ANALYTIC METHODS FOR ELECTRICITY PRICE FORECASTING: APPLICATION TO NEW ZEALAND ELECTRICITY MARKET

A THESIS SUBMITTED TO AUCKLAND UNIVERSITY OF TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Supervisor

Dr. Nuttanan Wichitaksorn

Dr. Wenjun Zhang

May 2024

By

Gaurav Kapoor

School of Engineering, Computer and Mathematical Sciences

Abstract

Electricity price forecasting has become a crucial focus for energy market participants in the last few decades. Its importance stems from the lack of efficient electricity storage options, and the uncertainty in its generation and ability to meet real-time demand. This thesis presents three independent forecasting studies for New Zealand electricity prices using diverse methods and models from electricity price forecasting literature. The models, methodologies, and variables considered in these studies are tailored for the New Zealand electricity market, and as such, the results are of direct interest to participants in the New Zealand electricity sector.

The first study forecasts daily electricity prices using Markov regime-switching (MRS) models and compares them to an extreme-value theory (EVT) framework. Due to the application of the generalised Pareto distribution for extreme prices, the EVT framework is able to perform better in in-sample density fitting than other models, despite its relative lack of complexity. However, the three-regime MRS model with time-varying transition probabilities presents the best out-of-sample price density fits.

The second study employs a mixed-frequency framework to forecast half-hourly electricity prices using hourly weather variables. A mixed-frequency vector-autoregressive (MF-VAR) model and the reverse unrestricted mixed-data sampling (RU-MIDAS) model are compared. LASSO regularization stands out as a key factor, consistently enhancing the forecasting performance when it is applied.

Inspired by LASSO's success, the third study further explores the impact of feature

selection on forecasting performance. This study compares statistical (GARCH and stochastic volatility) and machine learning models (LASSO-estimated auto-regressive, deep neural network (DNN), long short-term memory (LSTM), gated recurrent unit (GRU), extreme gradient boosting (XGBoost)) for daily electricity price forecasting. A meticulous comparison methodology involves a large number of external features and a variety of feature selection methods, including the LASSO, mutual information, and recursive feature selection. GARCH-t, SV-t, GARCH, and SV with LASSO-selected features consistently outperform benchmark models LEAR and DNN, showcasing performance increases of over 40% compared to GARCH and SV models with all features.

Contents

List of Tables

List of Figures

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning.

_____

Signature of candidate

Publications

This thesis contains material from the following manuscripts:

- Kapoor, G., Wichitaksorn, N., & Zhang, W. (2023). Analyzing and forecasting electricity price using regime-switching models: The case of New Zealand market. *Journal of Forecasting*, 1-16. https://doi.org/10.1002/for.3004.
- Kapoor, G., Wichitaksorn, N., Li, M., & Zhang, W. (2023). Forecasting Half-Hourly Electricity Prices using a Mixed-Frequency Structural VAR Framework. *Available at SSRN 4473100*.
- Kapoor, G., & Wichitaksorn, N. (2023). Electricity price forecasting in New Zealand: A comparative analysis of statistical and machine learning models with feature selection. *Applied Energy, 347*, 121446.

Acknowledgements

I would like to express my gratitude to my primary supervisor, Dr. Nuttanan Wichitaksorn, for his guidance and support throughout this project. Your willingness to help and your patience has been invaluable. I would also like to thank my secondary supervisor, Dr. Wenjun Zhang, for his feedback and support, as well as my mentor, Dr. Jiling Cao. Thank you to other members of the AUT postgraduate research group and to AUT staff members.

This project has been supported by a Callaghan Innovations fellowship grant (grant ID: GENED1801). I spent a significant amount of time during this PhD working at Genesis Energy in Auckland. I would like to thank Genesis Energy staff members, particularly Brice Valles and Julian McCree, for sharing their business and technical insights and guiding me through the project.

I would also like to thank the Editor-in-Chief, Associate Editor, and reviewers of Applied Energy and Journal of Forecasting journals for their feedback that greatly improved the manuscripts.

Finally, I would like to thank my family and friends for their support and encouragement throughout this journey.

Chapter 1

Introduction

1.1 Motivation and Contributions

Increasing demand for reliable energy across the globe has led to significant advancements in research pertaining to the energy markets, and particularly, the field of electricity price forecasting. The need for reliable energy is a product of the wave of deregulation which has encompassed global electricity markets since the 1980s.

The goal of deregulation of electricity markets globally was to increase competition and efficiency in the market, which would result in lower prices for consumers. However, deregulation has created new challenges for electricity markets, particularly highly renewable-focused markets like New Zealand. The inherent uncertainty and lack of reliability in using renewable energy sources, such as wind and hydroelectric power, has led to increased price volatility in the market. Without a proper forecasting framework, this volatility can create several inefficiencies in decision-making for generators and retailers in the electricity market, which further increases the cost of electricity for consumers.

The primary goal of this thesis is to develop a robust forecasting framework for electricity prices in New Zealand. This thesis presents three self-contained studies

which utilize statistical and machine learning techniques to forecast electricity prices in New Zealand. The first study implements Markov regime-switching (MRS) models to forecast daily electricity prices in New Zealand. The second study focuses on half-hourly electricity prices and utilizes mixed-frequency vector autoregression (MF-VAR) models to forecast prices. The third study provides an electricity price forecasting comparison of statistical models including GARCH and stochastic volatility, with machine learning models such as deep neural networks and recurrent neural networks.

Electricity price forecasting literature has grown significantly in the past two decades, with a wide variety of statistical and machine learning techniques being applied to electricity markets across the globe. However, there is a lack of literature on electricity price forecasting in New Zealand. The three studies presented in this thesis aim to fill this gap in the literature by providing a comprehensive overview of the application of statistical and machine learning techniques to the New Zealand electricity market. The techniques explored are competitive, and compared to state-of-the-art models being applied globally. Furthermore, a wide variety of fundamental variables related to the New Zealand electricity market are used in several contexts in the three studies. In this way, we are able to explore the impact of these variables on New Zealand electricity prices. The results of these studies provide novel insights into the characteristics of electricity prices, and the results of the studies can be used to make informed decisions by generators and retailers in the New Zealand electricity market.

1.2 The New Zealand Electricity Market

In New Zealand, the idea of deregulation stemmed in the late 1980s with the creation of a state-owned enterprise and corporatization of the electricity department in 1987. Prior to deregulation, the electricity sector was a state-owned monopoly. This structure provided a reliable supply of electricity, but at the cost of efficiency and

innovation, and doubts in the ability to meet growing demands. In 1993, the Energy Companies Act separated a singular enterprise for power generation into three separate entities. Furthermore, a company called Transpower was created to be responsible for the national electricity grid in New Zealand. This marked the beginning of structural changes aimed at introducing competition in the market. Later, the 1998 Electricity Industry Reform Act forced the separation of lines and energy businesses. This change ensured that the act of transmission of electricity was independent of the generation and retailing of electricity, fostering competition in the generation and retail sectors. Over the next decade, the electricity market in New Zealand was restructured with further changes to be a competitive market, with the Electricity Authority being established in 2010 to oversee the market. Further details on the chronology of deregulation in the New Zealand electricity market can be found in Ministry of Business, Innovation and Employment (2015).

Although the deregulation of the New Zealand electricity industry was intended to increase competition and lower prices, it has also led to price volatility. The market was subject to changes in the price of commodities on a global scale, particularly in the gas market, which had an effect on the cost of production. Consumers were frequently charged more for gas, which resulted in higher retail electricity rates. New prospects for renewable energy were created by the deregulation of the electrical industry, but there were drawbacks as well. The market was made more susceptible to the inherent volatility of these energy sources, despite the fact that increased competition encouraged innovation and the adoption of newer renewable technology.

For instance, because hydroelectric power is so prevalent in New Zealand, rainfall can have a significant impact on the availability of energy and, consequently, on electricity prices. Low water levels during dry periods can reduce hydroelectric power capacity, resulting in shortages and increased prices. The inherent variability and irregularity of many renewable sources also requires significant investment in technologies like battery

storage, demand response systems, and advanced grid management technologies to ensure the stability and reliability of the power system as the nation works to achieve its goal of generating all of its electricity from renewable sources.

The nationwide system of electricity transmission in New Zealand is known as the National Grid. This system is owned and operated by Transpower, a state-owned enterprise. The National Grid consists of 178 interconnected substations across all regions of New Zealand, including a transmission link between the North and South Islands through a high-voltage direct current (HVDC) link.

