshchenko (2011). Note that calibration refers to the consistency between forecast density and the actual data density. The PIT is defined as the value that the predictive cumulative distribution function attains at the observation, and is computed as:

$$z_n = \int_{-\infty}^{x_n} p(u)du, \tag{2.20}$$

where x_n is the actual observation and p is the one-step ahead forecast density function. If the forecasts are perfectly calibrated, i.e., the forecast density matches completely with the true density, then z_n would be Independent and Identically Distributed (IID) with a uniform density between 0 and 1. A visual inspection of PIT histogram is used to infer the deviation of PIT values from uniformity.

In the context of evaluating density forecasts from econometric models, the utility of PIT has been shown by Gencay and Selcuk (1998), who analyse the PIT histogram along with correlograms of different transformations of the PIT values. We use stochastic interpolation to calculate z_n, by drawing 10,000 sample values from the out-of-sample forecast distribution in order to construct the empirical cumulative density function, as done by Little *et al*. (2009). The PIT histogram (using 10 bins) for each of the parametric and nonparametric models is provided in Figure 2.9. Note that a U-shaped PIT histogram points towards the predictive distribution being too narrow (with respect to the actual distribution), whereas an inverted U-shaped (hump shaped) histogram corresponds to an over dispersed distribution, Gneiting *et al*. (2007). However, the PIT histogram is not a skill score that quantifies the relative performance of models, but is rather a visual aid that can be used in conjunction with different performance scores.

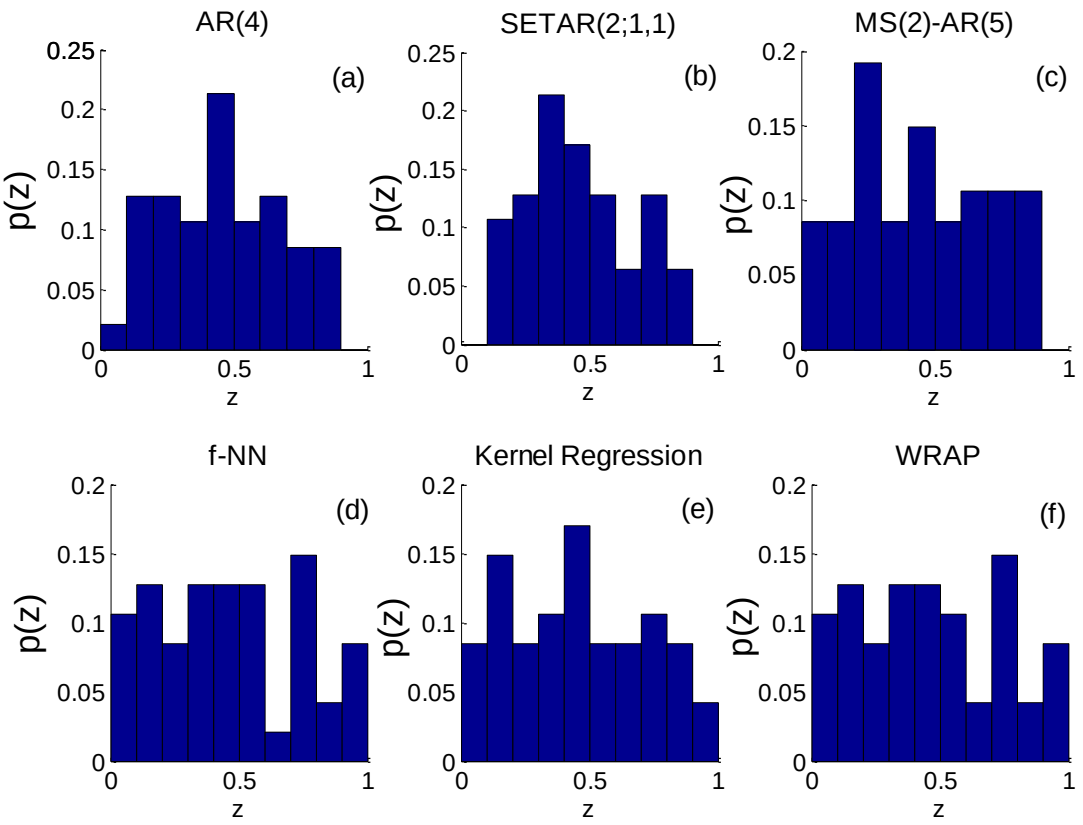

Figure 2.9— Probability integral transform (PIT) histogram for one-step-ahead out-of-sample forecasts obtained using (a) AR, (b) SETAR, (c) MS-AR, (d) *f*-NN, (e) kernel regression, and (f) WRAP.

We also plot the Autocorrelation Function (ACF) of PIT values (with 95% confidence intervals) to check for the presence of serial correlations. Note that if the structure in a time series is adequately captured by a model, this would be reflected in the corresponding z_n being serially independent. It can be seen from Figure 2.10 that the autocorrelation in PIT values across different models is not significant across most lags, showing the absence of serial correlations in z_n obtained from both parametric and nonparametric models.

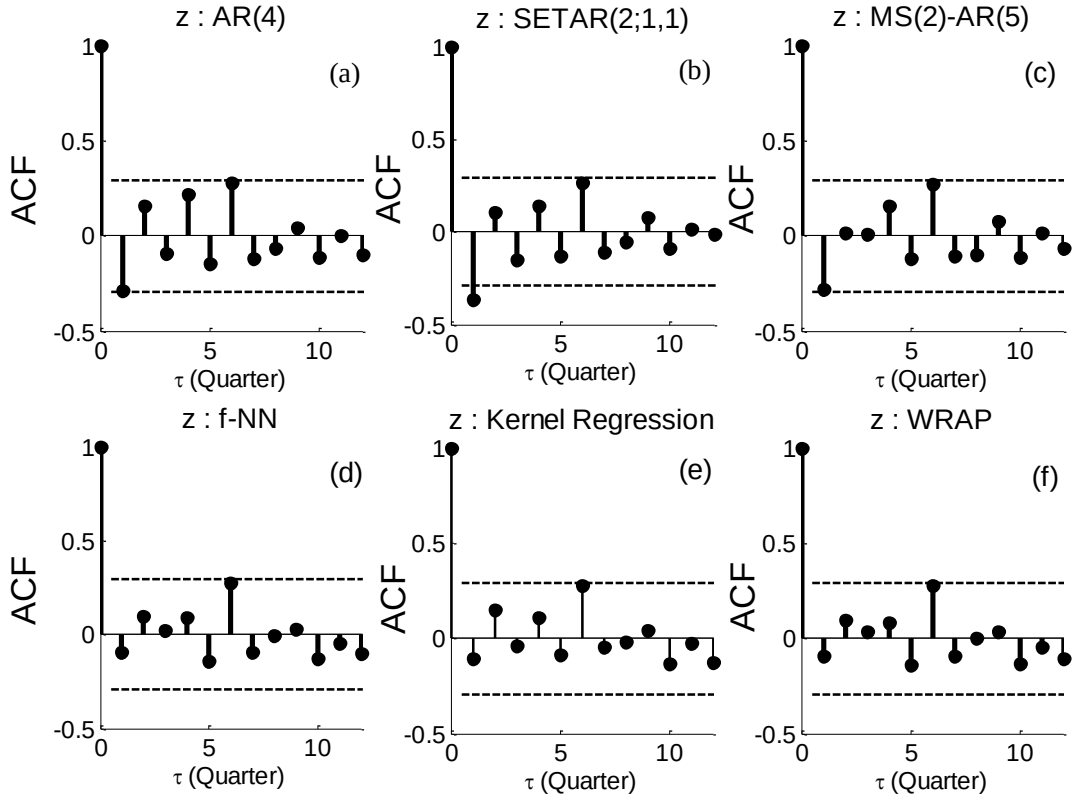

Figure 2.10— Autocorrelation function (ACF) for z-series (associated with the corresponding histogram shown above) for one-step ahead out-of-sample forecasts obtained using (a) AR, (b) SETAR, (c) MS-AR, (d) *f*-NN, (e) kernel regression, and (f) WRAP. The horizontal dashed lines correspond to the 95% confidence interval.

We use the two-sided Kolmogorov-Smirnov (KS test), proposed by Kolmogorov (1933) and Smirnov (1939), to investigate if the forecast errors for different models are drawn from the same underlying distribution. The two-sample Kolmogorov-Smirnov (KS test) test is a common choice to quantify the difference between two empirical cumulative distribution functions (ECDFs). The KS test is nonparametric, i.e., it does not assume a specific form of the distributions. Assuming a null hypothesis (denoted by H_0) that the forecast accuracy of different models is the same, the KS statistic and the corresponding p-values can be used to test the hypothesis. Specifically, the two sided KS statistic (KS_n) is defined as:

$$KS_n = sup_x|F_n(x) - G_n(x)|, \tag{2.21}$$

where sup (supremum) is the least upper bound, n is the length of the sample, while F and G are the two ECDFs under consideration. The KS test evaluates the maximum distance between the ECDFs. The higher value of KS_n signifies that the ECDFs are farther apart.

TABLE 2.5

KOLMOGOROV-SMIRNOV (KS) COMPUTED USING ONE-STEP AHEAD DENSITY FORECASTS ON THE OUT-OF-SAMPLE US GNP DATA. THE NULL HYPOTHESIS (H_0) IS THAT THE ACCURACY OF DENSITY FORECASTS GENERATED USING DIFFERENT MODELS IS THE SAME. CASES WHERE H_0 IS REJECTED (1% SIGNIFICANCE LEVEL) ARE REPORTED IN **BOLD**.

Model	*Unconditional*	*AR*	*SETAR*	*MS-AR*	*f-NN*	*Kernel Regression*
Unconditional	-	-	-	-	-	-
AR	**0.1137** (***)	-	-	-	-	-
SETAR	**0.1256** (***)	**0.0579** (**)	-	-	-	-
MS-AR	**0.1269** (***)	**0.0732** (***)	**0.0954** (***)	-	-	-
f-NN	**0.1934** (***)	**0.2270** (***)	**0.2289** (***)	**0.2107** (***)	-	-
Kernel Regression	**0.1361** (***)	**0.1859** (***)	**0.1882** (***)	**0.1752** (***)	**0.0899** (***)	-
WRAP	**0.1938** (***)	**0.2262** (***)	**0.2286** (***)	**0.2100** (***)	0.0132 (*)	**0.0877** (***)

*p>0.01; **p<0.01; ***p<0.001

We report the average KS statistic in Table 2.5. To compute the KS statistic, we first generate a sequence of one-step ahead density forecasts for the out-of-sample GNP time series (1997Q1-2008Q3), using all the parametric and nonparametric models and the unconditional benchmark. The KS statistic and the corresponding p-values are computed between the different models and the benchmark for each quarter, on the out-of-sample period. Based on a 1% significance level, the cases where the null hypothesis is rejected are reported in bold. It is evident from the Table 2.5, that the forecasts from nonparametric models (f-NN, kernel regression and WRAP) are significantly different from the parametric models (AR, SETAR and MS-AR), as also evident from the model evaluation scheme using different performance scores reported earlier. However, we note from Table 2.5 that forecasts generated using f-NN are not statistically different from WRAP.

In order to further validate our model, we compare the parametric and nonparametric models in terms of their accuracy in forecasting UK GDP. We employ seasonally adjusted UK GDP at market prices measured from 1955Q1 to 2009Q4 (units in million Pounds). The GDP time series was downloaded from the Office of National Statistics (ONS) in the UK.

The UK GDP log percentage return was divided from 1955Q2-1998Q4 for model training and from 1999Q1-2009Q4 for model testing. Note that the returns time series comprises one less observation compared to the original raw time series. The parametric and nonparametric models were re-estimated on this new dataset. Regarding classical models, the estimated models were AR(8), SETAR(2;2,2) and MS(2)-AR(3). The estimated dimension *m* for the nonparametric models was four across all three nonparametric models. For the f-NN method, we estimated f = 0.33. For kernel regression, we obtained β = 0.41, while for WRAP, we estimated f = 0.38 and β = 0.02.

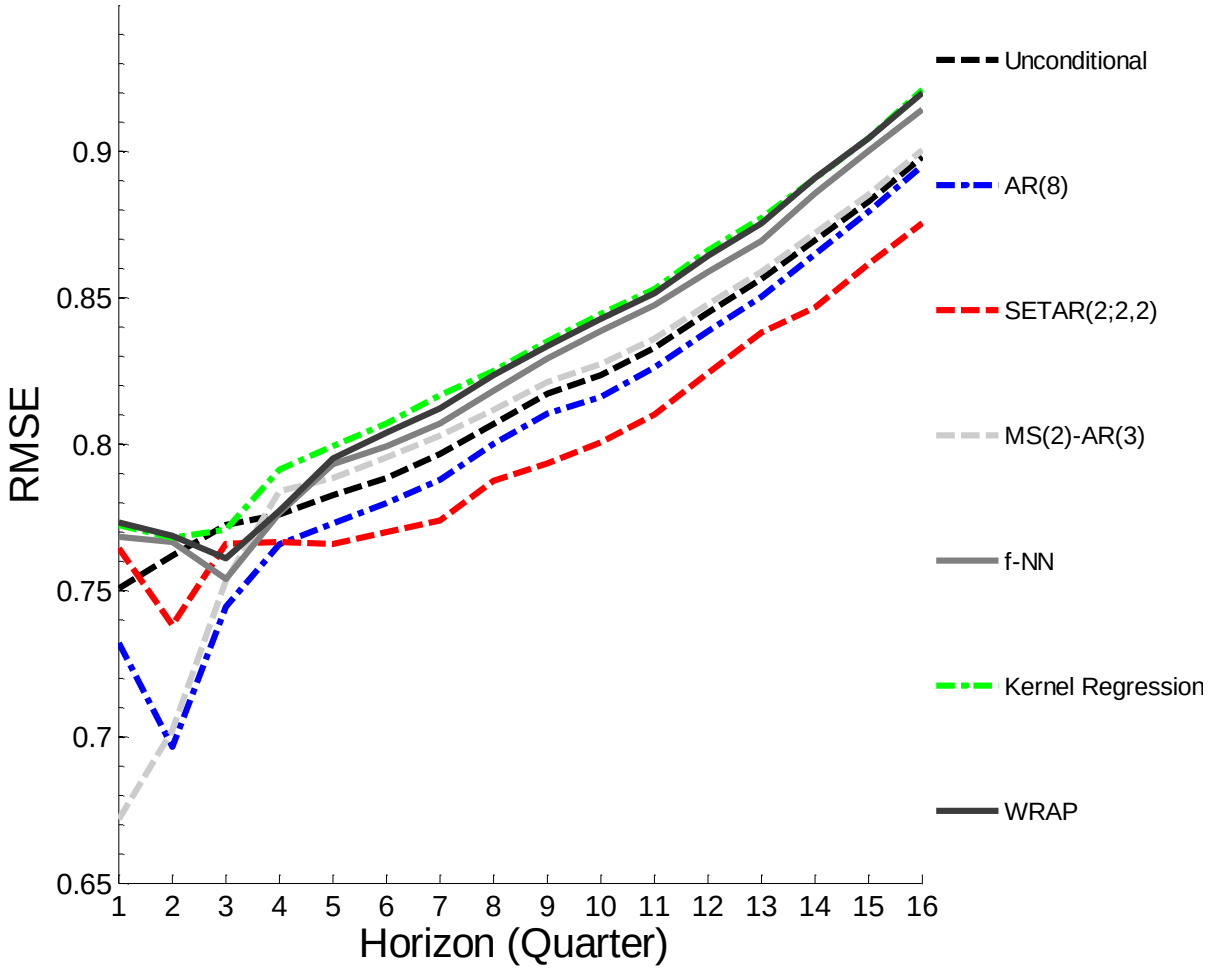

Figure 2.11— Out-of-sample point forecast performance of the parametric models (AR, SETAR and MS-AR), unconditional method and nonparametric models (*f*-NN, Kernel Regression and WRAP) for the longer UK GDP time series (evaluation period: 1999Q1-2009Q4). Model evaluation uses the root mean square error (RMSE).

We present point forecast comparison of different methods considered in this study based on the RMSE in Figure 2.11, while Figure 2.12 evaluates forecast performance based on the MAE. It is evident from Figures 2.11 and 2.12 that the point forecast performance of the parametric models (AR, SETAR, MS-AR) is superior compared to the nonparametric models (*f*-NN, kernel regression, WRAP), whereby SETAR is one of the most accurate models. It is, however,

encouraging to see from Figure 2.13, that the nonparametric models considered in this study are noticeably more accurate than the parametric models, across all forecast horizons, when the model evaluation is based on the comparison of post-sample density forecasts.

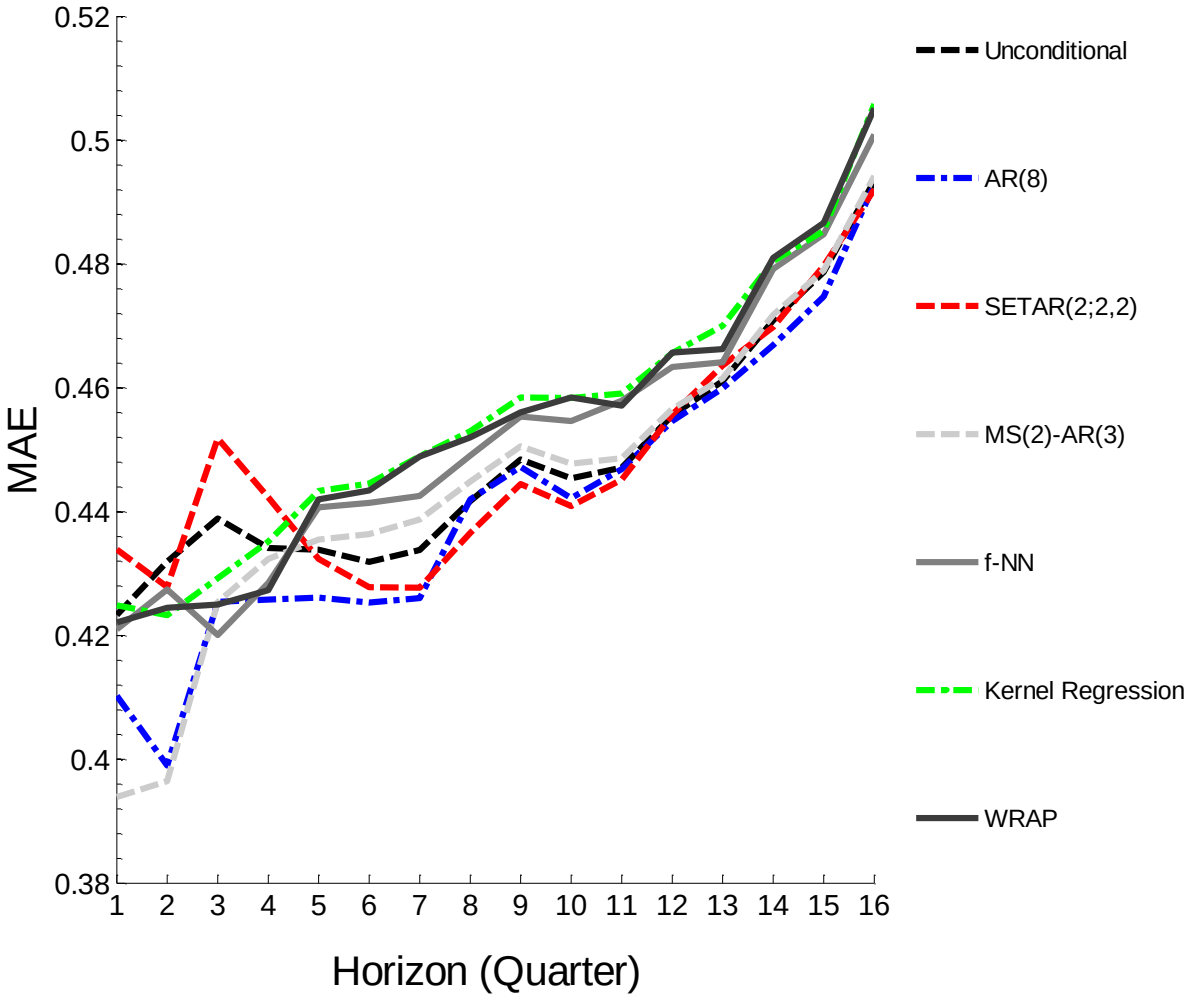

Figure 2.12— Out-of-sample point forecast performance of the parametric models (AR, SETAR and MS-AR), unconditional method and nonparametric models (*f*-NN, Kernel Regression and WRAP) for the longer UK GDP time series (evaluation period: 1999Q1-2009Q4). Model evaluation uses the mean absolute error (MAE).

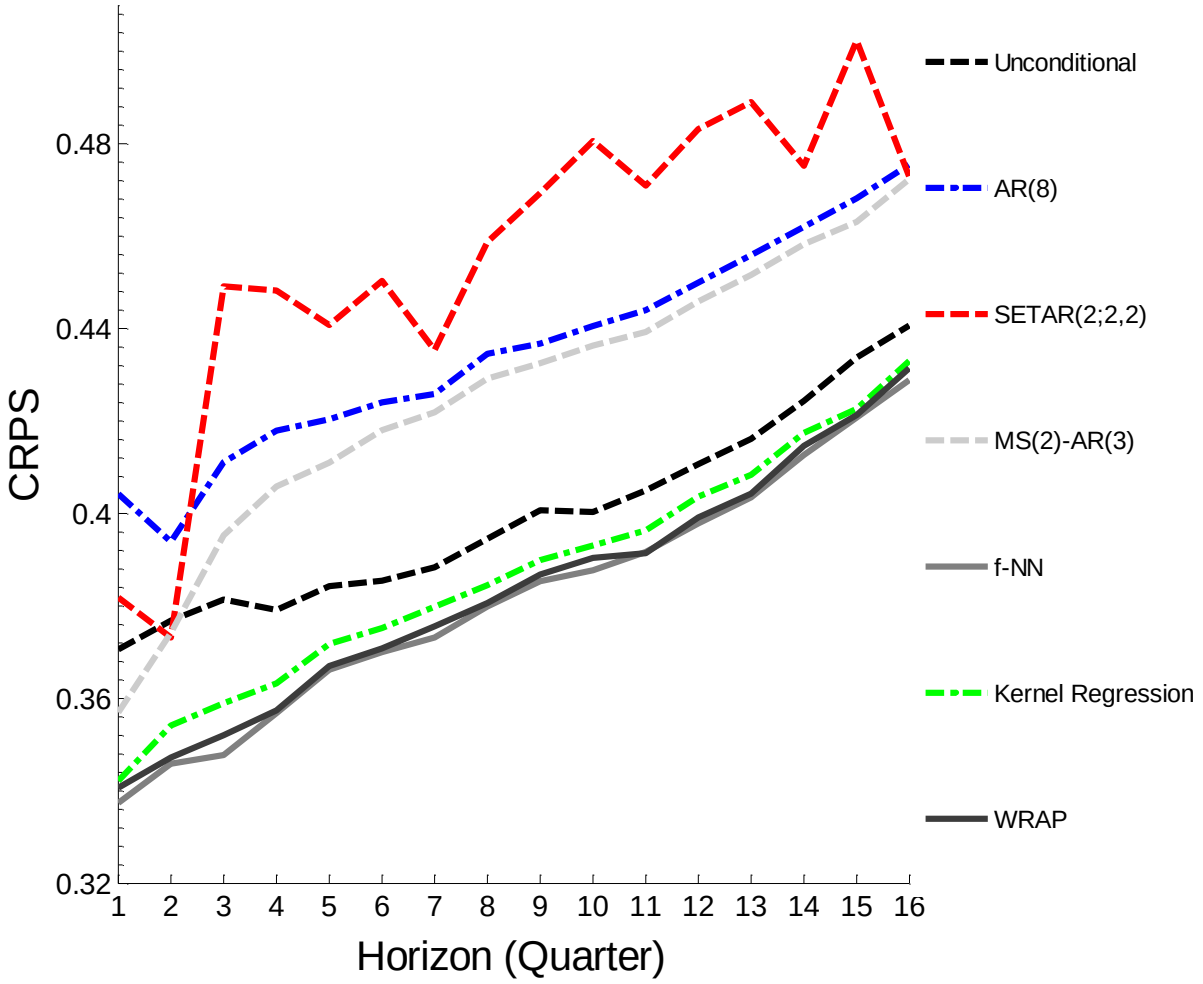

Figure 2.13— Density forecast performance of the parametric models (AR, SETAR and MS-AR), unconditional method and nonparametric models (*f*-NN, Kernel Regression and WRAP) for UK GDP time series (evaluation period: 1999Q1-2009Q4). Model evaluation based on the continuous ranked probability score (CRPS).

2.6 Summary and Concluding Remarks

We proposed a nonlinear and nonparametric model based on the estimation of few parameters that outperforms previously published models for forecasting the US GNP. The proposed model relies on very few assumptions regarding the functional form of the DGP to generate accurate forecasts, and can provide an efficient transition between the periods of growth

and recession. We find that neither the linear AR model, nor the classical regime switching models (SETAR and MS-AR) convincingly outperform a naïve point and density forecast based on the unconditional mean.

These results indicate the importance of employing the unconditional forecast as a benchmark instead of random walk benchmark for macroeconomic time series. Overall, a novel nonlinear and nonparametric method, WRAP, is found to provide the most accurate forecasts across multiple horizons on both GNP time series, when compared with the more complicated SETAR and MS-AR models, and simple benchmarks. Out of *f*-NN and kernel regression, we found that *f*-NN performs best and is highly competitive with the WRAP method. Despite its simplicity, the nonparametric models (*f*-NN, kernel regression and WRAP) outperform the more complex nonlinear models (SETAR(2;1,1) and MS(2)-AR(5) require 8 and 10 parameters, respectively) in terms of point and density forecasts.

Given the importance of quantifying uncertainty in the forecasts for an informed decision and policy-making, it is important to correctly assess candidate models, whether linear or nonlinear, based on their ability to generate probabilistic forecasts. We recommend that for generating accurate density forecasts, parameter estimation must be undertaken based on minimization of in-sample density forecast error, ensuring that the risk associated with decision-making based on such forecasts is adequately assessed.

Chapter 3

SHORT-TERM LOAD FORECASTING USING DISCOUNT WEIGHTED SVD-BASED EXPONENTIAL SMOOTHING

3.1 Introduction

The aim of this study is to propose novel methods for accurate short-term load (electricity demand) forecasting, based on a commonly used statistical *dimension reduction* technique, called singular value decomposition (SVD). The advantage of using SVD is that it transforms the data to an orthogonal space, and allows the focus to be on forecasting only the components that capture a major proportion of variance in the data, thereby reducing the dimensionality of the model to be considered. In this section, we discuss the motivation and challenges in modelling load, provide a literature review on the SVD-based methods used for forecasting intraday time series, and specify the contributions of this study.

3.1.1 Motivation and Challenges

Different operations within a power company require load forecasts for a range of different lead times. Load forecast can be classified into three types based on the lead times, namely: a) short-term, b) medium-term, and, c) long-term. Short-term load forecasts correspond to the lead times ranging from a few minutes up to one day ahead. Medium-term forecasts correspond to the lead times varying from a few days up to one year ahead, whereas for long-term forecasts, the lead times are longer than one year (Feinberg and Genethliou, 2005). Short-term forecasts of load are crucial for real-time scheduling of power systems, optimizing operational costs, and improving the reliability of distribution networks. Scheduling refers to the decision regarding which devices to operate at any given time and their power output. With the deregulation of

energy markets, short-term forecasts are needed by independent system operators for making informed market transactions, whereby inaccurate forecasts can have substantial financial implications (Charytoniuk and Chen, 2000; Bunn, 2000). Due to the impact of load forecasts on the reliability of power systems and its financial implications for energy markets, it is imperative to model load accurately. The medium-term load forecasts are needed by capacity markets and power systems, while long-term load forecasts are needed for planning new utilities (Feinberg and Genethliou, 2005).

Multivariate weather-based models have been employed previously for modelling load (Cottet and Smith, 2003; Dordonnat et al., 2008). The multivariate models utilize weather variables like temperature, wind speed, cloud cover, and humidity, along with the historical load observations. Univariate models, on the other hand, include only the historical load observations, and have been shown to be adequate for short-term load forecasting (Taylor, 2008). It has been argued that the weather variables tend to vary smoothly over short time scales, and this variation can be captured in the load data itself (Bunn, 2000). In an empirical study, Taylor (2008) reported that a univariate model outperformed a multivariate model for lead times up to about four hours ahead, while a combination of forecasts from the two methods was found to be the most accurate for lead times up to one day ahead. Hence, we employ univariate methods for short-term load forecasting in this study.

The presence of intraday and intraweek seasonal cycles is a prominent feature of load, and statistical time series methods aim to capture this 'double seasonality'. It has been shown recently by Taylor (2010a) that accommodating the intrayear seasonality in load, along with the intraday and intraweek seasonal cycles, leads to an improvement in short-term load forecast accuracy. The models proposed in this study are aimed at capturing the ‘triple seasonality’. The benefit of

accommodating the intrayear seasonality would potentially be more noticeable for medium and long-term load forecasting. As specified earlier, a major limitation with some of the existing methodology for load forecasting (Ramanathan *et al*., 1997; Cottet and Smith, 2003; Cancelo *et al*., 2008), however, is that it relies on building separate models for different periods of the seasonal cycle. Ramanathan *et al*. (1997) employ different models for each hour of the day. Using half-hourly load, Cottet and Smith (2003) build 48 regression-based models for each period of the day. Cancelo *et al*. (2008) use one daily model and 24 hourly models for load, whereby the daily model is used to forecast daily load for up to ten days ahead, while the hourly models are used to generate hourly forecasts for up to three days ahead. Building a separate model for each period of the seasonal cycle leads to a highly dimensional model, which is inefficient and less parsimonious. The aim of our study is to propose models that not only generate accurate short-term load forecasts, but also utilize the similarity between different periods in the seasonal cycle, leading to a reduction in the dimensionality of the model, such that the final modelling framework is both accurate and parsimonious.

For the half-hourly load data considered in this study, the intraday and intraweek cycle consists of 48 and 336 half-hourly periods, respectively. We represent the length of the intraday and intraweek seasonal cycles by m_1 and m_2, respectively. Hence, we have $m_1 = 48$ and $m_2 = 336$. For a non-leap year, the intrayear cycle consists of 17520 half-hourly periods, otherwise, it comprises 17568 half-hourly periods.

3.1.2 Overview of SVD

SVD is a commonly used statistical dimension reduction technique, which achieves dimension reduction by exploiting the correlations present between different variables in a multivariate dataset. To use SVD-based methods for modelling load, the data is first arranged in the form of a matrix. To exploit the similarity between different periods of the intraday seasonal cycle in the load data, we arrange the in-sample data (of length N) to form a matrix $\boldsymbol{Y}$ of size $(d \times m_1)$, where d is the total number of days in the in-sample data, and m_1 is the number of half-hourly periods in the intraday cycle. Hence, each row of $\boldsymbol{Y}$ consists of the daily load profile. The element in the ith row and j^{th} column of $\boldsymbol{Y}$, denoted by Y_{ij}, corresponds to the load observed on the ith day having an intraday period j.

Applying SVD to $\boldsymbol{Y}$ gives $\boldsymbol{Y'Y} = \boldsymbol{VSV'}$, where $\boldsymbol{V}$ is an orthogonal eigenvector matrix of size $(m_1 \times m_1)$, such that $\boldsymbol{V'V} = \boldsymbol{I} = \boldsymbol{VV'}$. Note that $\boldsymbol{Y'}$ denotes the transpose of matrix $\boldsymbol{Y}$. The matrix $\boldsymbol{S}$ is a positive definite diagonal matrix of size $(m_1 \times m_1)$, $\boldsymbol{S} = \text{diag}(s_1, s_2, \ldots, s_{m_1})$, where the diagonal elements, called *singular values*, are usually arranged in decreasing order, i.e., $s_1 \geq s_2 \geq \ldots \geq s_{m_1}$. The proportion of total variance in $\boldsymbol{Y}$ explained by a single column of $\boldsymbol{V}$, say v_j, is computed as $s_j / \sum_{i=1}^{m_1} s_i$. The data matrix $\boldsymbol{Y}$ is projected onto the $\boldsymbol{P}$ space via the columns of $\boldsymbol{V}$, using the relationship $\boldsymbol{P} = \boldsymbol{YV}$, where $\boldsymbol{P}$ is a matrix of size $(d \times m_1)$. Note that the original data matrix can be completely reconstructed using $\boldsymbol{Y} = \boldsymbol{PV'}$. Dimension reduction is achieved using only the first k columns of $\boldsymbol{V}$ and corresponding components ($k < m_1$), which capture a major proportion of the variance (or energy) in the data.

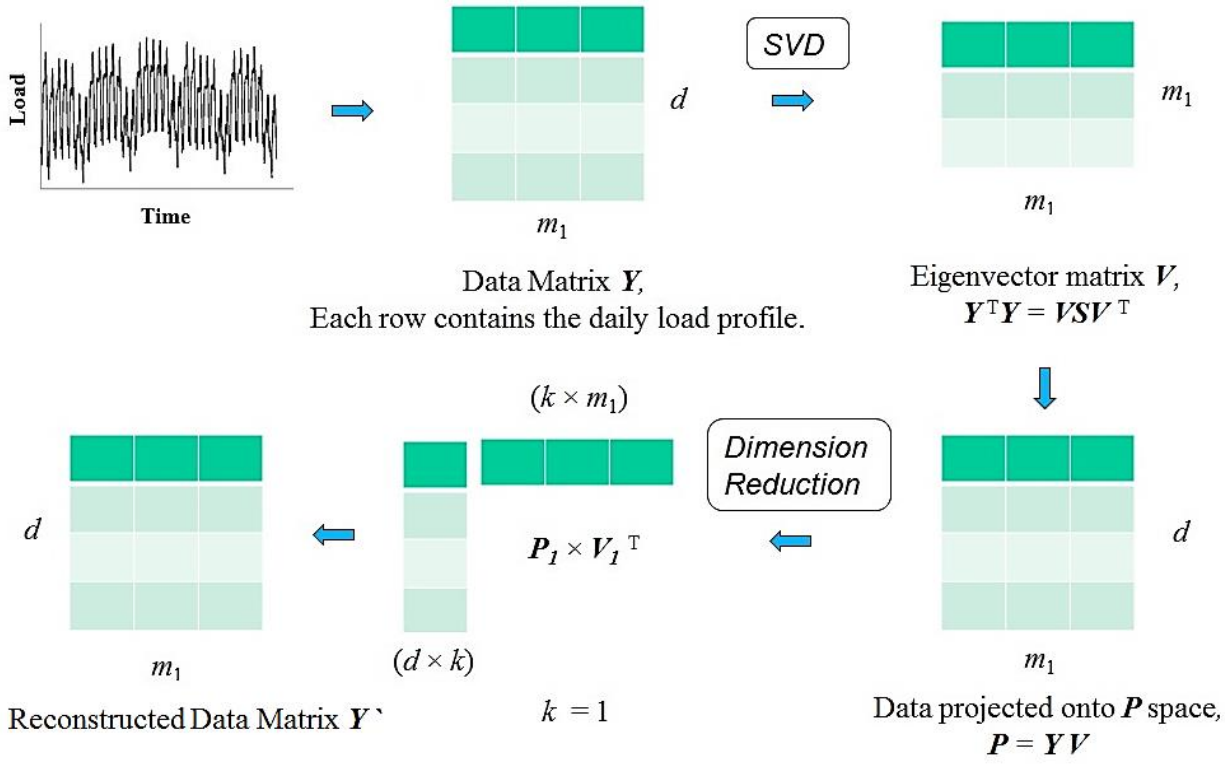

Figure 3.1— Pictorial representation of the different steps involved in applying SVD to the load data.

A pictorial representation of the above described SVD approach is provided in Figure 3.1. It can be seen from this figure, that the first step in applying SVD to load, involves converting the load time series to a data matrix $\boldsymbol{Y}$. We apply SVD to $\boldsymbol{Y}$ to obtain $\boldsymbol{V}$. Using columns of $\boldsymbol{V}$, $\boldsymbol{Y}$ is projected onto the $\boldsymbol{P}$ space. Using only the first columns of $\boldsymbol{P}$ and $\boldsymbol{V}$ (corresponding to the highest eigenvalue), we reconstruct an approximation of the original data matrix in the example shown in Figure 3.1. In this study, we select the number of columns of $\boldsymbol{P}$ and $\boldsymbol{V}$ to be used for modelling using cross-validation.

Note that when $\boldsymbol{Y}$ is arranged as a matrix of size $(d \times m_1)$, we refer to the columns of $\boldsymbol{V}$ as 'intraday feature vectors', while the columns of $\boldsymbol{P}$ are referred to as 'interday feature series'. On

the other hand, if $\boldsymbol{Y}$ is a matrix of size $(w \times m_2)$, the eigenvector matrix $\boldsymbol{V}$ is of size $(m_2 \times m_2)$, and the columns of $\boldsymbol{V}$ are referred to as 'intraweek feature vectors'. Using the relationship $\boldsymbol{P} = \boldsymbol{YV}$, we obtain matrix $\boldsymbol{P}$ of size $(w \times m_2)$, whereby the columns of $\boldsymbol{P}$ are referred to as 'interweek feature series'.

3.1.3 Historical Review

SVD-based methods have been employed previously for modelling the intraday data by Shen and Huang (2005, 2008a, 2008b), Taylor *et al*. (2006), and Taylor (2010b, 2012). Shen and Huang (2005) model the arrival rates for a call centre using SVD and a simple time series model (AR model with order one). A limitation of their proposed methodology is that it does not allow for intraday updating of the feature vectors and series based on the recent observations. To overcome this limitation, Shen and Huang (2008a) propose an extension of their previous model, whereby they employ dynamic intraday updating of the feature series using penalized least squares. Shen and Huang (2008b) use SVD for forecasting arrival volumes and arrival rates for call centres, assuming that the time series follows a Poisson process, and allow both the feature vectors and series to be updated with the inclusion of new observations. Taylor *et al.* (2006) apply SVD to the intraday load data, and build a regression model for the interday feature series using dummy variables for different days of a week.

A limitation of these SVD-based methods is that, in addition to SVD, the modelling framework relies on a regression-based model for updating the feature series. Taylor (2010b) proposes an intraday SVD-based exponential smoothing method, whereby the interday feature series are updated for all periods of the day using a single model, which makes it relatively easy to generate multi-step ahead forecasts. However, the model proposed by Taylor (2010b) requires

the initialization and updating of five different interday feature series, which are defined for different days of a week that exhibit a similar intraday cycle.

The intraweek SVD-based exponential smoothing model proposed in Taylor (2012) is simpler, since it employs a single interweek feature series to be initialized and updated for all periods of the day, across all days of the week. However, both the intraday and intraweek SVD-based methods do not allow for different periods of a seasonal cycle to be discounted at different rates.

3.1.4 Contributions

In this study, we propose two discount weighted (DW) SVD-based exponential smoothing methods that enable more recent observations to be assigned more weight in the dimension reduction. We propose the DW intraday and DW intraweek SVD-based exponential smoothing methods. We find that the proposed SVD-based methods lead to a substantial reduction in the dimensionality of the model to be considered, and are competitive with some of the most commonly used models for short-term load forecasting, namely, the seasonal Holt-Winters-Taylor (HWT) exponential smoothing, seasonal autoregressive moving average (SARMA), and the seasonal artificial neural network (ANN) method. A combination of the two proposed DW SVD-based methods generated the most accurate forecasts overall.

In this study, we use the models proposed by Taylor (2010b, 2012) as sophisticated benchmarks, and extend them by: (a) employing discounting in the calculation of the $\boldsymbol{V}$ matrix, and (b) using sequence updating for both feature series and feature vectors with the inclusion of new observations. Moreover, we accommodate triple seasonality in the existing model formulation of intraday and intraweek SVD-based exponential smoothing methods for the load data.

Though we only demonstrate the efficacy of the SVD-based methods using historical load observations, the proposed methods can be adapted to include weather variables, like temperature, wind speed, cloud cover and humidity. For a multivariate model, SVD-based approaches can be very useful in identifying and exploiting only the most relevant features of the dataset to be used for modelling. This can potentially result in substantial reduction in the dimension of the model, and the computation time associated with its implementation. It is noteworthy; however, that incorporating discount weighting and sequence updating within the SVD-based modelling framework can result in an increase in the computational complexity associated with the model.

3.2 Load Characteristics

In this study, we employ nine years of half-hourly load for Great Britain, stretching from 1 January 2001 to 31 December 2009. The data has been obtained from National Grid, which is the company responsible for the transmission of electricity in Great Britain. The data does not have any missing observations. We used the first eight years of data (consisting of 140,256 observations) for estimating the model parameters, and employed the final year (consisting of 17,520 observations) for model evaluation. Note that the data used for estimating the model parameters is referred to as the in-sample or training data. The data used for model evaluation is referred to as the out-of-sample or testing data. Using a rolling forecast origin through each period of the final year, we evaluate out-of-sample point forecasts, for lead times varying from one half-hour up to a day ahead.

The load for special days, e.g. public holidays and long weekends, witnesses a considerably lower load compared to other normal working days, around the same date. Following Taylor

(2010a), we smooth out the load for special days, by replacing it with the mean estimate of the normal load from the corresponding periods of the two adjacent weeks. Due to its distinct characteristics, load for special days need to be modelled separately than load for normal days. We develop models for modelling load for special days in the next chapter.

The complete smoothed load data is presented in Figure 3.2. It can be seen from this figure that load exhibits a recurring within-year pattern (due to seasonal effects), termed as the intrayear seasonality. The load in winter is higher than in summer, which is due to the increased use of electrical equipment for heating in Great Britain.

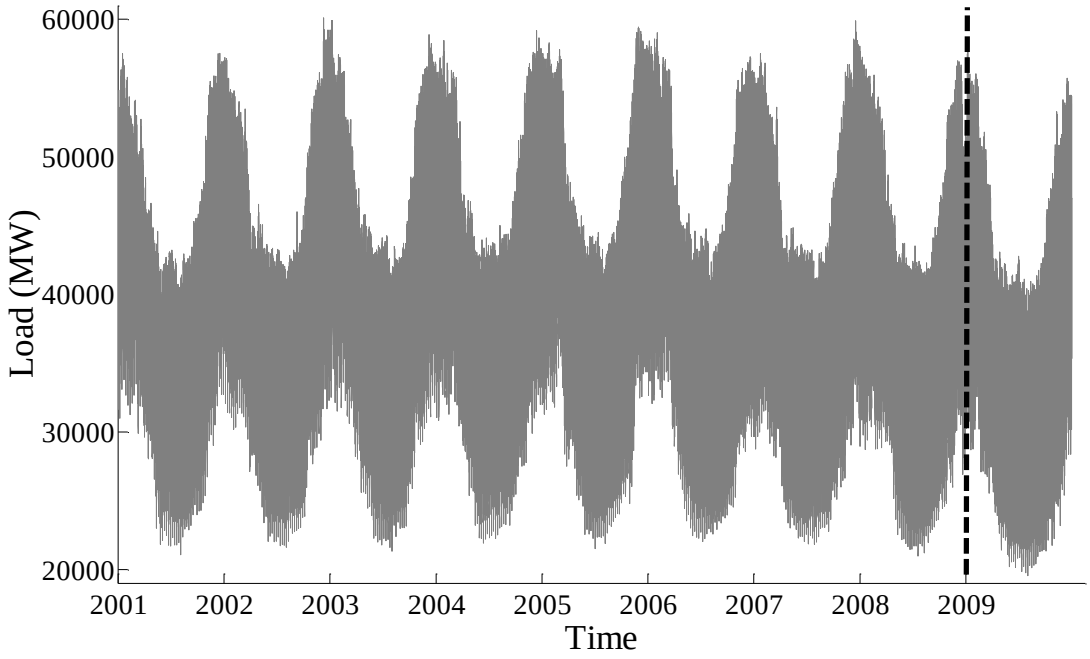

Figure 3.2— Half-hourly load for Great Britain stretching from 1 January 2001 to 31 December 2009. The vertical dashed line denotes the time index that divides the time series into non-overlapping estimation and evaluation periods.

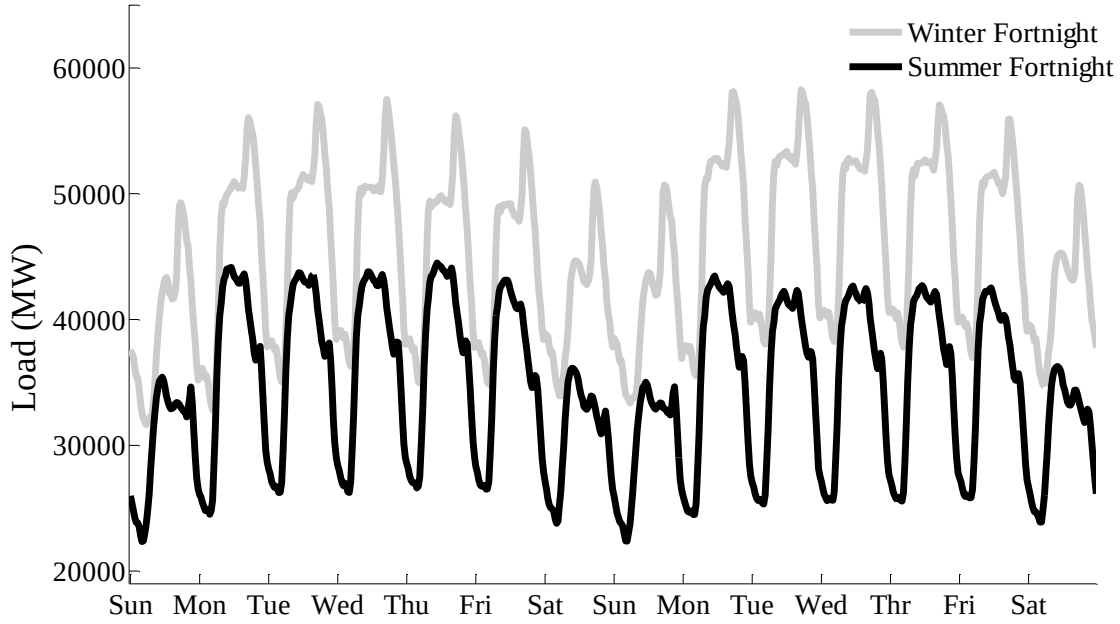

Figure 3.3— Load for a winter fortnight (16-29 January, 2005) and a summer fortnight (10-23 July, 2005).

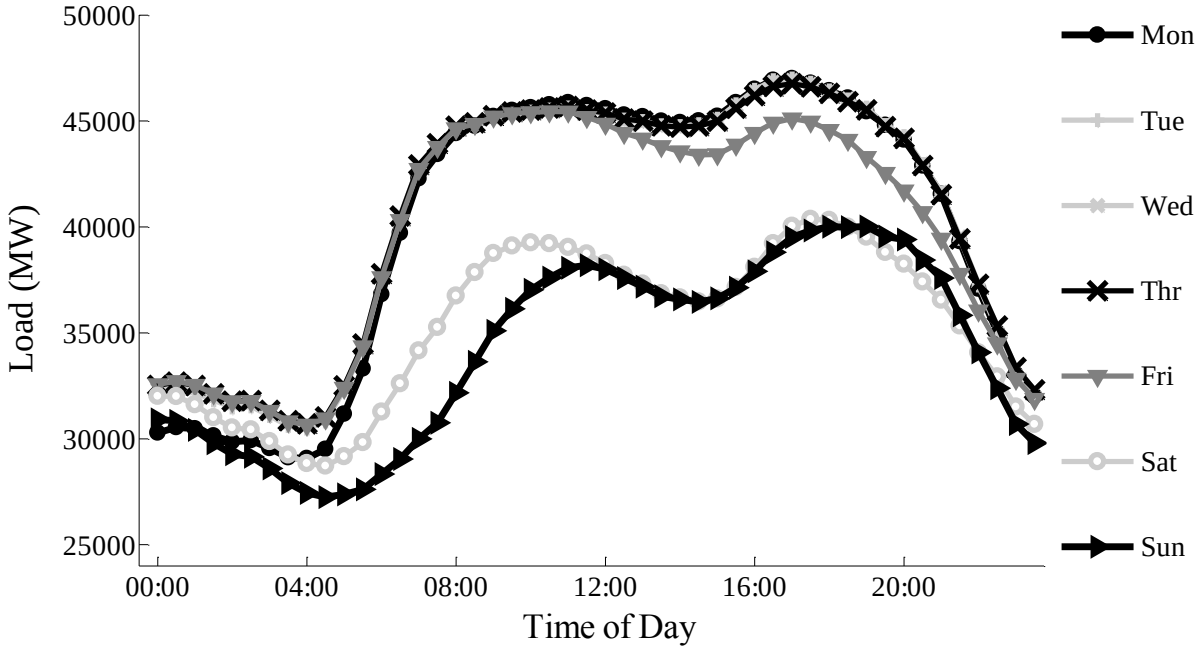

Figure 3.4— Average intraday profile for each day of the week, computed using only the estimation sample (2001-2008).

The seasonal difference in load is also evident in Figure 3.3, which presents load on a winter fortnight and a summer fortnight. The weekdays witness a higher load than the weekends for each fortnight. Within a fortnight, load is similar for the same day of the week. This within-week cycle is referred to as the intraweek seasonality, while the daily seasonal cycle is referred to as the intraday seasonality. The average intraday cycle for different days of the week is presented in Figure 3.4.

Figure 3.4 presents the average intraday load cycle for the different periods of a day, for each day of the week. This figure was calculated using only the estimation data. It can be seen from the figure that load on weekends is considerably lower than load on weekdays. Prior to modelling, we applied the natural log transformation to the data, in order to convert multiplicative effects to additive.

3.3 Forecasting Methods

In this section, we describe some of the commonly used univariate methods for short-term forecasting of load, namely, the HWT exponential smoothing, SARMA, and the ANN methods. We use these models as sophisticated benchmarks, against which the out-of-sample forecast accuracy of our SVD-based exponential smoothing methods is assessed.

3.3.1 HWT Exponential Smoothing Method

The single seasonal exponential smoothing method, referred to as the Holt-Winters (HW) exponential smoothing (Winters, 1960), was extended by Taylor (2003) in order to accommodate the intraday and intraweek seasonality in the intraday load data. We refer to this method as the

Holt-Winters-Taylor (HWT) exponential smoothing method. This method allows the inclusion of an intraday cycle to be nested within an intraweek cycle, and is represented as follows:

$$y_t = l_{t-1} + d_{t-m_1} + w_{t-m_2} + \phi e_{t-1} + \varepsilon_t \tag{3.1}$$

$$e_t = y_t - (l_{t-1} + d_{t-m_1} + w_{t-m_2}) \tag{3.2}$$

$$l_t = l_{t-1} + \lambda e_t \tag{3.3}$$

$$d_t = d_{t-m_1} + \delta e_t \tag{3.4}$$

$$w_t = w_{t-m_2} + \omega e_t \tag{3.5}$$

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$, and σ^2 denotes a constant variance. The smoothed level is given by l_t, while the log transformed load at time period t is represented as y_t. In our model formulations, we represent the length of the intraday and intraweek seasonal cycles by m_1 and m_2, respectively. We have $m_1 = 48$ and $m_2 = 336$. The term ϕ is an adjustment for first order autocorrelation in the error, denoted by e_t. This adjustment term was not included in the original HW exponential smoothing method. However, it was pointed out by Taylor (2003) that including the adjustment term leads to an improvement in the out-of-sample forecast performance of the HW exponential smoothing method. Hence, we include this adjustment term for the exponential smoothing methods considered in this study. The intraday seasonal index is denoted by d_t. The term w_t is defined as the seasonal index for the intraweek cycle, which remains after the intraday seasonality has been removed. The smoothing parameters λ, δ and ω determine the rate at which the level, intraday seasonal index and intraweek seasonal index are updated, respectively.

The double seasonal HWT exponential smoothing has recently been extended by Taylor (2010a) to accommodate the intraday, intraweek and intrayear seasonality in the load data. The incorporation of annual seasonality resulted in an improvement in short-term load forecast

accuracy. Hence, we consider the triple seasonal HWT exponential smoothing method in this study, which is represented as follows:

$$y_t = l_{t-1} + d_{t-m_1} + w_{t-m_2} + a_{t-m_3(t)} + \phi e_{t-1} + \varepsilon_t \tag{3.6}$$

$$e_t = y_t - (l_{t-1} + d_{t-m_1} + w_{t-m_2} + a_{t-m_3(t)}) \tag{3.7}$$

$$l_t = l_{t-1} + \lambda e_t \tag{3.8}$$

$$d_t = d_{t-m_1} + \delta e_t \tag{3.9}$$

$$w_t = w_{t-m_2} + \omega e_t \tag{3.10}$$

$$a_t = a_{t-m_3(t)} + \alpha e_t \tag{3.11}$$

where a_t is the additional term included in this method (compared to the double seasonal HWT method) that helps accommodate the intrayear seasonality in load. We closely followed Taylor (2010a) by setting the length of the intrayear seasonal cycle, for period t, to be $m_3(t) = 52 \times 336$, except for some years, where we set $m_3(t) = 53 \times 336$ for a few weeks around the clock-change. The seasonal terms d_t and w_t are defined in the same way as done for the double seasonal HWT method. The seasonal index a_t denotes the intrayear cycle, which remains after both the intraday and the intraweek seasonality have been removed. The smoothing parameters are denoted by λ, δ, ω and α.

In our implementation, the parameters for the seasonal exponential smoothing methods were estimated based on minimization of the one-step-ahead sum of squared errors (SSE) using only the in-sample data. Once estimated, the parameters were held fixed, and the estimated model was employed for generating forecasts for the out-of-sample data.

The advantage of the triple seasonal HWT exponential smoothing method is that it accommodates all three seasonal cycles in the load data, is conceptually simple to understand and easy to implement, and is competitive with some of the most commonly used models for short-term load forecasting, such as SARMA and ANNs, see Taylor (2010a). The disadvantage of this method, however, is that it requires the initialization and updating of the level, m_1 periods of the intraday cycle, m_2 periods of the intraweek cycle, and $m_3(t)$ periods of the intrayear cycle. This translates to initialization and updating of $(1 + m_1 + m_2 + m_3(\mathrm{t})) = 17857$ terms, for $m_3(t) = 52 \times 336$. This makes the triple seasonal HWT exponential smoothing method highly dimensional. Also, this method does not utilize the similarity between different periods in the seasonal cycle, which suggests that it is inefficient.

3.3.2 Seasonal ARMA

The seasonal ARMA belongs to the Box and Jenkins (1970) class of models, and has been commonly used for generating short-term forecasts of load (Soares and Medeiros, 2008; Taylor, 2010a). The single seasonal ARMA, with intraweek seasonality, is represented as follows:

$$\Upsilon_p(L)X_{P_2}(L^{m_2})(y_t - c) = \Omega_q(L)\Gamma_{Q_2}(L^{m_2})\varepsilon_t \tag{3.12}$$

where y_t is the log transformed load at time t, c is a constant; ε_t is the model error; L is the lag operator, Υ_p and X_{P_2} are polynomial functions for the AR terms having order p and P_2, respectively. The polynomial functions for the MA terms are denoted by Ω_q and Γ_{Q_2}, having corresponding model order given by q and Q_2. In the above model formulation (equation 3.12), we accommodate only the intraweek seasonality. It can be argued that if the intraweek seasonality is adequately modelled, there would be no need to separately model the intraday

seasonality. However, we also describe a model that incorporates both intraday and intraweek seasonality.

The above formulation of the single seasonal ARMA can be extended for double seasonality by including additional polynomial functions for the AR and MA terms. The double seasonal ARMA with the intraday and intraweek seasonality is represented as:

$$\Upsilon_p(L)\Phi_{P_1}(L^{m_1})X_{P_2}(L^{m_2})(y_t - c) = \Omega_q(L)\Theta_{Q_1}(L^{m_1})\Gamma_{Q_2}(L^{m_2})\varepsilon_t \tag{3.13}$$

where Φ_{P_1} and Θ_{Q_1} are the additional polynomial functions (compared to the single seasonal ARMA), which accommodate the intraday seasonality in the load data, and have order P_1 and Q_1, respectively. Taylor (2010a) proposed triple seasonal ARMA for modeling load, which is represented as:

$$\begin{aligned}\Upsilon_p(L)\Phi_{P_1}(L^{m_1})X_{P_2}(L^{m_2})\Psi\left(L^{m_3(t)}\right)(y_t - c) = \\ \Omega_q(L)\Theta_{Q_1}(L^{m_1})\Gamma_{Q_2}(L^{m_2})\Lambda\left(L^{m_3(t)}\right)\varepsilon_t\end{aligned} \tag{3.14}$$

where Ψ and Λ are the additional polynomial functions (compared to the double seasonal ARMA). The functions Ψ and Λ essentially capture the annual cycles, whereby Ψ is written as:

$$\begin{aligned}\Psi\left(L^{m_3(t)}\right) = 1 + \eta_1 L^{m_3(t)} + \eta_2 L^{m_3(t)+m_3\left(t-m_3(t)\right)} \\ + \eta_3 L^{m_3(t)+m_3\left(t-m_3(t)\right)+m_3\left(t-m_3\left(t-m_3(t)\right)\right)}\end{aligned} \tag{3.15}$$

where η_1, η_2 and η_3 are constant coefficients. The function Λ includes the same lags as Ψ, but comprises different coefficients. Using load for Great Britain, it was reported by Taylor (2010a) that employing log transforming of the data, as a pre-processing step for SARMA, did not lead to any improvement in the model's forecast accuracy, compared to the case when SARMA was applied to the actual data. Hence, in this study, we do not apply log transformation to the data for

SARMA. The model parameters for the SARMA models used in this study were estimated using maximum likelihood with a Gaussian error, employing only the in-sample data. We considered orders equal to or less than three. The model orders were selected using the Box and Jenkins (1970) methodology.

3.3.3 Seasonal ANN method

ANNs provide a nonlinear and nonparametric approach for time series modelling, as opposed to the HWT exponential smoothing and SARMA models, both of which rely on a linear and parametric modelling framework. The advantage of using ANNs for load forecasting is that they are able to model the complex nonlinear relationships between load and weather variables, without making prior assumptions about the functional form of the data generating process. The disadvantage of using ANNs is that there is no well-established systematic approach, or a consensus among researchers, for choosing a suitable ANN architecture, namely, the number of inputs, hidden layers, and units within each hidden layer, for a given dataset. To choose a suitable ANN method for this data, we closely follow the ANN architecture used by Taylor (2010a).

Specifically, we employ a feed-forward ANN method with a single hidden layer and a single output. ANNs are basically a weighted directional graph, connecting the input variables to the output via different links. If the weighted graph has no loops, it is called a feed-forward network; else, it is called a feed-back network. In the context of regression and classification problems, the inputs in an ANN method are basically the independent variables, while the output is the dependent variable. Since ANNs have been shown to be unsuitable for generating multi-step ahead forecasts (Atiya *et al*., 1999), we build a separate ANN model for each forecast horizon.

As a pre-processing step, we employ a differencing operator of the form $(1 - L^{m_1})(1 - L^{m_2})$, as used in Taylor (2010a), which led to an improvement in forecast accuracy. We difference the output using this operator, and normalize it by subtracting the mean, and dividing by the standard deviation. The input comprises lagged load observations, differenced using the above operator, and normalized to have zero mean and unit standard deviation. We select the lags for input variables to be consistent with the SARMA model. Specifically, for the ANN built for horizon h, we used load at the forecast origin, and at the following lags: $1, 2, m_1 - h, 2m_1 - h, 3m_1 - h, m_2 - h, 2m_2 - h, 3m_2 - h, LagA = m_3(t) - h, LagB = LagA + m_3(t - LagA - h),$ and $LagC = LagB + m_3(t - LagB - h)$, where $m_3(t)$ denotes the magnitude of the intrayear cycle length defined at the period t. This model can accommodate triple seasonality in load, and is represented as follows:

$$\hat{y}_t = f(\boldsymbol{x}_t, \boldsymbol{w}, \boldsymbol{v}) = \sum_{j=1}^{m} \boldsymbol{v}_j \left(\frac{1}{1 + e^{-(\sum_{i=1}^{k} w_{ji} \boldsymbol{x}_{it})}} \right) \tag{3.16}$$

where $\hat{y}_t$ is the corresponding prediction for the actual output y_t, $\boldsymbol{x}_t$ is the input vector of length k comprising lagged load observations (whereby the lags are selected as described above), $\boldsymbol{w}$ is a vector comprising weights for links that connect inputs to the units of the hidden layer, such that w_{ji} denotes the weight of a link that connects the input i to a unit j in the hidden layer, while m denotes the total number of units in the hidden layer. Similarly, $\boldsymbol{v}$ is a vector, comprising weights for links that connects m units in the hidden layer to the output. We use a sigmoid activation function that nonlinearly maps the inputs to the units in the hidden layer. For mapping the values in the units of the hidden layer to the output, we use a linear activation function. The weight

vectors $\boldsymbol{w}$ and $\boldsymbol{v}$ are estimated by minimizing the one-step ahead SSE on the in-sample data, using the following objective function:

$$min_{\boldsymbol{w},\boldsymbol{v}}\left(\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2+\lambda_1\sum_{j=1}^{m}\sum_{i=1}^{k}{w_{ji}}^2+\lambda_2\sum_{j=1}^{m}{v_j}^2\right) \tag{3.17}$$

where n is the total number of in-sample observations used for minimizing the objective function, λ_1 and λ_2 are the regularization parameters, which are included to avoid the model overfitting the data (Bishop, 1997). The regularization parameters prevent the network weights from becoming too large. Computing the output from the inputs is termed as forward-pass. Propagating the error from the output layer back to the input layer, and using it for updating the network weights is termed backpropagation.

We employ the backpropagation algorithm with learning rate η and momentum parameter μ to minimize the above objective function. The parameter η determines how fast or slow the weights change with respect to the output error. The parameter μ gives more weight to the more recent network weights, and is similar to the smoothing parameter used in the exponential smoothing methods. We estimate the model parameters using cross-validation, employing a hold-out sample corresponding to the last one year of the estimation sample.

3.4 SVD-based Exponential Smoothing Methods

We describe the intraday and intraweek SVD-based exponential smoothing methods, and propose new extensions through the use of discount weighting and sequence updating, and also via accommodating the intrayear seasonality in load.

3.4.1 Intraday SVD-based Exponential Smoothing

This method was proposed by Taylor (2010b) for forecasting the volume of call centre arrivals. Using this method, SVD is applied to the data arranged as a matrix $\boldsymbol{Y}$ of size $(d \times m_1)$. As we described earlier, the columns of matrix $\boldsymbol{V}$ (obtained by applying SVD to $\boldsymbol{Y}$) are referred to as 'intraday feature vectors', and the columns of matrix $\boldsymbol{P}$ (obtained using $\boldsymbol{P} = \boldsymbol{YV}$) are referred to as 'interday feature series'. Based on the intraday cycle for different days of a week (as shown in Figure 3.4), this method treats a week as consisting of five distinct intraday cycle types. Specifically, this method assumes Tuesday, Wednesday and Thursday to have a common intraday cycle, whereas Monday, Friday, Saturday and Sunday are each assumed to have a distinct intraday cycle. The first three intraday feature vectors and interday feature series are plotted in Figure 3.5 and Figure 3.6, respectively. The intraday feature vectors show component elements of the intraday seasonal patterns, such that the first intraday feature vector seems to exhibit the intraday cycle averaged across all days of the week, while the corresponding interday feature series shows the intraweek cycle. The second and third intraday feature vector show variations in the intraday patterns (due to, say, the weekend effect), and the corresponding interday feature series exhibit variations in the intraweek cycle over time.

The intraday SVD-based exponential smoothing method achieves dimension reduction by utilizing the similarity between different periods in the intraday cycle. This method is represented as follows:

$$y_t = \sum_{i=1}^{5} I_{it} \boldsymbol{p}_{t-1}^{(i)} \widetilde{\boldsymbol{v}}'_{[t \bmod m_1]} + \phi\, e_{t-1} + \varepsilon_t \tag{3.18}$$

$$e_t = y_t - \textstyle\sum_{i=1}^{5} I_{it} \boldsymbol{p}_{t-1}^{(i)} \widetilde{\boldsymbol{v}}'_{[t \bmod m_1]} \tag{3.19}$$

$$\boldsymbol{p}_t^{(i)} = \boldsymbol{p}_{t-1}^{(i)} + (\lambda \mathbf{1}_{m_1} \widetilde{\boldsymbol{V}} + (\delta + \omega I_{it}) \widetilde{\boldsymbol{v}}_{[t \bmod m_1]}) e_{\mathrm{t}} \tag{3.20}$$

$$I_{it} = \begin{cases} 1 \text{ if period } t \text{ occurs on a day of type } i \\ 0 \text{ otherwise} \end{cases} \tag{3.21}$$

where $\varepsilon_t \sim \mathrm{NID}(0, \sigma^2)$, $\boldsymbol{p}_t^{(i)}$ is a row vector of length k that consists of the elements in the first k interday feature series in period t, for a day having an intraday cycle of type i, $\widetilde{\boldsymbol{V}}$ is the matrix of size $(m_1 \times k)$ containing the first k intraday feature vectors, and $\widetilde{\boldsymbol{v}}_{[t \bmod m_1]}$ is a row of matrix $\widetilde{\boldsymbol{V}}$ having index $[t \bmod m_1]$. The index $[t \bmod m_1]$ identifies the period of the day on which t falls. The smoothing parameters are given by λ, δ and ω. The term ϕ is an adjustment term for autocorrelation in the error, denoted by e_{t}, while $\mathbf{1}_{m_1}$ is a row vector of ones having length m_1. The first term on the right side of equation 3.18, models y_t using the interday feature series $\boldsymbol{p}_t^{(i)}$ and the intraday feature vector $\widetilde{\boldsymbol{v}}_{[t \bmod m_1]}$. Note that, as already described, the data matrix $\boldsymbol{Y}$ can be reconstructed exactly using $\boldsymbol{Y} = \boldsymbol{P}\boldsymbol{V}'$. However, we employ only k $(k < m_1)$ feature vectors and feature series, which allows for dimension reduction. Equation 3.19 computes the error, which is used for projecting $\boldsymbol{V}$ onto $\boldsymbol{P}$ space.

The smoothing parameters in equation 3.20 are similar to the smoothing parameters used in the HWT exponential smoothing framework, and are used to update the interday feature series. The smoothing term λ adjusts the level of $\boldsymbol{p}_t^{(i)}$, the parameter δ updates the elements of $\boldsymbol{p}_t^{(i)}$ based upon their relationship with the period of the day on which t falls. This helps accommodate the intraday seasonality in load. Using equation 3.20, the parameter ω is used to capture the intraweek seasonal effects, and is used to update the elements of $\boldsymbol{p}_t^{(i)}$ only when t occurs on a day of type i. The advantage of this method is that it uses a single model to update the interday

feature series, as opposed to building separate time series models for the feature series, as done by Shen and Huang (2005, 2008a, 2008b).

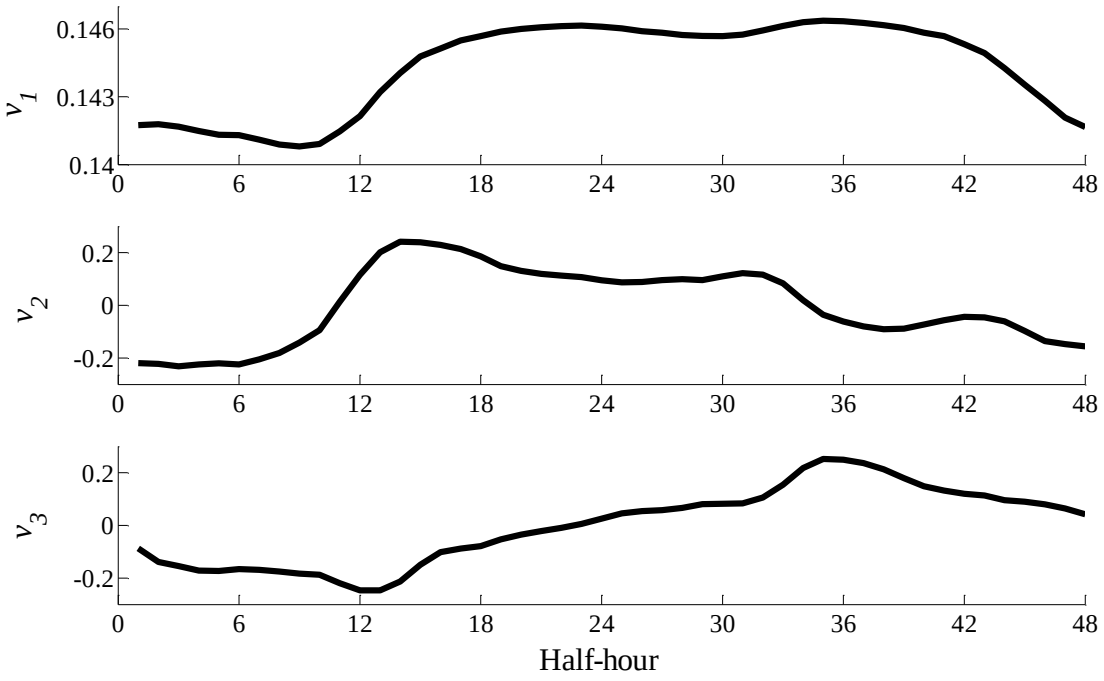

Figure 3.5— First three intraday feature vectors for the British log load time series.

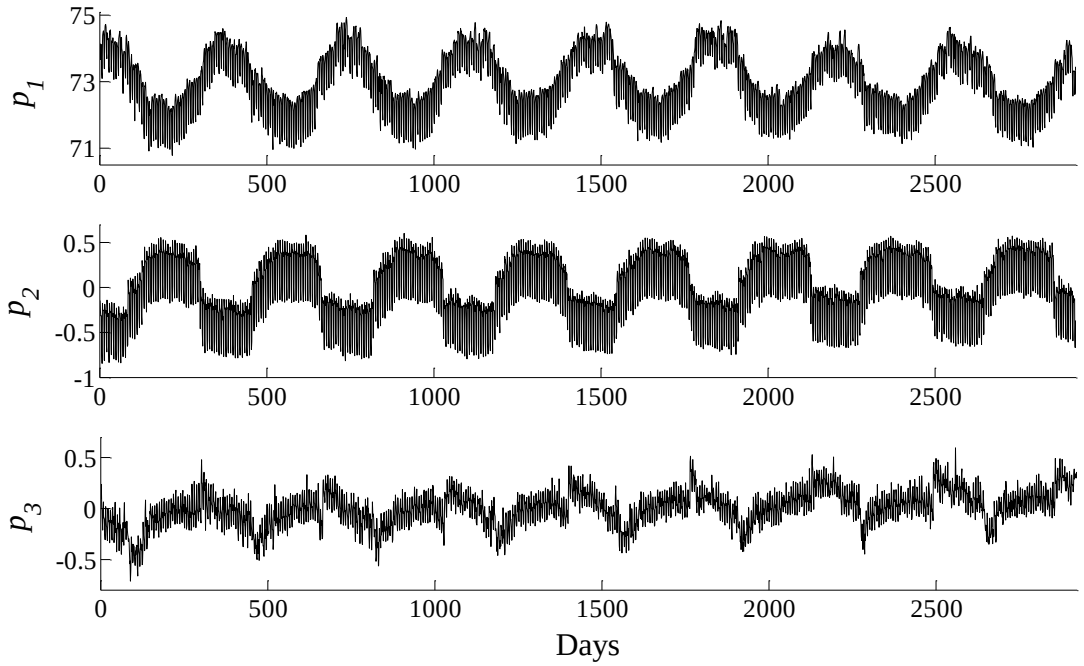

Figure 3.6— First three interday feature series for the British log load time series.

3.4.2 Intraweek SVD-based Exponential Smoothing

Recently, Taylor (2012) proposed the intraweek SVD-based exponential smoothing method for modelling short-term load. A limitation of the intraday SVD method is that it requires the initialization and updating of five different interday feature series $\boldsymbol{p}_t^{(i)}$ for different days of the week. Note that in the previous method, a week is treated as comprising five different intraday cycles. On the contrary, the intraweek SVD-based exponential smoothing method employs a single interweek feature series $\boldsymbol{p}_t$ for all days of the week, resulting in a much simpler model. In this method, the load data is arranged in the form of a matrix $\boldsymbol{Y}$ of size $(w \times m_2)$. Applying SVD to $\boldsymbol{Y}$ yields a matrix $\boldsymbol{V}$ of size $(m_2 \times m_2)$, the columns of which are referred to as 'intraweek feature vectors'. Using $\boldsymbol{Y}$ and $\boldsymbol{V}$, we get $\boldsymbol{P}$ which is a matrix of size $(w \times m_2)$, the columns of which are referred to as 'interweek feature series'. We select only the k feature vectors and series, where $k < m_2$, and project them back onto the $\boldsymbol{Y}$ space to generate forecasts. The first three feature vectors and series are plotted in Figure 3.7 and 3.8. The intraweek feature vectors show variations in the intraweek patterns, whereby the first intraweek feature vector shows the average intraweek cycle. The corresponding interweek feature series shows the intrayear cycle.

The intraweek SVD-based exponential smoothing method achieves dimension reduction by utilizing the similarity between different periods in the intraweek cycle. This method is represented as follows:

$$y_t = \boldsymbol{p}_{t-1}\widetilde{\boldsymbol{v}}'_{[t \bmod m_2]} + \phi\, e_{t-1} + \varepsilon_t \tag{3.22}$$

$$e_t = y_t - \boldsymbol{p}_{t-1}\widetilde{\boldsymbol{v}}'_{[t \bmod m_2]} \tag{3.23}$$

$$\boldsymbol{p}_t = \boldsymbol{p}_{t-1} + \left(\lambda \mathbf{1}_{m_2}\widetilde{\boldsymbol{V}} + \left(\delta \sum_{j=1}^{7} \widetilde{\boldsymbol{v}}_{[t \bmod m_1]+(j-1)m_1}\right) + \omega \widetilde{\boldsymbol{v}}_{[t \bmod m_2]}\right) e_t \tag{3.24}$$

where $\varepsilon_t \sim NID(0, \sigma^2)$, $\boldsymbol{p}_t$ is a row vector of length k comprising values in the first k interweek feature series in period t, $\widetilde{\boldsymbol{V}}$ is the matrix of size $(m_2 \times k)$ consisting of the first k intraday feature vectors, $\widetilde{\boldsymbol{v}}_{[s]}$ is a row of $\widetilde{\boldsymbol{V}}$ having index s. The parameter for error correction is denoted by ϕ. The smoothing parameters are λ, δ and ω, and $\mathbf{1}_{m_2}$ is a row vector of ones having length m_2. Equation 3.22 projects $\boldsymbol{p}_{t-1}$ onto the $\boldsymbol{Y}$ space using a column of $\widetilde{\boldsymbol{V}}$ having index $[t \bmod m_2]$, whereby this index selects the intraweek feature vector depending upon the period of the week on which t falls. In equation 3.24, the smoothing parameter λ updates the level of $\boldsymbol{p}_t$ for each period. The smoothing parameter δ takes into account the intraday seasonal effects on load by updating the elements of $\boldsymbol{p}_t$ based upon their relationship with the period of the day on which t falls, while ω accommodates the intraweek seasonal effects on load, and updates $\boldsymbol{p}_t$ based on the period of the week on which t falls.

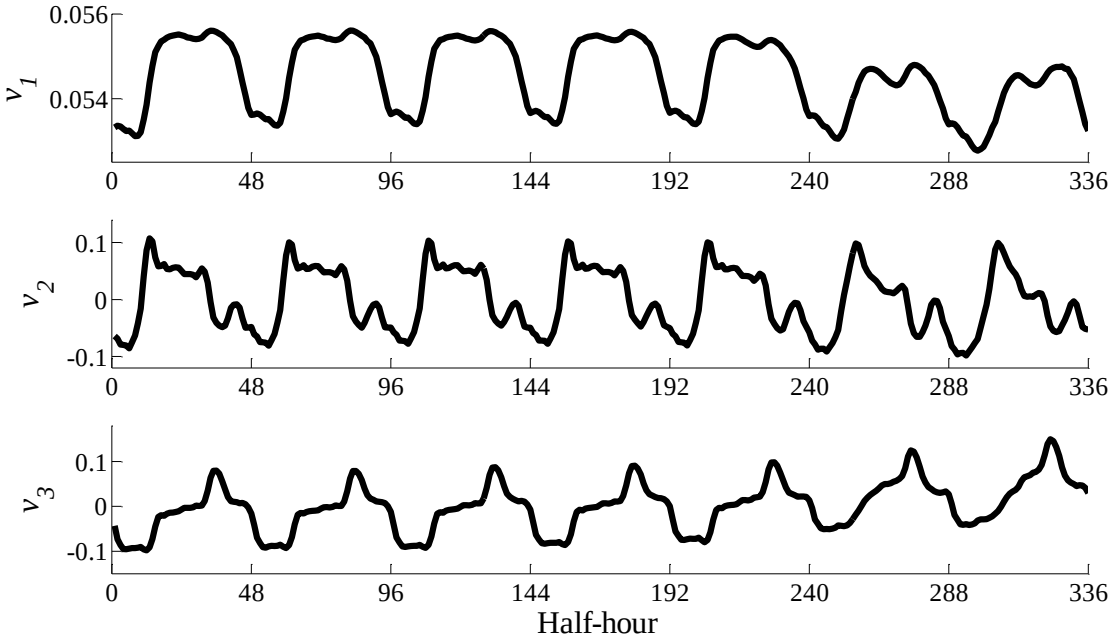

Figure 3.7— First three intraweek feature vectors for the British log load time series.

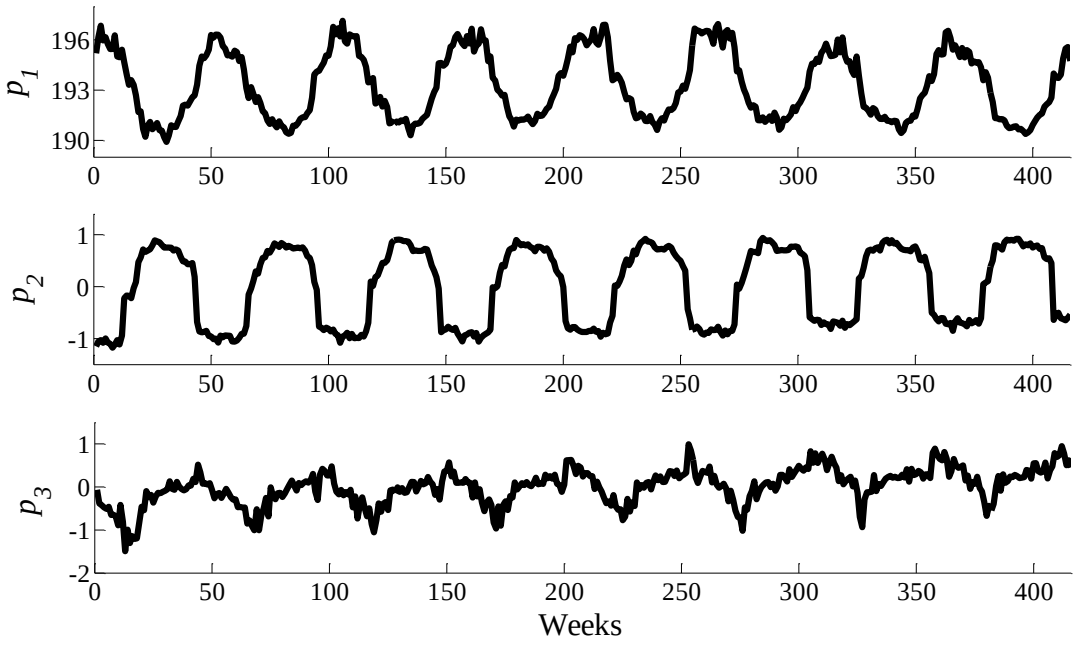

Figure 3.8— First three interweek feature series for the British log load time series.

The advantages of the intraday and intraweek SVD-based exponential smoothing methods are that they can lead to dimensionality reduction, whilst simultaneously accommodating the seasonal effects on load. The disadvantage, however, lies in the fact that only the feature series $\boldsymbol{p}_t$ is updated for different periods, while the matrix of feature vectors $\boldsymbol{V}$ is computed only once (using SVD), and then held fixed. The original intraday and intraweek SVD based methods are static, which means that they compute $\boldsymbol{V}$ just once, and hence, the model does not have the flexibility to adapt with the inclusion of new observations. Furthermore, the above discussed SVD-based methods do not accommodate the intrayear seasonality in the load data. To overcome these limitations, we propose extensions of the intraday and intraweek SVD-based methods. Specifically, we propose two methods that: (1) employ sequence updating for both feature vectors $\boldsymbol{V}$ and series $\boldsymbol{p}_t$, in order to adapt with the inclusion of new observations, and, (2) accommodate triple seasonality in the existing model formulation of intraday and intraweek SVD-based exponential smoothing methods. The two proposed models are referred to as the

discount weighted intraday SVD-based exponential smoothing method, and the discount weighted intraweek SVD-based exponential smoothing method.

3.4.3 Discount Weighted Intraday SVD-based Exponential Smoothing

As described earlier, the intraday SVD-based exponential smoothing method calculates $\boldsymbol{V}$ once by applying SVD to the matrix $\boldsymbol{Y}$, and updates only the feature series $\boldsymbol{p}_t^{(i)}$ using equation 3.20. Note that in the previous methods, SVD is applied to the data matrix, whereby $\boldsymbol{Y}$ was formed using the complete in-sample data. The rationale of our proposed method lies in updating both feature vectors $\boldsymbol{V}$ and feature series $\boldsymbol{p}_t$, and allowing more recent observations to be assigned higher weight through the use of discount factors. To update $\boldsymbol{V}$, a natural approach is to update $\boldsymbol{Y}'\boldsymbol{Y}$. Hence, instead of applying SVD to the whole matrix $\boldsymbol{Y}$, this method calculates $\boldsymbol{V}$ using $(\boldsymbol{Y}'\boldsymbol{Y})_j$, where $(\boldsymbol{Y}'\boldsymbol{Y})_j$ is computed using only the first j rows of the data matrix $\boldsymbol{Y}$. Once a new row of observations become available, we compute $(\boldsymbol{Y}'\boldsymbol{Y})_{j+1}$, and re-calculate $\boldsymbol{V}$ using the updated $(\boldsymbol{Y}'\boldsymbol{Y})_{j+1}$ matrix. Figure 3.9 plots the first column of $\boldsymbol{V}$ (first intraday feature vector), each time $\boldsymbol{V}$ is updated, with the inclusion of a new row of observations. This figure shows changes in the first intraday feature vector over time.

Figure 3.10 plots the first intraweek feature vector over time. It can be seen from Figures 3.9 and 3.10 that columns of $\boldsymbol{V}$ are not constant. This justifies our approach for updating both feature vectors and series with the inclusion of new observations.

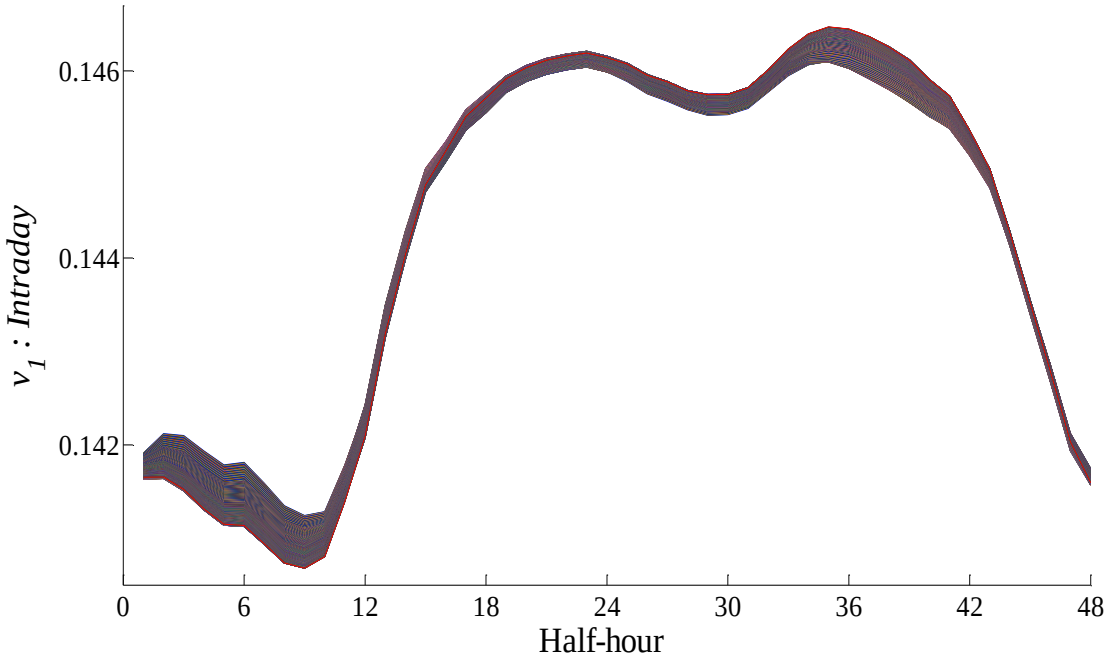

Figure 3.9— First intraday feature vector plotted over time.

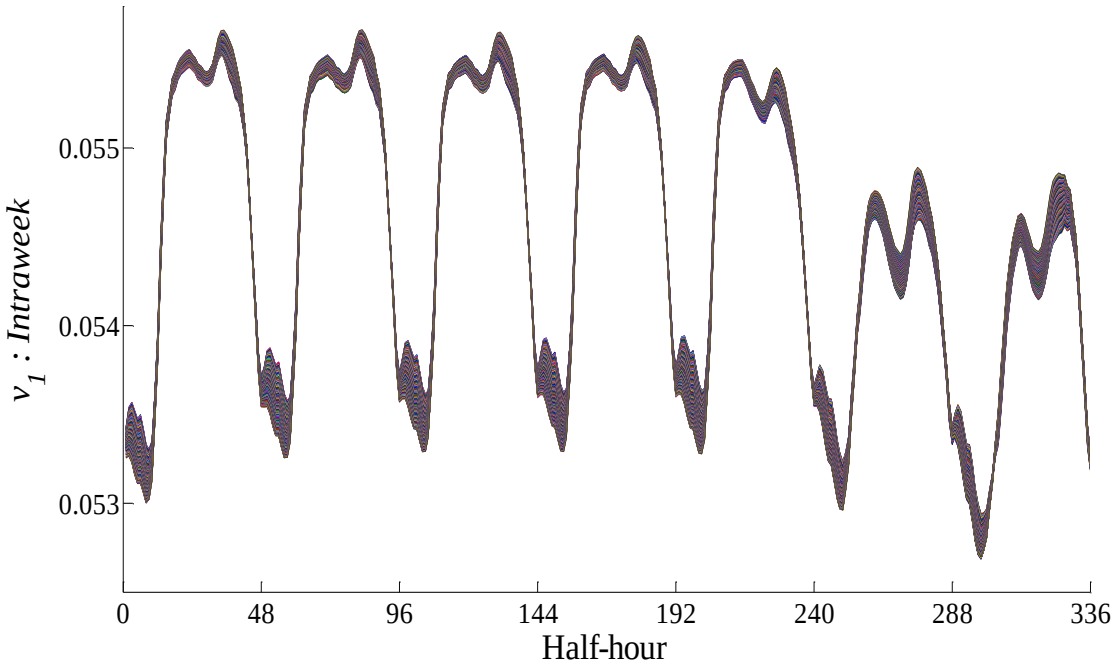

Figure 3.10— First intraweek feature vector plotted over time.

In order to place more weight on the more recent observations, Wold (1994) proposed exponentially weighted principal component analysis (EWPCA). The main idea of EWPCA is to use a smoothing factor in the computation of the matrix $\boldsymbol{Y'Y}$. However, a limitation of EWPCA is that it employs a single smoothing factor for all periods of the seasonal cycle. To overcome this limitation, we use discount weighting. The idea of discount weighting is based on discount weighted regression (DWR), see Ameen and Harrison (1984). Hence, we first discuss DWR, after which we describe the proposed methods.

The general idea behind DWR is to allow the model parameters to be time-varying (Ameen and Harrison, 1984). DWR can be seen as an extension of exponentially weighted regression (EWR), see Brown (1963). The limitation of EWR is that different model parameters are discounted at the same rate using a single discount factor. To overcome this limitation, Ameen and Harrison (1984) and Harrison and Johnston (1984) propose DWR, which employs different discount factors for each model parameter. In the context of this work, the use of DWR translates to different periods of a seasonal cycle being discounted at different rates. Specifically, using this method, we use the data matrix $\boldsymbol{Y}$ of size $(d \times m_1)$, and compute, $(\boldsymbol{Y'Y})_j = \boldsymbol{\gamma}^{\frac{1}{2}}(\boldsymbol{Y'Y})_{j-1}\boldsymbol{\gamma}^{\frac{1}{2}} + (\boldsymbol{y^T y})_j$, where $(\boldsymbol{Y'Y})_{j-1}$ is a matrix of size $(m_1 \times m_1)$, computed using only $j-1$ rows (and m_1 columns) of $\boldsymbol{Y}$. The vector $\boldsymbol{y}$ is the new row of m_1 observations that becomes available for updating the $\boldsymbol{Y'Y}$ matrix, while $\boldsymbol{\gamma}^{\frac{1}{2}}$ is a diagonal matrix of size $(m_1 \times m_1)$ consisting of discount factors, i.e. $\boldsymbol{\gamma}^{\frac{1}{2}}$ = diag$(\gamma_1^{\frac{1}{2}}, \gamma_2^{\frac{1}{2}}, \ldots, \gamma_{m_1}^{\frac{1}{2}})$, such that $\gamma_i^{\frac{1}{2}} \in (0, 1)$ for all i. The discount factors that lie along the diagonal of the discount factor matrix $\boldsymbol{\gamma}$, allow the different 'periods of the day' to be discounted at different rates.

In order to allow for sequence updating of the intraday feature vectors and series, we first compute the $\boldsymbol{Y'Y}$ matrix using the first year of the estimation sample, and then update this matrix with the inclusion of each new row of observations. Updating the matrix $(\boldsymbol{Y'Y})_j$ with the inclusion of new observations, and re-computing PCA to obtain a new set of feature vectors and series has previously been discussed previously (Wold, 1994; Dayal and MacGregor, 1997; Li *et al*., 2000; Lane *et al*., 2003). In addition to employing sequence updating of the feature vectors and series, we accommodate the triple seasonality in load by including an additional term within the framework of previously described methods based on dimension reduction. The proposed SVD-based method is represented as follows:

$$(\boldsymbol{Y'Y})_j = \boldsymbol{\gamma}^{\frac{1}{2}}(\boldsymbol{Y'Y})_{j-1}\boldsymbol{\gamma}^{\frac{1}{2}} + (\boldsymbol{y^T y})_j \tag{3.25}$$

$$(\boldsymbol{Y'Y})_j = \boldsymbol{VSV'} \tag{3.26}$$

$$y_t = \sum_{i=1}^{5} I_{it}\boldsymbol{p}_{t-1}^{(i)}\,\widetilde{\boldsymbol{v}}'_{[t\ mod\ m_1]} + \phi\, e_{t-1} + \varepsilon_t \tag{3.27}$$

$$e_t = y_t - \sum_{i=1}^{5} I_{it}\boldsymbol{p}_{t-1}^{(i)}\,\widetilde{\boldsymbol{v}}'_{[t\ mod\ m_1]} \tag{3.28}$$

$$\boldsymbol{p}_t^{(i)} = \boldsymbol{p}_{t-1}^{(i)} + \left(\lambda \mathbf{1}_{m_1}\widetilde{\boldsymbol{V}} + (\delta + \omega I_{it})\widetilde{\boldsymbol{v}}_{[t\ mod\ m_1]}\right)e_t + a_{t-m_{3(t)}+1} \tag{3.29}$$

$$a_t = a_{t-m_{3(t)}} + \alpha \mathbf{1}_{m_1}\widetilde{\boldsymbol{V}}e_t \tag{3.30}$$

$$I_{it} = \begin{cases} 1 \text{ if period } t \text{ occurs on a day of type } i \\ 0 \text{ otherwise} \end{cases} \tag{3.31}$$

where a_t accommodates the intrayear seasonality in load, and is referred to as the intrayear seasonal index. The intrayear seasonal index is added to the interday feature series for all periods of the day, across all day types. As described earlier, $(\boldsymbol{Y'Y})_j$ is a matrix of size $(m_1 \times m_1)$, computed using only $j-1$ rows and m_1 columns of $\boldsymbol{Y}$. The rest of the parameters used in this method have the same interpretation as the parameters employed in the intraday SVD-based

exponential smoothing method (equations 3.18-3.21). Using SVD, we compute $\boldsymbol{V}$ each time the $(\boldsymbol{Y}'\boldsymbol{Y})_j$ matrix is updated with the inclusion of a new row of observations (equation 3.26). The interweek feature series $\boldsymbol{p}_t^{(i)}$ is re-initialized completely once the newly estimated $\boldsymbol{V}$ matrix becomes available (using equation 3.29). Re-initializing $\boldsymbol{p}_t^{(i)}$ ensures that the interweek feature series are derived from the latest estimates of $\boldsymbol{V}$. Using this framework, it is possible to update the $(\boldsymbol{Y}'\boldsymbol{Y})_{j-1}$ matrix with the inclusion of each new observation, using appropriate normalizing factors. However, to avoid the high computational costs involved with frequent updating of the matrix $(\boldsymbol{Y}'\boldsymbol{Y})_j$ and re-initialization of $\boldsymbol{p}_t^{(i)}$, we update the $(\boldsymbol{Y}'\boldsymbol{Y})_j$ matrix once a new full row of observations becomes available.

3.4.4 Discount Weighted Intraweek SVD-based Exponential Smoothing

This method relies on estimating the $(\boldsymbol{Y}'\boldsymbol{Y})_j$ matrix from the data arranged as a matrix $\boldsymbol{Y}$ of size $(w \times m_2)$. The discount factor matrix $\boldsymbol{\gamma}$ of size $(m_2 \times m_2)$ allows for the different 'periods of the week' to be discounted at different rates. The matrix $(\boldsymbol{Y}'\boldsymbol{Y})_j$ is updated with the inclusion of a new row of observations, from which $\boldsymbol{V}$ and $\boldsymbol{P}$ are computed using SVD.