The challenges posed by New Zealand's deregulated and primarily renewable electricity market highlight the crucial need for reliable price forecasting. Accurate forecasting models need to be created in order to manage risk and guarantee steady operations for both generators and retailers in the electrical market. The New Zealand electricity sector is split into six distinct parts. These are generation, transmission, distribution, retail, consumption, and regulation. Generation companies generate electricity from various sources and inject electricity either into transmission or distribution lines. Presently, five companies dominate 92% of the total generation in New Zealand; these are Contact Energy, Genesis Energy, Meridian Energy, Mercury Energy, and Manawa Energy. Transmission of energy through the nationwide transmission network is managed solely by Transpower. Distribution companies operate medium and low-voltage networks at several grid exit points (GXPs) on the National Grid. There are currently 29 distribution companies, each serving a set of geographic locations. Retail companies are responsible for purchasing electricity from generators and on-selling to consumers. Numerous companies in New Zealand are in-charge of both the generation and retailing of electricity, and are commonly known as gen-tailers. 95% of the retail sectors is dominated by five generation companies; these are Contact Energy, Genesis Energy, Mercury Energy, Meridian Energy, and Trustpower. In regards to consumption, the industry sector leads in electricity consumptio, with approximately 44% of total usage,

followed by residential buildings at 33% and service sector buildings at 23%. The primary risk of electricity price fluctuations is borne by the generators, who are exposed to the volatility of the wholesale electricity market. Generators can manage this risk by entering into contracts with retailers and using the futures market to hedge against price fluctuations. Still, the risk of price fluctuations remains, and the ability to forecast prices accurately is crucial for managing this risk.

Electricity in New Zealand is traded at a wholesale level in a spot market through the New Zealand Electricity Market (NZEM). The NZEM is a real-time market that operates on a half-hourly basis. The market is operated by the Electricity Authority, which is responsible for ensuring that the market operates efficiently and fairly. The NZEM is a spot market, where prices are determined by the intersection of supply and demand. The price of electricity is determined by the marginal cost of the most expensive generator required to meet demand. Trading takes place at approximately 295 regional nodes on the National Grid every half-hour. Generators make offers to supply electricity, while retailers and major industrial users make bids to buy electricity. The market is cleared by the market operator, who matches supply and demand at each node and determines the price of electricity. The price of electricity can vary significantly between nodes, depending on factors such as the availability of generation, the level of demand, and the capacity of the transmission network. Bids and offers start 36 hours prior to actual real-time consumption. Up to four hours before the trading period, new forecasted prices are provided to guide market participants. During each half-hour period, Transpower publishes a real-time price every five-minutes and a time-weighted 30-minute average price. However, these prices are guidelines as the ex-post spot price is determined normally around 10:00 the following day, which reflects the true cost of electricity.

Statistical methods have proven to be a highly effective tool in electricity price forecasting. These methods can provide valuable insight into future price movements by

analysing historical data and identifying patterns. Time-series models, for example, can capture temporal patterns in price changes, whereas regression models can illuminate the relationship between prices and various influencing factors such as demand, supply, and weather conditions. Electricity price forecasting literature is filled with unique applications of statistical and machine learning techniques for global markets. Weron (2014); Hong et al. (2020); Lago et al. (2021) provide comprehensive reviews of advances in electricity price forecasting since the early 2000s.

The use of statistical techniques to forecast electricity prices can be particularly challenging and beneficial for a country like New Zealand, where renewable energy plays such a significant role. Due to the reliance of renewable energy generation on environmental variables, such as rainfall for hydroelectric power and wind speed for wind power, the accuracy of statistical models can be increased by incorporating meteorological forecasting data.

1.3 Thesis Structure

This thesis presents three self-contained forecasting studies based on techniques utilized in global electricity markets, as well as novel insights and results for forecasting electricity prices in regions across New Zealand. Each study was written as a manuscript and submitted to a peer-reviewed journal. Presently, the manuscripts in Chapter 2 and 4 have been published, and the other manuscript is being revised for submission to a journal. Details of the publications are available in the Publications section above. The manuscripts are presented in the order they were written, and each chapter contains an introduction to the study and a literature review, followed by the models and methodology, and concluding with the results and discussion.

The study presented in Chapter 2 implements Markov regime-switching (MRS) model framework for forecasting daily New Zealand electricity prices. In this study,

we focus on daily electricity spot prices, primarily due to its relevance in literature surrounding MRS models which leads to ease of comparison. Furthermore, many studies focus largely on daily prices particularly when making mid-term to long-term forecasts, for example, beyond one month. The choice between forecasting daily and half-hourly prices are also influenced by risk-management practices in the electricity market. Generators and retailers often hedge against price fluctuations using futures contracts, which are typically aggregated to a daily level, depending on the contract. Furthermore, building forecasting models for daily prices is an easier task than half-hour models, due to less significant jumps and volatility in the time-series. In this study, we obtain daily electricity prices as a simple average of half-hourly prices.

We assess the predictive capabilities of multiple models within the MRS framework and the extreme value theory (EVT) when applied to daily electricity prices in New Zealand's electricity market. The MRS approach encompasses up to five different regimes, including time-varying transition probabilities and including external market factors. To estimate parameters, we employ Hamilton's filter in conjunction with maximum likelihood estimation. We also explore the EVT peaks-over-threshold (EVT-PoT) framework and explore its relationship with the MRS models.

By building predictions for diverse market conditions, we find that the MRS models closely mirror true price densities when the market is steady. Interestingly, the simpler EVT-PoT model demonstrates robust performance, which we believe stems from its use of the generalized Pareto distribution to capture price spikes. While this study incorporates quarterly information about generation fuels, there was not a large focus on external variables. We believe that the inclusion of more significant features could improve the forecasting performance of the models. This led to the second study in this thesis presented in Chapter 3.

The study in Chapter 3 focuses on half-hourly electricity prices rather than daily. We focus our study on half-hourly prices since it reflects the complex price formation

process during intra-day trading. We want to analyze the behaviour of this price formation process with respect to current and past information about several weather variables. Furthermore, since half-hourly prices are the building blocks for lower-frequency prices, an ideal half-hourly price model can be implemented to characterize daily prices as well. Importantly, we have found that complex models generally are not capable of producing reliable forecasts for half-hourly prices, as they are more prone to spikes and exhibit significantly higher volatility. For this reason, modelin half-hour prices require several relevant external variables to explain these characteristics, rather than complex models. The main theme of this study is the use of mixed-frequency models to predict half-hourly prices using external variables available with hourly granularity. In this study, a mixed-frequency vector autoregression (MF-VAR) framework is employed to provide point forecasts of half-hourly electricity prices using several weather variables and electricity demand. Furthermore, we compare this approach to the reverse unrestricted mixed-data sampling (RU-MIDAS) model. We show that this model can be augmented into a VAR specification, which we denoted the RU-MIDAS-VAR model. This study also features a comparison between several parameter estimation techniques, in particular the least-squares method, Gibbs sampling, and mean-field variational Bayes (MFVB). MFVB approximates the posterior distribution of the parameters by minimizing the Kullback-Leibler divergence between the true posterior and a simpler distribution. Despite being an approximation method, we find MFVB estimates to be comparable to those obtained using Gibbs sampling, which is a more computationally intensive method.

Since we are primarily interested in forecasts for peak periods, we separate forecasts for peak and off-peak periods in a day. Our predictions, which take into account peak and off-peak data, demonstrate that demand and weather variables can replicate some significant characteristics of electricity prices. We also discover that the MF-VAR and RU-MIDAS-VAR models produce comparable forecast outcomes. As a

dimension-reduction and variable-selection method, the LASSO, adaptive LASSO, and random subspace regression are used to enhance forecasts. Random subspace methods succeed for large parameter sets while the LASSO significantly improves our forecasting outcomes in all scenarios. Seeing the success of LASSO with electricity related features stemmed the idea for the third study presented in Chapter 4.

Chapter 4 presents a study which focuses on the comparison between statistical and machine learning models for predictions in the New Zealand electricity market with a key interest in obtaining a large number of features, and using several feature selection methods to select the best subset which improves forecast results. In the last decade, many state-of-the-art machine learning models and methods have been proposed for electricity price forecasting. A general view in literature is presented that machine learning models outperform statistical models. However, we found that this is not always true. In fact, most comparisons in literature between the two classes have not been fair. In particular, while machine learning models undergo extensive feature selection and cross-validation, we found that statistical models did not receive the same treatment. As such, this study presents a fair comparison between the two classes of models.