This method is represented as follows:

$$(\boldsymbol{Y}'\boldsymbol{Y})_j = \boldsymbol{\gamma}^{\frac{1}{2}}(\boldsymbol{Y}'\boldsymbol{Y})_{j-1}\boldsymbol{\gamma}^{\frac{1}{2}} + (\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y})_j \tag{3.32}$$

$$(\boldsymbol{Y}'\boldsymbol{Y})_j = \boldsymbol{V}\boldsymbol{S}\boldsymbol{V}' \tag{3.33}$$

$$y_t = \boldsymbol{p}_{t-1}\tilde{\boldsymbol{v}}'_{[t\ mod\ m2]} + \phi\, e_{\mathrm{t}-1} + \varepsilon_{\mathrm{t}} \tag{3.34}$$

$$e_{\mathrm{t}} = y_t - \boldsymbol{p}_{t-1}\tilde{\boldsymbol{v}}'_{[t\ mod\ m_2]} \tag{3.35}$$

$$\boldsymbol{p}_t = \boldsymbol{p}_{t-1} + \left(\lambda \mathbf{1}_{m_2}\tilde{\boldsymbol{V}} + \left(\delta \sum_{j=1}^{7} \tilde{\boldsymbol{v}}_{[t\ mod\ m_1]+(j-1)m_1}\right) + \omega\tilde{\boldsymbol{v}}_{[t\ mod\ m_2]}\right)e_{\mathrm{t}} + a_{t-m_3(t)+1} \tag{3.36}$$

$$a_t = a_{t-m_{3(t)}} + \alpha \mathbf{1}_{m_2} \tilde{\boldsymbol{V}} e_{\mathrm{t}} \tag{3.37}$$

where the intrayear seasonal index a_t accommodates the annual seasonality in load (as explained previously). The rest of the parameters in this method have the same interpretation as the parameters used in the intraweek SVD-based exponential smoothing method (equations 3.22-3.24). The $(\boldsymbol{Y}'\boldsymbol{Y})_{j-1}$ matrix is updated using equation 3.32, such that the data matrix $\boldsymbol{Y}$ consists of $j-1$ rows and m_2 columns, and $\boldsymbol{y}$ is the new row of m_2 observations. For each new $\boldsymbol{V}$ that is computed with the inclusion of new observations, the interweek feature series are re-initialized completely using equation 3.36. The model parameters for all the above discussed SVD-based exponential smoothing methods were estimated by minimizing the one-step ahead SSE on the in-sample data. We used the in-build nonlinear optimization in Matlab.

3.5 Results and Observations

We generated point forecasts for the one-year post-sample period, and evaluated accuracy using the mean absolute percentage error (MAPE) and root mean squared percentage error (RMSPE) given by:

$$MAPE_h = 100 \times \frac{1}{N-T-h+1} \sum_{i=T+h}^{N} \left| \frac{\exp(y_i) - \exp(\hat{y}_i)}{\exp(y_i)} \right| \tag{3.38}$$

$$RMPSE_h = 100 \times \sqrt{\frac{1}{N-T-h+1} \sum_{i=T+h}^{N} \left(\frac{\exp(y_i) - \exp(\hat{y}_i)}{\exp(y_i)} \right)^2} \tag{3.39}$$

where $MAPE_h$ and $RMPSE_h$ denote the MAPE and RMSPE at horizon h, respectively, while $\exp(y_i)$ is the actual load, $\exp(\hat{y}_i)$ is the corresponding forecast, T is the forecast origin and N is

the length of the time series. The relative model rankings were similar for the two measures; hence we present results only for MAPE in this chapter.

The parameter estimates for the SVD-based exponential smoothing methods considered in this study are presented in Table 3.1. We find that the estimate for λ is low, whereas the value of the parameter ϕ is high for all SVD-based methods. Note that λ is the smoothing term for the level, whereas ϕ is the adjustment term for the first-order autocorrelations in the error. These estimates indicate that capturing autocorrelations in the error accommodates changes in the level to a large extent.

For the DW SVD-based methods, we investigated using two discount factors within the modelling framework. Specifically, for the DW intraday SVD-based exponential smoothing method, we used one discount factor for the half-hourly periods belonging to the first-half of the day, and another discount factor for the remaining periods. For the DW intraday SVD-based exponential smoothing method, we used one discount factor for the half-hourly periods falling on the weekdays, and another discount factor for the half-hourly periods occurring on the weekends. However, using multiple discount factors did not result in considerable improvement in the load forecast accuracy. Hence, we present results using one discount factor γ in this study. For both the DW intraday and DW intraweek SVD-based exponential smoothing methods, we updated $\boldsymbol{V}$ with the inclusion of each new row of observations.

Table 3.1 shows that the original and discount weighted intraday SVD-based exponential smoothing methods use only 23 and 18 feature vectors and series (out of 48), respectively. On the other hand, the original and DW intraweek SVD-based exponential smoothing methods employ only 35 and 23 feature vectors and series (out of 336), which implies a considerable

reduction in the dimensionality of the model. Compared to the intraday SVD-based models, the reduction in dimensionality is more pronounced for the intraweek SVD-based models. Similar results have been reported earlier by Taylor (2012). This may potentially be due to the presence of a higher percentage of more similar periods within the intraweek cycle, compared to the intraday cycle. Note that the model parameters, including the number of feature vectors and series (denoted by k), were estimated by minimizing the SSE on only the in-sample data.

TABLE 3.1

PARAMETER ESTIMATES FOR THE ORIGINAL AND DISCOUNT WEIGHTED VERSIONS OF THE INTRADAY AND INTRAWEEK SVD-BASED EXPONENTIAL SMOOTHING METHODS.

Parameters	***Orig. Intraday SVD***	***DW Intraday SVD***	***Orig. Intraweek SVD***	***DW Intraweek SVD***
λ	0.01	0.02	0.01	0.08
δ	0.31	0.28	0.56	0.55
ω	0.37	0.36	0.37	0.32
α	-	0.05	-	0.13
ϕ	0.94	0.95	0.94	0.90
γ	-	0.999	-	0.990
k	23	18	35	23

For the triple seasonal HWT exponential smoothing method, we estimated $\lambda = 0.02, \delta = 0.34, \omega = 0.01, \alpha = 0.32$ and $\phi = 0.94$. For the triple seasonal SARMA, the model order was selected using the Box and Jenkins (1970) methodology.

The estimated SARMA model is represented as:

$$(1 - 0.23L - 0.46L^2 - 0.24L^3)(1 - 0.18L^{m_1} - 0.13L^{2m_1} - 0.42L^{3m_1})(1 - 0.43L^{m_2} - 0.01L^{2m_2} - 0.02L^{3m_2})\Big(1 - 0.36L^{m_3(t)} - 0.24L^{m_3(t)+m_3(t-m_3(t))} - 0.16L^{m_3(t)+m_3(t-m_3(t))+m_3\left(t-m_3(t-m_3(t))\right)}\Big)(\exp(y_t) - 37872) = (1 + 0.98L + 0.52L^2 + 0.12L^3)(1 + 0.02L^{m_1} - 0.33L^{3m_1})(1 - 0.34L^{m_2} + 0.03L^{2m_2} + 0.01L^{3m_2})\Big(1 - 0.11L^{m_3(t)} - 0.11L^{m_3(t)+m_3(t-m_3(t))} - 0.08L^{m_3(t)+m_3(t-m_3(t))+m_3\left(t-m_3(t-m_3(t))\right)}\Big)\varepsilon_t \quad (3.40)$$

For the triple seasonal ANN method, we obtained $\lambda_1 = 0.001, \lambda_2 = 0.001, \eta = 0.1, \mu = 0.9$ and $m = 24$.

Figure 3.11 presents the MAPE obtained using different versions of the SVD-based exponential smoothing methods. The DW intraday SVD method is more accurate than its original counterpart at all lead times, whereas the DW intraweek SVD method is the most accurate method for lead times up to about six hours ahead, after which its performance worsens. Since, the DW intraweek SVD method is competitive at shorter lead times, while the DW intraday SVD method is one the best methods at longer lead times, we consider combining forecasts from these two DW SVD-based methods using a simple average. The results presented in Figure 3.11 demonstrate that a simple combination of the proposed methods generates the most accurate forecasts across all lead times.

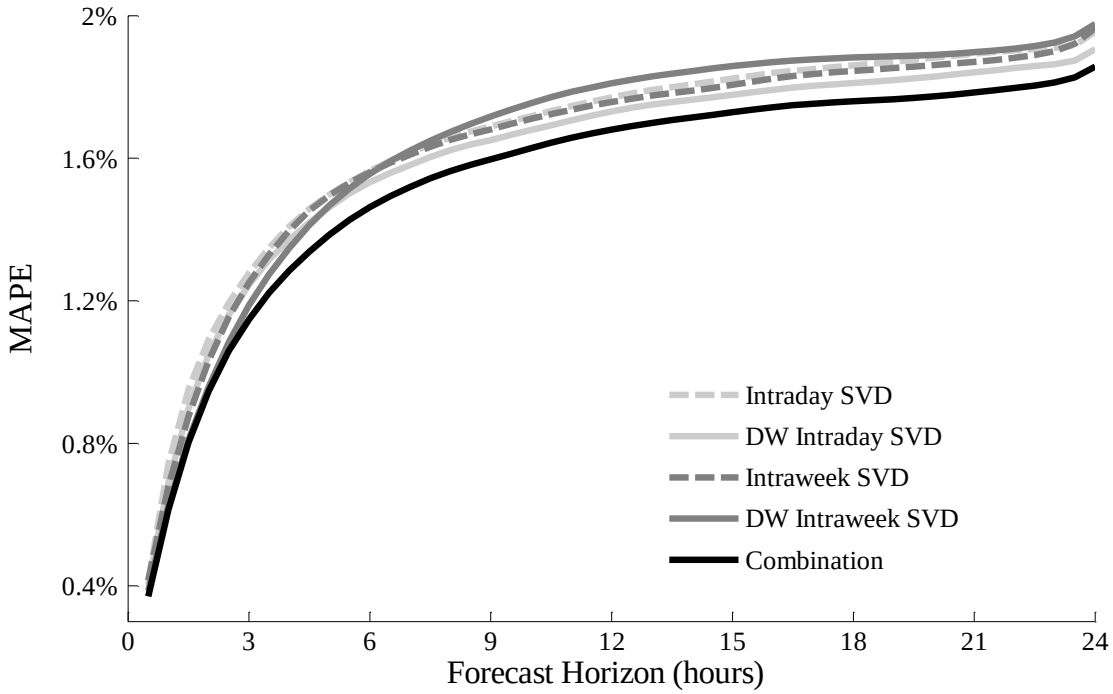

Figure 3.11— Out-of-sample MAPE for the original and discount weighted versions of the intraday and intraweek SVD-based exponential smoothing methods. This figure also presents the MAPE obtained using a combination of the proposed discount weighted SVD-based methods.

We also include two simple benchmarks in this study, namely, the seasonal random walk (SRW) and the seasonal moving average (SMA). For a given period using the SRW, we issue the observation from the corresponding period from the previous week as a forecast. Hence, a h-step ahead forecast (for $h \leq m_2$) using the seasonal random walk is computed as:

$$\hat{y}_{t+h} = y_{t+h-m_2} \tag{3.41}$$

Using the SMA, a simple average of the historical observations belonging to the corresponding period from the previous few weeks is issued as a forecast for a given period. In this study, we use the average of corresponding observations from the previous four weeks to compute the h-step ahead forecast, given as:

$$\hat{y}_{t+h} = (y_{t+h-m_2} + y_{t+h-2m_2} + y_{t+h-3m_2} + y_{t+h-4m_2})/4 \tag{3.42}$$

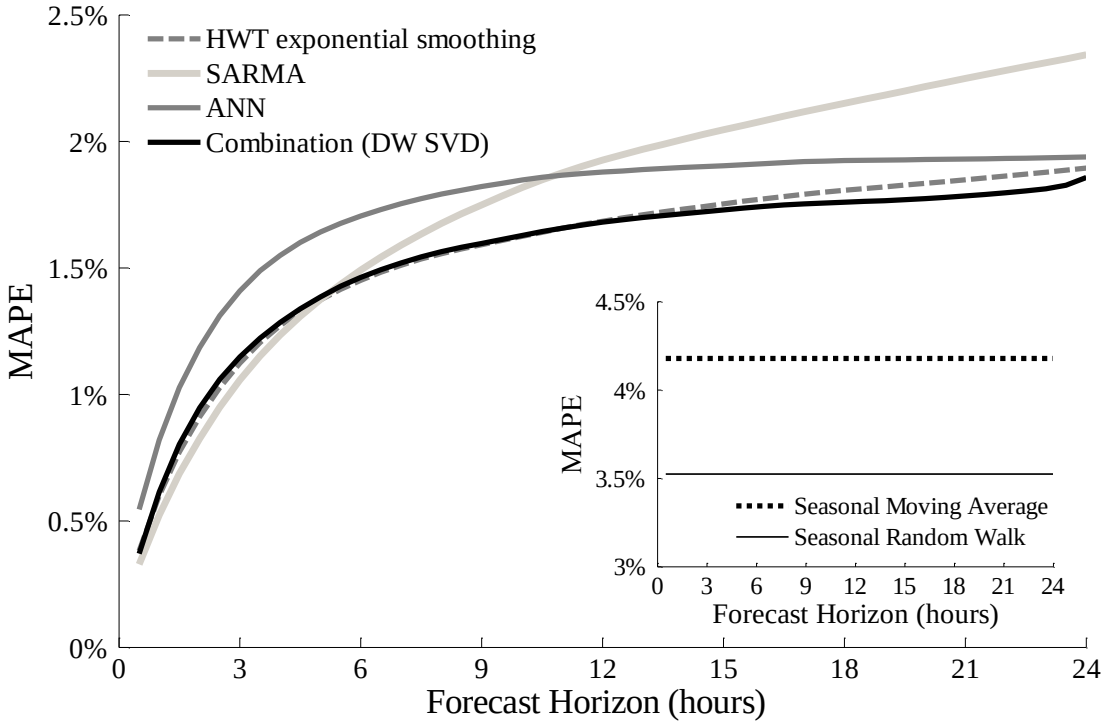

Figure 3.12— Out-of-sample MAPE for the triple seasonal versions of the HWT exponential smoothing, SARMA, the ANN method, and a combination of the two proposed methods, namely, discount weighted intraday and discount weighted intraweek SVD-based exponential smoothing methods. The MAPE for the simple benchmarks, namely, seasonal moving average and seasonal random walk, are presented in the inset.

Figure 3.12 presents the MAPE obtained using the triple seasonal variants of the HWT exponential smoothing, SARMA, the ANN method, simple benchmarks, and the combination of the DW intraday and the DW intraweek SVD-based exponential smoothing methods. The worst performing methods are the benchmarks (as shown in the inset, in Figure 3.12). For lead times up to about six hours, the forecast performance for SARMA, HWT exponential smoothing and a combination of the DW SVD-based methods is similar, with SARMA being slightly more accurate.

The ANN method is less accurate compared to both the HWT exponential smoothing method, and a combination of the DW SVD-based methods. This may be due to lack of nonlinearity in the structure of the load time series. For longer lead times (greater than twelve hours), SARMA is the worst performing method amongst the triple seasonal univariate methods considered in this study. The performance of HWT exponential smoothing and a combination of the DW SVD-based methods is very similar, with the latter being marginally more accurate at longer lead times.

In addition to SVD, we investigated the use of an alternate dimension reduction technique, the independent component analysis (ICA), see (Comon, 1994). It is noteworthy that SVD considers only the cross-correlations between data elements, and hence, completely ignores the higher order moments. On the other hand, ICA is based on the concept of mutual independence between different signals, and thus, it takes into account the higher order moments during modelling (Comon, 1994). Lu *et al*. (2009) used ICA and support vector regression for forecasting financial time series. However, we are not aware of any existing studies that have employed ICA for modelling load time series. Due to these reasons, we investigated using ICA for short-term load forecasting.

Specifically, we investigated intraday and intraweek ICA-based exponential smoothing, whereby we applied ICA (instead of SVD) to the data matrix in order to obtain the feature vector and series. For implementing ICA, we used the fastICA algorithm, which takes into account the higher order moments (up to fourth order moments), see (Hyvärinen and Oja, 2000). For updating the feature vector and series obtained using ICA, we used the same modelling framework as used for the SVD-based exponential smoothing methods. The forecasts obtained

using the ICA-based methods were found to be very similar with the SVD-based methods. We aim to further investigate the use of ICA for short-term load forecasting.

3.6 Summary and Concluding Remarks

Our aim has been to generate accurate load forecasts through parsimonious modelling approaches, which utilize the similarity in different periods of the seasonal cycle. The methodological contribution of this study lies in proposing extensions of the intraday and intraweek SVD-based exponential smoothing methods. The proposed methods led to a reduction in the dimension of the data, while a simple combination of the two proposed discount weighted SVD methods was competitive with the commonly used univariate modelling approaches for load. It is noteworthy that incorporating discount weighting and sequence updating in a model's framework can significantly improve its forecast accuracy, especially if there is considerable variability in the time series. Hence, for future work, we aim to investigate the efficacy of the proposed discounted weighted methods on datasets that exhibit both seasonality and noticeable variability, including time series for the volume of call center arrivals and hospital arrivals. Moreover, it would be worth employing the wavelet transform (Daubechies, 1992) as a tool for dimension reduction for the load data, in order to investigate if methods based on alternate dimension reduction techniques can generate more accurate short-term load forecasts, compared to the methods used in this study. Moreover, for horizons longer than a day, and in cases where weather predictions may be easily available, it would be interesting to extend the proposed univariate methods, to also include weather variables as additional explanatory variables, and employ dimension reduction techniques to select only the most relevant variables in the new feature space during the modelling of load data.

Chapter 4

SHORT-TERM FORECASTING OF ANOMALOUS LOAD USING RULE-BASED TRIPLE SEASONAL METHODS

4.1 Introduction

The aim of this study is to generate accurate forecasts for short-term load, under both normal and anomalous load conditions. The load observed on normal working days is referred to as *normal load*, whereas the load observed on special days is referred to as *anomalous load*. Occurring on special days, such as public holidays and long weekends, anomalous load conditions pose considerable modelling challenges due to their infrequent occurrence and significant deviation from normal load. Numerous methods have been proposed for forecasting load for normal working days, however, the modelling of anomalous load has often been ignored in the research literature. In this section, we discuss the motivation and challenges in modelling anomalous load, and provide the contributions of this study.

4.1.1 Motivation and Challenges

A variety of statistical methods have been proposed for short-term load forecasting, whereby the lead times under consideration vary from a few minutes up to a day ahead, see (Smith, 2000; Hippert *et al.*, 2005, Taylor, 2010a; Bunn, 2000). For short-term load forecasting, the conventional statistical time series methods aim to capture the 'double seasonality' in load, focussing only on the intraday and intraweek seasonality. Such methods have been shown to perform well for normal days. However, they are unlikely to be of use for special days with anomalous load, such as bank holidays and long weekends. The load *profile* (shape of intraday load curve) differs for different types of special day. Also, for a given special day, such as a

particular bank holiday, the profile may vary from one year to the next, based on the day of week and the precise time of the year (date) on which the special day occurs. Furthermore, the relatively infrequent occurrence of special days leads to the unavailability of an adequate number of observations required for sufficiently training the model. Due to these complexities, forecasting load under anomalous conditions has largely been omitted from time series forecasting methods (Smith, 2000; Hippert *et al.*, 2005), and has hence been left to the experience and judgment of the central controller of the electricity grid (Rahman and Bhatnagar, 1988; Hyde and Hodenett, 1994). The presence of intraday, intraweek and intrayear seasonal cycles is a prominent feature of load. Hence, the methods proposed in this study accommodate not only intraday and intraweek seasonal cycles, but also intrayear seasonality.

In this study, we propose a rule-based approach for load forecasting. It has been shown that, when domain knowledge is available and the time series has a consistent structure (seasonality in this case), rule-based subjective forecasting can outperform conventional extrapolation methods (Bunn and Wright, 1991; Collopy and Armstrong, 1992; Adyaa *et al.*, 2000). A discussion on the issues of using a rule-based approach with statistical forecasting methods has been provided by Bunn and Wright (1991). Collopy and Armstrong (1992) created a rule-based framework to combine forecasts from different extrapolation methods, and reported that rule-based methods reduced the forecast error by approximately half, compared to an equally-weighted combined forecast. Adyaa *et al.* (2000) emphasised that identification of relevant features or properties of a time series is crucial for the success of a rule-based forecasting method.

In this study, the rule essentially identifies the past special day that has daily load profile that will be most useful in enhancing the time series model's estimate of load for the future special day. The rule can be viewed as the prior information that is fed into the statistical model, along

with the historical load observations. The formulation of rules is done subjectively by inferring them directly from the data. Once developed, the rules are flexible enough to incorporate additional knowledge arising from the inclusion of new observations.

For exponential smoothing, autoregressive moving average (ARMA), and artificial neural network (ANN) methods, it has been shown recently that accommodating the intrayear seasonality in load, along with the intraday and intraweek seasonal cycles, leads to improved short-term load forecast accuracy (Taylor, 2010a). This triple seasonal framework is convenient for our modelling of anomalous load, because special days tend to reoccur annually. Our initial focus in this study is a new rule-based triple seasonal exponential smoothing approach, which models normal and anomalous load in a unified framework. After introducing this approach in Section 4.3.1, we also consider analogous rule-based ARMA, ANNs, intraday singular value decomposition (SVD)-based exponential smoothing method, and intraweek SVD-based exponential smoothing method, in Section 4.3.2, Section 4.3.3, Section 4.3.4, and Section 4.3.5, respectively.

As discussed in Chapter 3, the use of discounting and sequence updating within the SVD modelling framework led to an improvement in model's forecast accuracy across normal days. However, the discount weighted SVD methods were not noticeably better than the much simpler HWT exponential smoothing method. Moreover, given the unavailability of an adequate number of anomalous load observations, the efficacy of discounting and sequence updating in improving the model's forecast performance across special days is doubtful. Also, the estimation of discount factors and sequence updating of feature vector and series increases the computational complexity of the model under consideration. Hence, for modelling anomalous load, we do not include the discount weighted SVD-based methods in this chapter.

4.1.2 Historical Review

A plethora of methods have been proposed for forecasting load for normal days, see Bunn (2000). However, the modelling of anomalous load has often been ignored (Smith, 2000; Hippert *et al.* 2005; Taylor, 2010a). Smith (2000) substitutes load for special days by load observed on a recent Sunday during the modelling process. Hippert *et al.* (2005) avoid special days by replacing them in their modelling, by load observed on a corresponding day of the week from the last week. Prior to estimation and evaluation procedures, Taylor (2010a) smoothes out the load for special days, by replacing it with the mean estimate of the normal load from the corresponding periods of the two adjacent weeks. The choice of replacing or smoothing load for special days is subjective and not well-defined. The problem with ignoring load for special days is that it not only guarantees that the resulting model cannot be used for special days, but also results in considerably large forecast errors on normal days that lie in the vicinity of special days. In this study, we use the actual load time series for modelling, whereby the anomalous load observations are neither replaced, nor smoothed out.

Previous approaches for short-term forecasting of normal and anomalous load have mostly employed regression-based methods with dummy variables for special days (see Ramanathan *et al.*, 1997; Pardo *et al.*, 2002; Cottet and Smith, 2003; Cancelo *et al.*, 2008; Soares and Medeiros 2008; and Dordonnat *et al.*, 2008). Ramanathan *et al.* (1997) build a separate model for each hour of the day, and include special day and weather effects on load using dummy variables. Pardo *et al.* (2002) and Ramanathan *et al.* (1997) use dummy variables for different special days, and also employ additional dummy variables for the day following a special day. This helps accommodate the potential impact of anomalous load on a normal day following a special day. Cottet and Smith (2003) model load using a multivariate Bayesian model, whereby they employ

six dummy variables for different special days, and use 48 coefficients for each dummy variable to capture the intraday seasonality using half-hourly load. Cancelo *et al.* (2008) build a separate model for each hour of the day for Spanish load. They first issue a forecast for load assuming a normal day, and adjust it accordingly for special days using different dummy variables for different classes of special days. Soares and Medeiros (2008) build a two-stage model for each hour of the day for load, such that anomalous load is modelled in the first stage using dummy variables. Any unexplained component in load is modelled in the second stage using either an AR model or an ANN, however, Soares and Medeiros (2008) found that using an ANN did not lead to any improvement in forecast accuracy over a linear AR model. Dordonnat *et al.* (2008) build a regression model for each hour of the day, and include special day effects using dummy variables. They also include the effect of a variety of weather variables on load, such as temperature, cloud cover, and wind speed. In addition, they use extra dummy variables for *bridging days*. They define a day to be a bridging day if it is a Monday before a special day, or a Friday after a special day. The load on bridging days tends to be lower than normal, and hence, needs to be treated separately.

Apart from regression-based methods, some authors have proposed rule-based approaches for anomalous load forecasting (Rahman and Bhatnagar, 1988; Hyde and Hodnett, 1997), while others have used ANNs with fuzzy rules (Srinivasan *et al*., 1995; Kim *et al*., 2000, Song *et al*., 2000). Rahman and Bhatnagar (1988) propose a rule-based approach, whereby they formulate rules based on the logical and syntactical relationships between weather and load, and the daily load shapes, using load for an electric utility in Virginia, USA. Hyde and Hodnett (1997) formulate rules specifically for the Irish load data, whereby the rationale of their approach is to find the deviation of load for different special days from normal load for a given year, and use

this deviation as a correction term for the corresponding special day falling next year. Srinivasan *et al*. (1995) use ANNs with fuzzy rules, and find that the load profile on most special days is similar to the profile of a typical Sunday. Kim *et al*. (2000) classify special days into five different types, and employ an ANN for each special day type used in conjunction with fuzzy rules. They use a fuzzy rule base consisting of different rules inferred from the Korean load data. Song *et al*. (2000) propose a fuzzy linear regression method, and report that errors from their method were lower compared to an ANN and fuzzy inference method. Apart from fuzzy rules, ANNs have also been used in conjunction with self-organizing maps (SOM) for anomalous load forecasting (Lamedica *et al.*, 1996). It is noteworthy that most existing methods for anomalous load forecasting rely on classifying different special days as being the same (Cancelo *et al*., 2008; Soares and Medeiros, 2008; Dordonnat *et al*., 2008), while some require different models to be employed for normal and special days (Kim *et al*., 2000). We, however, treat each special day as having a unique profile, and adopt a unified modelling framework. Moreover, the existing rule-based methods are tailored only to the data at hand (Rahman and Bhatnagar, 1988; Hyde and Hodnett, 1997). This makes the existing models difficult to adapt to different datasets, and would require creating a completely new set of rules for the GB data. Due to the aforementioned reasons, we do not employ existing rule-based methods as benchmarks in this study.

The main strategies employed for short-term anomalous load forecasting have been to build separate models for normal and special days (Kim *et al*., 2000), or use a single model for both normal and special days, and incorporate the special day effects using a dummy variable, as explained above. Note that to model load adequately, the triple seasonal effects need to be included, which requires different dummy variables for identifying the period of the day, day of the week, and date, for each special day. Given that load for different special days is different,

dummy variables are also required for identifying the special day type. However, to include a correction coefficient with each dummy variable results in the final model being highly parameterized and difficult to interpret. To avoid the model becoming over-parameterized, different special days have often been classified as belonging to the same special day type (Cancelo *et al.*, 2008; Soares and Medeiros, 2008; Dordonnat *et al.*, 2008). This classification relies on the assumption that the load profile for different special days can be treated as similar, and would remain similar over the years. Given that the pattern of load for each special day tends to change each year, this assumption is too restrictive. Hence, we do not classify different special days as being the same, and instead treat each special day differently. For these reasons, we do not employ regression-based models with dummy variables as benchmarks. We identified eighteen special days in the one-year post sample data, each of which is treated as having a unique profile in this study.

As opposed to previously employed regression-based models, we incorporate prior information regarding anomalous load profiles into the modelling process using a rule-based approach. Furthermore, the methods proposed in this study accommodate not only intraday and intraweek seasonal cycles, but also intrayear seasonality.

4.1.3 Contributions

In this study, we employ triple seasonal variants of Holt-Winters-Taylor (HWT) exponential smoothing, SARMA, ANNs, intraday SVD-based exponential smoothing, and intraweek SVD-based exponential smoothing. These methods have been shown to be competitive for modelling intraday time series for normal days. The methodological contribution of this study is to demonstrate how these methods can be adapted to model load for special days, when used in

conjunction with a rule-based approach. Specifically, the contributions of this study lie in formulating a novel rule-based approach, such that the rules:

a. Allow for incorporation of prior expert knowledge into the modelling framework,
b. Allow every special day to have a unique load profile,
c. Take into account the triple seasonal effects on anomalous load,
d. Are sufficiently flexible that they can be updated with the inclusion of new observations, and
e. Can be easily adapted to different datasets, whereby the time series may exhibit seasonality and anomalous observations may pose modelling challenges.

In terms of methodology, we propose a novel rule-based modified HWT exponential smoothing method, and modified triple seasonal variants of the recently proposed intraday and intraweek SVD-based methods, allowing for:

a. Different discounting rates for normal and special days,
b. Reduction in the dimensionality of the model (applicable for SVD-based methods), and
c. Modelling normal and special days in a coherent and unified framework.

Along with HWT exponential smoothing and SVD-based methods, we adapt the conventional ARMA and ANN methods, to model both normal and special days, by using them in conjunction with the formulated rules.

Using nine years of half-hourly load for Great Britain, we evaluate point forecasts, for lead times varying from one half-hour up to a day ahead. We show that the out-of-sample forecast errors for anomalous load generated using our rule-based methods is about half the size of those obtained from the corresponding original models that were not rule-based. For both normal and

special days, the rule-based methods led to improvement in short-term load forecast accuracy, across all forecast horizons. The most accurate forecasts were generated using a combination of two different rule-based methods.

Although we demonstrate the applicability of the rule-based methods for load forecasting, the proposed methodology can easily be adapted for other applications. Some examples where this approach could potentially be useful for forecasting includes call centre arrivals, hospital admissions, and transportation counts, as the corresponding time series exhibit seasonality and anomalous conditions pose significant modelling challenges.

4.2 Anomalous Load Characteristics

In this study, we employ nine years of half-hourly load for Great Britain, stretching from 1 January 2001 to 31 December 2009. We used the first eight years of data (consisting of 140,256 observations) to estimate the model parameters, and employed the final year (consisting of 17,520 observations) for model evaluation. Using a rolling forecast origin through each period of the final year, we evaluate out-of-sample point forecasts, for lead times varying from one half-hour up to a day ahead.

The complete load data is presented in Figure 4.1. Note that this figure shows the actual load observed on normal and special days.

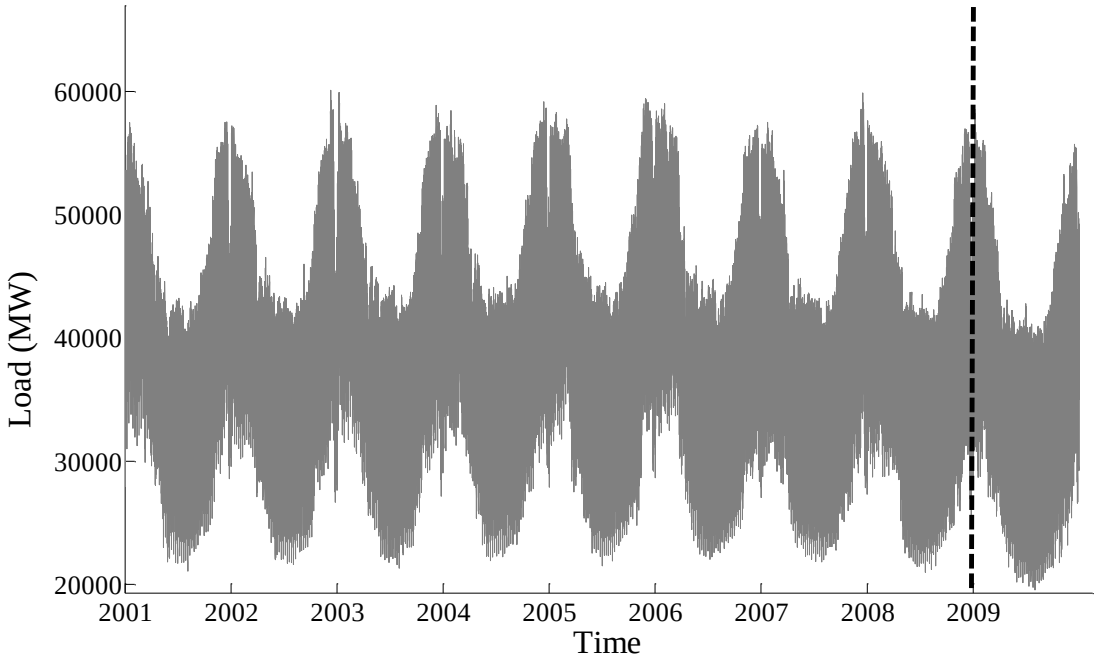

Figure 4.1— Half-hourly load for Great Britain stretching from 1 January 2001 to 31 December 2009. The vertical dashed line denotes the time index that divides the time series into non-overlapping estimation and evaluation set.

As discussed earlier, in order to model load adequately, the special day effects need to be taken into account with the seasonal patterns. Figure 4.2 plots load on a special day, namely a bank holiday Monday (a public holiday in Great Britain), and a normal working Monday from the preceding and following weeks. It can be seen from Figure 4.2, that load on a special day is lower compared to a normal day, and this is largely due to the closure of industries on a special day. It is due to this reason that anomalous load needs to be modelled separately from normal load.

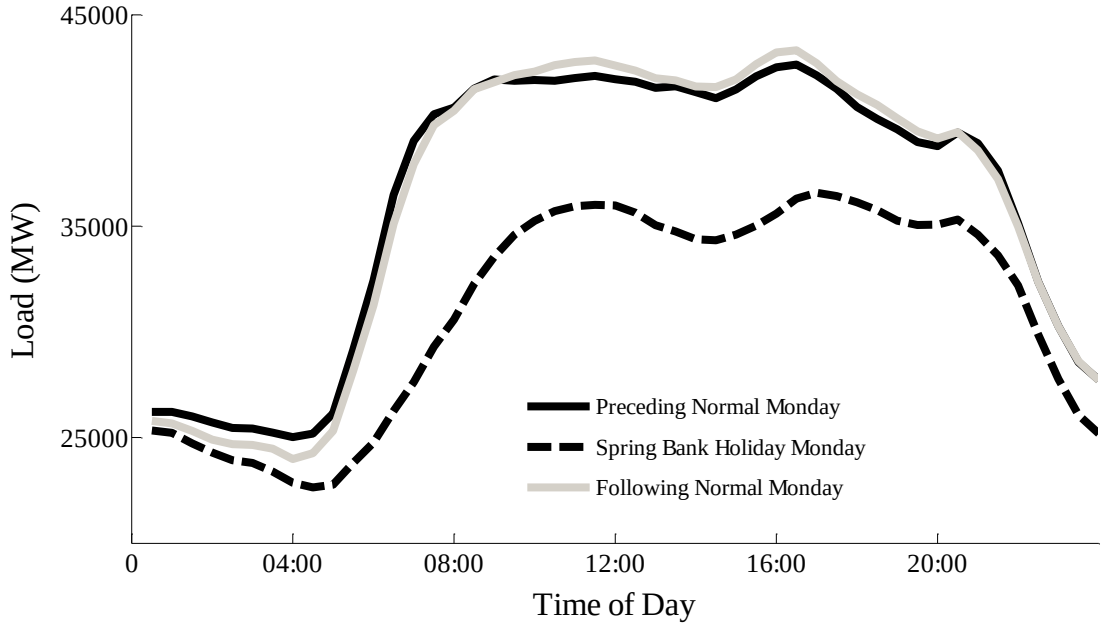

Figure 4.2— Load profile for a Spring Bank Holiday Monday (26 May 2008), a normal Monday (19 May 2008) from the preceding week, and a normal Monday (2 June 2008) from the following week.

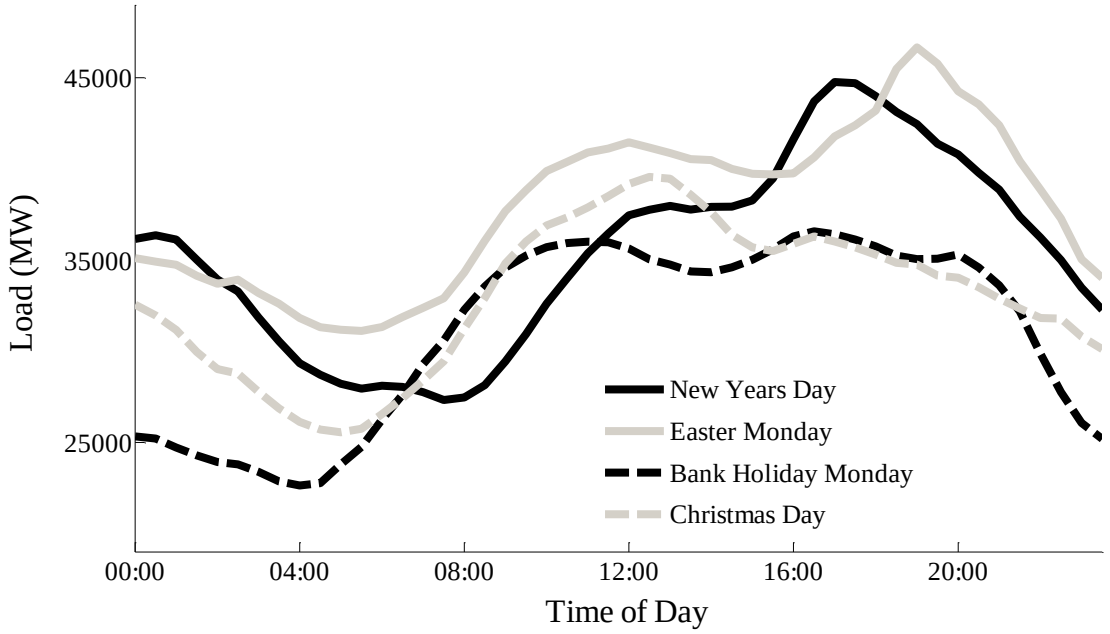

Figure 4.3— Load profile for New Year's Day (1 January), Easter Monday (24 March), Spring Bank Holiday Monday (26 May), and Christmas Day (25 December) observed in the year 2008.

Load for four different special days observed in the year 2008 is presented in Figure 4.3. Specifically, we plot the load observed on New Year's Day (1 January), Easter Monday (24 March), Spring Bank Holiday Monday (26 May), and Christmas Day (25 December). This figure shows that load for different special days is noticeably different. Also, even when two different special days fall at the same day of the week, e.g. Easter Monday and Spring Bank Holiday Monday, they exhibit different profiles.

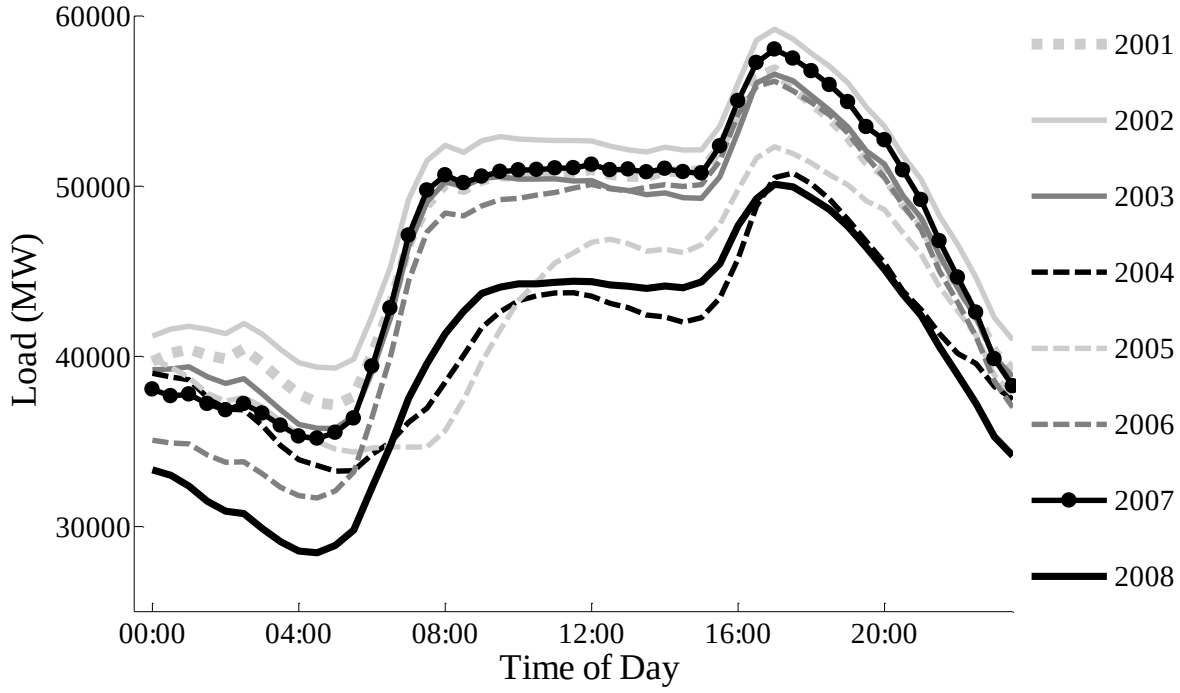

Figure 4.4— Load profile for Christmas Day (25 December), observed across all years within the estimation sample, stretching from 2001 to 2008.

Figure 4.4 plots load for a given special day, namely a Christmas Day (25 December), for all eight years within the estimation sample. It can be seen from this figure that the profile for a given special day changes every year. Christmas Day obviously falls on a different day of the week each year. This motivates us to treat each special day separately, whereby each special day is allowed to have a unique load profile, which may change over different years.

As defined in Section 3.3, we represent the length of the intraday, intraweek and intrayear seasonal cycles by m_1, m_2 and $m_3(t)$ respectively. We have $m_1 = 48$, and $m_2 = 336$. Following Taylor (2010a), we set the length of the intrayear seasonal cycle, for period t, to be $m_3(t) = 52 \times 336$, except for some years, where we set $m_3(t) = 53 \times 336$ for a few weeks around the clock-change.

Statistical methods designed for forecasting load on special days will always seek to learn from past special days. The proposed approach essentially does this for the exponential smoothing, SARMA, ANN and SVD-based methods through the specification of the intrayear cycle length $m_3(t)$. As discussed later in Section 4.4, this is a very natural way to accommodate the modelling of special days in these methods, and it is appealing because it involves a minimal change to the structure in these time series models.

4.3 Forecasting Methods

4.3.1 Rule-based Triple Seasonal Modified HWT Exponential Smoothing Method

An exponential smoothing method has recently been proposed by Taylor (2010a) for modelling the intraday, intraweek and intrayear seasonality in intraday load data. As described in Section 3.3.1, this method is referred to as the triple seasonal HWT exponential smoothing, and is represented in a single source of error state space form as follows:

$$y_t = l_{t-1} + d_{t-m_1} + w_{t-m_2} + a_{t-m_{3(t)}} + \phi e_{t-1} + \varepsilon_t \tag{4.1}$$

$$e_t = y_t - (l_{t-1} + d_{t-m_1} + w_{t-m_2} + a_{t-m_3(t)}) \tag{4.2}$$

$$l_t = l_{t-1} + \lambda e_t \tag{4.3}$$

$$d_t = d_{t-m_1} + \delta e_t \tag{4.4}$$

$$w_t = w_{t-m_2} + \omega e_t \tag{4.5}$$

$$a_t = a_{t-m_3(t)} + \alpha e_t \tag{4.6}$$

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$, and σ^2 denotes a constant variance. The model parameters were estimated by the maximization of a Gaussian likelihood function, which for this model is equivalent to minimizing the sum of squared one-step-ahead errors. Prior to modelling, we applied the natural log transformation to the data, in order to convert multiplicative effects to additive.

We extend the triple seasonal HWT exponential smoothing to model normal and anomalous load in a unified framework. We refer to this model as the rule-based modified HWT (RB-HWT) exponential smoothing, represented as follows:

$$y_t = l_{t-1} + d_{t-m_1} + w_{t-m_2} + a_{t-m_{3(t)}} + \phi e_{t-1} + I_{N_t}\varepsilon_t^{(N)} + I_{S_t}\varepsilon_t^{(S)} \tag{4.7}$$

$$e_t = y_t - (l_{t-1} + d_{t-m_1} + w_{t-m_2} + a_{t-m_{3(t)}}) \tag{4.8}$$

$$l_t = l_{t-1} + \lambda e_t \tag{4.9}$$

$$d_t = d_{t-m_1} + I_{N_t}\delta e_t \tag{4.10}$$

$$w_t = w_{t-m_2} + I_{N_t}\omega e_t \tag{4.11}$$

$$a_t = a_{t-m_3(t)} + I_{N_t}\alpha_1 e_t + I_{S_t}\alpha_2 e_t \tag{4.12}$$

where $\varepsilon_t^{(N)} \sim NID(0, \sigma_N^2)$ and $\varepsilon_t^{(S)} \sim NID(0, \sigma_S^2)$ are the model errors for normal and special days respectively, having corresponding variances σ_N^2 and σ_S^2. The binary indicator term I_{N_t} equals one if t occurs on a normal day, and zero otherwise, whereas I_{S_t} equals one for special days, and zero otherwise. Thus, at any given period t, we have $I_{S_t} = 1 - I_{N_t}$.

Figure 4.5 presents the intraday, intraweek and intrayear seasonal indices. As shown in Figure 4.5a, the intraday and intraweek seasonal indices, denoted by d_t and w_t, capture the daily and weekly patterns in load for normal days. In Figure 4.5b, the index a_t denotes the intrayear cycle, which remains after both the intraday and intraweek seasonality for only the normal days has been removed. Thus, a_t incorporates the intrayear seasonal pattern in load for all days, along with the intraday and intraweek seasonal load patterns for special days. Specifically, the

difference between a normal and anomalous load profile for the same day of the week is embedded and modelled via a_t.

In Figure 4.5b, the anomalous load periods are denoted in black. It is noteworthy that a_t on a special day is considerably lower compared to a normal day from the adjacent weeks. Inspection of the data reveals that load for the special days is noticeably lower than normal load for the same day of the week, around the same date. In Great Britain, the special days are New Year's day, the day following New Year's day, Good Friday, Easter Monday, three summer bank holiday Mondays, and the Christmas period, which stretches from 21 December to 31 December, inclusive. We refer to Good Friday and Easter Monday as Easter holidays.

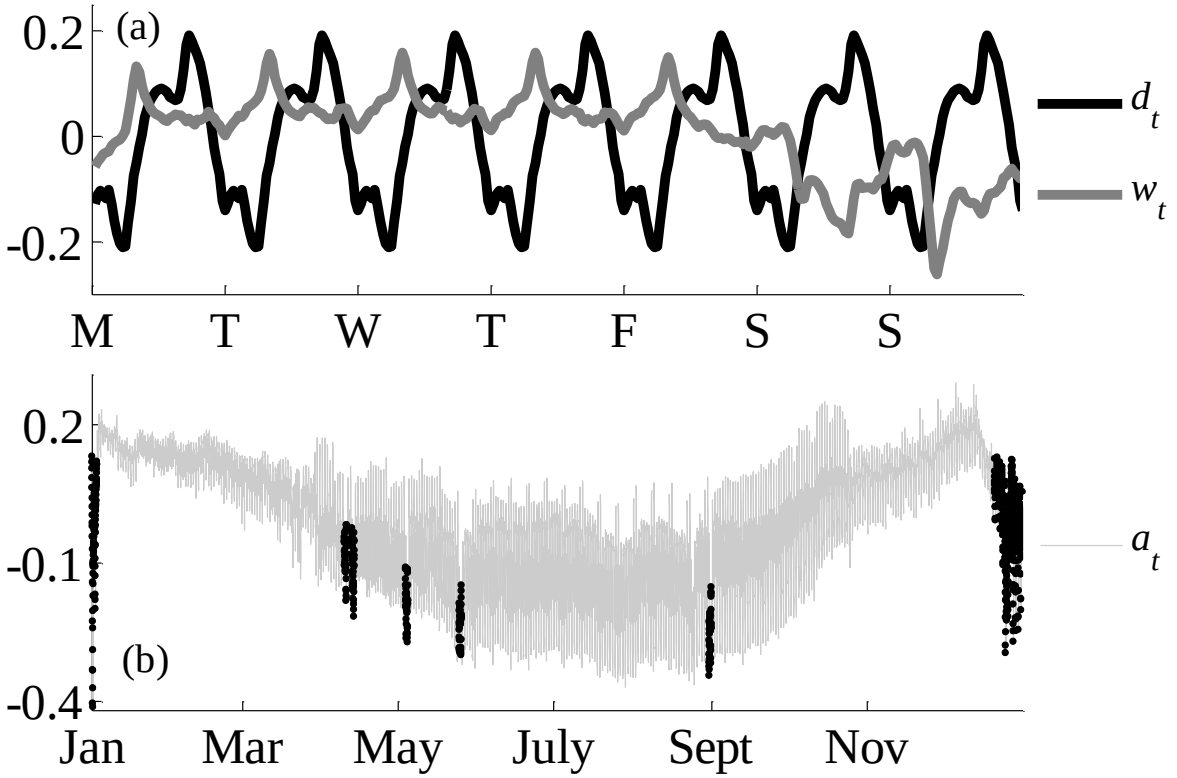

Figure 4.5— Seasonal indices for the intraday, intraweek, and intrayear cycle, denoted by d_t, w_t and a_t, respectively. In panel (b), the grey line denotes the index on normal days, while black dots correspond to anomalous periods.