The focus is on generalized autoregressive conditional heteroskedasticity (GARCH) and stochastic volatilty (SV) models for the statistical class, and the long short-term memory (LSTM), gated recurrent unit (GRU), extreme gradient boosting (XGBoost), deep neural network (DNN), and LASSO-estimated autoregressive model (LEAR) for the machine learning models. The latter two are considered benchmark models in our study. A key aspect of this study is the inclusion of a diverse set of explanatory features in all models. This includes, several weather variables, electricity demand, derivatives prices, among other features. With the exception of the LEAR model, all models are paired with feature selection techniques, including LASSO, mutual information, and recursive feature elimination.

A key finding in this study is the effectiveness of GARCH and SV models and their t-distribution variants when paired with feature selection techniques, including LASSO, mutual information. In most scenarios considered, these models outperform machine learning models, including the benchmark models, when coupled with LASSO feature selection.

Finally, Chapter 5 concludes the thesis with a final discussion on the motivations and contributions of this thesis, as well as future research directions for the individual studies.

Chapter 2

Analyzing and Forecasting Electricity Price using Regime-Switching Models: The Case of New Zealand Market

2.1 Introduction

Due to electricity non-storability, prices exhibit extreme volatility as a result of ever-changing market conditions. The real-time pricing mechanism on the national grid forces electricity prices to continuously vary as a function of instantaneous demand and supply driven by various external factors such as time, weather conditions, and the cost of electricity-generating fuel. The chaotic nature of weather conditions brings extreme volatility and spikes to electricity prices, more so in markets such as New Zealand, where hydro storage is a large fuel source and relies heavily on rainfall in specific regions. As New Zealand shifts away from fossil fuels and increases wind-farms production towards a 100% renewable energy scheme, the uncertainty of electricity generation increases as well, and as a result, so does volatility in prices. Understanding the behavior and forecasting of electricity prices with respect to external market conditions, such as

weather variables, fuel costs, and generator outages are equally important for market participants.

In this study, we fill a gap in the electricity price forecasting literature by applying the Markov regime-switching (MRS) framework of Hamilton (1989) to forecast daily electricity prices in New Zealand. We further explore the case of multiple regimes and time-varying transition probabilities. The MRS framework provides an ideal quantitative class of models to capture the distribution of actual electricity prices due to its ability to replicate various fundamental market conditions. Daily electricity prices in common literature are defined as the average spot price of electricity throughout the day. We follow this methodology and derive daily prices as the average of half-hourly prices throughout the day. We focus on daily electricity spot prices in this study, primarily due to its relevance in literature surrounding MRS models which leads to ease of comparison. The effectiveness of MRS models diminishes with half-hourly prices due to the significantly larger prices spikes and volatility characteristics in higher freqeuency data. Our focus in this study is on comparing density forecasts for one-month ahead prices, and as such, daily prices are suitable to work with.

Regime-switching models have gained a lot of traction for electricity price modeling since the early 2000s. Huisman & Mahieu (2003) initially applied the MRS framework to electricity prices in the Dutch, German and UK power markets. Weron et al. (2004) compare the performance of regime-switching models with jump-diffusion models and suggest that regime-switching models obtain a better fit to the Nordic power market. Bierbrauer et al. (2004) apply various two-regime specifications to prices from the Nordic power market and consider spike regimes with Gaussian, log-normal and Pareto distributions. They found that, while the simulated trajectories were satisfactory with these specifications, the simulations produced significantly more spikes than what was observed in empirical prices. See De Jong (2006), Mari (2008), Weron (2009), Janczura & Weron (2010), Paraschiv et al. (2015) and Mosquera-López, Manotas-Duque & Uribe

(2017) for more applications of the regime-switching model in global electricity markets. In this study, we estimate the MRS models using the Hamilton filter in conjunction with maximum likelihood, see Kim (1994) and Janczura & Weron (2012).

In our MRS framework, we incorporate external variables in the form of electricity generation fuel usage. Our primary focus is on the information held by the amount of electricity generated by various fuel sources in New Zealand, as such, we omit other popular variables, such as weather information and fuel prices. For electricity price forecasting using these factors, refer to Hagfors et al. (2016), Mosquera-López, Uribe & Manotas-Duque (2017), Paraschiv et al. (2014) and L. Wang et al. (2022).

A natural extension of MRS models is the application of time-varying transition probabilities (TVRS), introduced by Diebold et al. (1994). Mount et al. (2006) investigate the capability of MRS models to capture the volatility dynamics of the US PJM power market and show that TVRS models can capture volatility behaviour, if specified correctly. Becker et al. (2007) use a TVRS model with a generalised beta distribution to replicate the skewness observed in empirical prices and suggest that the transition of regimes is largely predictable by incorporating temperature and system load variables. Bazzi et al. (2017) apply the TVRS framework in a macroeconomic setting. Kanamura & Ōhashi (2008) study the transition probability matrix and incorporate demand and supply as factors. Their results suggest that transition probabilities should never be constant as they depend on both the current demand to supply-capacity ratio. Following this, we vary the transition probabilities as a function of demand and electricity generation using various fuels.

Another class of models similar to those of MRS models, are extreme value theory models, in particular the peak-over-threshold (EVT-PoT) model. Extreme value theory is the study of modeling probabilities of rare events and is popular in risk-management applications. While the usage of EVT-based models for electricity markets is sparse, some research shows that they are well-suited for modeling extreme tails of electricity

prices, see Byström (2005) and Paraschiv (2013). Paraschiv et al. (2016) apply the EVT-PoT method to the EPEX power market and study the price threshold beyond which the tail is considered to be extreme. Their results show realistic forecasts for the tail quantiles in electricity prices. To our knowledge, there has not been a study of the relationship between EVT-PoT and the MRS model. Both are considered mixture models, however, the switching mechanism varies. In this study, we form a relationship between the EVT-PoT model and the 3-regime MRS model using their likelihood functions. We show that the EVT-PoT model is a special case of the 3-regime MRS model when the transition probabilities are no longer Markovian and two of the regimes follow the generalized Pareto distribution.

While the MRS framework has seen significant usage for electricity price forecasting, newer models based on machine learning technology have become increasingly popular in the last decade. Meng et al. (2022) propose an LSTM network, which uses an attention mechanism to study the effect of renewable energy fuels on electricity price prediction for day-ahead prices in the Danish electricity market. Tschora et al. (2022) provide a comparison of several well-known machine learning models, such as support vector machines regression, random forest regression, deep neural network, and convolutional neural network while discussing the usage of LIME and SHAP techniques for interpreting these blackbox models. They predicted day-ahead prices for French, German, and Belgian electricity markets. Machine learning models are capable of significantly outperforming traditional models in short-term price prediction. However, our focus in this study is on the mid-term distribution of prices and the capability of replicating spikes, i.e., the upper-tail distribution of prices where we found our proposed models perform better than those of machine learning. As such, we rely on traditional models for ease of interpretation. That said, with the growing interest in machine learning concepts, their usability in electricity price distribution analysis is worth trying but we leave this as future work.

The contributions of our study are as follows. Firstly, we implement the MRS framework into an unobserved environment, i.e., the New Zealand electricity market. We provide a thorough study of the usefulness of multiple regimes in a highly renewable environment. New Zealand electricity price forecasting literature is sparse, and therefore this study provides useful results for market practitioners. We further incorporate time-varying transition probabilities (TVRS) as done by previous literature, however, we model the time-varying mechanism as a function of demand and fuel generation data. We believe incorporating several external variables into the time-varying mechanism will build a more realistic foundation to the theoretical models, as price behaviour is fundamentally affected by these variables. We provide in-sample and out-of-sample forecasts for various market scenarios.

Secondly, we implement the EVT-PoT model into an electricity price forecasting study. The EVT-PoT model has been applied to electricity markets to study extreme price behaviour, see Paraschiv et al. (2016). However, this framework has not been specifically tasked to forecast electricity prices altogether. In this study, we compare the forecast results of the EVT-PoT and MRS models. We also show that the EVT-PoT model is a special case of the 3-regime MRS model with certain restrictions.