We update a_t at different rates for the two day types. This allows the impact of the annual seasonal cycle on normal and anomalous load to be different. The indices d_t and w_t are not updated on special days, as they capture the seasonal patterns for only the normal days. Crucially, the intrayear cycle length, denoted by $m_3(t)$, is determined using a rule-based approach for special days, which is discussed in Section 4.4. This cycle length allows the selection of suitable historical anomalous load observations in the statistical model, to be extrapolated into the future for generating forecasts. For normal days, $m_3(t)$ is chosen to be the same as the intrayear cycle length used in the original triple seasonal HWT method.

For periods on normal days, the RB-HWT model defaults to the original HWT model. For special days, the RB-HWT model incorporates the daily and weekly seasonal effects in normal load, and adjusts it using the intrayear seasonal index, which captures the departure of anomalous load from normal load, for each special day. This allows the model to treat each special day as having its own unique load profile.

To initialize the seed values for d_t and w_t, we use the same averaging procedure as used for the original triple seasonal HWT exponential smoothing method. Specifically, we initialize d_t and w_t using observations for only the normal days. This ensures that d_t and w_t captures the daily and weekly patterns in load for normal days, but not for special days. However, to initialize a_t, we use load for both normal and special days, and remove the daily and weekly seasonal patterns for normal days from a_t, by subtracting d_t and w_t from it. This ensures that the initial seed of a_t constitutes the annual seasonal pattern in load for all days, as well as the intraday and intraweek seasonal load patterns for special days. We use the first three years of the estimation data to compute the initial seeds of all three seasonal indices.

The maximum likelihood estimation uses a Gaussian error, for which the variance is different on special days to that on normal days. The log-likelihood (LL) can be written as:

$$LL = -\frac{n_{ND}}{2}\log(2\pi\sigma_N^2) - \frac{n_{SD}}{2}\log(2\pi\sigma_S^2) - \sum_{t=365\times m_1}^{N}\left(\frac{I_{N_t}}{2\sigma_N^2}(\varepsilon_t^{(N)})^2 + \frac{I_{S_t}}{2\sigma_S^2}(\varepsilon_t^{(S)})^2\right) \quad (4.13)$$

where N is the size of the estimation sample, and n_{ND} and n_{SD} are the number of observations that fall on normal and special days, respectively, in the estimation sample, excluding observations from the first year.

4.3.2 Rule-based Triple Seasonal ARMA

In Taylor (2010a), the following triple seasonal ARMA model has been proposed for modeling load on normal days:

$$\Upsilon_p(L)\Phi_{P_1}(L^{m_1})X_{P_2}(L^{m_2})\Psi\left(L^{m_3(t)}\right)(y_t - c) = \Omega_q(L)\Theta_{Q_1}(L^{m_1})\Gamma_{Q_2}(L^{m_2})\Lambda\left(L^{m_3(t)}\right)\varepsilon_t \quad (4.14)$$

where c is a constant; ε_t is the model error; L is the lag operator, and Υ_p, Φ_{P_1}, X_{P_2}, Ω_q, Θ_{Q_1} and Γ_{Q_2} are polynomial functions of order p, P_1, P_2, q, Q_1, and Q_2, while the functions Ψ and Λ essentially capture the annual cycles, as explained previously in Section 3.3.2. We considered orders equal to or less than three. To extend the model for anomalous load, as we did for the HWT model, we used a rule-based approach to set the length of the lag $m_3(t)$ for the intrayear cycle for special days. We refer to the new model as rule-based SARMA (RB-SARMA), and represent it as:

$$\begin{aligned}&\Upsilon_p(L)\Phi_{P_1}(L^{m_1})X_{P_2}(L^{m_2})\left(I_{N_t}\Psi\left(L^{m_3(t)}\right) + I_{S_t}\theta\left(L^{m_3(t)}\right)\right)(y_t - c) = \\ &\Omega_q(L)\Theta_{Q_1}(L^{m_1})\Gamma_{Q_2}(L^{m_2})\left(I_{N_t}\Lambda\left(L^{m_3(t)}\right) + I_{S_t}K\left(L^{m_3(t)}\right)\right)\left(I_{N_t}\varepsilon_t^{(N)} + I_{S_t}\varepsilon_t^{(S)}\right)\end{aligned} \quad (4.15)$$

where $\varepsilon_t^{(N)} \sim NID(0, \sigma_N^2)$ and $\varepsilon_t^{(S)} \sim NID(0, \sigma_S^2)$ are the model errors for normal and special days, respectively, while θ and K are additional functions of the lag operator $L^{m_3(t)}$ used only for special days. Hence, we basically switch between different AR and MA polynomial functions, with different annual lag terms, depending on whether y_t belongs to a normal day, or a special day. For special days, the rule-based value for $m_3(t)$ allows the inclusion of load observations from three previous special days, which would be suitable for improving the model's forecast for a given special day. As with the RB-HWT method, the parameters of the RB-SARMA method are estimated using the maximum likelihood procedure as used for the RB-HWT method. The Box-Jenkins methodology was used to select polynomial function orders for AR and MA terms.

4.3.3 Rule-based Triple Seasonal ANN

We include the ANN method in this study as it has previously been employed for forecasting anomalous load (Srinivasan *et al*., 1995; Lamedica *et al*., 1996; Kim *et al*., 2000; Song *et al*., 2005; Fidalgo and Lopes, 2005). We extend the triple seasonal ANN method, as presented earlier in Section 3.3.3, using a rule-based approach to model load for normal and special days. Since ANNs have been shown to struggle with multi-step-ahead prediction (Atiya *et al*., 1999), we build a separate ANN model for each horizon. We specify the output to comprise load observations, differenced using the above operator, and normalized to have zero mean and unit standard deviation. We selected the lags for the input variables to be as consistent as possible with the SARMA model. Specifically, for the ANN built for horizon h, we used load at the forecast origin, and at the following lags: 1, 2, $m_1 - h$, $2m_1 - h$, $3m_1 - h$, $m_2 - h$, $2m_2 - h$, $3m_2 - h$, $LagA = m_3(t) - h$, $LagB = LagA + m_3(t - LagA - h)$, and $LagC = LagB + m_3(t - LagB - h)$. Using this model, $m_3(t)$ is defined in the same way as in the exponential

smoothing and SARMA models of the previous two subsections. We refer to this as the rule-based ANN, denoted by RB-ANN. We show that using ANN in conjunction with the formulated rules improves the model's accuracy in forecasting load for special days.

4.3.4 Rule-based Triple Seasonal Modified Intraday SVD-based Exponential Smoothing

We extend the triple seasonal formulation of the intraday (ID) SVD-based exponential smoothing method, as presented earlier in Section 3.4.1, and propose a rule-based modified intraday SVD (RB-ID SVD) exponential smoothing method. The proposed method can model normal and anomalous load in a unified framework. The proposed model is presented as follows:

$$y_t = \sum_{i=1}^{5} I_{it} \boldsymbol{p}_{t-1}^{(i)} \widetilde{\boldsymbol{v}}'_{[t \bmod m_1]} + a_{t-m_3(t)} + \phi\, e_{t-1} + I_{N_t}\varepsilon_t^{(N)} + I_{S_t}\varepsilon_t^{(S)} \tag{4.16}$$

$$e_t = y_t - \left(\sum_{i=1}^{5} I_{it} \boldsymbol{p}_{t-1}^{(i)} \widetilde{\boldsymbol{v}}'_{[t \bmod m_1]} + a_{t-m_3(t)}\right) \tag{4.17}$$

$$\boldsymbol{p}_t^{(i)} = \boldsymbol{p}_{t-1}^{(i)} + (\lambda \mathbf{1}_{m_1} \widetilde{\boldsymbol{V}} + I_{N_t}(\delta + \omega I_{it}) \widetilde{\boldsymbol{v}}_{[t \bmod m_1]}) e_t \tag{4.18}$$

$$a_t = a_{t-m_3(t)} + I_{N_t}\alpha_1 e_t + I_{S_t}\alpha_2 e_t \tag{4.19}$$

$$I_{it} = \begin{cases} 1 \text{ if period } t \text{ occurs on a day of type } i \\ 0 \text{ otherwise} \end{cases} \tag{4.20}$$

where $\varepsilon_t^{(N)} \sim NID(0, \sigma_N^2)$ and $\varepsilon_t^{(S)} \sim NID(0, \sigma_S^2)$ are the model errors for normal and special days, respectively. As defined earlier in Section 4.3.1, I_{N_t} equals one if t occurs on a normal day, and zero otherwise, whereas I_{S_t} equals one for special days, and zero otherwise. The intrayear seasonal index a_t is defined exactly the same way as defined earlier for RB-HWT. Specifically, a_t accommodates the intrayear seasonality in load for all days, along with the intraday and intraweek seasonal load patterns for special days. We select $m_3(t)$ using a rule-based approach, if t occurs on a special day. The rest of the parameters are the same as defined earlier in Section 3.4.1. In equation 4.18, for only the special days, we do not update the elements $\boldsymbol{p}_t^{(i)}$ based upon

their relationship with the period of day and period of week, on which t falls, as the intraday and intraweek seasonality for special days is modelled via a_t. The model parameters were estimated using the maximum likelihood procedure as used for the RB-HWT method, whereby the model errors had different variances for normal and special days. We estimated the number of feature vectors and feature series (denoted by k), using cross-validation.

4.3.5 Rule-based Triple Seasonal Modified Intraweek SVD-based Exponential Smoothing

For modelling anomalous load, we extend the triple seasonal variant of the intraweek (IW) SVD-based exponential smoothing method, as presented earlier in Section 3.4.2, and propose a rule-based modified intraweek SVD (RB-IW SVD) exponential smoothing method as follows:

$$y_t = \widetilde{\boldsymbol{p}}_{t-1}\widetilde{\boldsymbol{v}}'_{[t \bmod m_2]} + a_{t-m_3(t)} + \phi e_{t-1} + I_{N_t}\varepsilon_t^{(N)} + I_{S_t}\varepsilon_t^{(S)} \tag{4.21}$$

$$e_t = y_t - (\tilde{p}_{t-1}\widetilde{\boldsymbol{v}}'_{[t \bmod m_2]} + a_{t-m_3(t)}) \tag{4.22}$$

$$\widetilde{\boldsymbol{p}}_t = \widetilde{\boldsymbol{p}}_{t-1} + \left(\alpha 1_{m_2}\widetilde{\boldsymbol{V}} + I_{N_t}(\delta \textstyle\sum_{j=1}^{7} \widetilde{\boldsymbol{v}}_{[t \bmod m_1]+(j-1)m_1} + \omega\widetilde{\boldsymbol{v}}_{[t \bmod m_2]})\right)e_t \tag{4.23}$$

$$a_t = a_{t-m_3(t)} + I_{N_t}\alpha_1 e_t + I_{S_t}\alpha_2 e_t \tag{4.24}$$

where $\varepsilon_t^{(N)} \sim NID(0, \sigma_N^2)$ and $\varepsilon_t^{(S)} \sim NID(0, \sigma_S^2)$ are the model errors for normal and special days, respectively, while I_{N_t} and I_{S_t} are dummy variables, as defined earlier for RB-HWT. The index a_t accommodates the special day effects using a rule-based $m_3(t)$, as explained in the Section 4.3.1. The rest of the parameters are the same as defined earlier in Section 3.4.2. Since a_t accommodates the intraday and intraweek seasonal effects for special days, we do not update the elements of $\widetilde{\boldsymbol{p}}_t$ (equation 4.23), based upon their relationship with the period of day and period of week, on which t falls, for only the special days. The model parameters were estimated using the maximum likelihood procedure as used for the RB-HWT method. We estimated the number of feature vectors and series (denoted by k) using cross-validation.

4.4 Formulating Rules

A rule-based approach allows the encapsulation of domain knowledge and data characteristics into a statistical modelling framework (Bunn, and Wright, 1991; Collopy and Armstrong, 1996). The prior knowledge and data properties are quantified and expressed explicitly in the form of rules. In this chapter, as stated in Section 4.3, the rules have the sole purpose of defining the most suitable length of the annual lag $m_3(t)$ for special days, in the formulations of the HWT, SARMA, ANN and SVD models. In this section, we present four alternative rules for defining $m_3(t)$ for special days.

We present four alternative rules, which require subjective judgment. Using the data, we identified four distinct features that have a unique impact on anomalous load. These are the special day type, the period of day, the day of week, and the time of year. For a given special day, the load profile is treated as a function of the above features. In this study, we provide an empirical comparison between rule-based subjective forecasting methods and conventional extrapolation methods.

The special days are classified into two types. The first type includes those special days that fall at the same day of the week each year, e.g. Good Friday. The second type includes those days that occur on the same date each year, e.g. Christmas Day. We formulate different rules for the two special day types. The rules are formulated using only the estimation data. The calculation of intrayear cycle length $m_3(t)$ takes into account the inclusion of extra observations due to leap years, for all the four rules proposed in this study. Note that for a given rule to work, the length of the time series should either be equal or greater than the largest $m_3(t)$ selected using that rule.

4.4.1 Rule 1

To model load for a given special day, this rule refers to the historical load observed on the same special day from the previous year. For example, to forecast load for Good Friday in 2008 (21 March 2008), $m_3(t)$ is selected such that observations from the previous year's Good Friday (6 April 2007) are included in the model. This leads to $m_3(t) = 350 \times 48$ for each period t on the Good Friday in 2008. Similarly, to model load for Christmas Day, $m_3(t)$ refers to Christmas Day in the previous year. Thus, using this rule, $m_3(t)$ is defined such that y_t and $y_{t-m_3(t)}$ belong to:

a) same special day,

b) adjacent years, and,

c) the same period of the day.

To incorporate the intraday, intraweek, and intrayear seasonal effects on load, both y_t and $y_{t-m_3(t)}$ should belong to the same: a) period of the day, (b) day of the week, and, (c) date. A potential disadvantage of Rule 1 is that it ignores the day of the week on which the special day occurs. For example, Christmas Day in the current year obviously falls on a different day to Christmas Day in the previous year (e.g. Thursday in 2008 and Tuesday in 2007). Given that the average intraday cycle is not the same for each day of the week (as shown in Figure 3.4), it is worth investigating if different days of the week have a different impact on anomalous load. Also, it has been pointed out by Kim *et al*. (2000) that if a particular special day falls, in different years, on the same day of the week, then the load pattern will be very similar. Thus, in the next rule, we focus on incorporating the intraweek seasonal effects on special days.

4.4.2 Rule 2

This rule treats each day of the week as having a unique impact on anomalous load. Rule-based approaches for incorporating the day of the week effect on anomalous load have been employed by Hyde and Hodnett (1994, 1997) and Kim *et al.* (2000). Using this rule, $m_3(t)$ is selected such that y_t and $y_{t-m_3(t)}$ belong to the same:

a) special day,

b) day of the week, and,

c) period of the day.

For example, to model load for a New Year's day, $m_3(t)$ refers to a previous year's New Year's Day which occurred on the same day of the week as the current day. For example, consider New Year's Day in 2008. This was a Tuesday, and so $m_3(t)$ is defined to be such that we are referring back to the most recent occurrence of New Year's Day on a Tuesday, which was in 2002. Therefore, for New Year's Day in 2008, we set $m_3(t) = (5 \times 365 + 366) \times 48$. Using this rule, each day of the week is treated as having a distinct intraday cycle. Basically, this rule extends Rule 1 in order to simultaneously incorporate the triple seasonal effects on anomalous load. For special days that fall on the same day of the week each year, namely the three summer Bank holiday Mondays and Easter holidays, this rule is the same as Rule 1.

Treating each day of the week as having a unique intraday cycle allows the different days in a week to have a distinct impact on anomalous load, however, this also leads to a large $m_3(t)$. Note that $m_3(t)$ is selected to be six years in the above example for a New Year's day. Thus, a disadvantage of Rule 2 is that it can lead to the selection of a large value for $m_3(t)$, which implies long lags and hence the inclusion of less recent observations in the model. Also, we do not have enough observations for each special day that satisfies the criteria of this rule. Note that

in some cases, the magnitude of $m_3(t)$ tends to be larger than the total number of historical observations required to use this rule, in which case we revert to Rule 1. Furthermore, it has been argued that economic factors and weather conditions can alter the load patterns over time (Hyde and Hodnett, 1997; Kim *et al*., 2000). Thus, it might not be appropriate to consider lags beyond, say, a couple of years.

In situations where an insufficient number of observations are available, a special day has been treated as another special day type by Hyde and Hodnett (1994, 1997), while different special days have often been classified as being the same (Cancelo *et al*., 2008; Soares and Medeiros, 2008; Dordonnat *et al*., 2008). Hence, to reduce the magnitude of $m_3(t)$, for Rule 2, we classify special days within the Christmas period, as having the same profile. This also allows us to investigate the impact, on forecast accuracy, of treating different special days, as being the same. The special days that do not belong to the Christmas period, namely the New Year's day, day following New Year's day, Easter holidays, and bank holidays Mondays, are all treated separately, whereby each special day is allowed to have a unique profile.

4.4.3 Rule 3

Given that weekdays witness a relatively high load compared to weekends (Figure 3.4), this rule treats weekdays as having a distinct impact on anomalous load, in comparison with weekends. Specifically, for special days that fall on the same date each year, $m_3(t)$ is selected such that y_t and $y_{t-m_3(t)}$:

(a) belong to the same special day,

(b) either both fall on a weekday or both fall on a weekend, and,

(c) belong to the same period of the day.

For example, to forecast load for Christmas Day in 2009, which fell on a Friday, we set $m_3(t) = 365 \times 48$, because the previous year's Christmas Day also fell on a weekday. As another example, consider Boxing Day in 2009. This was a Saturday, and so $m_3(t)$ must be defined so that we are referring back to a previous Boxing Day that fell on a weekend, and this was Boxing Day in 2004, which feel on a Sunday. Therefore, for the periods on Boxing Day 2009, we set $m_3(t) = (4 \times 365 + 366) \times 48$.

For the three summer bank holiday Mondays, this rule is the same as Rule 1. For the Easter holidays, $m_3(t)$ is selected using Rule 1, with one extra condition, that both y_t and $y_{t-m_3(t)}$ occur either before, or after, the summertime clock-change in that year. For example, to forecast load for Good Friday in 2008 (21 March 2008), which occurs before the summertime clock-change, $m_3(t)$ refers to Good Friday in 2005 (25 March 2005), because this was the most recent Good Friday that had also occurred before the summertime clock-change. The summertime clock-changes in 2005 and 2008 occurred at 27 March and 30 March, respectively. This extra condition on Rule 1 was motivated by us observing that the daily load can change quite noticeably at clock-change.

We refer to the special days from 21 to 24 December, and from 27 to 30 December, inclusive, as proximity days. We use the notation *PD* for these days. They tend to witness lower load than normal days, but higher than other special days within the Christmas period. We define $m_3(t)$ differently for two separate classes of *PD*s, based on whether they precede a Christmas Day, or follow a Boxing Day. Furthermore, a *PD* is classified as a bridging proximity day, with the notation *B-PD*, if it is the only day occurring between a weekend and either a Christmas Day or a Boxing Day. Otherwise, it is treated as a non-bridging proximity day, *NB-PD*. Overall, the load on a PD tends to be lowered further if it is a *B-PD*, and hence we treat them as being

separate from *NB-PD*, as also done by Dordonnat *et al*. (2008). For a *PD*, Rule 3 sets $m_3(t)$ such that both y_t and $y_{t-m_3(t)}$:

(a) precede a Christmas Day or follow a Boxing day,

(b) occur either on a weekend or a weekday,

(c) either belong to the class *B-PD* or *NB-PD*, and,

(d) belong to the same period of the day.

For example, for Saturday 29 December 2007, we set $m_3(t) = 364 \times 48$, so that the lag refers to Saturday 30 December 2006. This is because both (a) followed a Boxing Day, (b) occurred at a weekend, and (c) belonged to the class *NB-PD*. As another example, consider Monday 24 December 2007. For this proximity day, we set $m_3(t) = 5 \times 365 + 366$ so that the lag refers to Monday 24 December 2001. This is because both (a) preceded a Christmas Day, (b) occurred on a weekday, and (c) belonged to the class *B-PD*. In cases where there are no historical load observations that satisfy this rule, we revert to Rule 1.

The advantage of this rule is that it incorporates recent information in the model by utilizing the similarity in the intraday cycle, models proximity days separately, and incorporates the effect of bridging days on anomalous load. However, this rule treats all weekdays as having the same intraday cycle. It can be seen from Figure 3.4, that not all weekdays have the same average intraday cycle. Also, the average intraday cycle for a Saturday is different from a Sunday. To overcome these limitations, we formulate Rule 4.

4.4.4 Rule 4

This rule treats different days of a week, which have a distinct intraday cycle, as having a distinct impact on anomalous load. A week is treated as comprising five different intraday cycles, as used in Taylor (2012). Specifically, Tuesday, Wednesday and Thursday are treated as

having a common intraday cycle, whereas Monday, Friday, Saturday and Sunday are each assumed to have a distinct intraday cycle.

This rule is similar to Rule 3, with the key difference that, for all special days that occur at the same date each year, instead of selecting an historical special day based on whether it occurs on a weekday or a weekend, this rule selects an historical special day of the same class of intraday cycle as the current special day. For example, consider Christmas Day in 2007. This was a Tuesday, and so $m_3(t)$ is defined to be such that we are referring to the most recent occurrence of Christmas Day with the same intraday cycle type as Tuesday, and this was Christmas Day in 2003, which fell on a Thursday. This is because Tuesday and Thursday are treated as having the same intraday cycle using this rule. Hence, for periods on Christmas Day 2007, we set $m_3(t) = (3 \times 365 + 366) \times 48$. For special days that fall on the same day of the week each year, such as Good Fridays, this rule is the same as Rule 3.

The advantage of this rule is that it allows the days with different intraday cycle types to have a distinct impact on anomalous load. Also, instead of classifying the days of a week as weekdays and weekends, this rule treats a week as consisting of five different intraday cycle types. The disadvantage of this rule, however, is that in order to find historical load observations that belong to the same intraday cycle type as the current day, the intrayear cycle length $m_3(t)$ tends to be large compared to Rule 3. A comparison of the post-sample forecast accuracy of different rule-based methods is presented in the next section.

4.5 Results and Observations

The main focus of our empirical analysis was the special days. In all, there were eighteen special days in the post-sample period. We generated point forecasts for the one-year post-

sample period, and evaluated accuracy using the mean absolute percentage error (MAPE) and root mean squared percentage error (RMSPE), as defined in equations 3.37 and 3.38, respectively. The relative model rankings were similar for the two measures; hence we present results only for MAPE in this chapter. We present the forecast accuracy of the five univariate methods presented in Section 4.3, when used in conjunction with the four different rules described in Section 4.4. We include the five univariate methods without any adjustment for special days as benchmarks. This amounts to treating the special days as normal days, with no domain knowledge incorporated in the models. Having estimated these benchmark models using the actual data, we evaluate their performance on special days. Since we evaluate the benchmark models across special days, we do not use any replacement or smoothing for special days for the original versions of the five univariate methods.

4.5.1 Results using load for Great Britain

In this study, we use the original load time series for Great Britain, as presented in Section 4.2, comprising the actual load observations for normal and special days. We apply the original versions of the five univariate methods described in Section 4.3, to the actual load time series, without incorporating any prior domain knowledge about anomalous load profiles, and evaluate their performance on special days. We then consider the proposed rule-based versions of all the five univariate methods considered in this study, whereby we estimate the model parameters using only the in-sample data, and compare their forecast performance on special days with their original counterparts. This allows us to gauge the scope of improvement in forecast accuracy offered using a rule-based approach, over the conventional statistical methods. We ensure that all models had access to the same observations during the training process, in order to provide a fair model comparison on the out-of-sample data.

TABLE 4.1

PARAMETER ESTIMATES FOR THE ORIGINAL AND RULE-BASED MODIFIED HWT (RB-HWT) EXPONENTIAL SMOOTHING USED IN CONJUNCTION WITH FOUR DIFFERENT RULES.

Parameters	***Orig. HWT***	***RB-HWT Rule 1***	***RB-HWT Rule 2***	***RB-HWT Rule 3***	***RB-HWT Rule 4***
λ	0	0.01	0.01	0.01	0.01
δ	0.34	0.16	0.16	0.16	0.16
ω	0.01	0.01	0.01	0.01	0.01
α_1	0.29	0.35	0.35	0.34	0.33
α_2	-	0.22	0.17	0.24	0.27
ϕ	0.98	0.97	0.97	0.97	0.97
σ_S/σ_N	-	2.84	2.81	2.48	2.63

The parameter estimates for the original and RB-HWT exponential smoothing method considered in this study are presented in Table 4.1. The smoothing parameters are denoted by λ, δ, ω, α_1 and α_2, while ϕ is the correction term for the first order autocorrelations in the error. The standard deviation of the model error for special days and normal days is denoted by σ_S and σ_N, respectively. Note that for the rule-based HWT exponential smoothing, we employ maximum likelihood estimation using a Gaussian error, for which the variance is different on special days to that on normal days. The ratio of error standard deviation is lower for Rule 3 compared to the other rules.

The rule-based SARMA model is represented as:

$$\left(1-\varsigma_1 L-\varsigma_2 L^2-\varsigma_3 L^3\right)\left(1-\beta_1 L^{m_1}-\beta_2 L^{2m_1}-\beta_3 L^{3m_1}\right)\left(1-\gamma_1 L^{m_2}-\gamma_2 L^{2m_2}-\gamma_3 L^{3m_2}\right)\Big(I_{N_t}(1-\delta_1 L^{Lag1}-\delta_2 L^{Lag2}-\delta_3 L^{Lag3})+I_{S_t}(1-\theta_1 L^{Lag1}-\theta_2 L^{Lag2}-\theta_3 L^{Lag3})\Big)(y_t-c)=(1-\lambda_1 L-\lambda_2 L^2-\lambda_3 L^3)\left(1-\rho_1 L^{m_1}-\rho_2 L^{2m_1}-\rho_3 L^{3m_1}\right)(1-\tau_1 L^{m_2}-\tau_2 L^{2m_2}-\tau_3 L^{3m_2})\Big(I_{N_t}(1-\upsilon_1 L^{Lag1}-\upsilon_1 L^{Lag2}-\upsilon_1 L^{Lag3})+I_{S_t}\big(1-\xi_1 L^{Lag1}-\xi_1 L^{Lag2}-\xi_1 L^{Lag3}\big)\Big)\left(I_{N_t}\varepsilon_t^{(N)}+I_{S_t}\varepsilon_t^{(S)}\right) \quad (4.25)$$

where $Lag1 = m_3(t)$, $Lag2 = Lag1 + m_3\big(t - m_3(t)\big)$, and $Lag3 = Lag2 + m_3\Big(t - m_3\big(t - m_3(t)\big)\Big)$. Note that for any period t, if $I_{N_t} = 1$, i.e. for a given period on a normal day, we set $m_3(t) = 52 \times 336$, except for some years, where we set $m_3(t) = 53 \times 336$ for a few weeks around the clock-change, as defined previously in Section 4.2. Similarly, if for any period t, $I_{S_t} = 1$, i.e. for a given period on a special day, we select $m_3(t)$ using a rule-based approach, as discussed in the previous section. The parameter estimates for best performing RB-SAMRA (with Rule 3), is represented as:

$$(1-0.22L-0.57L^2-0.15L^3)(1-0.18L^{m_1}-0.15L^{2m_1}-0.23L^{3m_1})(1-0.28L^{m_2}-0.03L^{2m_2}-0.04L^{3m_2})\Big(I_{N_t}(1-0.29L^{Lag1}-0.35L^{Lag2}-0.15L^{Lag3})+I_{S_t}(1-0.43L^{Lag1}-0.20L^{Lag2}-0.16L^{Lag3})\Big)(y_t-38399)=(1+1.04L+0.50L^2+0.14L^3)(1+0.06L^{m_1}-0.01L^{2m_1}-0.14L^{3m_1})(1-0.16L^{m_2}+0.03L^{2m_2})\Big(I_{N_t}(1-0.03L^{Lag1}-0.19L^{Lag2}-0.06L^{Lag3})+I_{S_t}(1-0.07L^{Lag1}-0.06L^{Lag2}-0.08L^{Lag3})\Big)\left(I_{N_t}\varepsilon_t^{(N)}+I_{S_t}\varepsilon_t^{(S)}\right) \quad (4.26)$$

For the triple seasonal ANN method, we obtained $\lambda_1 = 0.001, \lambda_2 = 0.001, \eta = 0.1, \mu = 0.9$ and $m = 24$. The parameter estimates for RB-ID SVD and RB-IW SVD exponential smoothing methods (along with their original counterparts), are presented in Table 4.2 and Table 4.3, respectively. The smoothing parameters are denoted by $\lambda, \delta, \omega, \alpha_1$ and α_2, while ϕ is the correction term for the first order autocorrelations in the error. The number of feature vectors and feature series used in the modelling is denoted by k. The standard deviation of the model error for special days and normal days is denoted by σ_S and σ_N, respectively.

TABLE 4.2

PARAMETER ESTIMATES FOR THE ORIGINAL AND RULE-BASED MODIFIED INTRADAY SVD (RB-ID SVD) EXPONENTIAL SMOOTHING METHOD USED IN CONJUNCTION WITH FOUR DIFFERENT RULES.

Parameters	***Orig. ID SVD***	***RB-ID SVD Rule 1***	***RB-ID SVD Rule 2***	***RB-ID SVD Rule 3***	***RB-ID SVD Rule 4***
λ	0.02	0.01	0.01	0.01	0.01
δ	0.41	0.18	0.18	0.18	0.18
ω	0.26	0.02	0.02	0.02	0.02
α_1	0.18	0.38	0.37	0.36	0.36
α_2	-	0.23	0.19	0.28	0.29
ϕ	0.96	0.97	0.98	0.97	0.97
k	20	32	33	33	33
σ_S/σ_N	-	2.67	2.74	2.36	2.54

Note that for RB-ID SVD, when used in conjunction with the different rules, we consistently estimated $k < m_1$, whereby the maximum value of k was thirty three (as shown in Table 4.2).

Similarly, for RB-IW SVD, we estimated $k \ll m_2$, for all four rules formulated in this study, whereby the maximum value of k was twenty nine (as shown in Table 4.3), which implies quite substantial reduction in the dimension of the model.

TABLE 4.3

PARAMETER ESTIMATES FOR THE ORIGINAL AND RULE-BASED MODIFIED INTRAWEEK SVD (RB-IW SVD) EXPONENTIAL SMOOTHING METHOD USED IN CONJUNCTION WITH FOUR DIFFERENT RULES.

Parameters	***Orig. IW SVD***	***RB-IW SVD Rule 1***	***RB-IW SVD Rule 2***	***RB-IW SVD Rule 3***	***RB-IW SVD Rule 4***
λ	0.02	0.01	0.01	0.01	0.01
δ	0.57	0.22	0.21	0.21	0.21
ω	0.31	0.05	0.06	0.06	0.05
α_1	0.12	0.44	0.44	0.43	0.43
α_2	-	0.24	0.20	0.29	0.31
ϕ	0.96	0.97	0.97	0.97	0.97
k	39	29	29	29	27
σ_S/σ_N	-	2.61	2.68	2.33	2.47

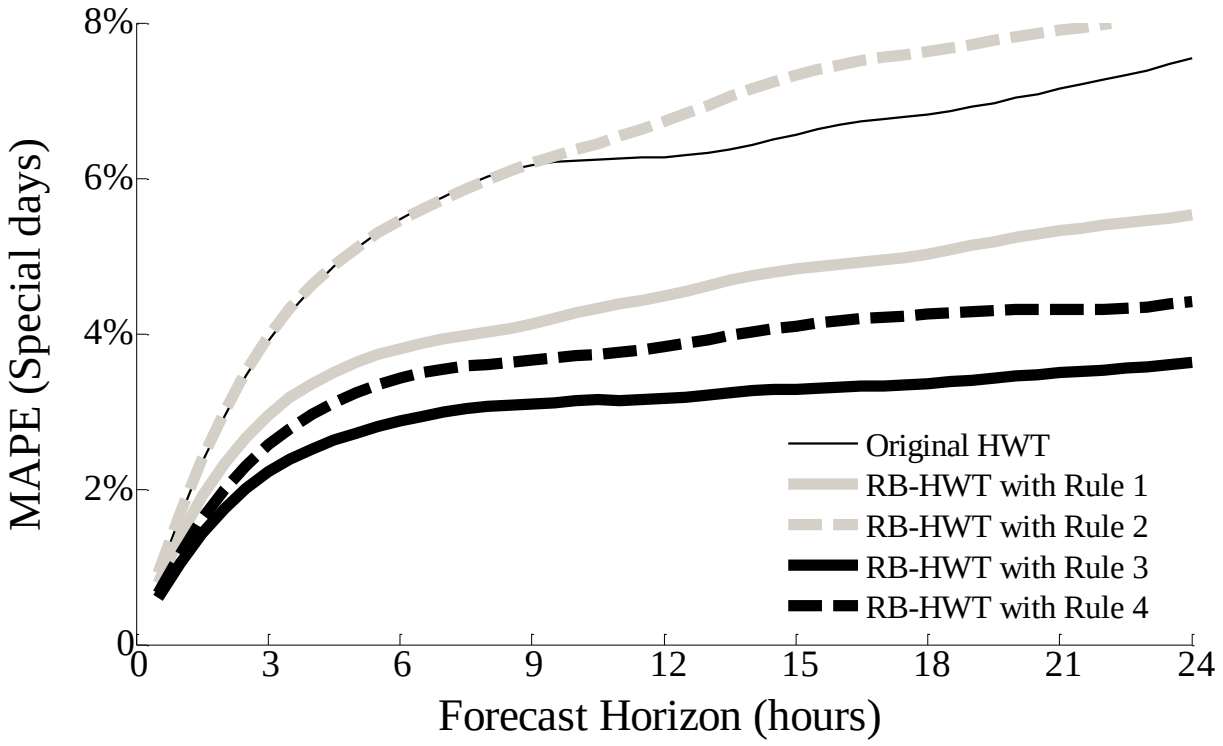

Figure 4.6— Out-of-sample MAPE across special days for the original and rule-based triple seasonal HWT exponential smoothing.

Figure 4.6 presents the HWT results for just the special days. As discussed earlier in Section 4.2, we generated point forecasts using a rolling forecast origin through the post-sample data, for horizons varying from one half-hour up to a day ahead. For each set of forecasts for a given horizon, we computed MAPE separately across both normal and special days. Figure 4.6 shows the original HWT method comfortably outperformed by all the RB-HWT approaches, apart from the one based on Rule 2, which performed poorly. The results for Rules 1 and 4 were similar, but Rule 3 led to substantially superior performance over all the other methods. The superior performance of the proposed methods over the original version highlights the value in using a subjective rule-based approach in situations where domain knowledge is available.

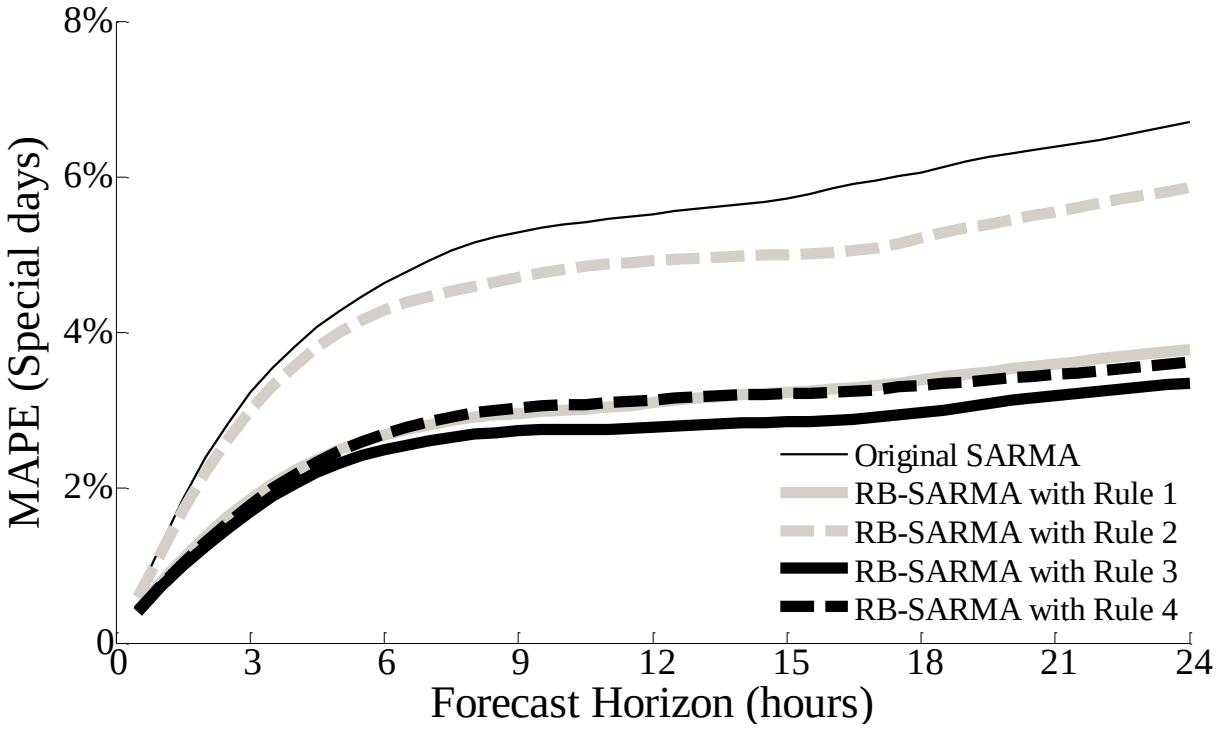

Figure 4.7— Out-of-sample MAPE across special days for the original and rule-based SARMA.

In Figure 4.7, we present the SARMA results for just the special days. The figure shows that, as with the HWT models, Rule 3 led to the best SARMA results, although its superiority over Rules 1 and 4 was relatively small. For both the HWT and SARMA methods, it is interesting to note from Figures 4.6 and 4.7 that the MAPE results for Rule 3 are about half the size of those for the corresponding original models that were not rule-based. As with RB-HWT, Rule 2 was disappointing for the RB-SARMA method. A possible explanation for Rules 2 and 4 being outperformed is that the length of lag $m_3(t)$ for these two rules tends to be larger than for the other rules. This leads to the incorporation of less recent information in the modelling.

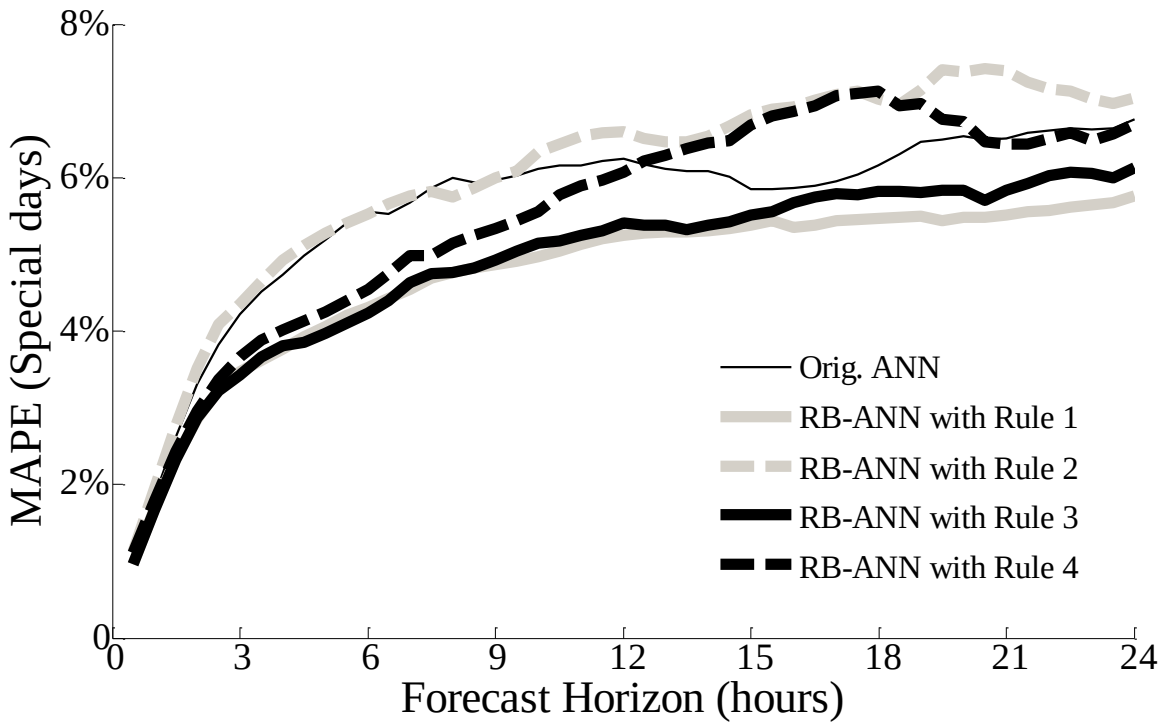

Figure 4.8— Out-of-sample MAPE across special days for the original and rule-based ANN.

Figure 4.8 shows the ANN results for just the special days. The most accurate results overall correspond to RB-ANN with Rules 1 and 3. Specifically, the original ANN method was less accurate than the RB-ANN methods with Rules 1 and 3. Across different lead times, the MAPE obtained using the ANN method is more volatile compared to the MAPE obtained using HWT exponential smoothing method and SARMA. This is perhaps due to the fact that we build a separate ANN model for each lead time, whereas for HWT exponential smoothing and SARMA, we generate multi-step ahead forecasts using a single model.

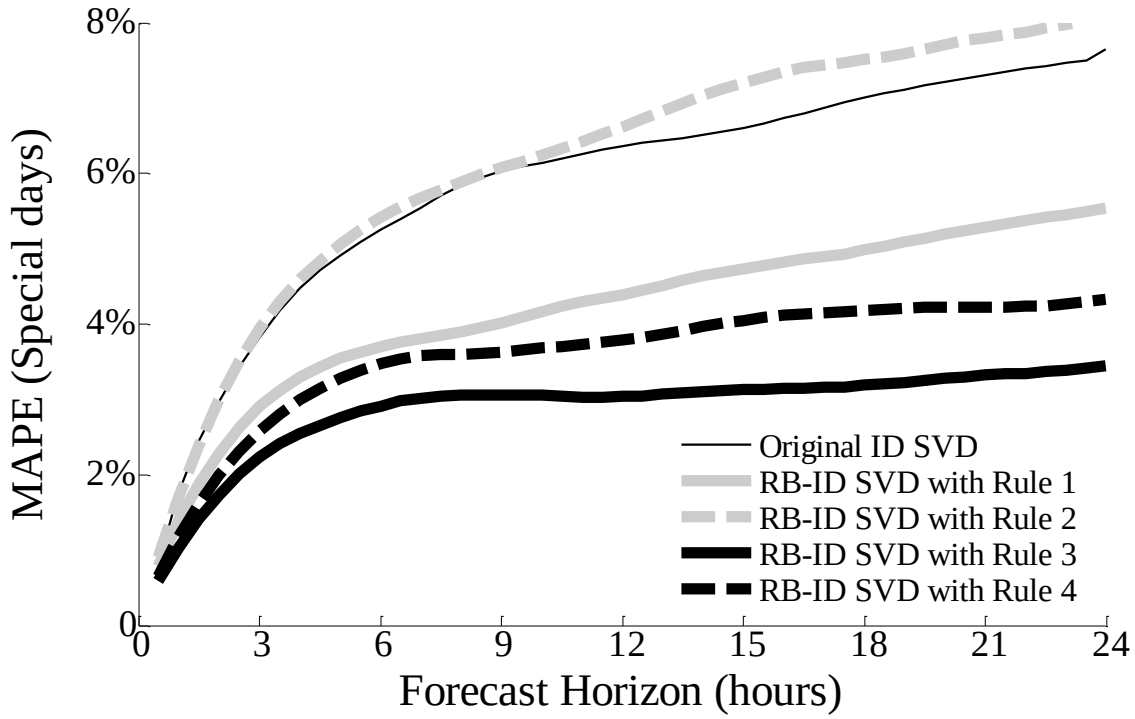

Figure 4.9— Out-of-sample MAPE across special days for the original and rule-based ID SVD.

In Figure 4.9, we present the results for ID SVD, across only the special days. The performance of the RB-ID SVD method is very similar to that of RB-HWT, when used in conjunction with different rules. As with RB-HWT, Rule 3 is the most accurate with RB-ID SVD. The performance of Rule 2 is the worst, while both Rule 1 and Rule 4 easily outperformed the original ID SVD method, whereby the performance of Rule 4 is considerably superior to Rule 1. It is noteworthy that all of the above four rule-based versions of ID SVD led to a reduction in the dimension of the model.

In Figure 4.10, we present the results for original and rule-based IW SVD, across only the special days. As with RB-HWT, RB-SARMA, and RB-ID SVD, Rule 3 is the best performing rule with RB-IW SVD. The performance of both Rule 1 and Rule 4 is very similar, with Rule 2 being the worst performing rule. Except for Rule 2, all other rules easily outperformed the original IW SVD method. Note that all of the above four rule-based versions of IW SVD led to

substantial reduction in the dimension of the model. We also investigated combining forecasts from the best performing SVD methods, namely RB-ID-SVD with rule 3 and RB-IW-SVD with rule 3. However, this did not result in a significant improvement in the post-sample forecast accuracy. Hence, we do not present results for this case in this chapter.