In addition, we explore two seasonality extraction techniques that are common in prior literature. Namely, the trigonometric and dummy variable techniques. We show that the latter approach is far better at estimating and extracting seasonality from electricity prices in the New Zealand market.

Our findings suggest that the 3-regime TVRS and EVT-PoT generally provide the best forecast density fits. Additional regimes do not improve forecasts despite their increased complexity. The EVT-PoT model performs reasonably well compared to the MRS and TVRS models, despite its lack of complexity. Though our proposed models perform equally well, the EVT-PoT seems to replicate the higher order moments of empirical price better than the other models. This results in a better density fit,

particularly at the tails. We can attribute this to the use of the generalized Pareto distribution (GPD) to model the tails.

The organization of this chapter is as follows. Section 2.2 examines two seasonality extraction techniques and their usefulness. Section 2.3 details the model-building and parameter estimation process for the MRS models, time-varying transition probability (TVRS) models, and the extreme value theory peak-over-threshold (EVT-PoT) model. Section 2.4 performs in-sample analysis and out-of-sample forecasts. Section 2.5 concludes by providing a discussion of the results and potential future work.

2.2 Seasonality Estimation

This section examines the trigonometric and dummy-variable seasonality models and justifies our use of the latter model. Lucia & Schwartz (2002) were the first to incorporate seasonality in an electricity spot price model as a deterministic function of time, focusing on the Nordic electricity market, and suggested two approaches: dummy variables and a trigonometric function. The first approach uses dummy variables to differentiate weekdays from weekends and holidays, and to differentiate months of the year, see Huisman & Mahieu (2003), Knittel & Roberts (2005), Borovkova & Permana (2006), Huisman et al. (2007) and Higgs & Worthington (2008), among others. The second approach allows seasonality to vary throughout the year in a continuous manner. This specification was extended by Geman & Roncoroni (2006) to incorporate an annual and semi-annual cycle. See Cartea & Figueroa (2005), Bierbrauer et al. (2007), Janczura et al. (2013), Mayer et al. (2015), Borovkova & Schmeck (2017) and Bennedsen (2017), for more details on this approach.

2.2.1 Price Characteristics

The New Zealand electricity national grid consists of several nodes, each with their own location-based electricity prices. We have chosen the Otahuhu node in the North Island of New Zealand, as it encompasses a location with high demand for electricity. Otahuhu is also one of two nodes whose price is referenced for derivatives contracts on the Australian Securities Exchange (ASX). Our data set consists of daily electricity prices from the Otahuhu node between January 01, 2008 and December 31, 2018.

Figure 2.1 shows the daily spot price of electricity. Notice the propensity of spikes in this figure. There are clearly several outliers in this dataset. However, there is a large spike up to nearly $1200/MWh. This outlier is strictly due to spot trading error and we do not attempt to model such human errors. However, we do not remove this

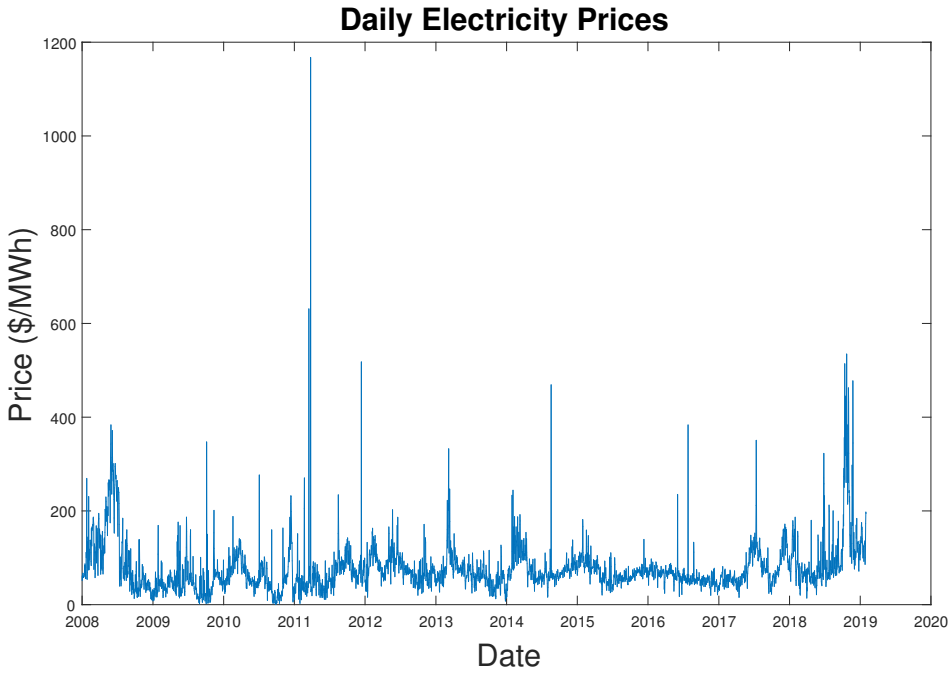


Figure 2.1: Daily electricity spot prices from the Otahuhu node in New Zealand (January 01, 2008 - December 31, 2018)

outlier from the training set. The regime-switching framework, given ideal parameter estimates, is capable of replicating all other spike behaviour. A characteristic that is not immediately obvious by observing Figure 2.1 is seasonality in electricity prices. Figure 2.2 displays the autocorrelation function (ACF) of electricity log-prices. We observe peaks in the ACF, most prominent at lags of multiples of 7, indicating the weekly seasonality. The peaks justify our proposed models of MRS and EVT-PoT. We also notice the long-term memory in electricity prices as observed by the persisting autocorrelation, as studied by Haldrup & Nielsen (2006) for the Nordic electricity market.

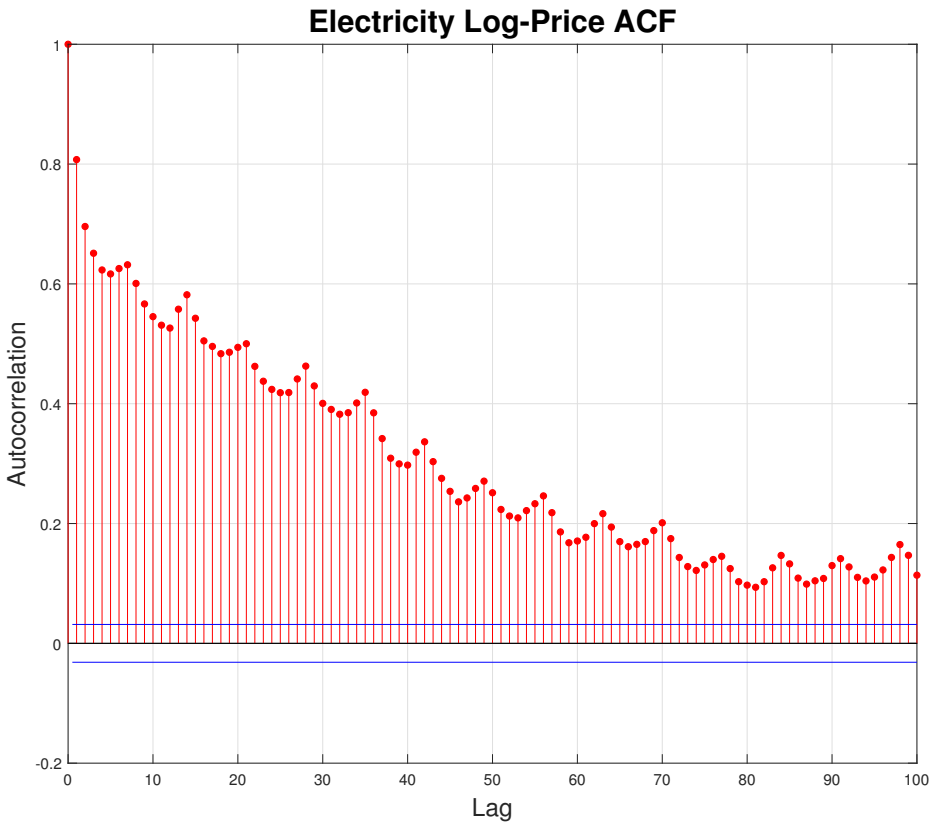


Figure 2.2: Autocorrelation of log-prices

2.2.2 Extracting Seasonality

Our electricity spot price models are broken down into two components given by the equation

$$E_t = \exp\{(s_t + y_t)\}, \tag{2.1}$$

where E_t is the spot price of electricity, s_t is a deterministic seasonal time-series and y_t is the stochastic time-series. In the model-building stage, our first objective is to estimate a seasonality function s_t using the electricity log-prices, after which we remove seasonality from the log-prices to obtain de-seasonalized electricity log-prices. The second objective is to build a stochastic model for de-seasonalized log-prices, which is denoted by y_t in Equation 2.1. Since electricity prices in New Zealand cannot be negative, we work with log-prices to avoid imposing additional constraints on our model parameters.