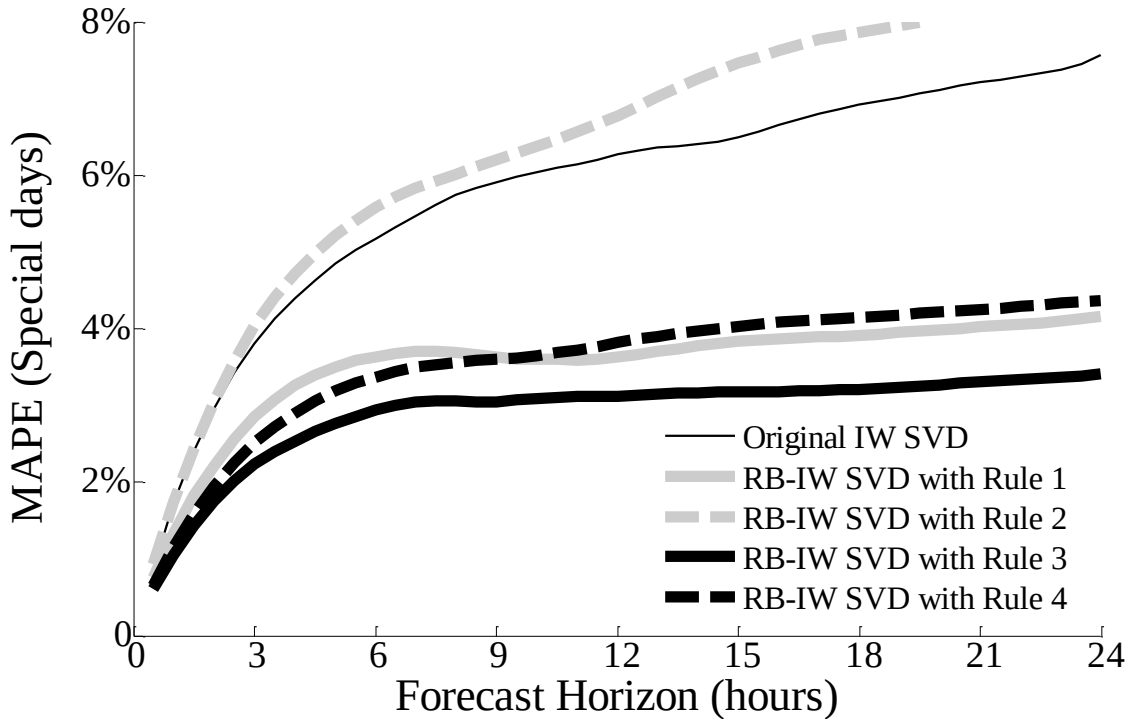

Figure 4.10— Out-of-sample MAPE across special days for the original and rule-based IW SVD.

It is noteworthy that for the rule-based adaptations of the five univariate methods considered in this study, Rule 2 is consistently the worst. A possible reason for this is that, unlike all the other rules, Rule 2 classifies days within the Christmas period as being the same. Hence, this result further justifies our approach to treat each special day as having a unique profile, which may change over the years.

In this study, we also include the following three simple benchmarks:

(i) *Seasonal random walk* – issue the load occurring on the same special day in the previous year as a forecast for a given special day.

(ii) *Seasonal moving average* – issue the mean of the load on the same special day in the previous four years as a forecast for a given special day.

(iii) *Recent Sunday* – issue the load on the most recent Sunday as a forecast for a given special day. This benchmark is used by Smith (2000).

We tested three further implementations of RB-ANN with Rule 1. (This rule was the most successful for RB-ANN.) As defined in Section 4.3.3, for the ANN built for horizon h, we used load at the forecast origin, and at the following lags: 1, 2, $m_1 - h$, $2m_1 - h$, $3m_1 - h$, $m_2 - h$, $2m_2 - h$, $3m_2 - h$, $LagA = m_3(t) - h$, $LagB = LagA + m_3(t - LagA - h)$, and $LagC = LagB + m_3(t - LagB - h)$. Each of the three further implementations of RB-ANN included a set of new inputs, in addition to the above described lags:

(1) *RB-ANN with calendar data* - We included as additional inputs: a counter running from 1 to 366 to indicate the day of the year; a counter running from 1 to 7 to indicate day of the week; a counter running from 1 to 336 to indicate half-hour of the week; and a counter running from 1 to 48 to indicate half-hour of day. Furthermore, following Fidalgo and Lopes (2005), we included sine and cosine function values of these 'counter variables'. We also included I_{N_t} as defined previously in Section 4.3.1.

(2) *RB-ANN with extra lags* – We used load at lags described in Section 4.3.3, whereby $m_3(t)$ was chosen to be the same as the intrayear cycle length used in the original ANN method. We also used, as extra inputs, load at *LagA*, *LagB*, and *LagC* (as defined in

Section 4.3.3), whereby for only the special days, $m_3(t)$ was selected using a rule-based approach, while for the normal days, $m_3(t)$ was the same as the intrayear cycle length, as described above. We also included I_{N_t} as an input.

(3) Separate ANN - Following Kim *et al*. (2000), we built separate ANNs for normal and special days. The results for implementations (1) and (2) were similar, with (3) being the worst. Hence, we only present results for implementation (1) in this chapter.

In Figure 4.11, we plot the MAPE results for the best performing method from Figures 4.6, 4.7, 4.8, 4.9 and 4.10, along with the following three simple benchmarks: (i) Seasonal random walk, (ii) Seasonal moving average, and (iii) Recent Sunday. The worst performing methods in Figure 4.11 are the benchmarks.

Of the three rule-based methods, the ANN results were the poorest, which is perhaps due to the absence of strong nonlinearity in the structure of the time series. Also, including calendar data as an additional input did not greatly improve the results of the RB-ANN method. Moreover, following Taylor *et al*. (2008), we incorporated an error correction model within the RB-ANN modelling framework. Specifically, if $E_T(h)$ denote the h step ahead forecast error, having corresponding forecast origin denoted by T, the error correction model is represented as $E_T(h) = \phi_h E_{T-h}(h)$. The error correction term ϕ_h is estimated separately for each horizon, using ordinary least squares regression, employing only the cross-validation hold out sample (corresponding to last one year of the estimation sample). However, using the error correction did not lead to a noticeable improvement in the forecast accuracy on the post-sample data. Hence, results for this case are not presented here. Future studies could investigate the efficacy of recurrent neural networks (RNNs) for modelling anomalous load. RNNs work as an

associative memory (Hopfield, 1982), and have been shown to be more accurate than the feedforward ANNs for nonlinear time series prediction (see, for example, Adam *et al*. 1993). As discussed by Tenti (1996), a problem with RNNs is that they require significantly more connections, and memory for simulation, compared to the ANNs based on backpropagation.

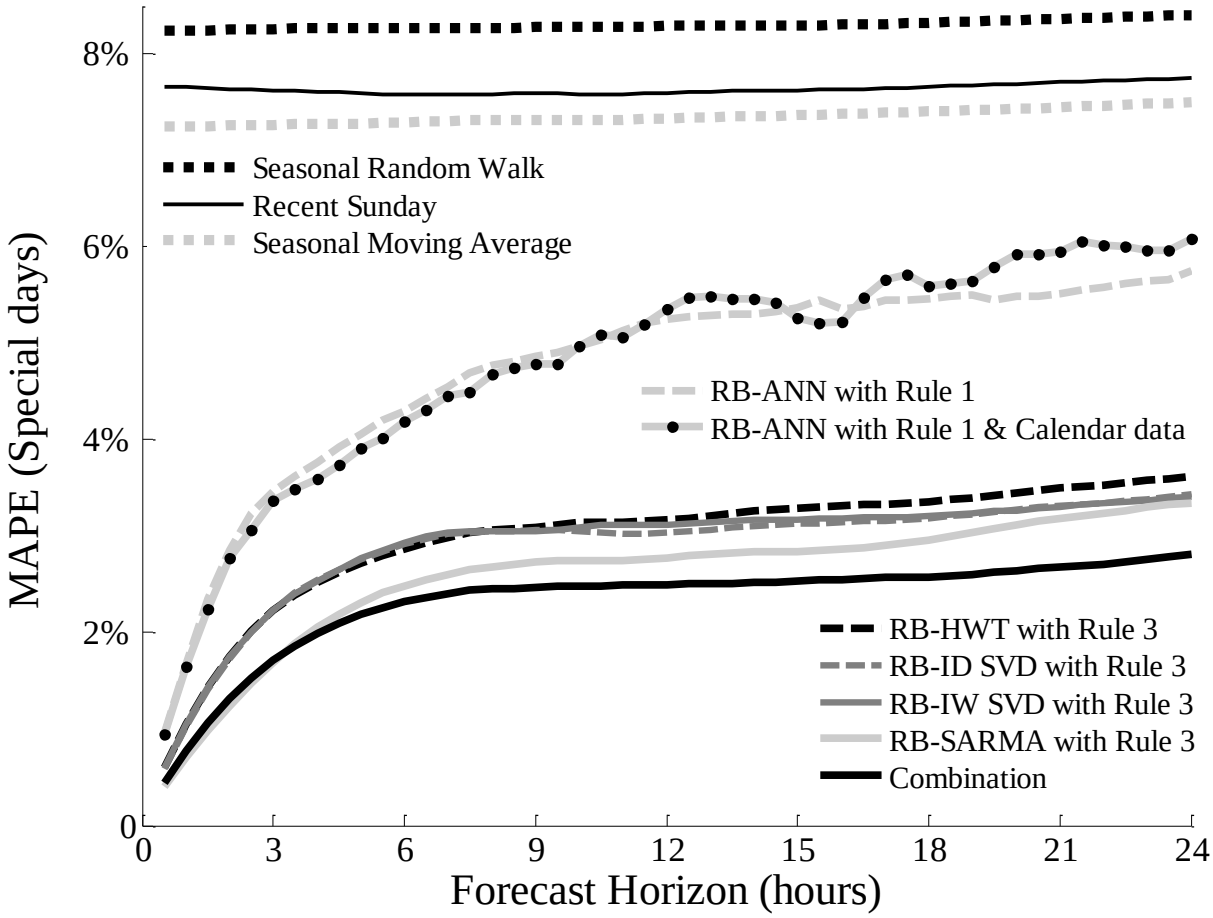

Figure 4.11— Out-of-sample MAPE across special days for the best performing method from Figures 4.6, 4.7, 4.8, 4.9 and 4.10, along with the simple benchmarks. This figure also presents the MAPE obtained using RB-ANN method with calendar data, and a combination of the two competitive rule-based methods, namely, RB-HWT with Rule 3, and RB-SARMA with Rule 3.

As shown in Figure 4.11, the results for RB-ID SVD and RB-IW SVD were similar with the former slightly preferable. Both SVD-based methods were competitive with RB-HWT, while RB-SARMA was the most accurate method overall. Given that SARMA and HWT are

competitive, and are based on different modelling approaches, we consider combining forecasts from these two methods using a simple average. The findings presented in Figure 4.11 shows that this led to greater accuracy than the individual models.

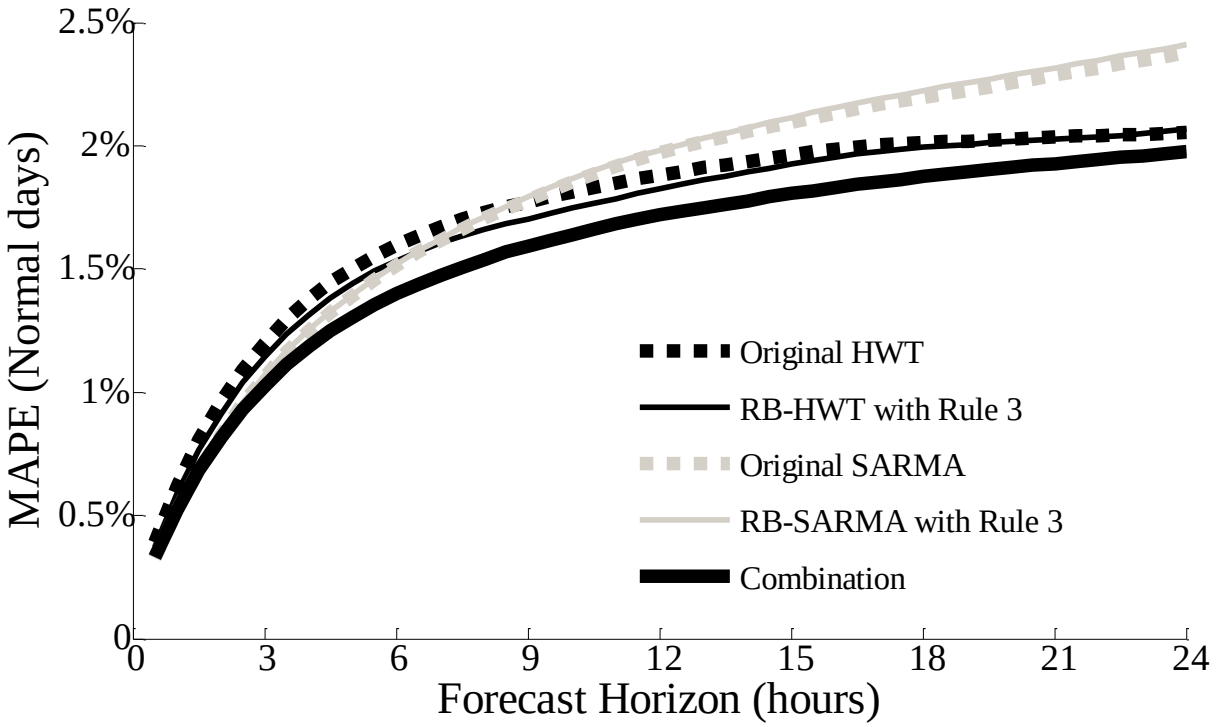

Figure 4.12— Out-of-sample MAPE across normal days, for the two competitive rule-based methods from Figure 4.11 (rule-based HWT and SARMA, both used with Rule 3), the corresponding original versions, and their combination.

In Figure 4.12, we report the forecasting performance of the original and rule-based versions of SARMA and HWT across normal days. Although we developed the rule-based modifications with the aim of improving the forecasting of special days, we would hope that these modifications would not lead to poor forecast accuracy for the normal days. Therefore, in Figure 4.12, it is reassuring to see that the results of the original and rule-based versions are similar for each method. A combination of the best performing rule-based versions of SARMA and HWT generated the most accurate forecasts across normal days, for all lead times considered in this study.

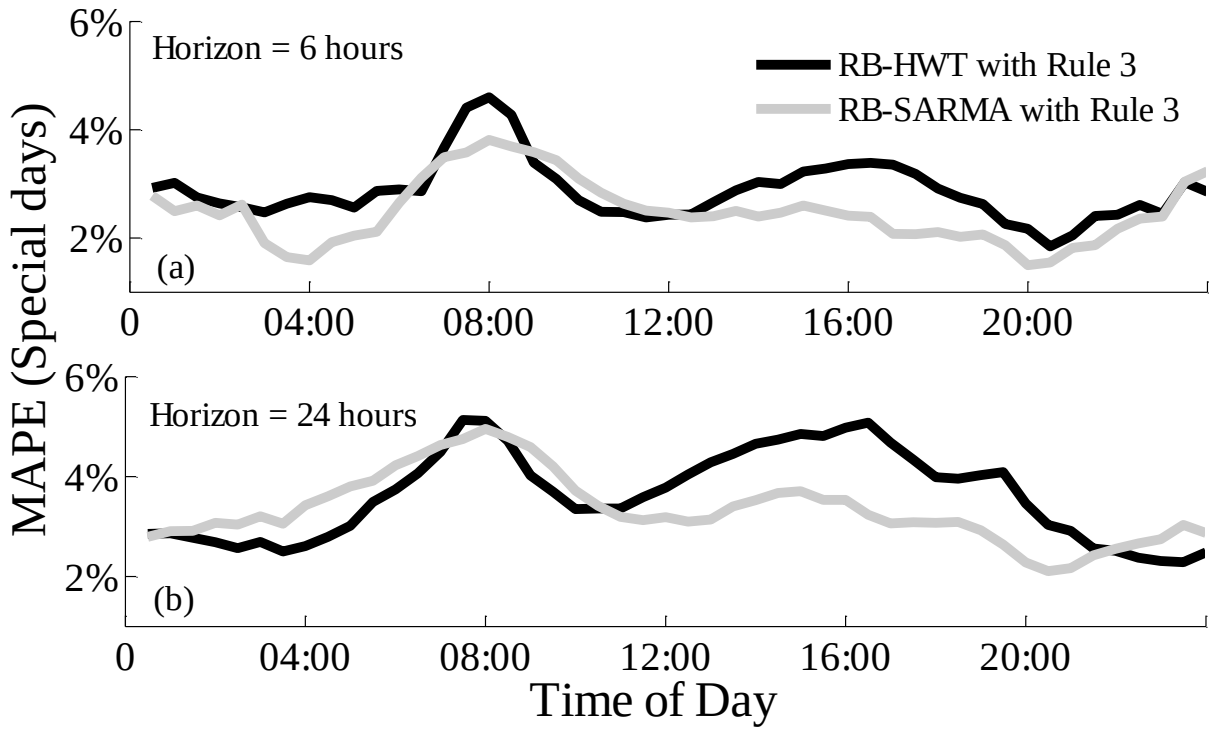

Figure 4.13— Out-of-sample MAPE across special days, obtained using RB-HWT and RB-SARMA with Rule 3, plotted against different periods of the day, for horizon equal to: (a) six hours, and (b) twenty four hours.

In Figure 4.13, we plot the out-of-sample MAPE results (across all special days) against each period of the day for the RB-HWT and RB-SARMA method, both used with Rule 3. Figure 4.13a shows the MAPE for six hour ahead forecasts, while Figure 4.13b plots the MAPE for one day ahead forecasts. As expected, the large MAPE values correspond to the periods of day when load witnesses a relatively large change. It can also be seen from these two figures that the MAPE obtained using RB-SARMA is lower than the RB-HWT method for most periods of the day.

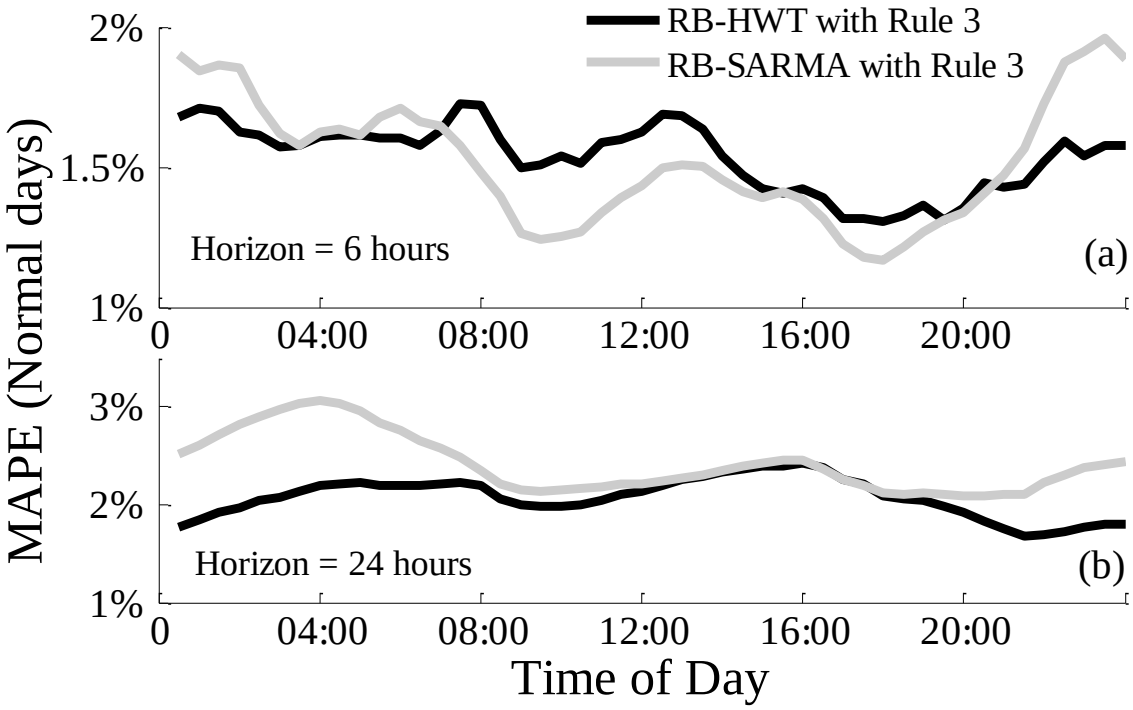

Figure 4.14— Out-of-sample MAPE across normal days, obtained using RB-HWT and RB-SARMA with Rule 3, plotted against different periods of the day, for horizon equal to: (a) six hours, and (b) twenty four hours.

In Figure 4.14, we plot the out-of-sample MAPE results across all normal days, against each period of the day for the RB-HWT and RB-SARMA method, both used with Rule 3. In Figure 4.14a, we plot MAPE for six hour ahead forecasts, while in Figure 4.14b, we show MAPE for one day ahead forecasts. For a given forecast horizon, as expected, the MAPE for rule-based approaches across normal days is considerably smaller compared to the MAPE obtained on special days. For one day ahead forecasts, we note that the MAPE for rule-based HWT is less than the MAPE for rule-based SARMA, for most periods of the day.

Given that anomalous load observations constitute approximately five percent of the time series, the reported reduction in error on special days can have significant practical implications. We identified eighteen special days in the one-year out-of-sample data.

4.5.2 Results using load for France

To investigate the efficacy of the proposed methodology for forecasting different load time series, we formulated rules using the French data. We used nine years of half-hourly load for France, stretching from 1 Jan. 2001 to 31 Dec. 2009. As done for the Great Britain data, we used the first eight years of the French data for estimating the model parameters, while observations from the final year were used for model evaluation. In all, we identified twenty four special days in the one-year post-sample period. Thus, load observations corresponding to special days constitute approximately 6.5% of the total available observations.

As done with GB load time series, we formulate four different rules for the French load. However, for a given rule-based method, we found that the forecast performance of all four rules was similar on the post-sample data. Hence, for France, we present the formulation of only the best performing rule (referred to as Rule 5). This rule is similar to Rule 3 formulated using the GB time series. The rationale of this rule is to treat weekdays as having a different impact on anomalous load, in comparison to weekends, while allowing each special day to have a unique profile. During the modelling, this rule requires that the current and historical anomalous load observations belong either to a weekday, or a weekend. Moreover, this rule does not classify different special days as being the same.

The advantage of Rule 5, is that it incorporates recent information in the model by utilizing the similarity in the intraday cycle, models proximity days separately, and incorporates the effect of bridging days on anomalous load. Let us now describe the formulation of this rule, by creating an ensemble of cases for modelling anomalous load.

Specifically, depending on the special day under consideration, we select $m_3(t)$ that adheres to some combination of the following cases (as specified in Table 1), such that both y_t and $y_{t-m_3(t)}$ are:

Case A: Same special day types, and, fall on the same period of the day.

Case B: Weekends.

Case C: Bridging proximity days.

Case D: Proximity days that precede a special day.

Case E: Proximity days that follow a special day.

Table 4.4 presents the complete list of special days observed in the one year post-sample period for France. For each special day observed in 2009, the table presents a selection of different cases (from the above defined ensembles of cases), which are applicable for the special day under consideration. The combination of cases determines the category to which a given special day belongs to, using which we compute the intrayear cycle length $m_3(t)$. The table also presents the corresponding historical special day, whose observations are included in the model to generate forecasts.

In Table 4.4, the seven days of the week are denoted by *M*. (Monday), *Tu. (Tuesday)*, *W*. (Wednesday), *Th*. (Thursday), *F*. (Friday), *Sa*. (Saturday) and *Su*. (Sunday). Moreover, it is noteworthy that Ascension Day in 2008 occurred on the same day as Labor day (1 May 2008), hence, while modelling load for this special day in 2009, we refer to the profile of Ascension Day in 2007 in order to avoid any discrepancy.

TABLE 4.4

LIST OF SPECIAL DAYS IN FRANCE OBSERVED IN 2009, AND THE CORRESPONDING HISTORICAL SPECIAL DAY REFERRED TO VIA A SELECTION OF DIFFEERENT CASES USING RULE 5.

Current Special Day Type : *Day of the Week*	**Date**	**Previous Special Day Referred via $m_3(t)$**	**Date**
Case A: Special days which occur on a weekday.			
1. New Year's Day : *Th.*	01/01/2009	New Year's Day : *Tu.*	01/01/2008
2. Easter Monday : *M.*	13/04/2009	Easter Monday : *M.*	24/03/2008
3. Labor Day : *F.*	01/05/2009	Labor Day : *Th.*	01/05/2008
4. WWII Victory Day : *F.*	08/05/2009	WWII Victory Day : *Th.*	08/05/2008
5. Ascension Day : *Th.*	21/05/2009	Ascension Day : *Th.*	17/05/2007
6. Whit Monday : *M.*	01/06/2009	Whit Monday : *M.*	12/05/2008
7. Bastille Day : *Tu.*	14/07/2009	Bastille Day : *M.*	14/07/2008
8. Virgin Mary Day : *Sa.*	15/08/2009	Virgin Mary Day : *F.*	15/08/2008
9. Remembrance Day : *W.*	11/11/2009	Remembrance Day : *Tu.*	11/11/2008
10. Christmas Day : *F.*	25/12/2009	Christmas Day : *Th.*	25/12/2008
11. New Year's Eve : *Th.*	31/12/2009	New Year's Eve : *W.*	31/12/2008
Case AB: Special days which occur on a weekend.			
12. All Saints Day : *Su.*	01/11/2009	All Saints Day : *Sa.*	01/11/2008
13. Boxing Day : *Sa.*	26/12/2009	Boxing Day : *Su.*	26/12/2004
Case ACD: Bridging proximity days that precede a special day.			
14. Day before Bastille Day : *M.*	13/07/2009	Day before Bastille Day : *F.*	15/07/2005
Case ACE: Bridging proximity days that follow a special day.			
15. Day after New Year's : *F.*	02/01/2009	Day after New Year's : *F.*	02/01/2004
16. Day after Ascension : *F.*	22/05/2009	Day after Ascension : *F.*	18/05/2007
Case AD: Proximity days that precede a special day.			
17. Christmas Week : *M.*	21/12/2009	Christmas Week : *M.*	22/12/2008
18. Christmas Week : *Tu.*	22/12/2009	Christmas Week : *M.*	22/12/2008
19. Christmas Week : *W.*	23/12/2009	Christmas Week : *Tu.*	23/12/2008
20. Christmas Week : *Th.*	24/12/2009	Christmas Week : *W.*	24/12/2008
Case ABE: Proximity days that occur on a weekend, and follow a special day.			
21. Christmas Week : *Su.*	27/12/2009	Christmas Week : *Sa.*	30/12/2006
Case AE: Proximity days that follow a special day.			
22. Christmas Week : *M.*	28/12/2009	Christmas Week : *M.*	29/12/2008
23. Christmas Week : *Tu.*	29/12/2009	Christmas Week : *M.*	29/12/2008
24. Christmas Week : *W.*	30/12/2009	Christmas Week : *Tu.*	30/12/2008

In Figure 4.15, we present MAPES for the best performing rule-based methods, and their original counterparts. It is encouraging to see that the error for rule-based adaptations of HWT, SARMA and ANN are less than their original counterparts which are not rule-based. Overall, RB-SARMA with Rule 5 generates the most accurate forecasts. In accordance with the GB results, the performance of RB-ANN is not competitive with RB-HWT and RB-SARMA for the French load. Moreover, RB-HWT is less accurate than RB-SARMA at all horizons. The performance of the rule-based SVD methods was not competitive with the RB-HWT method, hence, for conciseness, we do not present results for the SVD based methods in this thesis. We have shown that the rules formulated in this study can be easily adapted to different datasets, whereby the proposed methodology allows the modelling of both normal and anomalous load in a coherent and unified framework.

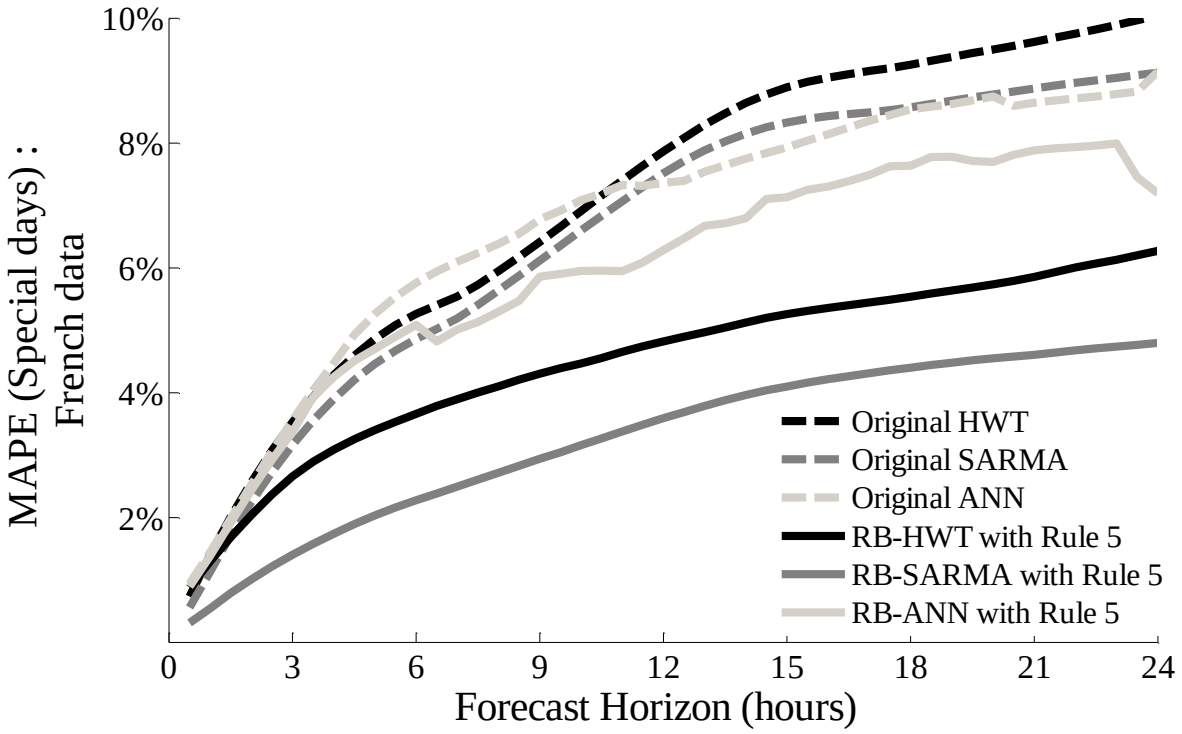

Figure 4.15— Out-of-sample MAPE across special days, for the original and rule-based HWT, SARMA, and ANNs, using the French load time series.

4.6 Summary and Concluding Remarks

In this study, we developed forecasting methods that can be used to effectively assist the central controller, and potentially be employed for generating real-time online forecasts in an automated framework under both normal and anomalous load conditions. Our empirical study showed that by using a number of empirically inferred rules, the accuracy of conventional methods can be significantly improved for anomalous load forecasting. We evaluated five different univariate modelling approaches, when used in conjunction with four different rules. Apart from formulating rules, the methodological contribution of this chapter lies in adapting some of the most commonly used univariate models to model normal and anomalous load. It is noteworthy that the formulated rules are flexible enough to adapt to new observations. We would suggest that the new rule-based HWT method has the appeal of simplicity when compared with rule-based versions of SARMA, ANN, and SVD methods, both in terms of the model structure and number of parameters. Overall, a combination of the rule-based HWT and SARMA methods was the most accurate across all horizons for both normal and special days.

An interesting line of future work would be to set the annual lag differently for different times of the same special day. This would amount to using observations from different parts of different past special days for a future special day. Also, it would be worth investigating the efficacy of the rule-based approaches developed in this work for different applications, such as forecasting call centre arrivals, hospital admissions, and transportation counts, as the corresponding time series exhibit seasonality and anomalous conditions pose significant modelling challenges.

Chapter 5

MODELLING ELECTRICITY SMART METER DATA BASED ON KERNEL DENSITY ESTIMATION

5.1 Introduction

The recent advent of smart meters has led to big micro-level datasets. A smart meter is an electronic device that measures electricity consumption at the installed facility, and transmits this information to the consumer and the energy supplier/operator on a near real-time basis. For the first time, the electricity consumption at individual sites is available on a near real-time basis. Efficient management of energy resources, electric utilities, and transmission grids, can be greatly facilitated by harnessing the potential of this data. The aim of this chapter it to generate probability density estimates for consumption recorded by individual smart meters. Such estimates can assist decision making by helping consumers identify and minimize their excess electricity usage, especially during peak times. For suppliers, these estimates can be used to devise innovative time-of-use (TOU) pricing strategies aimed at their target consumers.

We use kernel-based methods to enable nonparametric estimation, as it has been shown to be an efficient way of estimating the true underlying density (Silverman, 1986). The prominent intraday and intraweek variation in consumption prompt us to model consumption in terms of the period of week, and day of week, using a conditional kernel density (CKD) estimator. In this study, we propose methods based on kernel density (KD) estimation (see, for example, Rosenblatt, 1952; Silverman, 1986; Hastie *et al*., 2009) and conditional kernel density (CKD) (see, for example, Parzen, 1962; Rosenblatt, 1969; Hyndman *et al*., 1996; Jeon and Taylor, 2012). This method captures the seasonality in consumption, and enables a nonlinear and

nonparametric estimation of its conditional density. Within the CKD method, we use a discounting factor, which allows more emphasis to be placed on the more recent observations. In this section, we discuss the motivation and challenges in modelling consumption, provide a literature review on CKD estimator and studies undertaken using smart meter data, and specify the contributions of this study.

5.1.1 Motivation and Challenges

It is anticipated that based on smart meter information, significant energy and money savings can be achieved through detailed consumer feedback and tariffs designed for facilitating energy savings (National Consumer Council, 2008). It is estimated that just one degree Celsius turn-down of the gas central heating thermostat by all consumers could reduce domestic gas use by as much as ten percent (Domestic Metering Innovation, 2006). The large-scale installations of smart meters will generate massive amounts of data, offering unique insight into the consumption behaviour of different consumers. Over the coming years, smart meters are scheduled to replace the existing electronic meters. It is estimated that by 2019, approximately 60 million meters will be installed and operable in the United States alone (Edison Foundation Report, 2010). Similar installation plans are underway for Europe, whereby the electricity directive requires all member states to have smart meters installed for 80% of consumers by 2020, with full deployment to have been completed by 2022 (Directive 2009/72/EC).

Unlike conventional meters, smart meters provide site-specific information regarding electricity consumption over different periods of the day. This information can potentially change the landscape of energy markets, by allowing suppliers to make highly data-dependent decisions to develop innovative dynamic pricing strategies for their target consumers. Smart

meters, along with different time-of-use tariffs, can help consumers shift their consumption away from peak hours, which can result in significant efficiency savings (Cosmo *et al*., 2012). Moreover, smart meters can assist operators in decisions regarding scheduling, thereby facilitating safe and efficient operation of the utility. Given the potential of smart meters in facilitating efficient use of energy, and its financial implications for energy markets, it is imperative to develop methods for modelling electricity smart meter data accurately.

Smart meters would potentially play a major role in the preparation of smart grids (see, ERGEG, 2010). Given the massive scale of planned smart meter deployments across Europe and the US, there is an uncertainty associated with the impact that smart meters and smart grids would have on our societies and the environment. It has been estimated that the total cost of investment in smart electricity grids in Europe alone would be around €51 billion, whereby the benefits from smart grids could vary between €14 billion to €67 billion, depending mainly on the extent at which dynamic pricing is adopted, and also on the configuration of the grid (Faruqui *et al*., 2009). Moreover, modelling smart meter data could help improve the economics of demand-side aggregation products and promote larger structural market debates.

Regarding the motivation for this study, it is worth emphasizing that accurate future estimates of consumption can encourage consumers to identify and reduce their excess consumption, thereby facilitating an overall reduction in energy usage. For energy retailers, such estimates can help develop innovative TOU tariffs targeting consumers, thereby increasing the probability that the consumers will react to those tariffs, especially during peak times. For energy operators, future estimates of consumption can support informed decision-making regarding power scheduling, thereby facilitating safe and efficient operation of the electric utility.

Electricity consumption data from individual smart meters exhibits seasonality, comprising both intraday and intraweek seasonal cycles, along with considerable variability. Adequate assessment of the underlying variability in future consumption estimates can help both consumers and suppliers make informed decisions. Gardner and Rogers (1999) emphasize that electricity demand is a major source of uncertainty in planning electric power systems. We attempt to encapsulate the uncertainty in consumption by modelling its probability density, and not just a point estimate. The literature on modelling electricity smart meter data is very limited. There are few papers on short-term point forecasting of smart meter data (Ghofrani *et al.*, 2011; Ding *et al*., 2011) however, to the best of our knowledge, there are no existing studies on modelling the density of electricity consumption data from individual smart meters.

5.1.2 Historical Review

Most studies on smart meter data focus on estimating the impact of TOU pricing strategies on electricity consumption. Using Irish smart meter data for residential consumers, Cosmo *et al.* (2012) reported that TOU tariffs have a significant effect in reducing the consumption, especially during peak hours. Interestingly, Spector *et al*. (1995) found that most firms in the US and Israel did not react at all to newly introduced TOU tariffs, but the firms who responded made substantial changes. Alberini *et al*. (2011) analysed data for electricity and gas consumption for over 74,000 households in the US, and found a strong household response to energy prices. A similar study was undertaken by Bartush *et al*. (2011) using demand-based TOU tariffs for 500 households in Sweden. Although we do not focus on the impact of TOU tariffs on consumption, our work links to this area, because estimates of the probability density function of consumption for individual customers can potentially help suppliers devise innovative TOU pricing strategies.

In a recent trial involving more than 5000 electricity smart meter installations in Ireland, the deployment of TOU tariffs and information stimuli resulted in an overall reduction in electricity usage by 2.5% and peak usage by 8.8%, whereby 82% of the consumers reported change in their consumption patterns during the course of the trial (CER, 2011b). Interestingly, consumers who were also provided with a bi-monthly bill, energy statement, and an in-home display (IHD) device (in addition to TOU tariffs and information stimuli), were found to reduce their overall electricity usage by 3.2% and peak usage by 11.3%, while 91% of the consumers felt that electricity monitors were useful to help reduce peak usage. In a prepayment trial conducted by the CER, over 85% of the consumers said they were satisfied with prepayment as a mode for paying bills, and the consumers suggested that they would value a meter providing real-time data on their usage and credit status (CER, 2011a). To improve the quality of consumption feedback, information regarding future consumption estimates (and corresponding cost estimates) could potentially form a part of the detailed consumption feedback report, and the IHD device.

The literature on applications based on the CKD estimator is rather limited, especially in the context of time series forecasting. Hyndman *et al*. (1996) used the CKD estimator to model daily temperature, by conditioning current temperature on lagged temperature observations and a seasonal variable. Ruan (2010) employed the CKD estimator for real-time prediction of the respiratory response. Bessa *et al*. (2012) proposed a recursive time-adaptive CKD estimator, which adapts with the arrival of new observations without re-computing the entire density from scratch. Jeon and Taylor (2012) used CKD to predict the density of wind power conditional on wind velocity densities. It is noteworthy that kernel density estimators are used in machine learning and have been employed for forecasting electricity demand and price time series (see, for example, Nikolaev and de Menezes, 2008).

5.1.3 Contributions

The non-Gaussian and highly variable nature of individual smart meter electricity consumption data motivates nonparametric probability density estimation methods. In this study, we propose methods based on kernel density (KD) estimation and conditional kernel density (CKD) (see, for example, Rosenblatt, 1969). The conditioning in our CKD implementations aims to capture the seasonality in the smart meter data. Although this seasonality is usually far less clear than the seasonality in a time series of the total electricity demand for a country, both types of data tend to exhibit intraday and intraweek seasonal cycles. This prompts us to consider forms of KD and CKD that are inspired by the structure of models presented in the literature for modelling total national consumption, namely the exponential smoothing and autoregressive moving average (ARMA) models presented by Taylor (2003) and Gould *et al*. (2008), which were described in Chapter 3.

The KD and CKD methods that we propose to use only the historical consumption observations. The additional use of weather data would necessitate the availability and affordability of weather predictions for a location reasonably close to each smart meter. Furthermore, the use of weather data in a large-scale online prediction system raises issues of robustness. In view of these potential obstacles to implementation, we focus in this initial study on KD and CKD methods that do not incorporate weather data. We present the different KD and CKD methods in Section 5.3.

This chapter evaluates KD and CKD methods for estimating the density of electricity consumption using half-hourly data recorded from one thousand smart meters. The methodological contribution of this chapter lies in proposing two novel methods, namely KD estimation based on the type of intraday cycle (KD-IC) and CKD estimation conditional on the

type of intraday cycle (CKD-IC). Moreover, we incorporate a novel discounting method within the KD and CKD modelling framework, whereby weights decay exponentially in uniform steps, which ensures that observations belonging to the same week are given the same decay weight. We evaluate predictive accuracy for lead times from one half-hour up to one week ahead. The data corresponds to residential consumers, as well as small and medium-sized enterprises (SMEs). We evaluate the density, quantile, and point forecasts from the kernel estimation methods, using a range of different performance scores. The data-driven methods proposed in this study for smart electricity meter data can potentially be used for modelling data from smart gas or water meters. The evaluation of different methods considered in this study is presented in Section 5.4.

Recent advances in smart metering technology may pave the way for easy switching between suppliers, and between credit payment and prepayment schemes (Darby, 2012). Given the future prospect of a consumer to be able to switch between different suppliers and payment schemes, it is crucial to develop a framework through which the consumers may be able to compare different tariffs. In this study, we derive prediction intervals for electricity cost using the density estimates of electricity consumption, for different TOU tariffs. This would allow the consumers to compare different costs that would potentially be incurred in the future, for each available tariff, and select the tariff that would result in highest cost savings. For example, a consumer may want to switch tariff on a weekly basis, by selecting the tariff that would incur the lowest cost. Specifically, the consumers could use this information to appropriately switch between different tariffs provided by a given supplier, or switch between different suppliers. We present this study in Section 5.5.

5.2 Smart Meter Data Characteristics

We use eight months of half-hourly smart meter data for electricity consumption from 2 January 2010 to 31 August 2010, which amounts to a time series of 11616 observations from each meter. The data was recorded from 800 residential consumers, and 200 SMEs. There are no missing observations in the employed dataset. The data has been obtained from the Commission for Energy Regulation (CER) based in Ireland.

Figure 5.1 shows the smart meter consumption for a residential consumer, while Figure 5.2 presents consumption for an SME. As expected, consumption is lower for the residential consumer than for the SME. Furthermore, as can be clearly seen from Figures 5.1b and 5.2b, consumption for residential consumers is rather volatile, whereas consumption for the SME displays a prominent seasonal pattern.

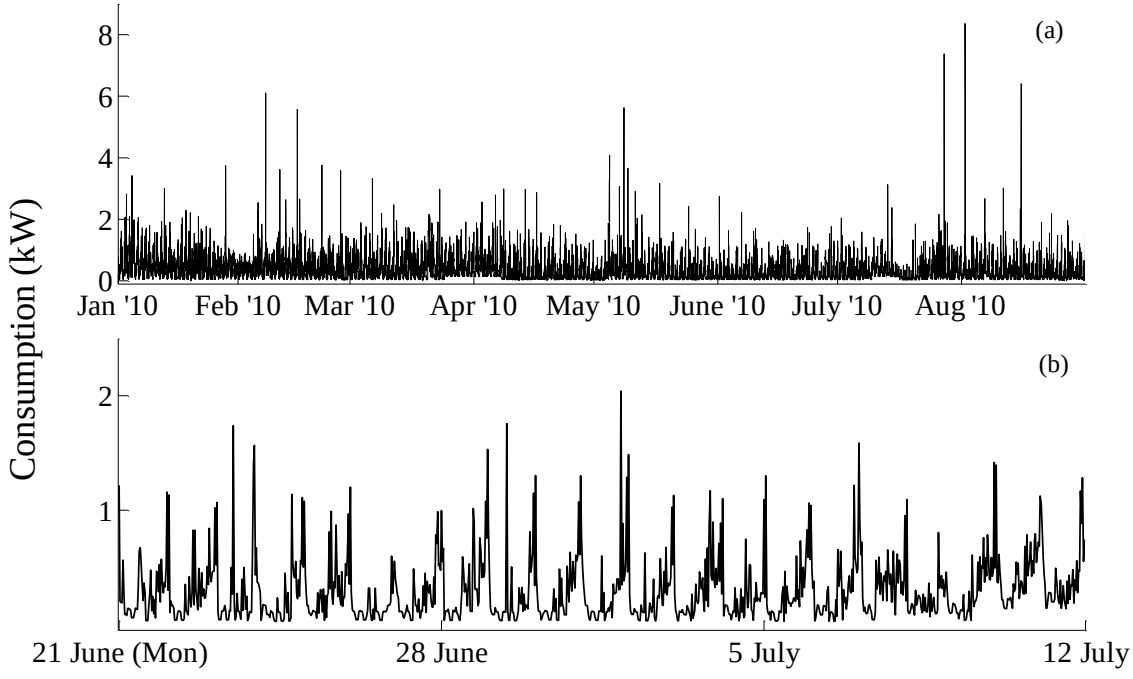

Figure 5.1— Consumption recorded for a residential consumer for: a) eight months, and b) three weeks.

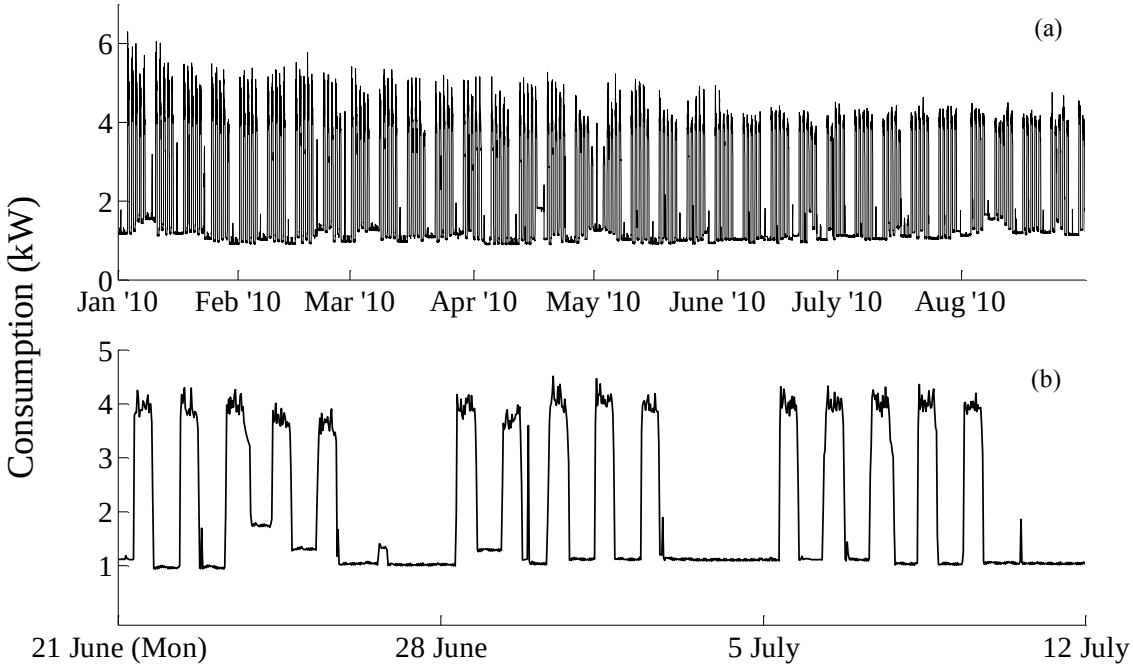

Figure 5.2— Consumption recorded for an SME for: a) eight Months, and b) three Weeks.

Figure 5.3 presents the average daily (intraday) consumption cycle for the same residential consumer and SME considered in Figures 5.1 and 5.2. We calculated the averages shown in Figure 5.3 using the first seven months of data, which as we discuss later, constitutes the estimation sample for all methods that we considered. For the residential consumer, consumption on weekends is not noticeably different to the consumption on weekdays, whereas for the SME, consumption is different on weekdays to the consumption on weekends. For the SME, consumption is significantly higher on weekdays, with the consumption being very low on weekends. Moreover, consumption for the residential consumer is high during the evenings on all days of the week, while for the SME, consumption is high only during the typical working hours of weekdays.