There are two main approaches to extracting seasonality from spot prices. The first is using a trigonometric function, and the second is using dummy variables and indicator functions. We discuss both methods, however, we use the latter because our results indicate it is superior at extracting seasonality.

Trigonometric Function

A trigonometric function for seasonality has the basic form

$$\begin{aligned} s_t =& a + b_1 \sin(2\pi t) + b_2 \cos(2\pi t) + b_3 \sin(4\pi t) + b_4 \cos(4\pi t) \\ &+ b_5 \sin(104\pi t) + b_6 \cos(104\pi t), \end{aligned}$$

where a is the mean seasonality level, and b_1 to b_6 are coefficients effecting the amplitude of each sub-function. This model is a combination of trigonometric functions

superimposed together with phases selected such that there are annual, semi-annual and weekly repeating cycles. Parameters for this model are estimated using the least squares technique, see Table 2.1.

a	b_1	b_2	b_3	b_4	b_5	b_6
4.1672	0.1468	0.0043	-0.0038	0.0643	-0.0032	-0.0042

Table 2.1: Parameter estimates for trigonometric seasonality function

However, the model does not accurately capture the seasonality in electricity prices. The autocorrelation function of de-seasonalised log-prices still show a significant amount of seasonality. The trigonometric approach is also extremely unreliable at recognizing weekly cycles.

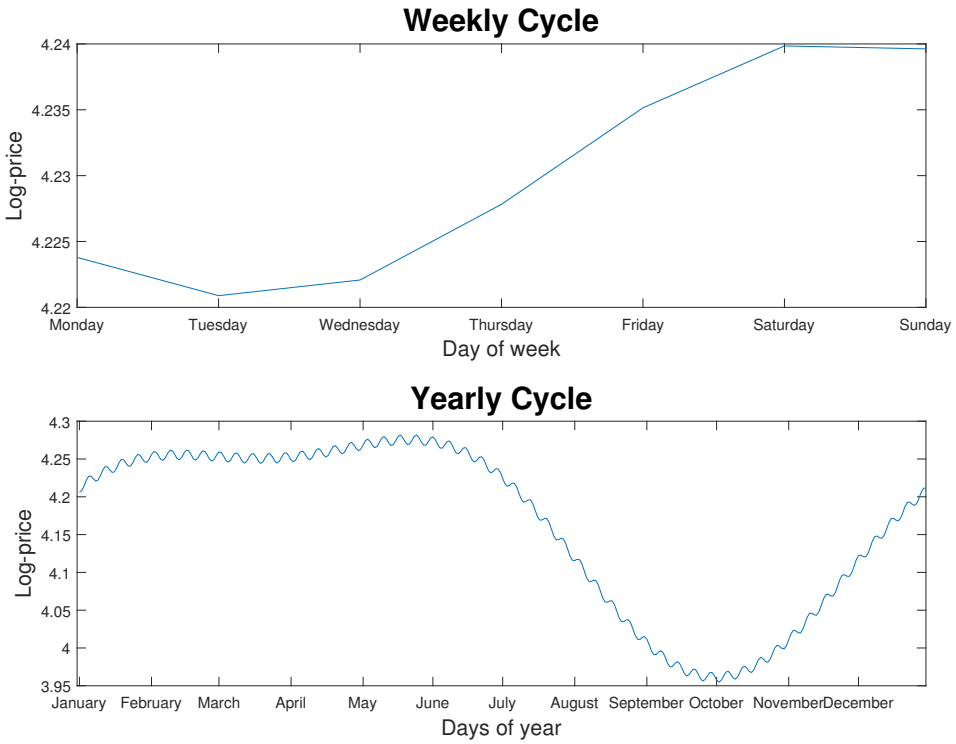


Figure 2.3: Weekly cycle in January and yearly cycle for trigonometric model

Figure 2.3 shows a weekly cycle in January on the top and the yearly cycle on the

bottom. We can see that the weekly cycle incorrectly depicts higher prices for weekends than weekdays. In the model, the incorrect estimation occurs during the months from October to June as the yearly cycle is trending upwards. This phenomenon occurs because the superimposition of multiple phases of incorrectly estimates the shorter cycles.

To further our reasoning to discard this model, electricity prices do not change smoothly in a continuous manner. Sudden changes in supply and demand, especially in shorter cycles, e.g., from Sunday to Monday, lead to rapid changes in seasonality. Essentially, the trigonometric model depicts a moving average of prices, rather than the actual seasonal effect.

Figure 2.4 compares the ACF of electricity log-prices and the ACF of de-seasonalized prices from the trigonometric seasonality model. The 7-lag cycles still persist in the ACF, implying that the seasonality has not been removed in the de-seasonalized prices.

Dummy Variable Approach

The dummy variables model relies on a sum of indicator functions with regression coefficients. The model has the form

$$s_t = \sum_{i=1}^{7} a_i D_{i,t} + \sum_{j=1}^{12} b_j M_{j,t}.$$

Here, $D_{i,t}$ is an indicator function that denotes the day of the week at time t, where $i = 1, 2, ...7$, and $M_{j,t}$ is an indicator function that denotes the month at time t, where $j = 1, 2, ..., 12$. Parameter estimates in Table 2.2 are again obtained using the least squares technique, however the estimation procedure is longer with this model due to a larger parameter-set.

The weekly and yearly cycles can be observed through the dummy variable approach in Figure 2.5. The weekly cycle is estimated well using this model and maintains its

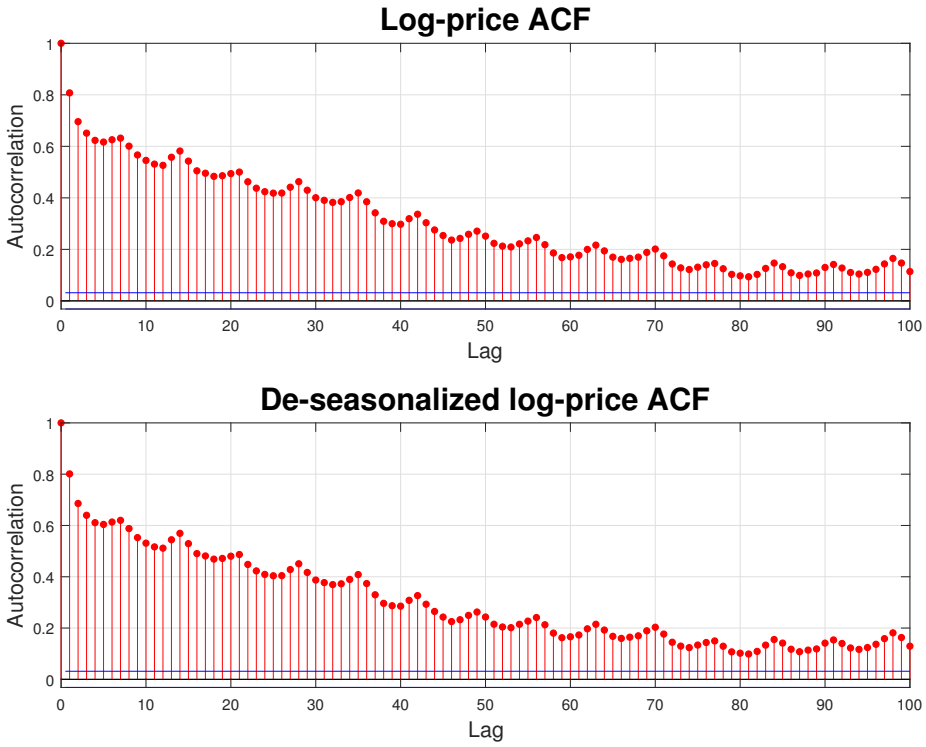


Figure 2.4: ACF of log-prices and de-seasonalized log-prices using trigonometric model

shape over the yearly cycle, unlike the trigonometric model. Furthermore, we observe that the weekly cycle is relatively more emphasized when observing the yearly cycle.