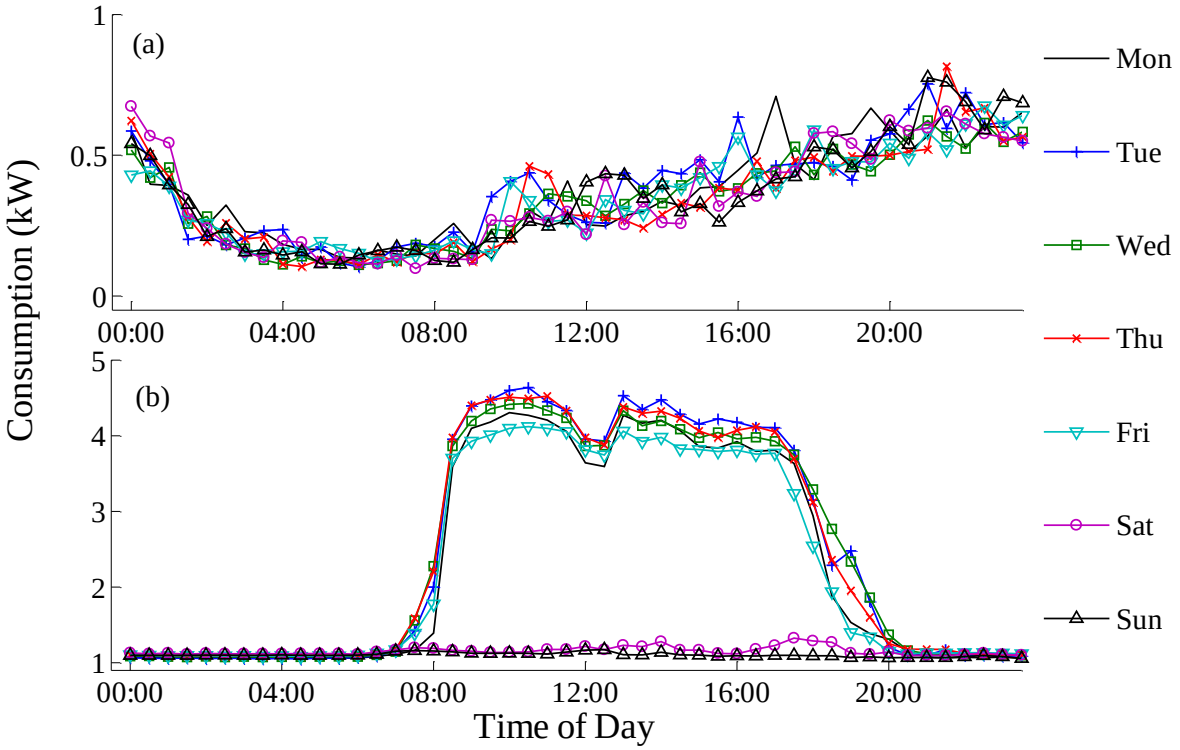

Figure 5.3— Average intraday profile for each day of the week for: (a) a residential consumer, and (b) an SME.

Figures 5.1 to 5.3 indicate that the series exhibit, to varying degrees, repeating intraweek seasonal cycles, and also, at least for the weekdays, repeating intraday cycles. The density estimation methods proposed in this study aim to capture this double seasonal variation in consumption.

As we explained earlier, our dataset consisted of 1000 time series, each consisting of eight months of half-hourly consumption observations. We used the first seven months of the data, comprising 10128 observations (*in-sample*), for optimizing method parameters, while the final month, constituting 1488 observations (*post-sample*), was used to evaluate estimation accuracy. We used the final month of the in-sample period as the cross-validation hold-out sample. Using a fixed moving window of length six months, we generated a sequence of density forecasts, for

lead-times ranging from one-half hour up to one week ahead, by using as forecast origin each mid-night in the post-sample data.

Prior to modelling, consumption observations for all consumers were normalized to be less than or equal to one. This allows, for each method, averaging of accuracy metrics across consumers. The lengths of the intraday and intraweek cycles are represented by m_1 and m_2, respectively. Since the smart meter data is recorded half-hourly, we have m_1 = 48 and m_2 = 336. We identified seven special days, e.g. public holidays, using the Irish smart meter data. Since most SMEs do not operate on public holidays, the consumption on special days was found to be considerably lower compared to normal working days, for the same day of the week, around the same date. However, not all SMEs witnessed considerable reduction in consumption on special days. For special days lying within the estimation sample (for both residential consumers and SMEs), we smoothed out the anomalous consumption observations, by replacing them with recent historical observations belonging to a normal working day. For special days falling in the evaluation sample, namely a bank holiday Monday (2 August) in Ireland, we did not smooth out the anomalous observations. For each consumer, we first compared the actual consumption for the last special day observed in the estimation sample (namely a bank holiday Monday, 7 June), with the consumption observed on a: a) corresponding normal day of the week, from previous week, and b) recent Sunday. If the consumption observed on the bank holiday Monday in June was more similar to a normal working day than a Sunday (as quantified using the mean absolute error (MAE)), we modelled the August bank holiday Monday (in evaluation period) as a normal working day. Otherwise, we treated the August bank holiday Monday as a Sunday (as done by Smith, 2000). This approach helps model consumption by appropriately taking into account the different impact of special days on each consumer. It is noteworthy that to accommodate special

day effects, a rule-based approach (as proposed in Chapter 4) is not feasible for the existing smart meter data, as developing a rule-based framework would require much longer consumption time series compared to the one used in this study. Also, it is not clear from the current dataset, if a rule-based approach would indeed be the appropriate strategy for accommodating special day effects for consumption.

Electricity retailers will have many thousands, and quite possibly millions, of smart-metered customers. It is unlikely to be practical to estimate separately the parameters of a density estimation method for the time series of consumption recorded by each meter. Estimation of method parameters based on just a sample of the series is, therefore, required. It seems natural to surmise that method parameters will be more similar for consumers with similar consumption characteristics, than for those with very different consumption patterns. This motivates the categorization of the series, according to consumption characteristics, followed by the estimation of method parameters for a sample of the consumers from each category. In this study, we used the categorization scheme employed by the CER, which was based on principal component analysis (PCA) of customer characteristics, namely, demographic, household and socio-economic parameters (for example, income level, size of households, region, overall usage level, mode of payment etc.) (CER, 2011a). This categorization dictated the allocation of the consumers to different tariffs and stimuli. For each consumer, we only had information regarding tariff and stimulus type; however, future studies could investigate more sophisticated schemes for categorizing consumers using additional relevant characteristics. Moreover, since model parameters are estimated separately across different categories, it would also be worth investigating the impact of alternate clustering schemes on the model's forecast performance. Let us now briefly describe the tariffs and stimuli.

During the trial conducted by the CER, consumers were divided into different control and treatment groups. The bills for the control group were based on their normal tariff, whereas consumers within the treatment groups were subjected to different TOU tariffs. Specifically, for the residential consumers in the treatment group, there were four different tariff types. As specified in CER (2011a), the different tariffs correspond to different TOU prices, which refer to a day period (8am-5pm and 7pm-11pm on weekdays, and 5pm to 7pm on weekends and public holidays), peak period (5pm-7pm on weekdays, excluding bank holidays), and a night period (11pm-8am). The residential electricity costs (in cents per kilowatt-hour (kWh)), for the four tariff types and a weekend tariff, are presented in Table 5.1. The weekend tariff was introduced for some consumers, whereby night rates were applicable for all periods on weekends, with a different day, peak, and night rates being applicable for weekdays (excluding bank holidays).

TABLE 5.1
TIME-OF-USE TARIFFS (Cents per kilowatt hour (kWh)) FOR RESIDENTIAL CONSUMERS.

	Day	Peak	Night
Tariff A	14	20	12
Tariff B	13.5	26	11
Tariff C	13	32	10
Tariff D	12.5	38	9
Weekend Tariff	14	38	10

In addition to being allocated a specific tariff, consumers within the treatment group were provided with different stimuli, i.e., information regarding their electricity usage. Consumers within the control group did not receive any stimulus. There were four different stimuli for residential consumers (see CER, 2011a; 2011b), namely, a) bi-monthly billing; b) monthly

billing; c) bi-monthly billing with an in-home device that provides real-time consumption information; and, d) bi-monthly billing with overall load reduction, which offers a financial incentive if consumption is reduced by a certain target. The number of residential consumers as per their tariff and stimulus is presented in Table 5.2.

TABLE 5.2

NUMBER OF RESIDENTIAL CONSUMERS ALLOCATED AS PER DIFFERENT TARIFFS AND STIMULI.

	Bi-monthly detailed bill	Monthly detailed bill	Bi-monthly detailed bill with in-home device	Bi-monthly detailed bill with overall load reduction	Total
Tariff A	50	50	50	50	200
Tariff B	20	20	20	20	80
Tariff C	50	50	50	50	200
Tariff D	20	20	20	20	80
Weekend	-	-	-	-	20
Control	-	-	-	-	220
					800

For SMEs, CER allocated two different tariffs, which are outlined in Table 5.3. The SMEs were provided with four stimulus plans, a) monthly detailed bill; b) bi-monthly detailed bill with in-home device; c) bi-monthly detailed bill; and d) bi-monthly detailed bill with consumption and cost information also made available via a web account. Although the SMEs were allocated two different tariffs, we only had information regarding the stimulus that was allocated to the SMEs. Hence, in this study, for the purpose of model estimation, we categorize SMEs based on different stimulus. The number of SMEs as per their tariff and stimulus is presented in Table 5.4.

TABLE 5.3

TIME-OF-USE TARIFFS (Cents per kWh) FOR SMEs.

	Day 8am-5pm & 7pm-11am	Peak 5pm-7pm, Mon-Fri, excl. holidays	Night 11pm-8am
Tariff A	15	22	14
Tariff B	16	22.5	7.5

TABLE 5.4

NUMBER OF SMEs ALLOCATED AS PER DIFFERENT STIMULI.

Stimulus	Monthly detailed bill	Bi-monthly detailed bill with in-home device	Bi-monthly detailed bill with web-access	Bi-monthly detailed bill	Control	Total
	20	20	35	40	85	200

In our study, we categorize consumers according to the different tariff and stimulus types described in Tables 5.2 and 5.4. This results in a total of eighteen categories for residential consumers, and five categories for SMEs. To reduce the computational time, and allow scalability of the proposed modelling approach, we estimated method parameters for up to ten percent of the consumers within each category. The median of the optimized values for each parameter was then used to generate post-sample forecasts for all consumers in the corresponding category. The same process was then repeated for all different categories separately, in order to generate post-sample forecasts for all one thousand consumers. Fildes *et al*. (1998) adopt a similar approach for parameter optimization, and report that using commonly occurring parameter values leads to better forecasting results, compared to the case when parameters are either chosen arbitrarily, optimized once, or optimized at each forecast origin.

5.3 Modelling based on the Kernel Density Estimator

5.3.1 Unconditional and Conditional Kernel Density Estimation

Nonparametric density estimation has received considerable attention in the research community due to its widespread applications in statistical data analysis (Hastie *et al*., 2009). Kernel density (KD) estimation allows the nonparametric estimation of a density f for a set of observations $\{Y_1, Y_2, \dots, Y_n\}$. The standard Rosenblatt (1952) and Parzen (1962) kernel density estimator is represented as:

$$\hat{f}(y) = \sum_{t=1}^{n} K_{h_y}(Y_t - y) \tag{5.1}$$

where $\hat{f}(y)$ denotes the local density estimate, at point y; $K_{h_y}(\bullet) = K\left(\bullet/h_y\right)/h_y$ denotes the kernel function; and h_y is the bandwidth of the kernel function. A kernel is basically a weighting function centered at a given point, whereby the rate of weight decay is controlled by the kernel bandwidth.

We use a Gaussian kernel for all density estimation methods in this chapter, normalizing the kernel to add up to one. To generate a full density, we repeat the density estimation for a range of different values of y. Given that consumption observations for most of the consumers are more concentrated towards zero, we divide the range of y into one hundred non-uniform increments. Specifically, for each consumer, we determine the first ninety values of y by dividing the range between zero and the 90th percentile of the observations into ninety uniform increments. The remaining values of y are determined by dividing the interval between the 90th percentile and one, into ten uniform increments.

In this study, in addition to KD estimation, we also investigate the efficacy of a *conditional* kernel density (CKD) estimator for modelling electricity consumption, which involves kernelling in both the x and y-directions (Rosenblatt, 1969). The standard KD estimation involves only one kernel in the y-direction, and we refer to this as *unconditional* KD estimation. The CKD estimator produces an estimate of the conditional density $f(y|x)$ of a variable y, conditional on the variable x. The essential idea is that, for a given x, the density function at the value y is constructed by applying kernel density estimation to the set of observations $\{Y_1, Y_2, \dots, Y_n\}$, with each Y_t value weighted in accordance with the closeness of the corresponding X_t relative to the value x. As opposed to a classical linear regression based approach, which focusses on modelling the conditional expectation $E(y|x)$ in a parametric framework, CKD allows a nonlinear and nonparametric modelling of the whole density. With CKD, the conditional density of a dependent variable Y_t, denoted by $f(y|x)$, for $X_t = x$, (X_t being the independent variable) is represented as:

$$\hat{f}(y|x) = \frac{\sum_{t=1}^{n} K_{h_x}(X_t - x)\, K_{h_y}(Y_t - y)}{\sum_{t=1}^{n} K_{h_x}(X_t - x)} \tag{5.2}$$

where h_x is the additional bandwidth in the x-direction. The CKD estimator does not make strong prior assumptions regarding the form of the data distribution, and allows the flexibility for the estimated density to be, for example, multi-modal, fat-tailed or skewed.

There are two challenges with using KD estimation: a) the inappropriate selection of the kernel bandwidth can result in over/under-smoothing of the true density, and b) the presence of boundary effects in kernels may introduce bias in the final density estimate. In Section 5.3.2, we discuss the strategy adopted in this study to overcome these two challenges, and in Sections 5.3.3-5.3.9, we present the different kernel-based methods.

5.3.2 Kernel Bandwidth Selection and Boundary Correction

It is crucial to estimate the kernel bandwidth appropriately, as a large bandwidth leads to over-smoothing of the main underlying features of the time series, whereas a small bandwidth leads to under-smoothing, such that the estimated density is too rough (Jones *et al*., 1996).

There are two main approaches proposed in the literature to estimate the kernel bandwidths: a) rule-based, and b) data-based. Rule-based approaches select the bandwidths using a pre-determined rule, which is optimal under certain density type assumptions (see, for example, Sheather and Jones, 1991; Bashtannyk and Hyndman, 2001). On the other hand, data-based approaches make no strong prior density assumptions, and estimate the bandwidths using the available data (Fan and Yim, 2004). Note that compared to the rule-based approaches, data-based approaches require more computation time. It has been shown by Fan and Yim (2004) that estimating the bandwidth using a cross-validation procedure results in a better performance of the method compared to the rule-based approaches. Hence, we use a data-based approach for bandwidth selection. Specifically, we estimate the bandwidths by minimizing one-step-ahead prediction error of the density estimates, using the one month cross-validation hold-out sample. As discussed further in Section 5.4, we quantify density forecast accuracy using the conditional ranked probability score (CRPS) (see Gneiting *et al*., 2007).

A potentially serious problem with nonparametric density estimation is the presence of boundary effects (Jones, 1993). Boundary effects correspond to the inclusion of only one-sided information at the boundaries, which results in the biased estimation of the final density. For our estimation of the density of electricity consumption, estimation near the lower boundary of zero is a particular concern, because we obviously only have observations above zero, and none

below. Since we normalize consumption to be less than or equal to one, the lower and upper boundaries in our application correspond to zero and one, respectively. We denote the lower boundary by L ($L = 0$) and the upper boundary is denoted by U ($U = 1$). For all different values of y, which are smaller than the bandwidth h_y, i.e. for $L \leq y < h_y$, the employed symmetric kernel will erroneously allocate considerable non-zero weights in the negative y-direction, resulting in the under-estimation of the true density. Similarly for y lying close to the upper boundary U, such that $U - h_y < y \leq U$, the kernel will tend to over-estimate the true density.

A plethora of different approaches have been proposed for boundary correction in kernel density estimation (see, for example, Jones, 1993; Zhang and Karunamuni, 1998; Yang, 2000). In this study, we adopt the boundary correction approach proposed by Dai and Sperlich (2010). We use this approach because it has been shown to have advantages over the other well-known boundary correction methods proposed in the literature, and does not incur considerable computational costs. This approach involves the adjustment of the bandwidth for density estimation near the boundaries:

$$h_{yFinal} = \begin{cases} \max(y - L, \varepsilon) & if\ L \leq y < (L + h_y), \\ \max(U - y, \varepsilon) & if\ (U - h_y) < y \leq U, \\ h_y & otherwise \end{cases} \tag{5.3}$$

where h_{yFinal} denotes the final local bandwidth used for estimating the density centered at y. The term ε was included for numerical reasons to avoid the bandwidth becoming zero. We specify $\varepsilon = 0.001$. The rationale of this method is to reduce the bandwidths in the boundary area, in order to prevent the kernel from allocating non-zero weights outside the upper and lower boundary limits. The extent of bandwidth reduction is determined by the closeness of y to the boundary limits.

We incorporated boundary correction during both the estimation and the evaluation processes, for all kernel-based methods considered in this study.

5.3.3 Unconditional Kernel Density Estimation (KD-U)

To issue a density forecast using this benchmark method, we apply the standard KD estimation to a moving window of the most recent historical observations. This method, which we refer to as KD-U, is represented as:

$$\hat{f}(y) = \sum_{t=m}^{n} K_{h_y}(Y_t - y) \tag{5.4}$$

where $m = n - l + 1$, such that l denotes the length of the moving window, n corresponds to the forecast origin, while Y_t denotes the consumption at time t. In this study, we use a six month moving window. Note that this benchmark method makes no attempt to capture the potential seasonality in the consumption data. All of the remaining methods that we considered, and which we describe in Section 5.3, aim to accommodate the seasonality in the consumption time series.

5.3.4 Separate Kernel Density Estimation for Each Period of the Week (KD-W)

To estimate the density for a given period of the week, this method applies KD estimation to historical observations that belong to the same period of the week. We refer to the method as KD-W. Specifically, to estimate the density for a period of week, denoted by $W (W \in [1 \ldots m_2])$, we apply KD estimation to historical observations (lying within the moving window of l periods) that belong to the period W. Furthermore, this method incorporates an exponential time decay parameter, also referred to as the *discount factor*, which allows more emphasis to be placed on the more recent observations within the density estimation framework. This method is represented as:

$$\hat{f}(y) = \sum_{t \in [m,n] \mid t \bmod m_2 = W} \lambda^{\lfloor (n-t)/m_2 \rfloor} K_{h_{y_week}}(Y_t - y) \tag{5.5}$$

where λ ($0 < \lambda \leq 1$) denotes the discount factor; h_{y_week} denotes the bandwidth for observations belonging to the same period of week; mod denotes the modulus operator; and $\lfloor \bullet \rfloor$ denotes the floor operator. Note that we have m_1 = 48 and m_2 = 336. The discount factor decays exponentially in uniform steps of size m_2. This ensures that observations belonging to the same week are given the same decay weight. We note that this seasonal discounting approach differs from the more standard discounting incorporated in the CKD implementation of Jeon and Taylor (2012). A low discount factor leads to a fast decay in weights, thereby placing more emphasis on more recent observations. In addition to the bandwidth, we optimized the discount factor using cross-validation. We included the seasonal discounting in all of the methods that we describe in the remainder of Section 5.3.

5.3.5 CKD Estimation Conditional on the Period of Week (CKD-W)

This method estimates the kernel density by conditioning consumption on the period of the week. This method, which we term CKD-W, is represented as:

$$\hat{f}(y|w) = \frac{\sum_{t=m}^{n} \lambda^{\lfloor (n-t)/m_2 \rfloor} K_{h_{x_week}}(W_t - w) K_{h_y}(Y_t - y)}{\sum_{t=m}^{n} \lambda^{\lfloor (n-t)/m_2 \rfloor} K_{h_{x_week}}(W_t - w)} \tag{5.6}$$

where $K_{h_{x_week}}$ is the bandwidth defined in the x-direction for the period of week, and for any period t, W_t equals $t \bmod m_2$, while consumption is conditioned on the period of week w ($w \in [1 \ldots m_2]$). This method can be viewed as applying the CKD estimator to observations plotted against the period of week at which they occur, as in Figure 5.4, which shows this for twenty four weeks of data.

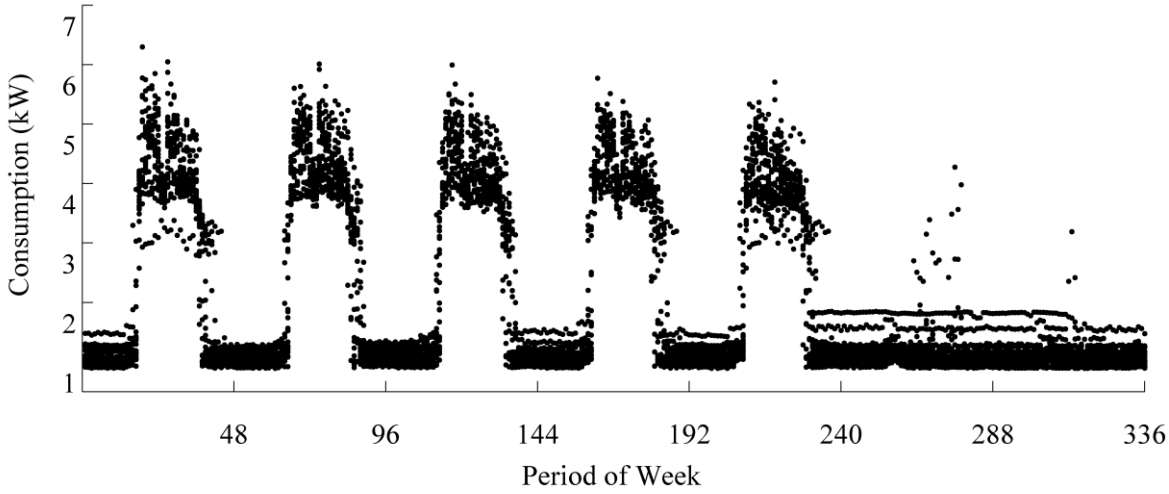

Figure 5.4— SME consumption across different periods of the week.

Note. Figure shows twenty four weeks of data. First two hundred and forty half-hourly periods correspond to weekdays, while the rest correspond to weekends. The SME considered in this figure is the same SME considered in Figure 5.2.

The rationale of this method is that, during the density estimation, consumption observations are weighted in proportion to how close their period of occurrence was to the current period of the week. Note that, in this method, periods one and m_2 are treated as being one period apart, and not $(m_2 - 1)$ periods apart. Hence, kernelling in the *x*-axis treats periods of the week *j* and *k* as being a distance apart equal to $\min(|j-k|, m_2 - |j-k|)$. This metric has been used by Hyndman *et al*. (1996) for measuring the difference between two periods within a year.

5.3.6 CKD Estimation Conditional on the Period of Week and Period of Day (CKD-WD)

Time series of intraday electricity load, at the country level, have been observed to possess both intraweek seasonal cycles and, at least for the weekdays, repeating daily cycles. Parametric time series models have been proposed that aim to accommodate both of these forms of cycles (see Taylor, 2003; De Livera *et al*., 2011). With this same aim, we extended the CKD-W

estimator of Section 5.3.5 so that it is conditional on both the period of week, and the period of the day. This method, which we term CKD-WD, is represented as:

$$\hat{f}(y|w,d) = \frac{\sum_{t=m}^{n} \lambda^{\lfloor (n-t)/m_2 \rfloor} K_{h_{x_week}}(W_t - w)\, K_{h_{x_day}}(D_t - d) K_{h_y}(Y_t - y)}{\sum_{t=m}^{n} \lambda^{\lfloor (n-t)/m_2 \rfloor} K_{h_{x_week}}(W_t - w)\, K_{h_{x_day}}(D_t - d)} \tag{5.7}$$

where h_{x_day} is the bandwidth defined for conditioning on the period of day d ($d \in [1 \ldots m_1]$). For any given period t, we specify D_t as $t \bmod m_1$. The kernelling that conditions on the period of the day treats periods of the day j and k as being a distance apart equal to $\min(|j-k|, m_1 - |j-k|)$. This method can be viewed as being the nonparametric analogue to the double seasonal (HWT) exponential smoothing (Taylor, 2003).

5.3.7 KD Estimation based on Type of Intraday Cycle (KD-IC)

This method treats a week as being composed of two different types of intraday cycle. Specifically, the method treats all five weekdays as having the same intraday cycle and it treats Saturdays and Sundays as having the same intraday cycle. The classification of different days of the week based on the similarity in their intraday cycle has been considered by Gould *et al*. (2008). Their time series model has been termed intraday cycle (*IC*) *exponential smoothing*, and it is this that prompts us to label the method of this section as KD-IC. This method is represented as:

$$\hat{f}(y) = \sum_{t \in [m,n]} \lambda^{\lfloor (n-t)/m_2 \rfloor} K_{h_{y_IC}}(Y_t - y) \tag{5.8}$$

where h_{y_IC} denotes the bandwidth in the y-direction. This method is similar to the KD-W method, the difference being that instead of performing KD estimation on just the past observations belonging to the same period of the week (as done in the KD-W method), the KD-

IC method performs KD estimation using past observations belonging to the same period on past days of the same day type.

5.3.8 CKD Estimation Conditional on Type of Intraday Cycle (CKD-IC)

This method, which we refer to as CKD-IC, can be viewed as a synthesis of the CKD-W and KD-IC methods, such that: a) observations belonging to weekdays are modelled separately from weekends, and b) two different bandwidths (in the x-direction) are used, with one for weekdays and a distinct one for weekends. This method has the appeal of allowing greater number of similar observations to be included during the density estimation.

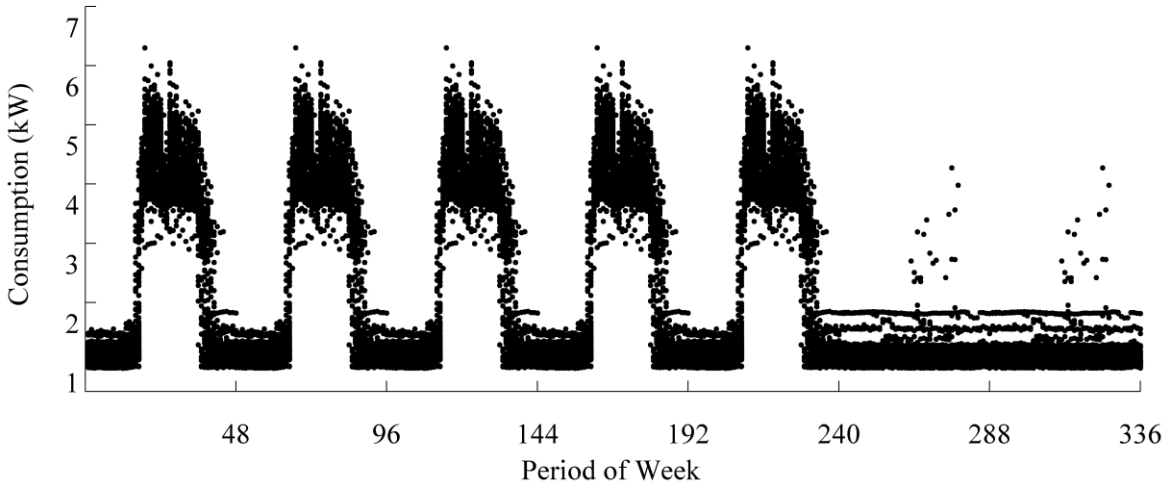

Figure 5.5— Consumption across different periods of the week for an SME. Each period includes observations belonging to that period, plus observations belonging to periods with the same intraday cycle.

Note. Figure shows twenty four weeks of data. First two hundred and forty half-hourly periods correspond to weekdays, while the rest correspond to weekends. The SME considered in this figure is the same SME considered in Figure 5.2.

In Section 5.3.5, we explained that, if based on twenty four weeks of data, the CKD-W method can be viewed as applying the CKD estimator to the plot in Figure 5.4, which shows twenty four observations plotted against each period of the week. For the CKD-IC method, the analogous

plot is presented in Figure 5.5. In this plot, each period on a weekday has the twenty four observations for the corresponding period on that particular weekday, plus the twenty four observations for the corresponding period on each of the other weekdays, resulting in 120 observations plotted for each period on each weekday. Similarly, the observations for Saturday and Sunday are pooled so that 48 observations are plotted for each weekend period. The inclusion of two different bandwidths in the x-direction allows observations to be smoothed differently for weekdays, compared to weekends.

5.3.9 CKD Estimation Conditional on Lagged Consumption (CKD-Lag)

Using this CKD method, the density is estimated by conditioning consumption for a given period, on the consumption observed during the same period from the previous week. Hyndman *et al*. (1996) proposed a similar approach using CKD, where they modelled current daily temperature, conditional on the temperature observed the previous day. This method can be seen as being the nonparametric analogue to seasonal autoregressive modelling, as both methodologies incorporate historical observations at lags corresponding to the period of the seasonal cycle during the modelling.

5.4 Empirical Analysis

This section evaluates post-sample forecast accuracy of the methods presented in Section 5.4. Section 5.4.1 evaluates density forecasts; Section 5.4.2 evaluates quantile forecasts; and Section 5.4.3 evaluates point forecasts. As mentioned in Section 5.2, we estimated parameters for up to ten percent of the consumers within each tariff/stimulus category. As described in Section 5.2, we classified the 800 residential consumers into 18 different categories, based on their corresponding tariff and stimulus, while the 200 SMEs were classified into 5 different categories.

The weighted mean of model parameters estimated across different categories are presented in Table 5.5. The weights were specified to be proportional to the number of consumers or number of SMEs in each category.

TABLE 5.5

MODEL PARAMETERS (BANDWIDTHS AND DISCOUNT FACTOR) FOR DIFFERENT METHODS AVERAGED ACROSS RESIDENTIAL CONSUMERS AND SMES.

	Parameters	***Residential consumers***	***SMEs***
KD-U	h_y	0.014	0.061
KD-W	h_y	0.012	0.038
	λ	0.942	0.926
CKD-W	h_{x_week}	0.909	0.488
	h_y	0.014	0.044
	λ	0.944	0.917
CKD-WD	h_{x_day}	0.651	0.354
	h_{x_week}	0.553	0.354
	h_y	0.013	0.045
	λ	0.994	0.925
KD-IC	h_y	0.014	0.039
	λ	0.998	0.917
CKD-IC	$h_{x_weekday}$	0.704	0.354
	$h_{x_weekend}$	0.825	1.042
	h_y	0.015	0.045
	λ	0.977	0.938
CKD-Lag	h_{x_lag}	0.017	0.045
	h_y	0.017	0.045
	λ	0.958	0.929

The notations for bandwidths used in Table 5.5 are exactly the same as specified earlier in Section 5.3, whereby h_{x_week} denotes the bandwidth in the x-direction for the period of week,

h_{x_day} corresponds to the bandwidth for the period of day, $h_{x_weekday}$ and $h_{x_weekend}$ denote the bandwidths for weekdays and weekends (used in IC-CKD), respectively, while h_{x_lag} is the bandwidth for lagged load. As shown in Table 5.5, the bandwidths in the y-direction are quite small, for different methods considered in this study, which is expected as a substantial proportion of observations are concentrated around zero. Note that the model parameters were optimized based on in-sample density forecast performance, as quantified using the CRPS, for all methods based on KD and CKD estimation used in this study.

We plot the discounting weights, for different values of the discount factor (λ) in Figure 5.6. As specified in Section 5.3, λ decays exponentially in uniform steps of size m_2. This figure shows that a low discount factor leads to a fast decay in weights, thereby placing more emphasis on more recent observations.

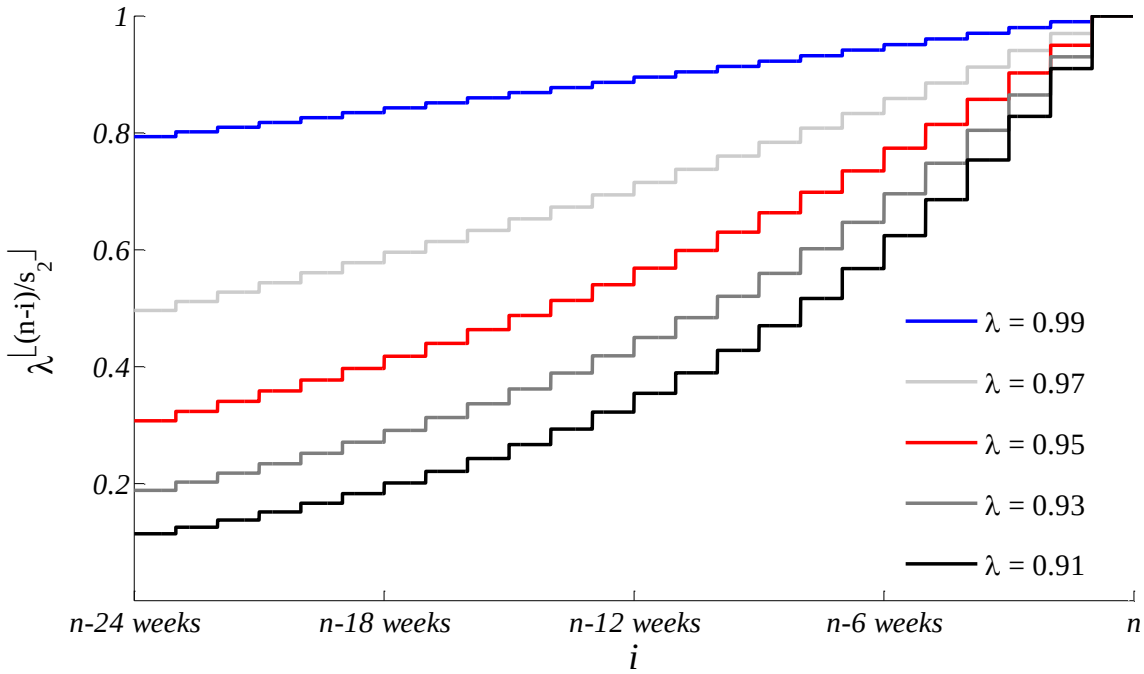

Figure 5.6— Discounting weights for different values of the discount factor λ.

5.4.1 Evaluating Density Forecasts

This section presents post-sample density forecast of different methods based on KD and CKD estimation considered in this study. In Figure 5.7, we plot the post-sample density forecasts for up to a week ahead for a residential consumer and an SME. The forecast origin was the final half-hour on a Sunday mid-night (15 August) belonging to the post-sample period. Hence, horizons for up to five days ahead correspond to weekdays, while the remaining horizons correspond to weekends.

For the residential consumer, we used the KD-IC method to generate forecasts (Figure 5.7a), as it was one of the most accurate method across residential consumers (as explained later in this section). Similarly, we used the CKD-W method to generate density forecasts for an SME (Figure 5.7b). Note that the point forecasts were generated by taking the median of the density forecast.

For the SME, it is encouraging to see in Figure 5.7b that the CKD-W method accommodates the seasonality in consumption, and correctly distinguishes weekdays from weekends.

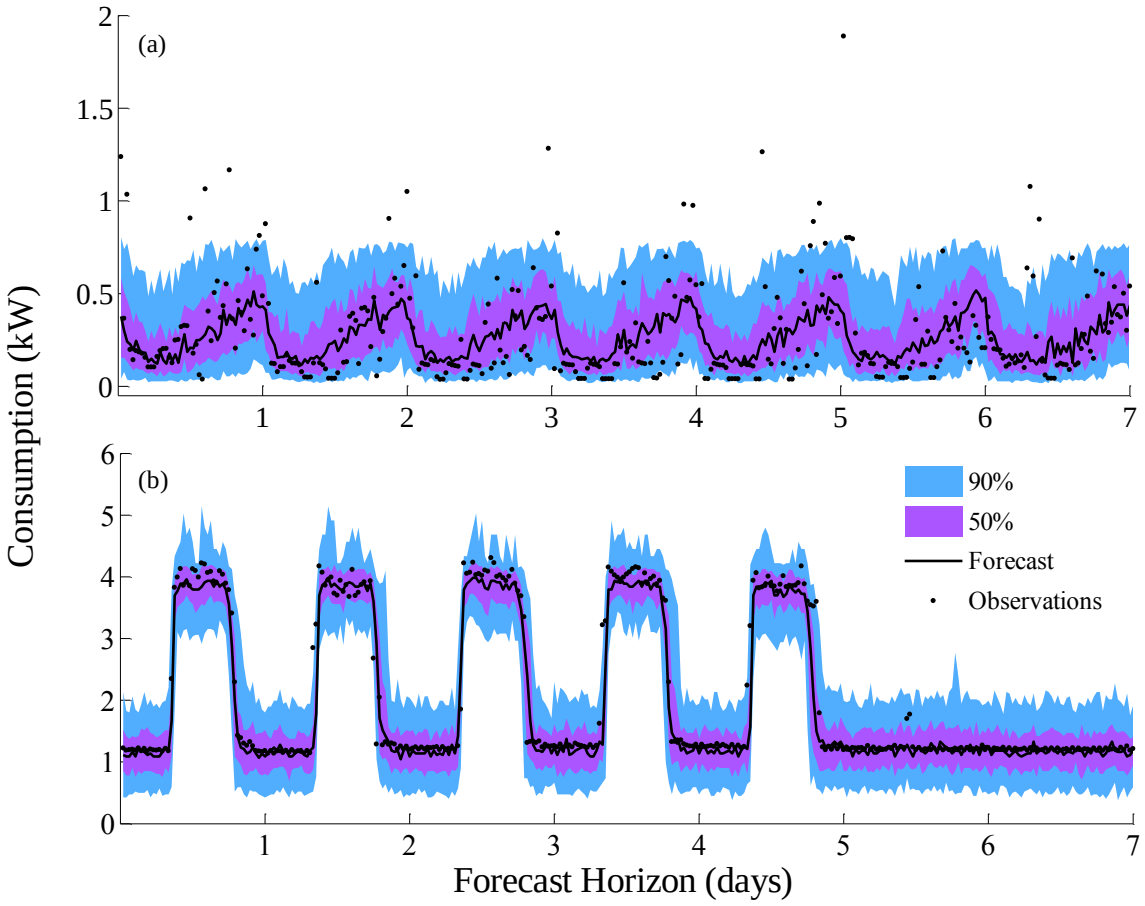

Figure 5.7— Density forecasts and corresponding actual observations for: a) a residential consumer, and b) an SME.
Note. Point forecasts correspond to the median of the predicted density. Horizon corresponding to the first five days refers to weekdays, while the rest correspond to weekends.

For density forecast evaluation, we use the CRPS, as it quantifies both *calibration* and *sharpness*, (see Gneiting *et al*., 2007), as described in Section 2.5. The CRPS can be defined in terms of the predicted cumulative distribution function F as follows:

$$\mathrm{CRPS}(F, y) = \int_{-\infty}^{+\infty} (F(z) - \mathbf{1}\,(z \geq y))^2 dz \tag{5.9}$$

where y denotes the actual observation, while $\mathbf{1}$ denotes the indicator function. In this study, we used CRPS for both model estimation and evaluation.

Figure 5.8 presents the post-sample CRPS results averaged across all 800 residential consumers, for lead times at half-hourly intervals, from one half-hour up to one week ahead. The forecast performance was very similar for CKD-W, CKD-WD, KD-IC and CKD-IC. For simplicity, we present results only for the KD-IC method, as it was one of the most accurate methods, and is also one of the simplest methods (in terms of the total number of model parameters it employs). It is interesting to see that the KD-U method is outperformed by the other methods, at all lead times. As expected, the CRPS is high during periods of the week that witness a relatively large change in consumption. Among the CKD methods, the performance of CKD-Lag is the worst. The performance of the KD-W was competitive with the best performing CKD methods.

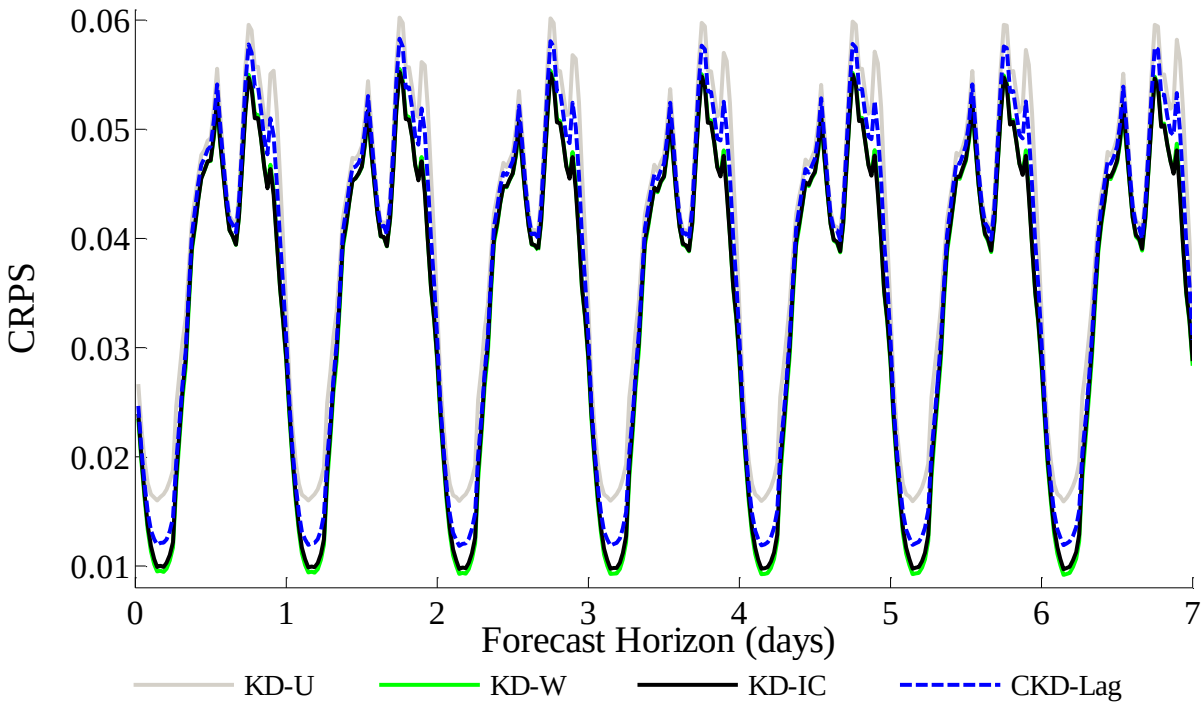

Figure 5.8— Mean CRPS for density forecasting of the 800 residential consumers. Note that lower CRPS values are better.

In Figure 5.9, we present the post-sample CRPS results averaged across the 200 SMEs. The performance of the KD-IC and CKD-IC methods were very similar, however, both methods were outperformed by the CKD-W and KD-W methods. Hence, for conciseness, we do not present results for the KD-IC and CKD-IC methods. Furthermore, we do not present the CKD-WD method, as it was slightly less accurate than the CKD-W method. All methods considerably outperformed the KD-U method. The two best performing methods, overall, were KD-W and CKD-W, both of which are presented in Figure 5.9. Interestingly, for the SMEs, the KD-U benchmark was substantially outperformed. This was not the case for residential consumers. A possible explanation for this result is that the residential consumption exhibits much greater volatility, making it relatively hard to model, whereas the consumption for SMEs typically displays prominent seasonality. Note that the KD-U benchmark method does not account for seasonality.

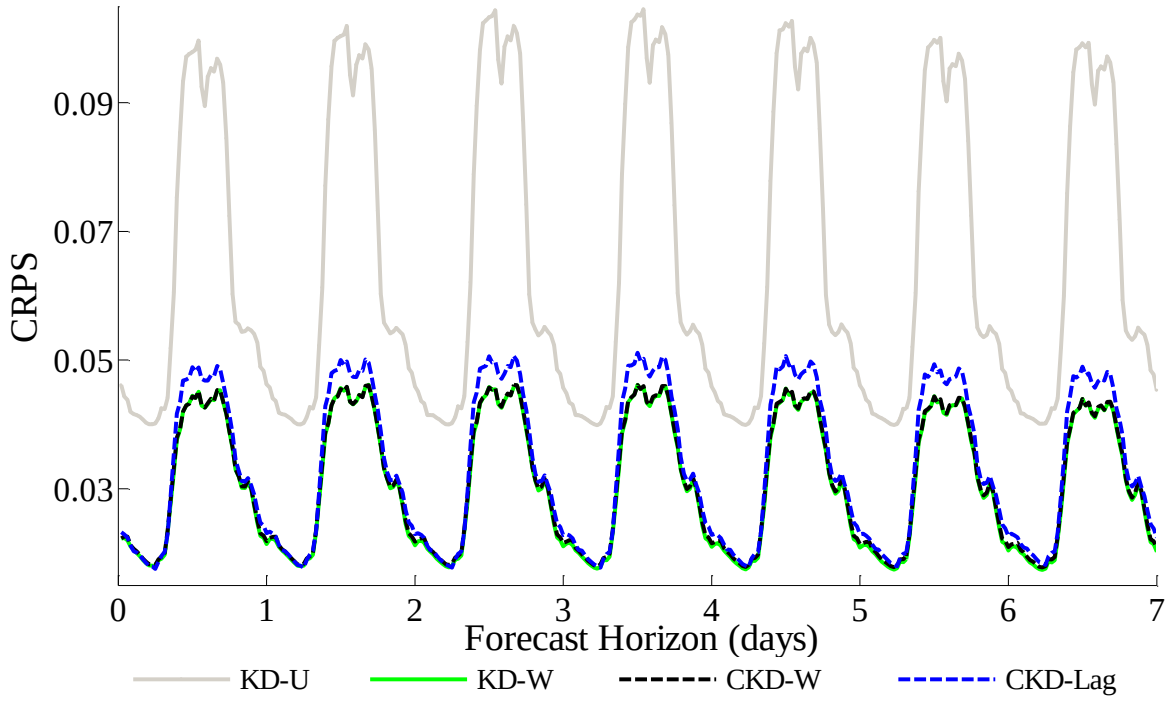

Figure 5.9— Mean CRPS for density forecasting of the 200 SMEs. Note that lower CRPS values are better.

In Figure 5.10, we plot the post-sample density forecasts for up to a week ahead for an SME. We consider the same SME as used in Figure 5.7b. This SME witnessed a considerable reduction in consumption on special days. However, instead of accommodating special day effects during the modelling, as done for this SME in Figure 5.7b, we completely ignore the special day effects while generating forecasts for this SME in Figure 5.10. This allows us to gauge the impact of ignoring special day effects on the forecast accuracy of a model. For Figure 5.10, the forecast origin was the final half-hour on a Sunday mid-night, belonging to the post-sample period. Specifically, the forecast origin for Figure 5.10 was chosen to be 1 August.

In Figure 5.10, forecasts for up to one day ahead are highly inaccurate. Note that one day-ahead forecasts correspond to forecasts for different periods on a Monday. A careful inspection of the data revealed that this was due to the occurrence of a public holiday in the estimation period, namely the bank holiday on Monday (2 August) in Ireland. In Figure 5.10, the CKD-W method wrongly treats a bank holiday Monday as a normal working Monday. Furthermore, we find that the confidence intervals are quite large for Mondays that follow the bank holiday Monday.

Some studies have focused on accommodating the unusual electricity demand patterns due to holiday effects during the modelling process (see, for example, Kim *et al.*, 2000). Using a longer time series than the one considered in this study, future work could investigate accommodating such holiday effects in the modelling of smart meter data, by treating each special day as having a unique profile, as done previously in Chapter 4 for anomalous load.