The plots in Figure 2.6 show the autocorrelation function (ACF) for actual log-prices and the de-seasonalized log-prices obtained from the dummy variables seasonality model. It is clear that the majority of seasonality has been removed in the de-seasonlized log-prices. These results justify the use of dummy variables over a trigonometric function for our seasonality model.

Month Effects			Week-day Effects		
Month	**Estimate**	**Standard Error**	**Week-day**	**Estimate**	**Standard Error**
January	1.4668	1.1546	**Monday**	2.7240	1.1887
February	1.6832	2.3476	**Tuesday**	2.7071	1.1529
March	1.7117	1.6571	**Wednesday**	2.7226	1.1846
April	1.6076	1.2946	**Thursday**	2.6663	1.0814
May	1.5756	1.0797	**Friday**	2.5237	1.1914
June	1.6302	1.3870	**Saturday**	2.4096	1.5724
July	1.6083	1.1657	**Sunday**	2.6513	1.0680
August	1.4722	1.1360			
September	1.2107	2.0379			
October	1.3014	2.0080			
November	1.5797	1.2007			
December	1.5566	1.0496			

Table 2.2: Monthly and weekly parameter estimates and standard error for dummy variables seasonal model

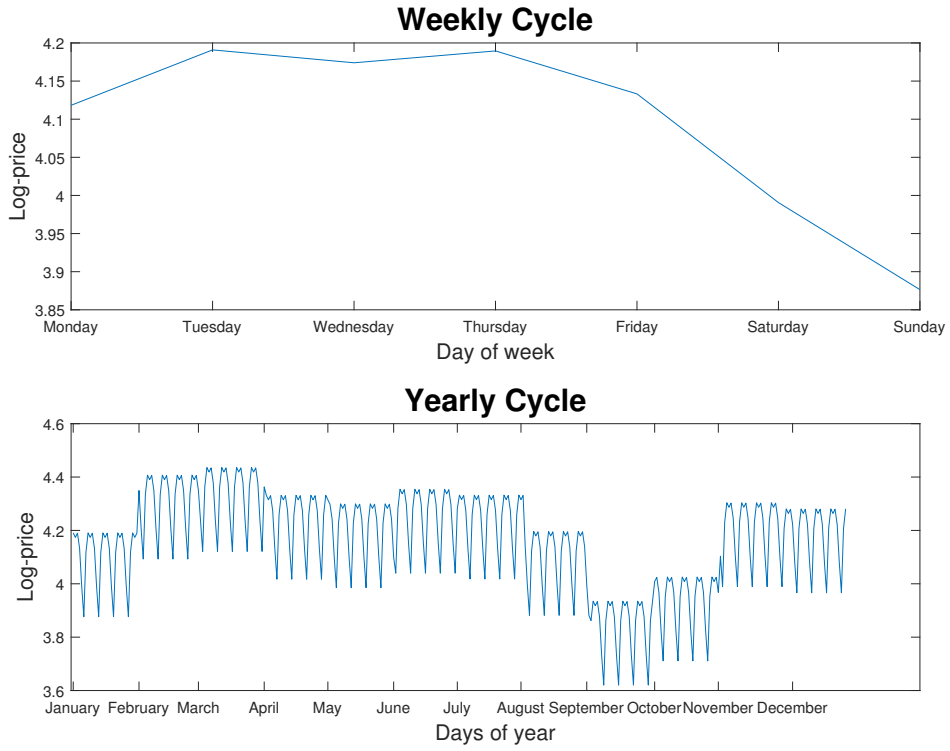


Figure 2.5: Weekly and yearly cycle for dummy variable model

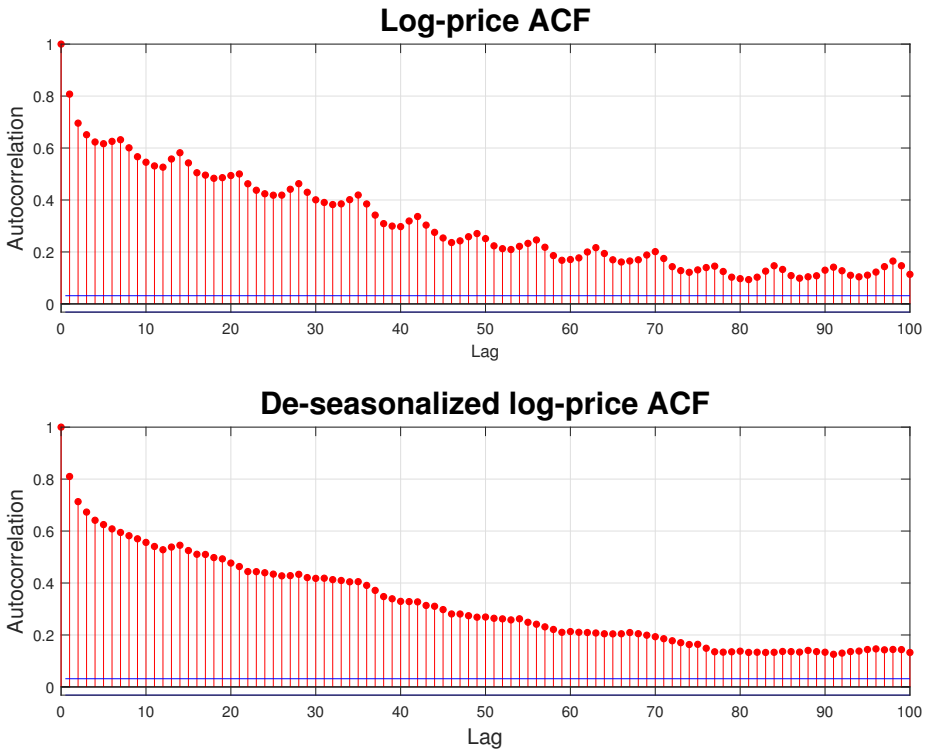


Figure 2.6: ACF of log-price and de-seasonalized log-prices using dummy variables model

2.3 Model and Estimation

In this section, we first introduce the standard 2-regime Markov regime-switching (MRS) model. We then introduce external market variables and incorporate them into our model. Next, we explain the parameter estimation method. The framework is then extended to include additional regimes and time-varying transition probabilities. We apply Hamilton's filter and use maximum likelihood for parameter estimation, which is able to adapt to the model extensions. Finally, we discuss the extreme-value-theory peaks-over-threshold (EVT-PoT) model and compare it with the MRS models where the EVT-PoT can be represented as a special case of the three-regime MRS model.

2.3.1 2-Regime MRS Model

The idea of regime-switching is extremely intuitive for electricity prices due to the fundamental nature of the market. The transition between regimes is governed by a Markov process, i.e. the probability that price is in a specific regime at the current time is dependent on the regime that the price time-series was in at the previous time step only. When demand and supply are in equilibrium, prices are generally stable with relatively low volatility, whereas, when a short-term mismatch of demand and supply occurs, prices respond erratically. To capture this effect, MRS models typically specify a base regime and a spike regime. A simple form of a 2-regime MRS model is defined by two stochastic processes, which can be represented as continuous-time Ornstein-Uhlenbeck processes.

Henceforth, we refer y_t to price for simplicity, however, y_t strictly denotes de-seasonalized log-price. Since our focus is on parameter estimation of these models, and ultimately providing simulations and forecasts of price, we prefer to work with

discrete-time models given by

$$y_t = \alpha_1(\mu_1 - y_{t-1}) + \sigma_1\epsilon_t \text{ when } R_t = 1,$$

$$y_t = \alpha_2(\mu_2 - y_{t-1}) + \sigma_2\epsilon_t \text{ when } R_t = 2,$$

where R_t denotes the regime at time t, and $\epsilon_t \sim N(0,1)$. Note that the error terms (ϵ_t) have an implicit assumption of independently and identically distributed, which is applied to all models we implement. While this discrete form is not exactly equal to the continuous form, the underlying properties of the model remain the same, while focusing on price y_t itself rather than price returns.

If the last known price y_{t-1} is less than μ the process has a positive skew forcing price upwards, and vice versa. In the above equations, the mean price level μ is a constant. The mean-reversion rate and volatility parameters may also be time-varying. Mean-reversion is the most fundamental aspect of commodities, including electricity, and is the basis for nearly every stochastic model in this field.