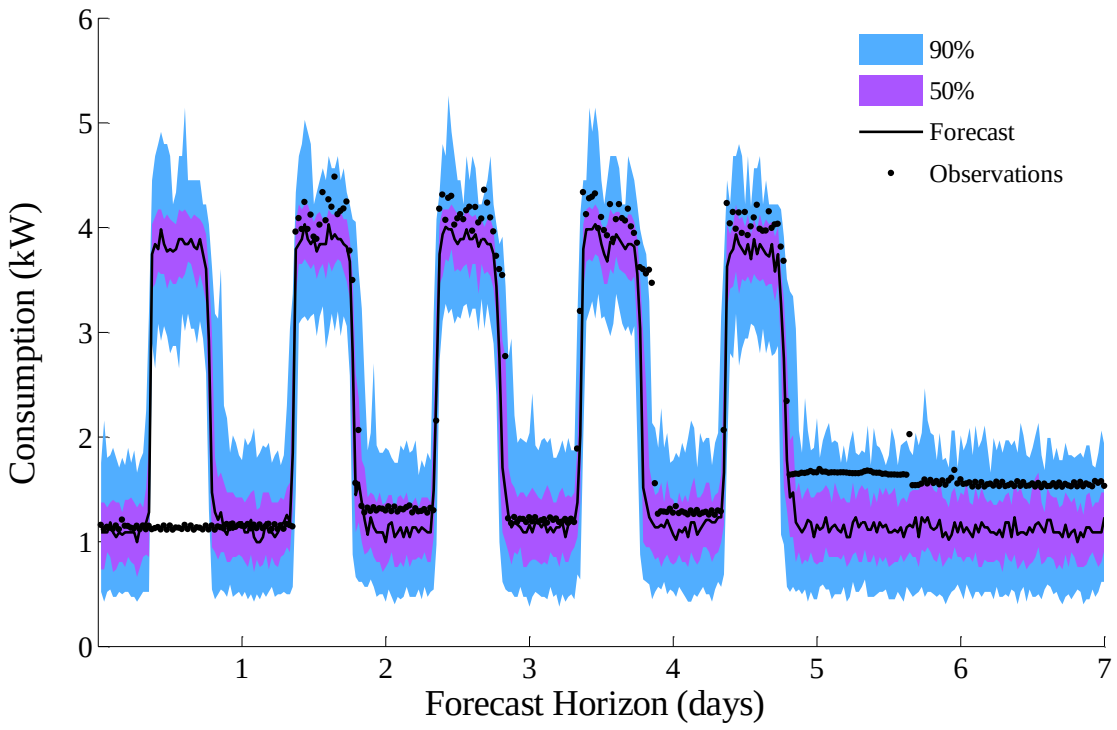

Figure 5.10— Density forecasts and corresponding actual observations for an SME, for forecast origin belonging to a Sunday mid-night on 1 August.

Note. Point forecasts correspond to the median of the predicted density. Horizon corresponding to the first five days refers to weekdays, while the rest correspond to weekends. Monday 2 August in Ireland was a bank holiday. The SME considered in this figure is the same SME considered in Figure 5.7b.

The methods in this chapter have been implemented using a fixed moving window of length six months. To investigate the effect of window length on the post-sample forecast performance, we consider three implementations of the KD-IC method, with moving windows of length: a) two months, b) four months, and c) six months. We chose to focus here on the KD-IC method as it was competitive with other CKD-based methods, and involves estimation of only a few parameters. For a given window length, the model parameters were re-estimated across different consumers, using the same procedure described in Section 5.2. Note that the hold-out cross-validation sample and estimation sample were common for all three implementations of the KD-

IC method, in order to allow for a fair comparison of post-sample forecast accuracy for different window lengths. Figure 5.11 presents the CRPS for the three different versions of the KD-IC method. We present the CRPS averaged across all SMEs. As expected, the CRPS improves when increasing the window length, as more historical information gets included in the model. Hence, we use a six month moving window in this study. In a practical setting, the decision regarding the length of moving window will depend on the availability of computational resources, amount of data available, and acceptable levels of accuracy for the task at hand.

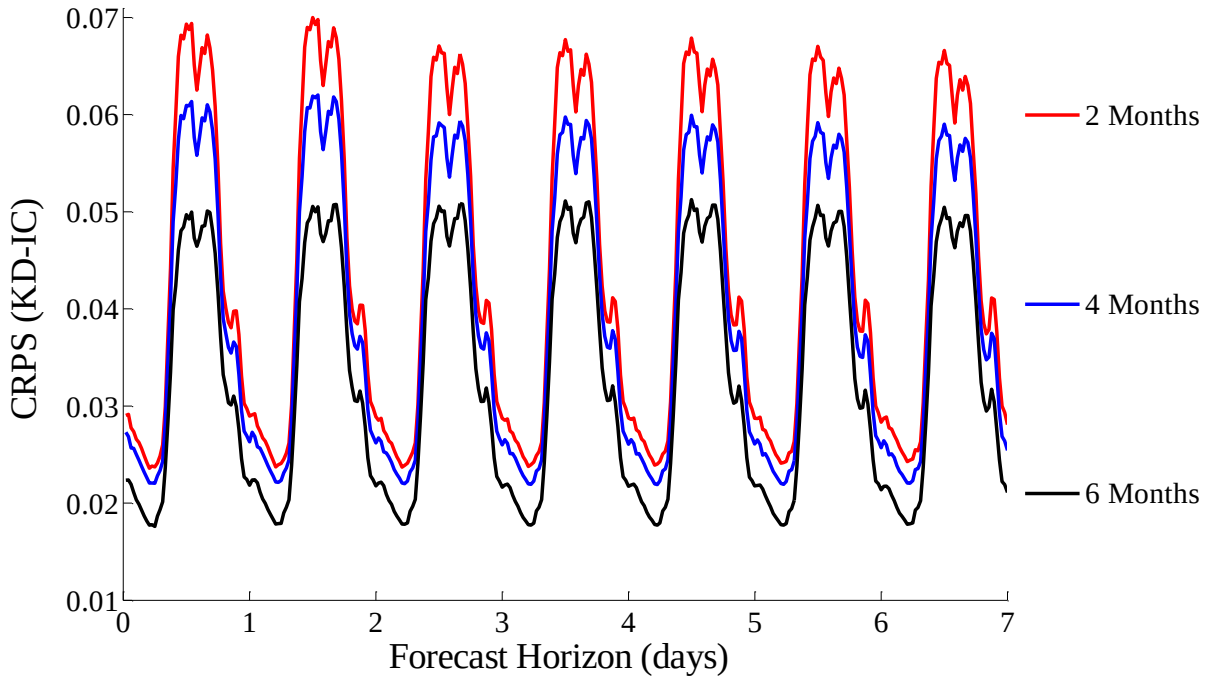

Figure 5.11— Mean CRPS for density forecasts of 200 SMEs, generated using the KD-IC Method for moving window of lengths: a) two months, b) four months, and c) six months. Note that lower CRPS values are better.

5.4.2 Evaluating Quantile Forecasts

We evaluate quantile forecasts using the unconditional coverage, which measures the percentage of observations that are lower than the θ quantile forecast. Ideally, this percentage

should be θ. The difference between the expected and obtained coverage is known as bias. Figure 5.12 presents the unconditional coverage averaged across the 800 residential consumers. We evaluate forecasts for the 5%, 15%, 25%, 35%, 45%, 50%, 55%, 65%, 75%, 85%, and 95% quantiles.

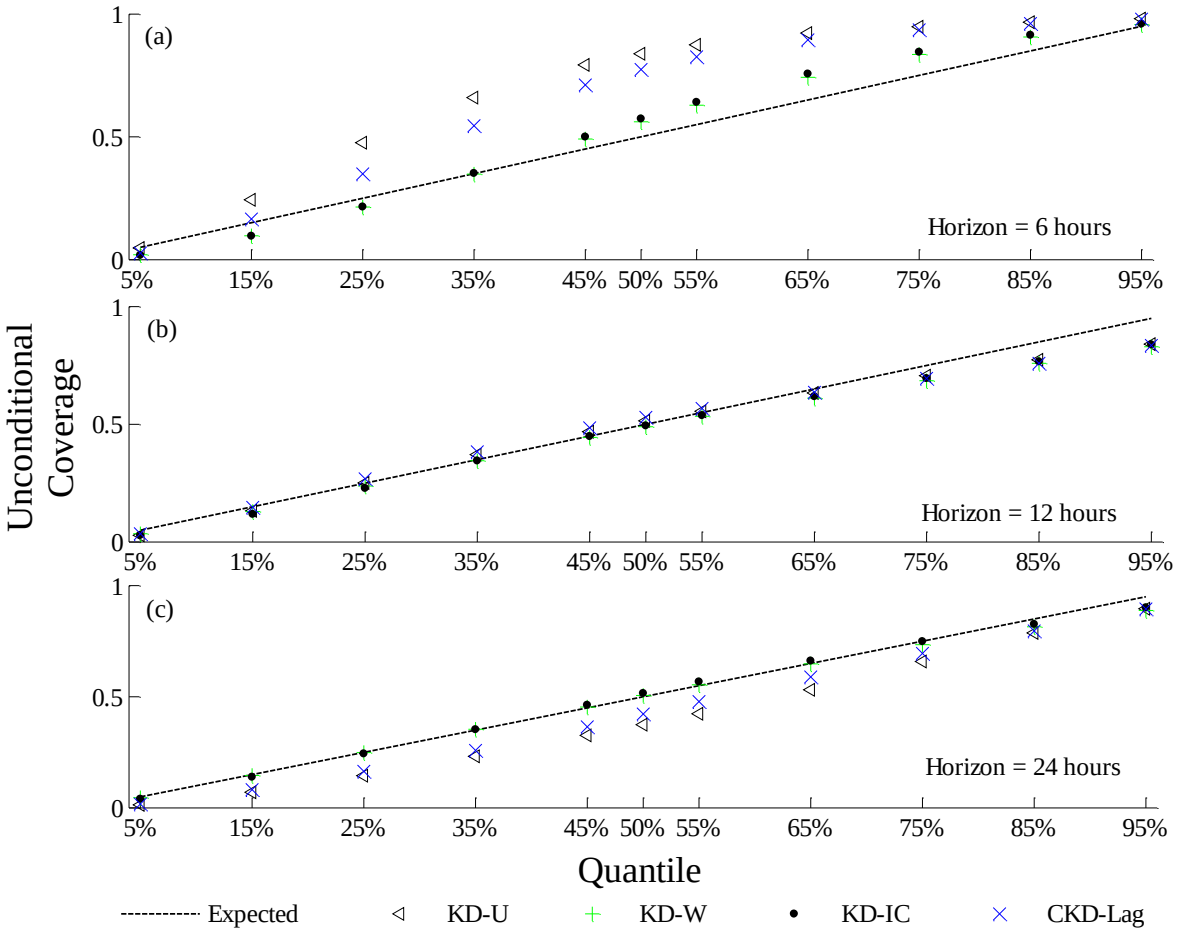

Figure 5.12— Mean unconditional coverage for quantile forecasting across the 800 residential consumers for horizons corresponding to: a) 6 hours, b) 12 hours, and c) 24 hours.
Note. Values closer to the diagonal (dashed line) are better.

Figure 5.12a shows unconditional coverage for six hour ahead forecasts. Compared to the KD-U method, the unconditional coverage obtained using KD-IC and KD-W is much closer to perfect coverage. Among the CKD based methods, we identified CKD-Lag as the worst performing method. Note that in both Figures 5.12a and 5.12b, values closer to the diagonal line

are better. As shown in Figure 5.12b, the unconditional coverage for different methods is rather similar, for twelve hour ahead forecasts. For one-day ahead forecasts, the unconditional coverage obtained using KC-IC and KD-W is very close to the perfect coverage, for all quantiles considered in this study, as shown in Figure 5.12c.

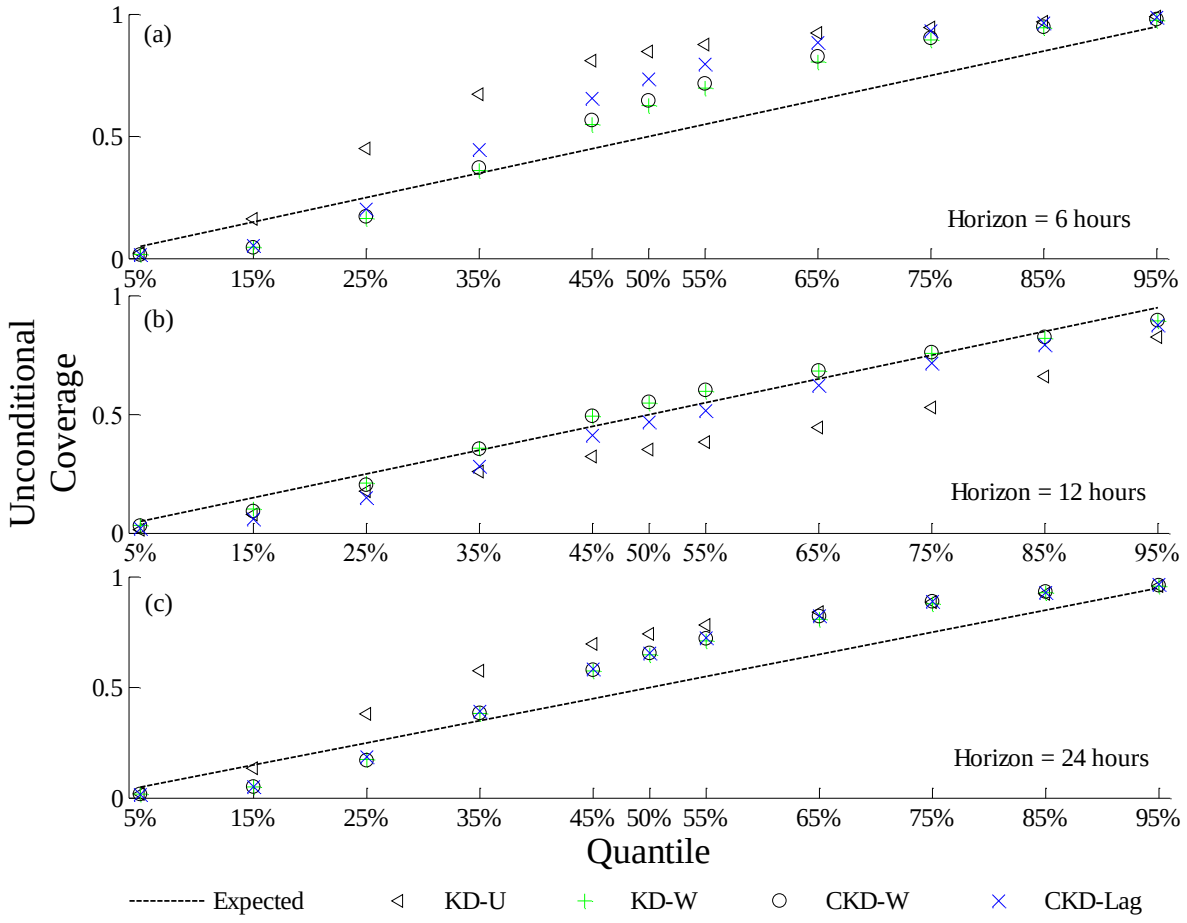

Figure 5.13— Mean unconditional coverage for quantile forecasting across the 200 SMEs for horizons corresponding to: a) 6 hours, b) 12 hours, and c) 24 hours.
Note. Values closer to the diagonal (dashed line) are better.

Figure 5.13 presents the unconditional coverage averaged across the 200 SMEs. The CKD-W method is one of the best performing methods, while the KD-U and CKD-Lag are the worst performing methods.

Figure 5.14a plots the mean hit percentage for the 800 residential consumers, for six-hour ahead forecast. It is noteworthy that the hit percentage measures the unconditional coverage, while an exhaustive forecast evaluation process would require a measure of the conditional coverage, see (Christoffersen, 1998). Both KD-IC and KD-W are more accurate than the KD-U and CKD-Lag methods. As shown in Figures 5.14b and 5.14c, both KD-IC and KD-W perform relatively well compared to the other methods, for both twelve hour and one-day ahead forecasts, respectively.

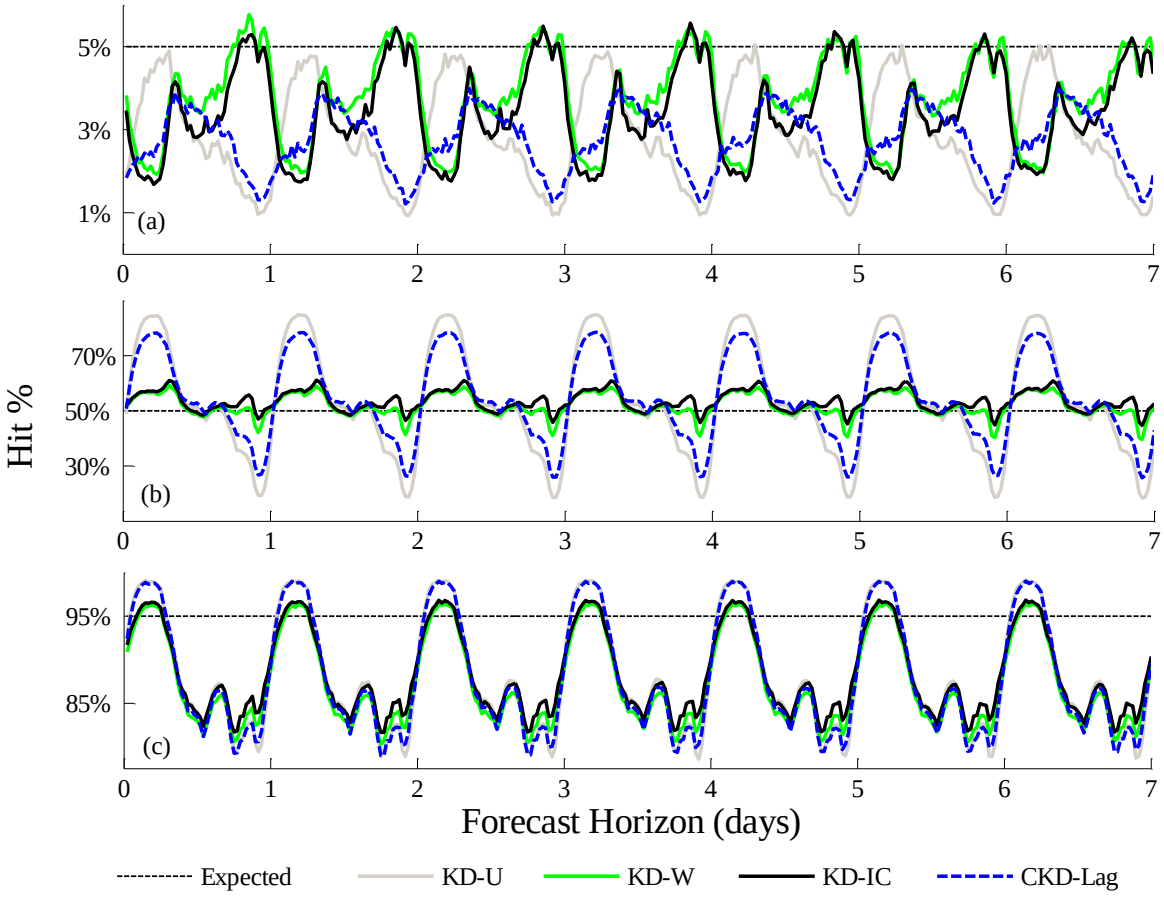

Figure 5.14— Mean Hit % for across the 800 residential consumers for forecasting of the: a) 5% quantile, b) 50% quantile (median), and c) 95% quantile.
Note. Values closer to the horizontal black dashed line are better.

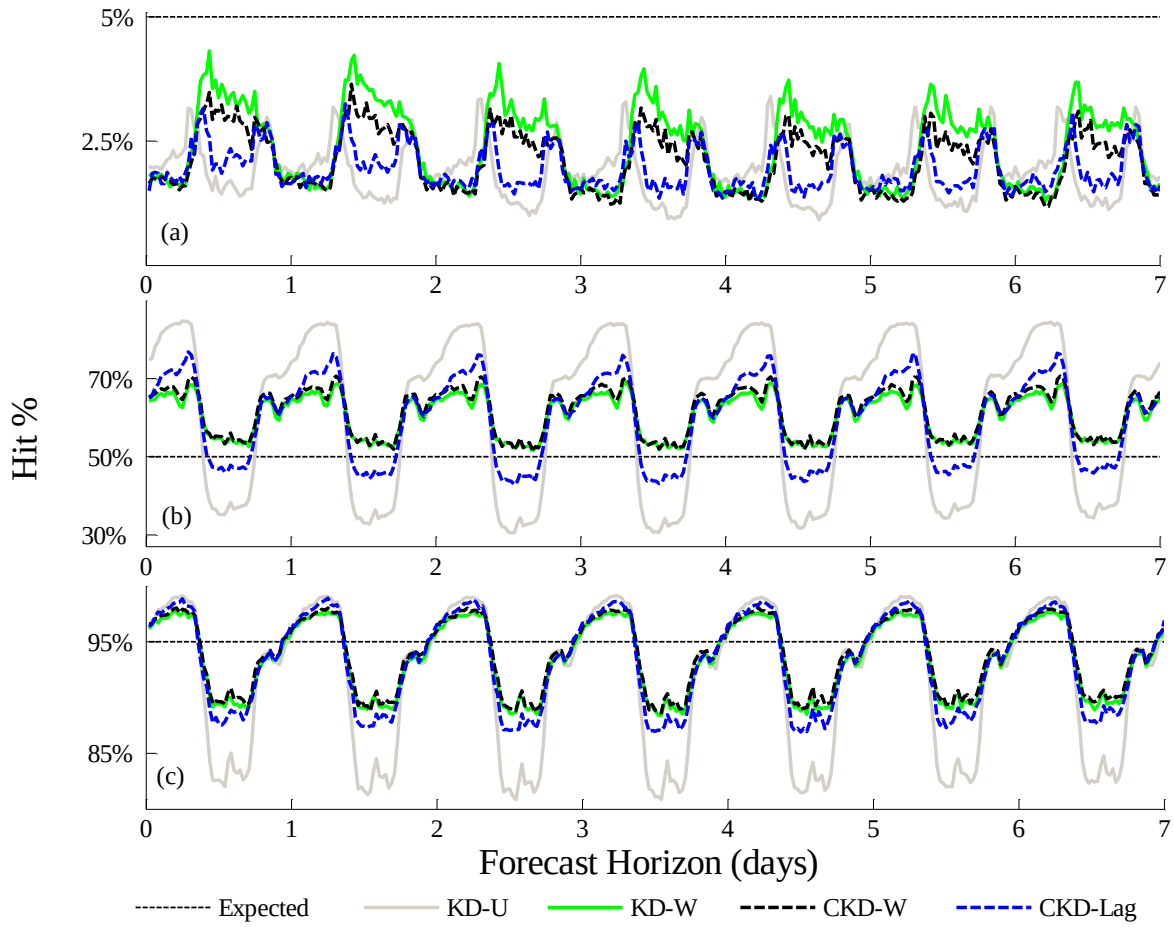

Figure 5.15— Mean hit % for across the 200 SMEs for forecasting of the: a) 5% quantile, b) 50% quantile (median), and c) 95% quantile.
Note. Values closer to the horizontal black dashed line are better.

Figure 5.15 presents the hit percentage of the 200 SMEs. For the SMEs, CKD-W is one of the most accurate methods overall, while both KD-U and CKD-Lag are the worst performing methods. These rankings obtained using the unconditional coverage are reasonably consistent with the rankings obtained earlier using the CRPS.

5.4.3 Evaluating Point Forecasts

To evaluate point forecasts, we used the mean absolute error (MAE) and the root mean square error (RMSE) as performance scores. It has been pointed out by Gneiting (2011c), that the median of the density forecast is the optimal forecast if the loss function is symmetric piecewise linear, whereas for a quadratic loss function, the mean is the optimal forecast. Hence, we issue the median of the density forecast as a point forecast, and evaluate it using the MAE. Similarly, we also issue the mean of the density forecast as a point forecast, and use RMSE for evaluation. In Figure 5.16, we present the post-sample MAE results averaged across the 800 residential consumers, while Figure 5.17 shows the MAE results averaged across the 200 SMEs, for the best performing methods. The MAE rankings of the different models across residential consumers and SMEs are very similar to the rankings obtained using the CRPS.

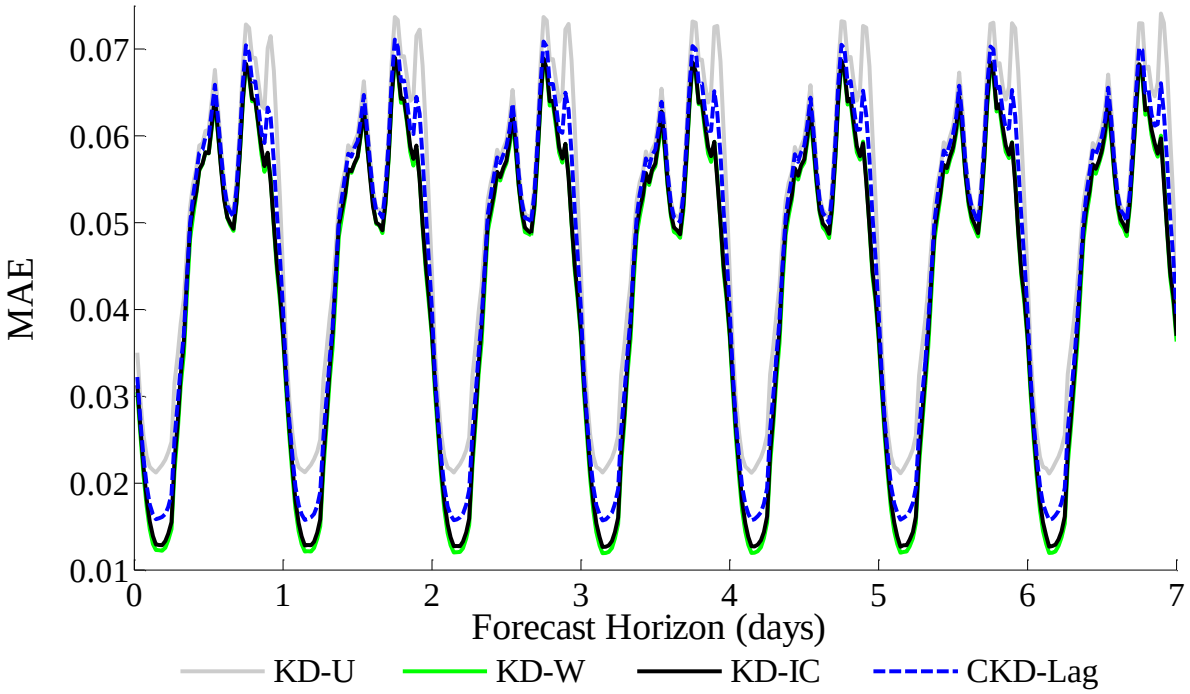

Figure 5.16— Mean MAE for point forecasting of the 800 residential consumers. Note that lower MAE values are better.

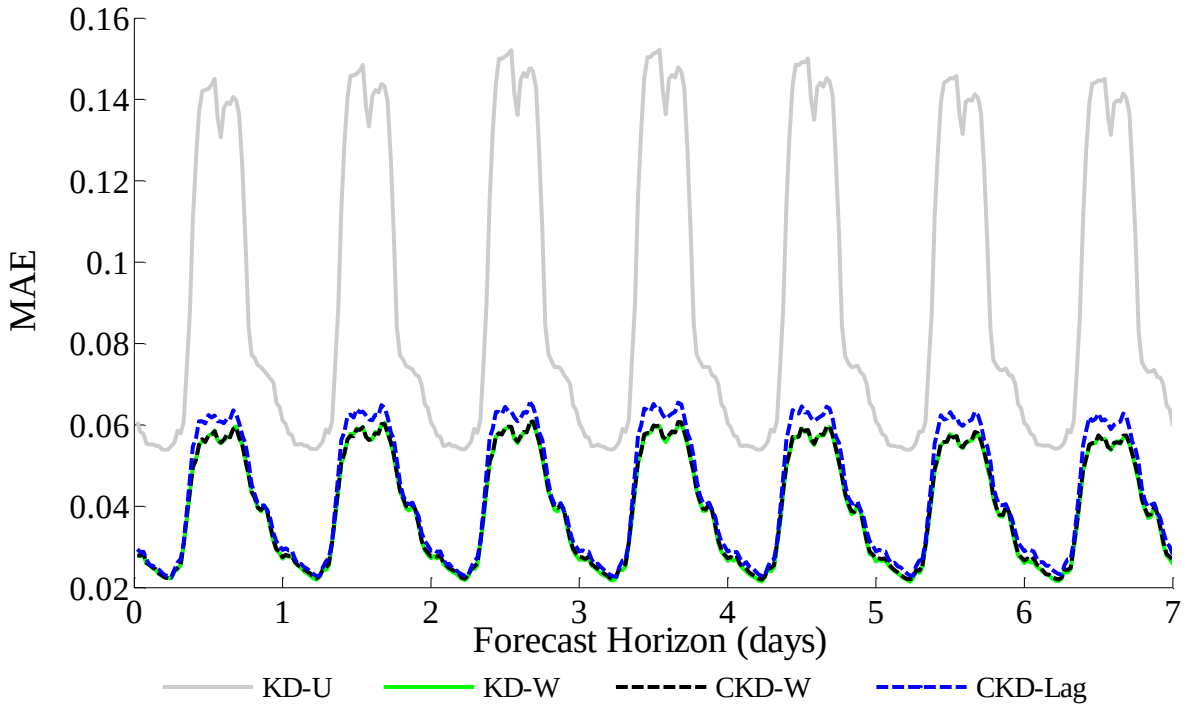

Figure 5.17— Mean MAE for point forecasting of the 200 SMEs. Note that lower MAE values are better.

5.5 Deriving Prediction Intervals for Electricity Cost

We derive density estimates of electricity cost for different TOU tariffs, in order to have several density forecasts for cost. This would allow the consumers to choose between the different cost densities, e.g. a density with low mean and high variance versus a density with a higher mean but lower variance. Moreover, energy retailers could devise innovative and effective pricing strategies using these estimates. Furthermore, such estimates can help fuel-poor household select appropriate payment schemes. As defined by DSDNI (2012), "A household is in fuel-poverty if, in order to maintain an acceptable level of temperature throughout the home, the occupants would have to spend more than 10% of their income on all household fuel use."

According to a recent report by DECC (2012), 43.7% households in Northern Ireland are fuel-poor (2009 estimate), while there are approximately 4.5 million fuel-poor households in the

UK (2010 estimate), out of which 4 million are also classified as being vulnerable, i.e. a household with elderly people, children or someone with a long-term illness. It has been reported, based on a survey, that 90% of the households in the lowest income group use prepayment as a mode for paying bills, as opposed to using credit payment, see Waddams (2002). It has been pointed out by Darby (2012) that although unit energy prices are higher under prepayment schemes, customers opt for this mode of payment because it gives them greater control over their consumption. It is thus crucial, especially for prepayment consumers, to be able to gauge the potential energy costs that would be incurred in the future. In this regard, future cost estimates can potentially help the consumers: a) choose between credit and prepayment schemes, b) identify and select the tariff that would incur the lowest cost, and, c) adjust the energy usage to ensure that the cost does not exceed stipulated budget.

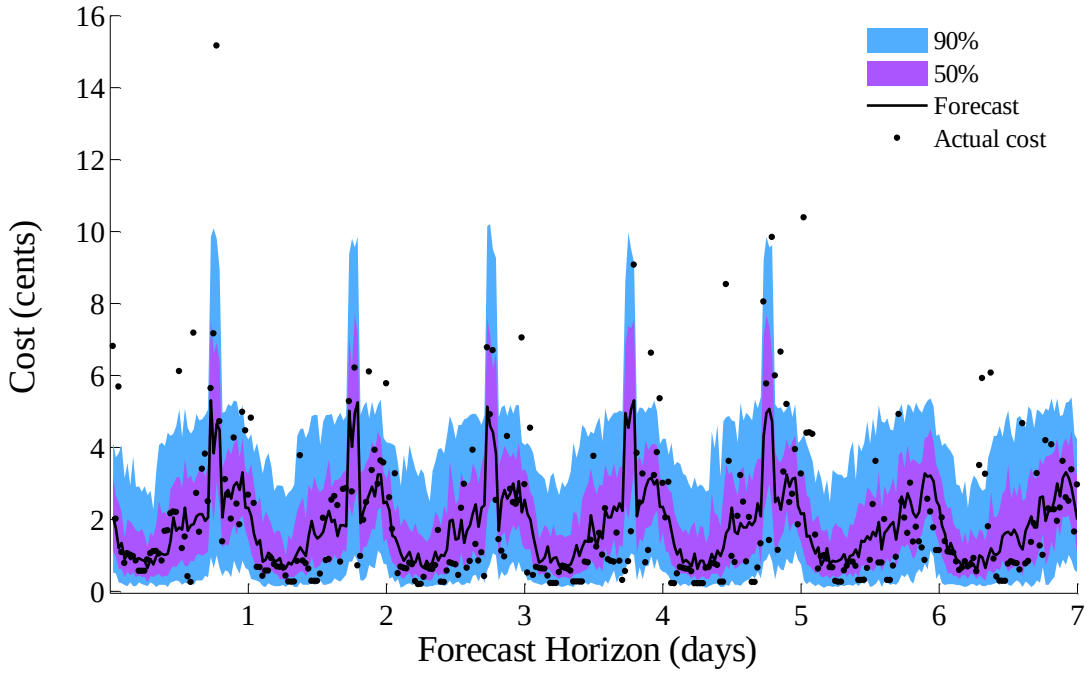

Figure 5.18— Prediction intervals and corresponding actual cost for a residential consumer.
Note. Point forecasts correspond to the median of the predicted density. Horizon corresponding to the first five days refers to weekdays, while the rest correspond to weekends. Cost intervals were derived using Tariff B.

We derive prediction intervals for electricity cost based on density forecasts of electricity consumption. In Figure 5.18, we plot the post-sample prediction intervals for cost, for horizons up to a week ahead, for the same residential consumer considered in Figure 5.7a. The forecast origin in Figure 5.18 was chosen as the final half-hour on a Sunday mid-night (15 August) belonging to the post-sample period, which is the same as the forecast origin used in Figure 5.7a. Hence, horizons for up to five days ahead correspond to weekdays, while the remaining horizons correspond to weekends. For a given period, we first sample realizations from the consumption density. We then multiply the sampled consumption values (in kW), with the electricity cost (cents per kWh) associated for the period under consideration (for a given TOU tariff), to obtain prediction intervals for cost. In Figure 5.18, we compute prediction intervals for cost for a residential consumer using Tariff B (as described in Table 5.1), as this consumer was allocated Tariff B by the CER. The cost associated with a given tariff is highest during peak-times (5pm-7pm) on weekdays (Table 5.1). It is due to this reason, that there are five distinctive peaks observed during peak-times on weekdays, as shown in Figure 5.18.

Out of the 800 residential consumers considered in this study, 220 consumers belonged to the control group (as specified in Table 5.2). The consumers in the control group were not allocated TOU tariffs by the CER, and did not receive any stimulus for the whole duration of the smart meter trial (CER, 2011a). Instead, the control group was billed on their normal electricity supplier tariff. We do not have information regarding the tariff applicable for each consumer belonging to the control group. Hence, we do not use smart meter data for the control group consumers. Specifically, for this particular study, we employ data from 580 (800-220) residential consumers that were allocated TOU tariffs, and were also provided with stimuli by the CER.

In addition to the tariffs described in Table 5.1, we investigate a non-TOU tariff, denoted by Tariff E, which incurs the same cost of 13.25 cents per kWh, for each period of the day. We devised Tariff E by computing the weighted mean of the cost associated with different periods of the day, using Tariff D, whereby the weights were proportional to the number of half-hours occurring during the day, peak and night period (as described in Section 5.2). Hence, for a given density forecast of consumption, we derive six density estimates of cost, using the five TOU tariffs described in Table 5.1, and an additional Tariff E.

We derive consumption densities for horizons up to a week ahead, by rolling the forecast origin through each Sunday mid-night in the post-sample data. Using consumption densities for the coming week, we derive cost densities for different tariffs, and select a tariff that incurs the lowest total cost for the whole week. Thus, each Sunday midnight, we select a suitable tariff for the coming week, for each consumer. Note that there are four Sunday mid-nights in the post-sample data. Since we use data from 580 residential consumers, we have a total of 2320 (4×580) instances where, in each instance, we generate and compare a cost density for each of the six tariffs, and on this basis select the optimal tariff to be used for the next week.

In order to compare cost densities for different tariffs, we select a tariff whose associated cost density has a lower: (1) Expectation (mean), (2) Median (50%), (3) 25% Quantile, (4) 75% Quantile, (5) 5% Quantile and (6) 75% Quantile. Furthermore, we also include stochastic dominance as a tariff selection criterion. Specifically, the rationale of this criterion is to compare the different CDFs of cost, denoted by, say, F_A and F_B, corresponding to tariffs A and B, respectively, and select Tariff A, if $\int_{-\infty}^{x}(F_A(t) - F_B(t))dt \geq 0$, else, select Tariff B.

Table 5.6 presents the total number of times different tariffs were selected using the seven different selection criteria. Interestingly, only two of the six tariffs, namely Tariff D and Tariff E, were selected, with Tariff E selected more often than Tariff D. This result highlights that Tariff A, B, C and the weekend tariff were not competitive with other tariffs. Based on future estimates of cost, for each half-hour over the coming week, we calculate the density for the total cost for one week.

TABLE 5.6

NUMBER OF TIMES DIFFERENT TARIFFS WERE SELECTED (USING TARIFFS FROM TABLE 5.1, AND AN ADDITIONAL TARIFF E), FOR THE 580 RESIDENTIAL CONSUMERS, OVER THE FOUR WEEKS OF THE POST-SAMPLE, USING SEVEN DIFFERENT CRITERIA.

Tariff/Criterion	Expectation (Mean)	Median (50%)	25%	75%	5%	95%	Stochastic Dominance
Tariff D	1089	1102	1209	1002	1369	830	1089
Tariff E	1231	1218	1111	1318	951	1490	1231

We investigate if it is more cost-effective for a consumer to: a) switch between different available tariffs, or, b) use only one tariff at all times. To compute the cost associated with tariff switching, we first derive prediction intervals for cost using density estimates of consumption, for different available tariffs. We then identify the tariff that incurs the lowest cost over the coming week, by comparing cost densities for different tariffs. For a given consumer, having identified the optimum tariff (based on a particular selection criterion), we compute the cost associated with that tariff, for the whole week under consideration, using the actual consumption observations. This procedure allows us to get an estimate of the actual cost for the week (for an optimum tariff), for all four weeks in the post-sample period. We sum the total weekly cost for

all four weeks, to compute the total cost incurred over the whole post-sample period, for each consumer. Moreover, we also compute the actual cost associated with using a single tariff. Specifically, for a given consumer, instead of switching tariffs based on the comparison of cost densities (as done above), we simply use the tariff allocated to each consumer by the CER, and compute the actual total cost associated with the allocated tariff for the whole post-sample period. Note that the CER allocated only one TOU tariff to each consumer during the whole duration of the smart meter trial.

Figure 5.19 plots the difference between the total cost associated with switching tariffs, and the total cost associated with using a single tariff (as allocated by the CER), for the whole post-sample period. We plot the cost difference across the 580 residential consumers, for all seven tariff selection criteria. Hence, in Figure 5.19, negative values correspond to the instances when switching tariffs resulted in a lower total cost compared to using a single tariff. It can be seen from Figure 5.19 that switching tariffs results in a much lower cost overall, compared to the case when consumers were allocated a single TOU tariff for all periods.

The number of residential consumers as per their tariff selection schemes is presented in Table 5.7. Specifically, Table 5.7 presents the number of consumers who were able to reduce their total electricity cost by switching tariffs, and the number of consumers who incurred lower costs by using a single allocated tariff. It is evident from Table 5.7 that switching tariffs results in lower electricity costs for a majority of consumers, while using a single tariff was preferable for only less than ten percent of the total residential consumers, across all seven different criteria considered in this study. Compared to the other selection criteria, choosing a tariff based on the 95% criterion was preferable for the highest proportion of consumers (530 out of 580), to reduce the total electricity cost using tariff switching. Interestingly, switching tariffs based on the 95%

criterion also resulted in the highest average savings per consumer (as presented in Table 5.7), indicating that cost densities for most consumers considered in this study exhibit a heavy right tail. These results highlight the scope of potential savings that can be achieved by switching between different tariffs, compared to the case when customers are allocated one TOU tariff for all periods.

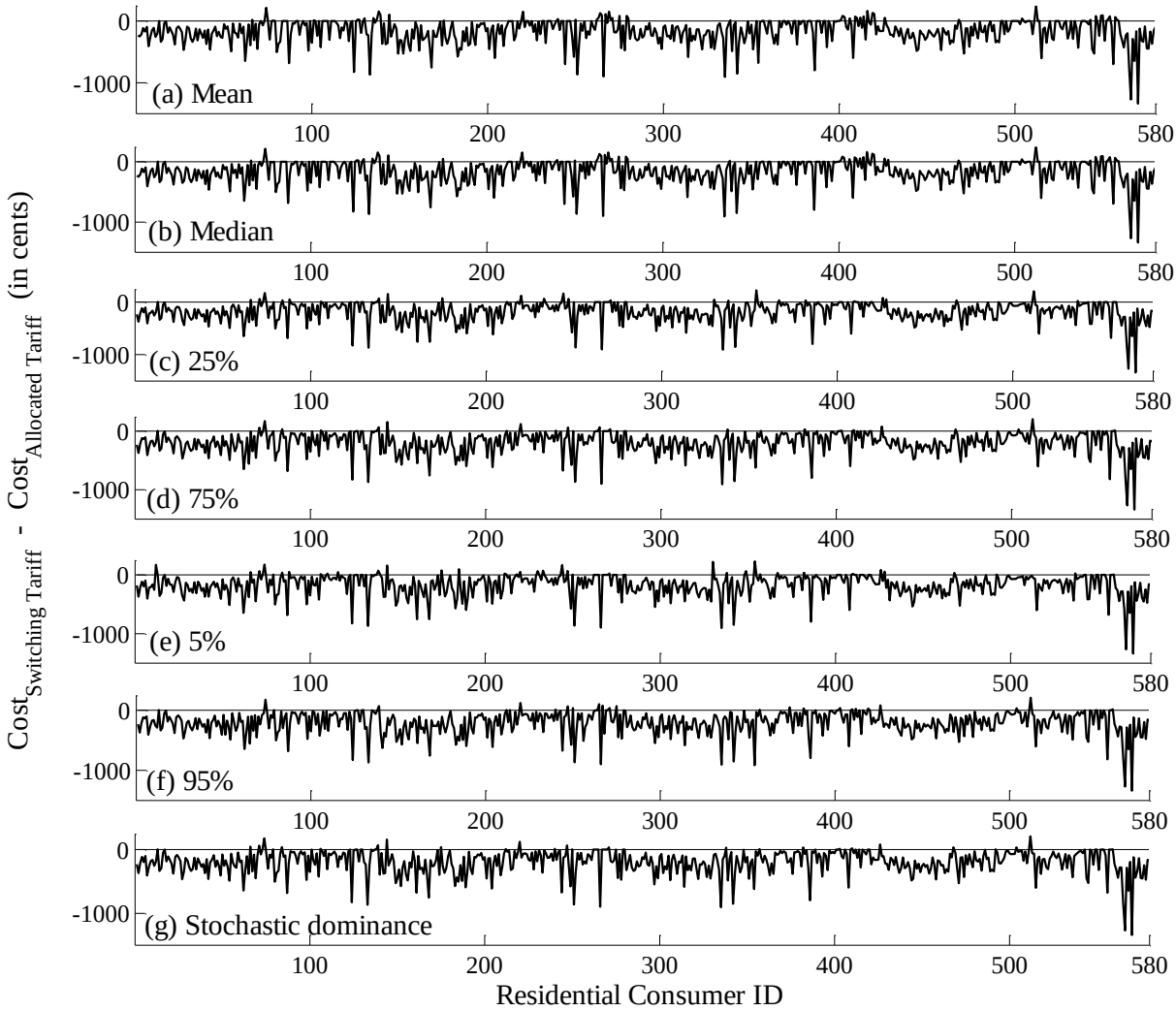

Figure 5.19— Difference between the Total Actual Cost Associated with Switching Tariffs and the Total Actual Cost Associated with Using a Single Allocated Tariff, Across 580 Residential Consumers, for the Post-sample Period, Using Seven Different Criteria.

TABLE 5.7

NUMBER OF WHO ACHIEVED LOWER COSTS BY: A) SWITCHING TARIFFS, AND, B) USING A SINGLE ALLOCATED TARIFF, AND WHO FOUND NO COST DIFFERENCE BETWEEN THE TWO SCHEMES. TABLE ALSO PRESENTS THE AVERAGE SAVINGS ACHIEVED VIA SWITCHING TARIFFS (IN CENTS PER CONSUMER) ACROSS ALL 580 RESIDENTIAL CONSUMERS.

Scheme/ Criterion	Expectation (Mean)	Median (50%)	25%	75%	5%	95%	Stochastic Dominance
Switching Tariff	432	432	506	521	495	530	515
Allocated Tariff	58	57	34	27	40	26	28
No Difference	90	91	40	32	45	24	37
Average Savings	169	167	187	201	176	210	198

5.6 Summary and Concluding Remarks

In this chapter, we have considered modelling the density of electricity smart meter data. To enable density forecasting, we used methods based on KD estimation and CKD estimation. The methods considered in this work were aimed at accommodating the double seasonality in consumption, along with the underlying variability. We considered different implementations of the CKD estimator, where we conditioned consumption on the period of week, period of day, and lagged consumption. We used a discount factor to place more emphasis on the more recent observations. The evaluation of post-sample density and quantile forecasts for SMEs showed that the methods considered in this study convincingly outperformed the unconditional KD estimator. Encouragingly, the proposed methods were able to accommodate the seasonality in consumption, and clearly distinguished weekdays from weekends. We concluded that for forecasting the total consumption time series, building a single model for the total consumption leads to better

forecasting accuracy, compared to the case where each consumption time series is modelled individually. Moreover, we derived prediction intervals for electricity cost based on density forecasts of electricity consumption, for six different tariffs. Using seven different tariff selection criteria, our empirical study showed that switching between different tariffs results in considerable cost savings overall, compared to the case when the consumers are allocated a single tariff for all periods.

A potentially useful line of future work would be to adequately accommodate the holiday effects during the density estimation of electricity consumption, where each special day is modelled separately. In addition, it would also be worth investigating the use of different classification techniques for identifying consumers with similar electricity consumption patterns. This information could potentially be useful for scaling the modelling framework to a large number of consumers, without incurring considerable computational costs.

Chapter 6

SUMMARY AND DIRECTIONS FOR FUTURE RESEARCH

In this thesis, we develop novel forecasting methodologies, having applications in two different areas: a) macroeconomics, and b) energy. For each application, we provide a comparison of the proposed methods, with a range of different state-of-the art models, using real datasets. Our comprehensive empirical comparison of the existing and proposed methods was undertaken using multiple performance scores. The results show strong potential for the models proposed in this thesis. This chapter provides a summary of the thesis, along with directions for future research. Specifically, Sections 6.1, 6.2, 6.3 and 6.4 provide summary and conclusions for Chapters 2, 3, 4, and 5, respectively. A discussion on potential interesting lines of future work is provided in Section 6.5.

6.1 Nonlinear and Nonparametric Modelling Approaches for Forecasting the US GNP

The GNP of a country is a commonly used macroeconomic indicator. Accurate future estimates of GNP are required for well-informed policy and decision-making. The GNP time series exhibit regime shifts, and existing models for forecasting GNP are structured to be able to characterise regimes of growth and recession in a business cycle. As presented in Chapter 2, some of the most commonly used methods for forecasting GNP time series are: a) autoregressive (AR) model, b) self-exciting threshold autoregressive (SETAR) model, and c) Markov-switching autoregressive (MS-AR) model (see, for example, Hamilton, 1989; Terasvirta and Anderson,

1992; Tiao and Tsay, 1994; Potter, 1995; Krolzig, 2001; Clements *et al.*, 2003; Clements *et al.*, 2004).