Considering the two-regime scenario, we would expect the base regime to have a lower mean-reversion rate α_1, and lower volatility σ_1 compared to the spike regime. The spike regime is expected to have a higher mean-reversion rate α_2 because prices are expected to revert back to the long-term mean μ_2 more quickly. The spike regime is also expected to have higher volatility σ_2 because prices are expected to fluctuate more erratically. Furthermore, the base regime would have a smaller long-term mean μ_1 compared to that of the spike regime.

The change of regimes is governed by a Markov process through a transition matrix

$$\mathbf{P} = (p_{ij}) = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix},$$

where $p_{ij} = \mathbb{P}(R_t = j | R_{t-1} = i)$.

In addition to the above model, we also incorporate some explanatory variables that might help explain the variation in the electricity price. The model regimes are then given by

$$\begin{aligned} y_t &= \boldsymbol{\gamma_1 x_{t-1}} + \alpha_1(\mu_1 - y_{t-1}) + \sigma_1\epsilon_t \text{ when } R_t = 1, \\ y_t &= \boldsymbol{\gamma_2 x_{t-1}} + \alpha_2(\mu_2 - y_{t-1}) + \sigma_2\epsilon_t \text{ when } R_t = 2, \end{aligned} \tag{2.2}$$

where $\epsilon_t \sim N(0,1)$. The 6×1 vector $\boldsymbol{x_t}$ incorporates the external factors, in particular

$$\boldsymbol{x_t} = \left[\text{Hydro}_t, \text{Geothermal}_t, \text{Wind}_t, \text{Coal}_t, \text{Gas}_t, \text{Demand}_t\right]^\top, \tag{2.3}$$

where Hydro_t represents the amount of electricity (GWh) generated using hydro at time t, and similarly for the other fuels. Demand_t is the total demand (GWh) in the North Island of New Zealand for electricity at time t. The parameters $\boldsymbol{\gamma_1}$ and $\boldsymbol{\gamma_2}$ are 1×6 vectors of regression coefficients for the explanatory variables. Due to the data availability restrictions the fuel sources in $\boldsymbol{x_t}$ have quarterly frequency, versus the daily for electricity prices and demand. The data for external variables is pre-processed such that $t-1$ refers to the quantity from the previous quarter for fuel sources and the previous day for demand.

2.3.2 Parameter Estimation

The estimation procedure for MRS models is not straightforward since the regimes are unknown and not directly observable. The log-likelihood function of a 2-regime MRS model is given by

$$\ln L = \sum_{t=1}^{T} \ln \sum_{i=1}^{2} \left(f(y_t | R_t = i, \theta) \mathbb{P}(R_t = i) \right), \tag{2.4}$$

where y_t is the de-seasonalized log-price of electricity and R_t is the regime at time t. Equation 2.4 is essentially the log-likelihood of the processes in both regimes weighted according to their probability of occurrence. The marginal probability of being in regime i, i.e., $\mathbb{P}(R_t = i)$ is not observed but inferences are made recursively using all past information.

Hamilton (1989) applied the Expectation-Maximization (EM) algorithm of Dempster et al. (1977) for estimation of regime-switching models. The EM algorithm calculates filtered and smoothed probabilities for each time step to calculate the likelihood based on some initial parameter set $\theta^{(0)}$. In our 2-regime MRS model the parameter set $\theta = \{\alpha_1, \alpha_2, \mu_1, \mu_2, \sigma_1, \sigma_2, \boldsymbol{\gamma_1}, \boldsymbol{\gamma_2}, p_{11}, p_{22}\}$ has a total of 20 parameters.

We follow the estimation methodology used in Perlin (2015) who applies Hamilton's filter for probability inference. Consider ϕ_t as the collective set of all information available at time t. The algorithm for the estimation of a two-regime model works as follows:

Step 1. Specify an initial probability for time t=0 We must first specify the initial probability for the process being in either regime, i.e, $\mathbb{P}(R_0 = i)$ for $i = 1, 2$. This probability can be arbitrarily set, e.g., we can choose $\mathbb{P}(R_0 = i) = 0.5$ for $i = 1, 2$. We can also use the unconditional probabilities

$$\mathbb{P}(R_0 = 1|\phi_0) = \frac{1 - p_{11}}{2 - p_{11} - p_{22}},$$
$$\mathbb{P}(R_0 = 2|\phi_0) = \frac{1 - p_{22}}{2 - p_{11} - p_{22}},$$

where p_{11} and p_{22} are transition probabilities specified in our parameter set θ.

Step 2. Calculate the filtered probability at next time step using information from previous time step Set $t = 1$ and determine the filtered probabilities of both regimes

given all information available at time $t-1$

$$\mathbb{P}(R_t = i|\phi_{t-1}) = \sum_{j=1}^{2} p_{ij}(\mathbb{P}(R_{t-1} = j|\phi_{t-1})),$$

where p_{ij} is again the transition probability, i.e. $p_{ij} = \mathbb{P}(R_t = j|R_{t-1} = i)$.

Step 3. Update the filtered probability using new information at time t The following formula is applied to update the filtered probability for time t

$$\mathbb{P}(R_t = i|\phi_t) = \frac{f(y_t|R_t = i, \phi_{t-1})\mathbb{P}(R_t = i|\phi_{t-1})}{\sum_{i=1}^{2} f(y_t|R_t = i, \phi_{t-1})\mathbb{P}(R_t = i|\phi_{t-1})},$$

where

$$f(y_t|R_t = i, \phi_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-(y_t - \alpha_i(\mu_i - y_{t-1}) - \boldsymbol{\gamma_i x_{t-1}})^2}{2\sigma_i^2}\right),$$

i.e., the densities of our two processes from Equation 2.2.

Step 4. Repeat the process for all observations We set $t = t + 1$ and repeat steps *2* and *3* until $t = T$ and we have calculated filtered probabilities for all time periods.

Step 5. Maximize the log-likelihood function We then proceed to maximize the log-likelihood function given in Equation 2.4 using any maximization algorithm and obtain the optimal parameter set.

Refer to Hamilton (1989), C.-J. Kim et al. (1999) or Perlin (2015) for detailed information on the algorithm. The same algorithm can be applied to extensions of our MRS framework, however, increasing the number of regimes or incorporating time-varying transition probabilities requires additional constraints to the probabilities in the transition matrix and increases computational times.

2.3.3 Multiple-Regime MRS Model

In this section, we extend the standard 2-regime switching model to the multiple regimes and generalise the models to k regimes. The purpose of this extension is to study the behaviour of electricity prices in New Zealand under multiple scenarios. We believe that the different regimes of the models and their behaviour provide information about the different characteristics of electricity prices.

A k-regime MRS model has the following transition probability matrix

$$\mathbf{P} = (p_{ij}) = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{bmatrix}. \tag{2.5}$$

The dynamics of each regime are defined as

$$y_t = \boldsymbol{\gamma_i x_{t-1}} + \alpha_i(\mu_i - y_{t-1}) + \sigma_i \epsilon_t \text{ when } R_t = i,$$

where $i = 1, 2, ..., k$ and the parameters are identical to those in Equation 2.2.

The parameter estimation methodology is identical to the 2-regime MRS model, however, the log-likelihood function has a general form

$$\ln L = \sum_{t=1}^{T} \ln \sum_{i=1}^{k} \left(f(y_t | R_t = i, \theta) \mathbb{P}(R_t = i) \right).$$

Furthermore, the parameters in the model increase in a quadratic manner. The quadratic increase arises because the number of transition probabilities needed to be estimated in a k-regime model is $k(k-1)$. Since there are 9 other parameters in each regime, a k-regime MRS model has $k(k-1) + 9k = k^2 + 8k$ parameters.

2.3.4 Time-Varying Transition Probability Regime-Switching Model

The time-varying regime-switching (TVRS) model extends the standard MRS framework by using logistic functions for the transition probabilities, making them dependent on explanatory variables. There is prevalent research in electricity markets suggesting that incorporating time-varying transition probabilities can significantly improve the model and its forecasting capabilities, see Diebold et al. (1994), Mount et al. (2006) and Bazzi et al. (2017).

We first consider the case of 2-regime TVRS models and then extend the model to incorporate k-regimes because implementing more than two regimes requires additional constraints on the transition probabilities so that the rows of the transition probability matrix may sum to 1.