It is noteworthy that SETAR and MS-AR models are extensions of the linear AR model, and are also referred to as regime switching models. Both SETAR and MS-AR are parametric models, which make strong prior assumptions about the total number of regimes, the exact time period of regime shifts, and the behaviour of GNP in each separate regime, in advance of seeing the data. In this study, we estimate a SETAR(2;1,1) model comprising eight parameters, and a MS(2)-AR(5) model including ten parameters. It is noteworthy that we have few quarterly post-war GNP observations. Specifically, in this study, we have a total of 247 quarterly observations, out of which 199 observations were used for model estimation, while the remaining observations were employed for model evaluation. Several researchers have reported that regime-switching models are not considerably superior to their linear counterparts, in terms of post-sample forecast accuracy (Goodwin, 1993; Dueker and Neely, 2007), indicating that the regime-switching models may potentially be over-fitting the in-sample data. It is due to these reasons, that we investigate parsimonious nonparametric methods, comprising only a few parameters, for forecasting the US GNP.

In Chapter 2, we investigate two nonparametric methods: a) *fraction nearest neighbour* (f-NN), and b) *kernel regression*. The contribution of this study lies in proposing a novel method, called the *weighted random analogue prediction* (WRAP) method. The rationale for employing f-NN, kernel regression and WRAP, is that these methods: a) cope with nonlinearities in the data generating process, b) require the estimation of a few parameters (which allows us to optimize model parameters based on the in-sample density forecast performance), and c) provide

transition between regimes, without making prior assumptions about the presence and number of regimes in the time series.

As simple benchmarks, we use: a) unconditional mean, and b) random walk benchmark (also referred to as persistence). We evaluate point forecasts using the root mean squared error (RMSE) and mean absolute error (MAE), while for density forecast evaluation, we used the continuous ranked probability score (CRPS) and probability integral transform (PIT). Moreover, we also use the two-sided Kolmogorov-Smirnov (KS) test statistic, to investigate if the forecast distributions from two different models were statistically different.

Using the above performance scores, we found that SETAR and MS-AR were competitive with AR for lead times up to only two quarters ahead. Interestingly, Clements and Krolzig (1998), and Galbraith (2003) have reported similar results based on the comparison of out-of-sample point macroeconomic forecasts obtained from different parametric models. Moreover, we found that the unconditional benchmark was superior to the parametric models (AR, SETAR, MS-AR) considered in this study, for both point and density forecasts. The performance of the nonparametric models, namely f-NN, kernel regression, and WRAP, was superior compared to the unconditional benchmark and the parametric models. The superiority of nonparametric models relative to the parametric models was enhanced when the comparison was undertaken using the CRPS. We also demonstrate the efficacy of the nonparametric methods, over the benchmarks and parametric models, using time series for the UK gross domestic product (GDP). The results of this study indicate the importance of using the unconditional forecast as a benchmark instead of the random walk benchmark for macroeconomic time series. Encouragingly, the performance of the proposed WRAP method was superior to the parametric models, across both US GNP and UK GDP time series, for both point and density forecasts.

6.2 Short-term Load Forecasting using Discount Weighted SVD-based Methods

Short-term forecasts of electricity demand are crucial for the safe and efficient operations of a power system. A major limitation with some of the existing methodology for load forecasting, however, is that it relies on building separate models for different periods of the seasonal cycle (Ramanathan *et al*., 1997; Cottet and Smith, 2003; Cancelo *et al*., 2008). This leads to the modelling approach being of high dimension. To overcome this limitation, Shen and Huang (2005, 2008a, 2008b), Taylor *et al*. (2006), and Taylor (2010b, 2012), propose methods based on singular value decomposition (SVD) for modelling the intraday time series. The rationale of using SVD-based methods is that it allows the focus to be on forecasting only the components that capture a major proportion of variance in the data, thereby reducing the dimensionality of the model to be considered. However, most of the previously proposed SVD-based methods employ sequence updating of only the feature series, whereby feature vector is computed only once. This is a limitation as both feature series and feature vector change over time, as observed using the load data. Furthermore, none of the existing SVD-based methods take into account the intrayear seasonal effects while modelling load.

In Chapter 3, the contribution of our work lies in proposing two novel SVD-based methods, namely: a) *discount weighted* (DW) *intraday SVD-based exponential smoothing*, and b) *DW intraweek SVD-based exponential smoothing*. The above two methods can be viewed as extensions of the SVD-based methods used by Taylor (2010b, 2012). The idea of discount weighting is based on discount weighted regression (DWR). The general idea behind DWR is to allow the model parameters to be time-varying, whereby different discount factors are employed for each model parameter. In the context of modelling load, discount weighting is appealing as it

allows different periods of a seasonal cycle to be discounted at different rates, whereas SVD is worth investigating for load, as it aims to reduce the dimension of the model by utilizing similarity between different periods in a seasonal cycle.

The rationale of the proposed DW SVD-based methods is that they: a) achieve dimension reduction, b) incorporate discount factors, which enable more recent observations to be assigned more weight in the dimension reduction, c) adapt with the inclusion of new observations via sequence updating of both feature series and feature vectors, and d) accommodate triple seasonality in load.

As sophisticated benchmarks, we employ some of the most commonly used methods for forecasting intraday time series, specifically, we use triple seasonal variants of: a) Holt-Winters-Taylor (HWT) exponential smoothing, b) seasonal autoregressive moving average (ARMA), c) artificial neural networks (ANNs), d) intraday SVD-based exponential smoothing, and e) intraweek SVD-based exponential smoothing methods. Furthermore, as simple benchmarks, we use: a) seasonal moving average (SMA), and b) seasonal random walk (SRW).

We found that all the SVD-based methods considered in this study led to a reduction in the dimensionality of the model. The two simple benchmarks (SMA and SRW) were outperformed by all the other methods, at all lead times. Encouragingly, a combination of the two proposed DW SVD-based methods outperformed the SARMA and ANN methods, and were found to be competitive with the HWT exponential smoothing method overall.

6.3 Short-term Forecasting of Anomalous Load using Rule-based Methods

In Chapter 4, we propose a novel rule-based approach for forecasting short-term load, with emphasis on special days. Existing studies have often ignored anomalous load observations

during the modelling process (Smith, 2000; Hippert et al., 2005). Most studies on short-term forecasting of normal and anomalous load have employed regression-based methods with dummy variables for special days (see Cancelo *et al*., 2008; Soares and Medeiros 2008; and Dordonnat *et al*., 2008). A limitation with the regression-based models for special days is that they are highly parameterized. Furthermore, the existing studies on anomalous load forecasting often rely on classifying different special days as having the same profile.

The methodological contribution of our study is to adapt some of the most commonly used statistical forecasting methods, to model both normal and anomalous load in a coherent and unified framework, using a rule-based approach. As opposed to previously employed regression-based models, a rule-based approach allows the incorporation of prior information regarding anomalous load profiles into the modelling process. In this study, a rule essentially identifies the past special day that has daily load profile that will be most useful in enhancing the time series model's estimate of load for the future special day. The rule can be viewed as the prior information that is fed into the statistical model, along with the historical load observations. The rationale of the proposed rule-based modelling framework is that it: a) allows for incorporation of prior expert knowledge into the modelling framework, b) allows every special day to have a unique load profile, c) takes into account the triple seasonal effects on anomalous load, d) is sufficiently flexible that they can be updated with the inclusion of new observations, and e) can be easily adapted to different datasets, whereby the time series may exhibit seasonality and anomalous observations may pose modelling challenges.

In Chapter 4, the methodological contribution of our work lies in proposing triple seasonal variants of the following modified *rule-based* methods: a) HWT exponential smoothing, b) SARMA, c) ANNs, d) intraday SVD-based exponential smoothing, and e) intraweek SVD-based

exponential smoothing. Furthermore, we formulate four different rules subjectively by inferring them directly from the data. In addition to the above described rule-based methods, we investigate following three implementations of ANNs: a) rule-based ANN with calendar data, b) rule-based ANN with extra lags, and c) separate ANNs. As simple benchmarks, we employ: a) SRW, b) SMA, and c) recent Sunday, whereby we replace anomalous load, with load observed on the most recent Sunday, as done by Smith (2000).

The MAPE results for the rule-based based methods (for the best performing rules, across special days) were about half the size of those for the corresponding original models that were not rule-based. Our results were consistent across load for both Great Britain and France.

The results of this study indicate that when domain knowledge is available and the time series has a consistent structure (seasonality in this case), rule-based subjective forecasting can outperform conventional extrapolation methods, as also shown by Bunn and Wright (1991), Collopy and Armstrong (1992), and Adyaa *et al.* (2000).

6.4 Modelling Electricity Smart Meter Data Based on Kernel Density Estimation

Most existing studies on smart meter data focus on estimating the impact of time-of-use (TOU) pricing strategies on electricity consumption (Cosmo *et al*., 2012, Spector *et al*., 1995; Alberini *et al*., 2011; Bartush *et al*., 2011). However, to the best of our knowledge, there are no studies which focus on generating density estimates of electricity consumption using smart meter data. Given the fact that smart meters are going to replace conventional electronic meters, both in the US and Europe, and its substantial potential impact in facilitating energy savings, we consider modelling electricity consumption for residential consumers and SMEs, as recorded from smart

meters.

The non-Gaussian and highly variable nature of individual smart meter electricity consumption data motivate us to employ nonparametric probability density estimation methods. Since kernel-based methods enable nonparametric estimation, and have been shown to be an efficient way of estimating the true underlying density (Silverman, 1986), we investigate methods based on *kernel density* (KD) estimation and *conditional kernel density* (CKD). The prominent intraday and intraweek variation in consumption prompt us to model consumption in terms of the period of week, and day of week, using a CKD estimator. This method captures the seasonality in consumption, and enables a nonlinear and nonparametric estimation of its conditional density. Within the CKD method, we use a discounting factor, which allows more emphasis to be placed on the more recent observations. Specifically, in this study, we employ the following seven different kernel-based methods: a) *unconditional kernel density estimation*, b) *separate kernel density estimation for each period of the week*, c) *CKD estimation conditional on the period of week*, d) *CKD estimation conditional on the period of week and period of day*, e) *KD estimation based on type of intraday cycle*, f) *CKD estimation conditional on type of intraday cycle*, and g) *CKD estimation conditional on lagged consumption*. Encouragingly, the KD-U method (which did not accommodate seasonality in the smart meter data), was easily outperformed by other methods that took the seasonal effects into account during the modelling, across both residential consumers and SMEs.

We also investigate if it is more suitable to: a) model consumption time series for each consumer individually, and then aggregate the individual forecasts to get an estimate of the total consumption, or b) sum each consumption time series to compute the total consumption first, and then model the total consumption using a single model. The results suggested that building a

single model for the total consumption leads to better forecasting accuracy, compared to the case where each consumption time series is modelled individually.

The results of this study indicate that the density of electricity smart meter data can be modelled reasonably well using methods based on KD and CKD estimation. The post-sample results show that the methods employed in this study accommodate the intraday and intraweek seasonality in consumption, and correctly distinguish weekdays from weekends.

6.5 Directions for Future Research

We employ a range of different methods to generate density forecasts in Chapter 2 for GNP, and in Chapter 5 for smart meter data. However, despite using a range of different models in the above chapters, which rely on very different modelling approaches, we did not consider combining density forecasts from these models, for any given application. It has been argued that forecasts from multiple models are more accurate, compared to the case when only one model is employed (Palmer, 1988). Hence, a potentially interesting line of future work would be to combine density forecasts obtained using the different models considered in this study, and create a composite forecast based on the aggregate of different forecasts.

The rationale for combining forecasts from different models lies in utilizing the individual strengths of different models, such that forecasts from the final composite model are at least as accurate as the best performing individual model (Stock and Watson, 2004). Combining forecasts from different models also allows one to diversify the risk associated with using a single model.

There is a vast literature on combining point forecasts (Clemen, 1989; Taylor and Majithia, 2000; Stock and Watson, 2004). Recently, there has been some interest in combining density forecasts, see Timmermann (2005), Wallis (2005), Gneiting and Ranjan (2011). A review of

guidelines for the use of forecast combination has been provided by de Menezes *et al*. (2000), focussing on different properties of the forecast error. In an empirical study, Taylor (2008) reported that for forecasting electricity demand for Great Britain, a univariate time series model outperformed a multivariate weather model for lead times up to about four hours ahead, while a combination of forecasts from the two methods was found to be the most accurate for lead times up to one day ahead. Hence, it would potentially be worth combining the short-term load forecasts, generated using univariate models in Chapter 3, with forecasts from a multivariate weather model.

In Chapter 3, we use SVD as a tool for dimension reduction. It would be worth investigating if methods based on alternate dimension reduction techniques (e.g. independent component analysis, wavelet transform) can generate more accurate short-term load forecasts, compared to the methods used in this study. Moreover, it would be interesting to extend the proposed univariate methods, to also include weather variables as additional explanatory variables and employ dimension reduction techniques to select only the most relevant variables in the new feature space during the modelling of load data.

In Chapter 4, we model anomalous load using a rule-based approach, such that for a given special day, we referred to only one historical special day. A simple and potentially interesting line of future work would be to set the annual lag differently for different times of the same special day. This would amount to using observations from different parts of different past special days for a future special day.

It would also be worth investigating the efficacy of the rule-based approaches developed in this work for different applications, such as forecasting call center arrivals, hospital arrivals, and transportation counts, as the corresponding time series exhibit seasonality and anomalous

conditions pose significant modelling challenges. For example, the time series for number of hospital arrivals (*attendances*) exhibits both seasonality and considerable volatility (Jones *et al*., 2002; Au-Yeung *et al.*, 2008), which may require a GARCH model to be used along with the seasonal methods. Furthermore, it would be worth explicitly modelling the attendances on special days, such as bank holidays, which exhibit noticeably different patterns of arrivals to the norm. Note that methods considered in Chapters 3 and 4 for load data have not been employed previously for the attendances data.

In Chapter 5, we apply the different methods based on KD and CKD estimation to electricity smart meter data. It would be worth investigating the efficacy of these kernel-based methods for modelling data from smart gas or water meters, as both these datasets exhibit intraday and intraweek seasonal patterns. Moreover, it would be worth investigating the efficacy of a spatiotemporal CKD estimator for modelling wind power. It is noteworthy that spatiotemporal modelling approaches have previously been used for modelling wind speed (Tastu *et al*., 2010), while CKD has been employed for modelling the density of wind power (Jeon and Taylor, 2012). In this regard, a multivariate spatiotemporal CKD model could be designed to take into account the correlation between different relevant weather variables at neighbouring wind farm sites during the modelling of wind data. Moreover, the CKD estimator used in Chapter 5 can potentially be adapted to model the global temperature anomaly. Hyndmann *et al*. (1996) use a CKD estimator to model temperature for a given day, by conditioning it on the temperature observed on the previous day. It would be worth constructing density forecasts for the global temperature anomaly using CKD, by utilizing information regarding the historical temperature, CO2 emissions, precipitation, and other relevant climate variables. We refer to this method as the multivariate CKD estimator, whereby temperature at a given time is conditioned on past

temperature, and other relevant climate variables. Recently, Fildes and Kourentzes (2011) have reported that a class of simple time series models generates more accurate point forecasts for the global temperature anomaly compared to a global climate model (GCM). We are not aware of any studies that have compared density forecasts obtained from a GCM with time series models. Fildes and Kourentzes (2011) employed ANNs to enable nonlinear and nonparametric modelling of global temperature time series, focussing only on point forecasts. Given that the CKD estimator enables nonlinear and nonparametric modelling of the data, it would be worth using CKD for modelling the density of global temperature anomaly time series. Furthermore, since GCMs and time series models rely on completely separate sources of information for generating forecasts, it would be worth investigating the scope of combining density forecasts generated using these two modelling approaches.

REFERENCES

Adam, O., Zarader, J.L. and Milgram, M. (1993). 'Identification and prediction of non-linear models with recurrent neural network,' *Lecture Notes in Computer Science*, 686, 531–535.

Adyaa, M., Collopy, F., Armstrong, J.S. and Kennedy, M. (2000). 'An application of rule-based forecasting to a situation lacking domain knowledge,' *International Journal of Forecasting*, 16, 471–484.

Akaike, H. (1973). 'Information theory and an extension of the maximum likelihood principle', in Second *International Symposium on Information Theory*, Budapest: Akademiai Kiado, 267–281.

Alberini, A., Gans, W. and Velez-Lopez, D. (2011). 'Residential consumption of gas and electricity in the US: the role of prices and income', *Energy Economics*, 33, 870–881.

Ameen, J.R.M. and Harrison, P.J. (1984). 'Discounted weighted estimation,' *Journal of Forecasting*, 3, 285–296.

Ang, A., Piazzesi, M. and Wei, M. (2006). 'What does the yield curve tell us about GDP growth?', *Journal of Econometrics*, 131, 359–403.

Asber, D., Lefebvre, S., Asber, J., Saad, M. and Desbiens, C. (2007). 'Non-parametric short-term load forecasting', *Electrical Power and Energy Systems*, 29, 630–635.

Atiya, A.F., El-Shoura, S.M., Shaheen, S.I. and El-Sherif, M.S. (1999). 'A comparison between neural-network forecasting techniques – case study: river flow forecasting,' *IEEE Transactions on Neural Networks*, 10, 402–409.

Atkeson, C.G., Moore, A.W. and Schaal, S. (1997). 'Locally weighted learning', *Artificial Intelligence Review*, 11, 11–73.

Au-Yeung, S.W.M., Harder, U., McCoy, E.J. and Knottenbelt, W.J. (2009). 'Predicting patient arrivals to an accident and emergency department,' *Emergency Medical Journal*, 26, 241–244.

Bacon, D.W. and Watts, D.G. (1971). 'Estimating the transition between two intersecting straight lines', *Biometrika*, 58 (3), 525–534.

Barhoumi, K., Benk, S., Cristadoro, R., Reijer, A.D., Jakaitiene, A., Jelonek, P., Rua, A., Rünstler, G., Ruth, K. and Nieuwenhuyze, C.V. (2008). 'Short-term forecasting of GDP using large monthly datasets : A pseudo real-time forecast evaluation exercise', European Central Bank, *Occasional Paper Series*, Number 84 / April 2008.

Barkoulas, J., Baum, C.F. and Chakraborty, A. (2003). 'Nearest-neighbour forecasts of U.S. interest rates', *Boston College Working Papers in Economics*, Number 313.

Bartusch, C., Wallin, F., Odlare, M., Vassileva, I. and Wester, L. (2011). 'Introducing a demand-based electricity distribution tariff in the residential sector: demand response and customer perception', *Energy Economics*, 39, 5008–5025.

Bashtannyk, D.M. and Hyndman, R.J. (2001). 'Bandwidth selection for kernel conditional density estimation,' *Computational Statistics and Data Analysis*, 36, 279–298.

Bessa, R.J., Miranda, V., Botterud, A., Wang, J. and Constantinescu, E.M. (2012). 'Time adaptive conditional kernel density estimation for wind power forecasting,' *IEEE Transactions on Sustainable Energy*, 3, 660–669.

Bessec, M. and Bouabdallah, O. (2005). 'What causes the forecasting failure of Markov-Switching models? A Monte Carlo study,' *Studies in Nonlinear Dynamics and Econometrics*, 9(2), Article 6.

Beven, K. (2002). 'Towards a coherent philosophy for modelling the environment,' *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 458, 2465–2484.

Billah, M. B., King, M.L., Snyder, R.D. and Koehler, A.B. (2006). 'Exponential smoothing model selection for forecasting,' *International Journal of Forecasting*, 22, 239–249.

Binner, J.M., Tino, P., Tepper, J., Anderson, R.G., Jones, B. and Kendall, G. (2009). 'Does money matter in inflation forecasting?,' Federal Reserve Bank of St. Louis, *Working Paper*, 2009–030A.

Bishop, C.M. (1997). 'Neural networks for pattern recognition,' Oxford University Press, Oxford.

Bliefernicht, J. and Bardossy, A. (2007). 'Probabilistic forecast of daily areal precipitation focusing on extreme events,' *Natural Hazards Earth Systems Science*, 7, 263–269.

Boero, G. and Marrocu, E. (2002). 'The performance of SETAR models: A regime conditional evaluation of point, interval and density forecasts,' *Working Paper*, 2008–02.

Box, G.E.P. and Jenkins, G.M. (1970). 'Time series analysis, forecasting and control,' San Francisco: Holden–Day.

Brock, W.A. and Sayers, C.L. (1988). 'Is the business cycle characterized by deterministic chaos?,' *Journal of Monetary Economics*, 22, 71–90.

Brown, R.G. (1963). ‘Smoothing, forecasting and prediction of discrete time series,’ Englewood Cliffs, NJ: Prentice-Hall.

Buchmann, M. and Hubrich, K. (2010). ‘Forecasting output and inflation in the US: regime changes and uncertainty,’ European Central Bank, *Working Paper*, 2010.

Bunn, D.W. (2000). ‘Forecasting loads and prices in competitive power markets,’ *Proceedings of the IEEE,* 88, 163–169.

Bunn, D. and Farmer, E. (1985). ‘Comparative models for electrical load forecasting,’ New York: Wiley, 3–11.

Bunn, D.W. and Wright, G. (1991). ‘Interaction of judgmental and statistical forecasting methods: issues and analysis,’ *Management Science,* 37, 501–518.

Cancelo, J.R., Espasa, A. and Grafe, R. (2008). ‘Forecasting from one day to one week ahead for the Spanish system operator,’ *International Journal of Forecasting,* 24, 588–602.

Charytoniuk, W. and Chen, M.-S. (2000). ‘Very short-term load forecasting using artificial neural networks,’ *IEEE Transactions on Power Systems*, 15, 263–268.

Clark, T.E. and McCracken, M.W. (2002). ‘Forecast-based model selection in the presence of structural breaks,’ Federal Reserve Bank of Kansas City, RWP, 02–05.

Clemen, R.T. (1989). ‘Combining forecasts: a review and annotated bibliography,’ *International Journal of Forecasting,* 5, 559–583.

Clements, M.P. and Hendry, D.F. (1998). ‘Forecasting economic processes,’ *International Journal of Forecasting*, 14, 111–131.

Clements, M.P. and Krolzig, H,-M. (1998). ‘A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP,’ *Econometrics Journal*, 1, C47-C75.

Clements, M.P. and Smith, J. (1999). ‘A Monte Carlo study of the forecasting performance of empirical SETAR models,’ *Journal of Applied Econometrics*, 14, 123–141.

Clements, M.P., Franses, P.H., Smith, J. and Dijk, D.V. (2003). ‘On SETAR non-linearity and forecasting,’ *Journal of Forecasting*, 22, 359–375.

Clements, M.P., Franses, P.H. and Swanson, N.R. (2004). ‘Forecasting economic and financial time-series with non-linear models,’ *International Journal of Forecasting*, 20, 169–183.

Cleveland, W.S., Devlin, S.J. and Grosse, E. (1988). ‘Regression by local fitting: methods, properties and computational algorithms,’ *Journal of Econometrics*, 37, 87–114.

Collopy, F. and Armstrong, J.S. (1992). ‘Rule-based forecasting: development and validation of expert systems approach to combining time series extrapolations,’ *Management Science*, 38, 1394–1414.

Comon, P. (1994). ‘Independent component analysis: a new concept?,’ *Signal Processing*, Elsevier, 36, 287–314.

Commission for Energy Regulation (2011a). Electricity smart metering customer behaviour trials findings report. Dublin: Commission for Energy Regulation.

Commission for Energy Regulation (2011b). Results of electricity cost-benefit analysis, customer behaviour trials and technology trials. Dublin: Commission for Energy Regulation.

Cottet, R. and Smith., M. (2003). 'Bayesian modelling and forecasting of intraday electricity load,' *Journal of the American Statistical Association*, 98, 839–849.

Cosmo, V.D., Lyons, S. and Nolanab, A. (2012). 'Estimating the impact of time-of-use pricing on Irish electricity demand,' Economic and Social Research Institute, Dublin, Ireland, *Working paper*.

Daubechies, I. (1992). 'Ten Lectures on Wavelets,' CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics.

Dai, J. and Sperlich, S. (2010). 'Simple and effective boundary correction for kernel densities and regression with an application to the world income and Engel curve estimation,' *Computational Statistics and Data Analysis*, 54, 2487–2497.

Dayal, B.S. and MacGregor, J.F. (1997). 'Recursive exponentially weighted PLS and its application to adaptive control and prediction,' *Journal of Process Control*, 7, 169–79.

de Menezes L.M., Bunn, D. and Taylor, J.W. (2000), 'Review of practical guidelines for combining forecasts,' *European Journal of Operational Research*, 120, 190–204.

Darby, S.J. (2012). 'Metering: EU Policy and implications for fuel poor households,' *Energy Policy*, 49, 98–106.

DECC (2012). Annual report on fuel poverty statistics, Department of Energy and Climate Change, A National Statistics Publication.

De Livera, A.M., Hyndman, R.J. and Snyder, R.D. (2011). 'Forecasting time series with complex seasonal patterns using exponential smoothing,' *Journal of the American Statistical Association*, 106, 1513–1527.

Diebold, F.X. and Nason, J.A. (1990). 'Nonparametric exchange rate prediction,' *Journal of International Economics*, 28, 315–332.

Directive 2009/72/EC of the European Parliament and of the Council, 2009. *Official Journal of the European Union*, L 211/55–L 211/92.

Ding, Ni., Besanger, Y., Wurtz, F., Antoine, G., and Deschamps, P. (2011). 'Time series method for short-term load forecasting using smart metering in distribution systems,' *PowerTech, 2011 IEEE Trondheim*.

Domestic Metering Innovation (2006). Office of Gas and Electricity Markets (OFGEM) Report.

Dordonnat, V., Koopman, S.J., Ooms, M., Dessertaine, A. and Collet, J. (2008). 'An hourly periodic state space model for modelling French national electricity load,' *International Journal of Forecasting*, 24, 566–587.

DSDNI (2012). Ending Fuel Poverty: A Strategy for Northern Ireland. Department for Social Development Northern Ireland, Belfast.

Dueker, M. and Neely, C.J. (2007). 'Can Markov switching models predict excess foreign exchange returns?,' *Journal of Banking and Finance*, 31, 279–296.

Engle, R.F. (1982). 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation,' *Econometrica*, 50(4), 987–1007.

ERGEG (2010). Position paper on smart grids. An ERGEG conclusions paper, Ref: E10-EQS-38-05.

Fan, J. and Yim, T.H. (2004). 'A cross-validation method for estimating conditional densities,' *Biometrika*, 91(4), 819–834.

Faruqui, A., Harris, D. and Hledik, R. (2010). 'Unlocking the €53 billion savings from smart meters in the EU: how increasing the adoption of dynamic tariffs could make or break the EU's smart grid investment,' *Energy Policy*, 38, 6222–6231.

Feinberg, E.A. and Genethliou, D. (2005). 'Load forecasting,' Applied Mathematics for Restructured Electric Power Systems: Optimization, Control, and Computational Intelligence, Chow, J.H., Wu, F.F., Momoh, J.J., (eds.), Springer.

Feng, H. and Liu, J. (2002). 'A SETAR model for Canadian GDP: Nonlinearities and forecast comparisons,' *University of Victoria Working Paper*, EWP 0206.

Fernández-Rodríguez, F., Sosvilla-Rivero, S. and Andrada-Félix, J. (1999). 'Exchange-rate forecasts with simultaneous nearest-neighbour methods: evidence from the EMS,' *International Journal of Forecasting*, 15, 383–392.

Fildes, R., Hibon, M., Makridakis, S. and Meade, N. (1998). 'Generalising about univariate forecasting methods: Further empirical evidence,' *International Journal of Forecasting*, 14, 339–358.

Fildes, R. and Kourentzes, N. (2011). 'Validation and forecasting accuracy in models of climate change,' *International Journal of Forecasting*, 27, 968–995.

Gardner, D.T. and Rogers, J.S. (1999). 'Planning electric power systems under demand uncertainty with different technology lead times,' *Management Science*, 45, 1289-1306.

Ghofrani, M., Hassanzadeh, M., Etezadi-Amoli, M., and Fadali, M. S. (2011). 'Smart meter

based short-term load forecasting for residential customers,' *North American Power Symposium* (NAPS), 1–5.

Gould, P.G., Koehler, A.B., Ord, J.K., Snyder, R.D., Hyndman, R.J. and Vahid-Araghi, F. (2008). 'Forecasting time-series with multiple seasonal patterns,' *European Journal of Operational Research*, 191, 207–222.

Gençay, R. and Selçuk, F. (1998). 'A visual goodness-of-fit test for econometric models,' *Studies in Nonlinear Dynamics and Econometrics*, 3, 157–167.

Gneiting, T., Balabdaoui, F. and Raftery, A.E. (2007). 'Probabilistic forecasts, calibration and sharpness,' *Journal of the Royal Statistical Society*, 69, 243–268.

Gneiting, T. and Raftery, A.E. (2007). 'Strictly proper scoring rules, prediction, and estimation,' *Journal of the American Statistical Association*, 102, 359–378.

Gneiting, T. and Thorarinsdottir, T.L. (2010). 'Predicting inflation: professional experts versus no-change forecasts,' arXiv:1010.2318v1.

Gneiting, T. and Ranjan, R. (2011a). 'Comparing density forecasts using threshold and quantile-weighted scoring rules,' *Journal of Business and Economic Statistics*, 29, 411–422.

Gneiting, T. and Ranjan, R. (2011b). 'Combining predictive distributions,' arXiv:1106.1638v1.

Gneiting, T. (2011c). 'Quantiles as optimal point forecasts,' *International Journal of Forecasting*, 27, 197–207.

Goldfeld, S.M. and Quandt, R.E. (1973). 'A Markov model for switching regressions,' *Journal of Econometrics*, 1, 3–16.

Goodwin, T.H. (1993). ‘Business-cycle analysis with a Markov-switching model,’ *Journal of Business and Economic Statistics,* 11, 331–339.

Gooijer, J. G. D. and Kumar, K. (1992). ‘Some recent developments in non-linear time series modelling, testing and forecasting,’ *International Journal of Forecasting,* 8, 135–156.

Granger, C.W.J. and Terasvirta, T. (1993). ‘Modelling nonlinear economic relationships,’ Oxford University Press.

Gupta, R. and Modise, M.P. (2011). ‘Macroeconomic variables and South African stock return predictability,’ University of Pretoria, *Working Paper*, 2011–07.

Hamilton, J.D. (1989). ‘A new approach to the economic analysis of nonstationary time series and the business cycle,’ *Econometrica,* 57(2), 357–384.

Hamilton, J.D. (1990). ‘Analysis of time series subject to changes in regime,’ *Journal of Econometrics,* 45, 39–70.

Hansen, B.E. (1992). ‘The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP,’ *Journal of Applied Econometrics,* 7, S61–S82.

Hansen, B.E. (1996a). ‘Erratum: The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP,’ *Journal of Applied Econometrics*, 11, 195–198.

Hansen, B.E. (1996b). ‘Inference when a nuisance parameter is not identified under the null hypothesis,’ *Econometrica,* 64, 413–430.

Hansen, B.E. (1999). ‘Threshold effects in non-dynamic panels: Estimation, testing, and inference,’ *Journal of Econometrics*, 93(2), 345–368.

Hansen, B.E. (2001). 'The new econometrics of structural change : dating breaks in U.S. labor productivity,' *The Journal of Economic Perspectives*, 15, 117–128.

Harrison, P.J. and Johnston, F.R. (1984). 'Discount weighted regression,' *Journal of the Operational Research Society*, 35, 923–932.

Harvey, A. and Oryshchenko, V. (2011). 'Kernel density estimation for time series data,' *International Journal of Forecasting*, 28(1), 3–14.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). 'The elements of statistical learning: data mining, inference, and prediction,' Springer, 2nd edition.

Hippert, H.S., Bunn, D.W. and Souza, R.C. (2005). 'Large scale neural networks for electricity load forecasting: Are they overfitted?,' *International Journal of Forecasting*, 21, 425–434.

Hopfield, J.J. (1982). 'Neural networks and physical systems with emergent collective computational abilities,' *Proceedings of the National Academy of Sciences*, 79, 2554–2558.

Hyde, O. and Hodnett, P.F. (1994). 'Rule-based procedures in short-term electricity load forecasting,' *IMA Journal of Mathematics Applied in Business and Industry*, 5, 131–141.

Hyde, O. and Hodnett, P.F. (1997). 'An adaptable automated procedure for short-term electricity load forecasting,' *IEEE Transactions on Power Systems*, 12, 84–94.

Hyndman, R.J., Bashtannyk, D.M. and Grunwald, G.K. (1996). 'Estimating and visualizing conditional densities,' *Journal of Computational and Graphical Statistics*, 5, 315–336.

Hyvärinen, A. and Oja., E. (2000). 'Independent component analysis: algorithms and applications,' *Neural Networks*, 13, 411–430.

Jeon. J. and Taylor, J.W. (2012). 'Using conditional kernel density estimation for wind power density forecasting,' *Journal of the American Statistical* Association, 107 (497), 66–79.

Jones, M.C. (1993). 'Simple boundary correction in kernel density estimation,' *Statistics and Computing*, 3, 135–146.

Jones, M.C., Marron, J.S. and Sheather, S.J. (1996). 'A brief survey of bandwidth selection for density estimation,' *Journal of the American Statistical Association*, 91, 401–407.

Jones, S.S., Thomas, A. and Evans, R.S. (2008). 'Forecasting daily patient volumes in the emergency department,' *Academy of Emergency Medicine*, 15, 159–170.

Kantz, H. and Schreiber, T. (2004). 'Nonlinear time series analysis,' Cambridge University Press.

Kapetanios, G. and Tzavalis, E. (2005). 'Nonlinear modelling of autoregressive structural breaks in a US diffusion index dataset,' Bank of England, *Working Paper*, Number 537.

Kapetanios, G., Labhard, V. and Price, S. (2007). 'Forecast combination and the Bank of England's suite of statistical forecasting models,' Bank of England, *Working Paper*, Number 323.

Kim, C-J. and Nelson, C.R. (1970). 'State-space models with regime switching,' Massachusetts, MIT Press.

Kim, K-H., Youn, H-S. and Kang, Y-C. (2000). 'Short-term load forecasting for special days in anomalous load conditions using neural networks and fuzzy inference method,' *IEEE Transactions on Power Systems*, 15, 559–565.

Kolmogorov, A.N. (1933). 'On the empirical determination of a distribution function,' Giornale dell'Instituto Italiano degli Attuari, 4, 83–91.

Koutsoyiannis, D., Efstratiadis, A., Mamassis, N. and Christofides, A. (2008). 'On the credibility of climate predictions,' *Hydrological Sciences*, 53, 671–684.

Krolzig, H.-M. (1997). 'Markov switching vector autoregressions: modelling, statistical inference and application to business cycle analysis,' *Lecture Notes in Economics and Mathematical Systems*, 454. Berlin: Springer–Verlag.

Krolzig, H.-M. (2001). 'Business cycle measurement in the presence of structural change: International evidence,' *International Journal of Forecasting*, 17, 349–368.

Lamedica, R., Prudenzi, A., Sforna, M., Caciotta, M. and Ceccellli, V.O. (1996). 'A neural network based technique for short-term forecasting of anomalous load periods,' *IEEE Transactions on Power Systems*, 11, 1749–1756.

Lane, S., Martin, E.B., Morris, A.J. and Gower, P. (2003). 'Application of exponentially weighted principal component analysis for the monitoring of a polymer film manufacturing process,' *Transactions of the Institute of Measurement and Control*, 25, 17–35.

Li, W., Yue, H.H., Valle-Cervantes, S. and Qin, S.J. (2000). 'Recursive PCA for adaptive process monitoring,' *Journal of Process Control*, 10, 471–84.

Little, M.A., Rodda, H.J.E. and McSharry, P.E. (2008). 'Bayesian objective classification of extreme UK daily rainfall for flood risk applications,' *Hydrology and Earth System Sciences*, 5, 3033–3060.

Lorenz, E.N. (1969). 'Atmospheric predictability as revealed by naturally occurring analogues,' *Journal of the Atmospheric Sciences*, 26, 636–646.

Loungani, P. (2001). 'How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth,' *International Journal of Forecasting*, 17, 419–432.

Lu, C-J., Lee, T-S. and Chiu, C-C. (2009). 'Financial time series forecasting using independent component analysis and support vector regression,' *Decision Support Systems*, 47(2), 115–125.

Marcellino, M., Stock, J.H. and Watson, M.W. (2003). 'Macroeconomic forecasting in the Euro area: country specific versus area-wide information,' *European Economic Review*, 47, 1–18.

Marcellino, M., Stock, J.H. and Watson, M.W. (2006). 'A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series,' *Journal of Econometrics*, 135, 499–526.

Mays, J.E., Birch, J.B. and Starnes, B.A. (2001). 'Model robust regression: combining parametric, nonparametric, and semiparametric methods,' *Journal of Nonparametric Statistics*, 13(2), 245–277.

McCracken, M.W. (2004). 'Asymptotics for out-of-sample tests of Granger causality,' *Journal of Econometrics*, 140, 719–752.

Morley, J. and Piger, J. (2004). 'The importance of nonlinearity in reproducing business cycle features,' Federal Reserve Bank of St. Louis, *Working Paper*, 2004–032A.

Mulhern, F.J. and Caprara, R.J. (1994). 'A nearest neighbor model for forecasting market response,' *International Journal of Forecasting*, 10, 191–207.

Nadaraya, E.A. (1964). 'On estimating regression,' *Theory of Probability and Its Applications*, 9, 141–142.

National Consumer Council Report (2008). 'The consumer implications of smart meters,' prepared by Owen, G. and Ward, J., Sustainability first.

Nikolaev N., de Menezes, L.M. (2008), 'Sequential Bayesian kernel modelling with non-Gaussian noise,' *Neural Networks*, 21, 36–47.

Noel, N. and Newbold, P. (1994). 'Late forecasts and early revisions of United States GNP,' *International Journal of Forecasting*, 10(3), 455–460.

Obled, C., Bontron, G. and Garçon, R. (2002). 'Quantitative precipitation forecasts: a statistical adaptation of model outputs through an analogues sorting approach,' *Atmospheric Research*, 63, 303–324.

Öcal, N. (2000). 'Nonlinear models for U.K. macroeconomic time series,' *Studies in Nonlinear Dynamics and Econometrics*, 4, 123–135.

Palmer, T.N. (1988). 'Medium and extended range predictability and stability of the Pacific/North American mode,' *Quarterly Journal of the Royal Meteorological Society*, 114, 691–713.

Panagiotidis, T. (2002). 'Testing the assumption of linearity,' *Economics Bulletin*, 3, 1−9.

Paparella, F., Provenzale, A., Smith, L.A., Taricco, C. and Vio, R. (1997). 'Local random analogue prediction of nonlinear processes,' *Physics Letters A*, 235, 233–240.

Pardo, A., Meneu, V. and Valor, E. (2002). 'Temperature and seasonality influences on Spanish electricity load,' *Energy Economics*, 24, 55–70.

Parzen, E. (1962). 'On estimation of a probability density function and mode,' *Annals of Mathematical Statistics,* 33, 1065–1076.

Pesaran, M.H. and Potter, S.M. (1997). 'A floor and ceiling model of US output,' *Journal of Economic Dynamics and Control*, 21, 661–695.

Potter, S.M. (1995). 'A nonlinear approach to US GNP,' *Journal of Applied Econometrics*, 10, 109–125.

Rahman, S. and Bhatnagar, R. (1988). 'An expert system based algorithm for short term load forecast,' *IEEE Transactions on Power Systems*, 3, 392–399.

Ramanathan, R., Engle, R., Granger, C.W.J., V-Araghi, F. and Brace, C. (1997). 'Short-run forecasts of electricity load and peaks,' *International Journal of Forecasting*, 13, 161–174.

Regonda, S.K., Rajagopalan, B., Lall, U., Clark, M. and Moon, Y.-I. (2005). 'Local polynomial method for ensemble forecast of time series,' *Nonlinear Processes in Geophysics*, 12, 397–406.

Rosenblatt, M. (1952). 'Remarks on a multivariate transformation,' *Annals of Mathematical Statistics,* 23, 470–472.

Rosenblatt, M. (1969). 'Conditional probability density and regression estimators,' Krishnaiah, P.R. (eds.), *Multivariate Analysis II*. New York: Academic Press, 25–31.

Ruan, D. (2010). 'Kernel density estimation-based real-time prediction for respiratory motion,' *Physics in Medicine and Biology*, 55, 1311–1326.

Schwarz, G.E. (1978). 'Estimating the dimension of a model,' *Annals of Statistics*, 6, 461–464.

Shen, H. and Huang, J.Z. (2005). 'Analysis of call center arrival data using singular value decomposition,' *Applied Stochastic Models in Business and Industry*, 21, 251–263.

Shen, H. and Huang, J. Z. (2008a). 'Interday forecasting and intraday updating of call center arrivals,' *Manufacturing and Service Operations Management*, 10, 391–410.

Shen, H. and Huang, J.Z. (2008b). 'Forecasting time series of inhomogeneous Poisson processes with application to call center workforce management,' *The Annals of Applied Statistics*, 2, 601–623.

Sheather, S.J. and Jones, M.C. (1991). 'A reliable data-based bandwidth selection method for kernel density estimation,' *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 53**,** 683–690.

Silverman, B.W. (1986). 'Density estimation for statistics and data analysis,' Chapman & Hall.

Smirnov, N.V. (1939). 'On the estimation of the discrepancy between empirical curves of distribution for two independent samples,' *Bulletin of Moscow University*, 2, 3–16.

Smith, M. (2000). 'Modelling and short-term forecasting of New South Wales electricity system load,' *Journal of Business, Economics and Statistics*, 18, 465–478.

Soares, L.J. and Medeiros, M.C. (2008). 'Modelling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data,' *International Journal of Forecasting*, 24, 630–644.

Song, K-B., Baek, Y-S., Hong, D.H. and Jong, G. (2005). 'Short-term load forecasting for the holidays using fuzzy linear regression method,' *IEEE Transactions on Power Systems*, 20, 96–101.

Spector, Y., Tishler., A. and Ye, Y. (1995). 'Minimal adjustment costs and the optimal choice of inputs under time-of-use electricity rates,' *Management Science*, 41, 1679–1692.

Srinivasan, D., Chang, C.S. and Liew, A.C. (1995). 'Demand forecasting using fuzzy neural computation, with special emphasis on weekend and public holiday forecasting,' *IEEE Transactions on Power Systems*, 10, 1897–1903.

Stock, J.H. and Watson, M.W. (1999). 'A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series,' Ch. 1 in Cointegration, Causality and Forecasting: A Festschrift for Granger, C.W.J., Engle, R., White, H., (eds), Oxford University Press, Oxford, 1–44.

Stock, J.H. and Watson, M.W. (2004). 'Combination forecasts of output growth in a seven-country data set,' *Journal of Forecasting*, 23, 405–430.

Stone, C.J. (1977). 'Consistent nonparametric regression,' *The Annals of Statistics*, 5(4), 595–620.

Sun, S. and Chen, Q. (2008). 'Kernel regression with a Mahalanobis metric for short-term traffic flow forecasting,' *Lecture Notes in Computer Science*, 5326, 9–16.

Swanson, N.R. and White, H. (1995). 'A model selection approach to assessing the information in the term structure using linear models and artificial neural networks,' *Journal of Business and Economic Statistics*, 13, 265–275.

Swanson, N.R. and White, H. (1997). 'A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks,' *Review of Economics and Statistics*, 79, 540–550.

Tastu, J., Pinson, P., Trombe, P-J. and Madsen H. (2013). 'Probabilistic forecasts of wind power generation accounting for geographically dispersed information,' *IEEE Transactions on Smart Grid*, 5, 480–489.

Taylor, J.W. (2003). 'Exponential smoothing with a damped multiplicative trend,' *International Journal of Forecasting*, 19, 715–725.

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

Taylor, J.W. (2008). 'Exponentially weighted information criteria for selecting among forecasting models,' *International Journal of Forecasting*, 24, 513–524.

Taylor, J.W. (2010a). 'Triple seasonal methods for short-term load forecasting,' *European Journal of Operations Research*, 204, 139–152.

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

Taylor, J.W. (2012). 'Short-term load forecasting with exponentially weighted methods,' *IEEE Transactions on Power Systems*, 27(1), 458–464.

The Edison Foundation IEE Report (2012). Utility-scale smart meter deployments, plans, and proposals.

Tenti, P. (1996). 'Forecasting foreign exchange rates using recurrent neural networks,' *Applied Artificial Intelligence*, 10, 567–582.

Tiao, G.C. and Tsay, R.S. (1994). 'Some advances in non-linear and adaptive modelling in time-series,' *Journal of Forecasting*, 13, 109–131.

Timmermann, A. (2005). 'Forecast combinations,' Elliot, G., Granger, C.W.J., Timmermann, A., (eds.), Handbook of Economic Forecasting, Elsevier, Amsterdam (2006).

Terasvirta, T. (1995). 'Modelling nonlinearity in U.S. Gross National Product 1889-1987,' *Empirical Economics*, 20, 577–597.

Terasvirta, T. and Anderson, H.M. (1992). 'Characterizing nonlinearities in business cycles using smooth transition autoregressive models,' *Journal of Applied Econometrics*, 7, S119–S139.

Tong, H. (1978). 'On a threshold model,' In Chen, C. H., (eds.), *Pattern Recognition and Signal Processing*, 101–141. Amsterdam: Sijhoff and Noordoff.

Tong, H. (1983). 'Threshold models in nonlinear time series analysis,' New York, Springer-Verlag.

Tong, H. and Lim, K.S. (1980): 'Threshold autoregression, limit cycles and cyclical data,' *Journal of the Royal Statistical Society*, 42, 245–292.

Tsay, R.S. (1991). 'Detecting and modelling nonlinearity in univariate time series analysis,' *Statistica Sinica*, 1, 431–451.

Waddams, C. (2002). 'Prepayment meters: the consumer perspective,' *Energy Action*, 86, 14–15.

Wall, M.E., Rechtsteiner, A. and Rocha, L.M. (2003). 'Singular value decomposition and principal component analysis'. A Practical Approach to Microarray Data Analysis. Berrar, D.P., Dubitzky, W., Granzow, M., (eds.), pp. 91–109, Kluwer: Norwell.

Wallis, K.F. (2005). 'Combining Density and Interval Forecasts: A modest proposal,' *Oxford Bulletin of Economics and Statistics*, 67, 0305–9049.

Watier, L. and Richardson, S. (1995). 'Modelling of an epidemiological time series by a threshold autoregressive model,' *The Statistician*, 44 (3), 353–364.

Winters, P.R. (1960). 'Forecasting sales by exponentially weighted moving averages,' *Management Science*, 6, 324–342.

Woodcock, F. (1979). 'On the use of analogue forecasts to improve regression forecasts,' *Monthly Weather Review*, 108, 292–297.

Wold, S. (1994). 'Exponentially weighted moving principal component analysis and projection to latent structures,' *Chemometrics and Intelligent Laboratory Systems*, 23, 149–61.

Yang, L. (2000). 'Root-n convergent transformation-kernel density estimation,' *Journal of Nonparametric Statistics*, 12, 447–474.

Yang, Y. (2004). 'Combining forecasting procedures: some theoretical results,' *Econometric Theory*, 20, 176–222.

Zhang, S. and Karunamuni, R.J. (1998). 'On kernel density estimation near endpoints,' *Journal of Statistical Planning and Inference*, 70, 301–316.

Zheng, I.Y. and Rossiter, J. (2006). 'Using monthly indicators to predict quarterly GDP,' Bank of Canada, *Working Paper*, Number, 2006–26.