The transition probability matrix for a 2-regime TVRS model has the form

$$\mathbf{P_t} = (p_{ij,t}) = \begin{bmatrix} p_{11,t} & p_{12,t} \\ p_{21,t} & p_{22,t} \end{bmatrix}, \tag{2.6}$$

where we use the logistic functions specification in Diebold et al. (1994) for the transition probabilities

$$\begin{aligned} p_{11,t} &= \frac{\exp(\boldsymbol{\beta_1 x_{t-1}})}{1+\exp(\boldsymbol{\beta_1 x_{t-1}})}, \\ p_{12,t} &= 1 - p_{11,t}, \\ p_{22,t} &= \frac{\exp(\boldsymbol{\beta_2 x_{t-1}})}{1+\exp(\boldsymbol{\beta_2 x_{t-1}})}, \\ p_{21,t} &= 1 - p_{22,t}, \end{aligned} \tag{2.7}$$

where $\boldsymbol{x_t}$ is the (6×1) vector of explanatory variables described in Equation 2.3. $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ are each (1×6) vectors of regression coefficients. The remaining dynamics of the model such as the individual regime processes remain the same as those in Equation

2.2.

We apply the same algorithm for parameters estimation as that in Section 4.2 however the parameter set changes to $\theta = \{\alpha_1, \alpha_2, \mu_1, \mu_2\, \sigma_1, \sigma_2, \boldsymbol{\gamma_1}, \boldsymbol{\gamma_2}, \boldsymbol{\beta_1}, \boldsymbol{\beta_2}\}$ for a total of 30 parameters. We no longer specify the initial transition probabilities but rather determine them as functions of $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ according to Equation 2.7.

We now consider the k-regime specification of time-varying transition probabilities. The transition probability matrix from Equation 2.6 changes to

$$\mathbf{P_t} = (p_{ij,t}) = \begin{bmatrix} p_{11,t} & p_{12,t} & \cdots & p_{1k,t} \\ p_{21,t} & p_{22,t} & \cdots & p_{2k,t} \\ \cdots & \cdots & \cdots & \cdots \\ p_{k1,t} & p_{k2,t} & \cdots & p_{kk,t} \end{bmatrix}. \tag{2.8}$$

Since the row-wise sum of the transition probability matrix must be equal to 1, we need to apply a multinomial logit specification for the transition probabilities, see Bazzi et al. (2017). For each row in Equation 2.8, we need to estimate $k-1$ probabilities. Therefore, for a k-regime specification there are $k(k-1)$ probabilities that need to be determined.

The transition probabilities are calculated as follows

$$p_{ij,t} = \frac{\exp(\boldsymbol{\beta_{ij}} \boldsymbol{x_{t-1}})}{1 + \sum_{j=1}^{k-1} \exp(\boldsymbol{\beta_{ij}} \boldsymbol{x_{t-1}})}, \tag{2.9}$$

where $\boldsymbol{\beta_{ij}}$ for $i = 1, ..., k$ and $j = 1, ..., k-1$ is still a 1×6 vector of coefficients. Equation 2.9 ensures that the $k-1$ probabilities in each row sum to less than or equal to 1. We then calculate the final probability in each row according to the equation

$$p_{ik,t} = 1 - \sum_{j=1}^{k-1} p_{ij,t},$$

for $i = 1, ..., k$, such that the sum of probabilities in each row is equal to 1.

The estimation methodology remains largely the same as the 2-regime TVRS case, with the only difference being that the transition probabilities are calculated using the multinomial logit function in Equation 2.9. The computational time for a k-regime TVRS specification increases significantly due to the increased number of parameters as k increases. For a k-regime TVRS model there are $6k(k-1)$ transition probabilities that need to be computed and $9k$ parameters from the regime processes. Therefore, a k-regime TVRS model has a total of $6k^2 + 3k$ parameters, significantly greater than its non-time-varying counterpart.

2.3.5 EVT Peaks-Over-Threshold Model

In this section, we discuss how an extreme value theory (EVT) specification can be used to model the tails of electricity prices. In particular, we study the EVT peaks-over-threshold (EVT-PoT) model to replicate the upper and lower tails of the prices and compare them with the MRS framework. We define peaks and troughs for the price values above and below a certain threshold i.e. we separate the dataset into three categories; the upper tail, the lower tail and the mid-section. We then model the mid-section using a normal distribution and the tails using a generalized Pareto distribution (GPD), where the GPD probability density function has the form

$$f(y) = \frac{1}{\sigma}\left(1 + \frac{\kappa(y-\theta)}{\sigma}\right)^{\left(-\frac{1}{\kappa}-1\right)}, \tag{2.10}$$

where $y \geq \theta$ when $\kappa \geq 0$, and $\theta \leq y \leq \theta - \sigma/\kappa$ when $\kappa < 0$. In Equation 2.10, σ dictates the scale, θ specifies the threshold, and κ controls the shape of the function.

Let u and $1-u$ denote the lower threshold and upper threshold quantile levels, respectively. Let $F(y)$ denote the cumulative distribution function of our electricity de-sesasonalized log-price dataset y. Then $F^{-1}(u)$ denotes the lower price threshold. In other words, all values of y, for $y \leq F^{-1}(u)$ are sub-categorized into a new dataset y_{low}.

Similarly, all values of y, for $y \geq F^{-1}(1-u)$ are sub-categorized into a new dataset y_{up}. The remaining prices in y are sub-categorized into a new dataset y_{mid}.

Parameter estimation of this model is relatively simple as it is a mixture model with pre-determined quantile levels. We obtain estimates using the maximum likelihood estimation technique for each of the sub datasets. The GPD log-likelihood function is maximized for $-y_{low}$ and y_{up}, and the Gaussian log-likelihood for y_{mid}. We transform y_{low} to $-y_{low}$ because the data must be positive.

With the model setting above, the EVT-PoT model can be represented as a special case of the 3-regime MRS model. This can be shown using the likelihood function of both models and applying certain constraints to the likelihood function of the MRS model to make it equivalent to the EVT-PoT model. The EVT-PoT model is a type of mixture model and therefore its likelihood function is defined as the sum of the likelihoods of each regime weighted according to the selected quantile level.

Mathematically, let $f(y_t|s_t = i)$ denote the density of the process y_t at time t given that the process is in regime i, where $i = 1, 2$ or 3 if the process is the lower tail regime, middle regime, or the upper tail regime, respectively. Then, the likelihood L of the EVT-PoT process is defined as

$$L = \prod_{t=1}^{T} \Big[f(y_t|s_t = 1)u + f(y_t|s_t = 2)(1-2u) + f(y_t|s_t = 3)u \Big],$$

where by denoting the lower tail quantile u and the upper tail quantile $1-u$, we get $\mathbb{P}(s_t = 1) = u$, $\mathbb{P}(s_t = 2) = 1-2u$ and $\mathbb{P}(s_t = 3) = u$ for all t, which forms a relationship between the MRS framework and the EVT-PoT framework, and can be re-written as

$$L = \prod_{t=1}^{T} \Big[f(y_t|s_t = 1)\mathbb{P}(s_t = 1) + f(y_t|s_t = 2)\mathbb{P}(s_t = 2) + f(y_t|s_t = 3)\mathbb{P}(s_t = 3) \Big], \tag{2.11}$$

where $\mathbb{P}(s_t = i)$ for $i = 1, 2,$ or 3 is the probability that the process at time t is in state i.

2.3.6 Relationship Between 3-Regime MRS and EVT-PoT Models

This section shows how the EVT-PoT Models can be nested into our 3-regime MRS framework. The 3-regime MRS model is similar to the EVT-PoT model, however, the regime-switching process in the Markovian model follows a Markov process rather than the pre-defined probability of switching in the EVT-PoT model.

The transition probability matrix for a 3-regime MRS model can be observed in Equation 2.5 for $k = 3$. In the transition matrix, p_{ij} denotes the probability of the process being in regime j at time t given that the process was in regime i at time $t - 1$. Mathematically, $p_{ij} = \mathbb{P}(s_t = j|s_{t-1} = i)$.

Let $f(y_t|s_t = i)$ denote the density of a 3-regime MRS process y_t at time t given that the process is in regime i, where $i = 1, 2$ or 3. Then the likelihood L of the process is

$$\begin{aligned} L = \prod_{t=1}^{T} \Big[& f(y_t|s_t = 1) \left[p_{11}\mathbb{P}(s_{t-1} = 1) + p_{21}\mathbb{P}(s_{t-1} = 2) + p_{31}\mathbb{P}(s_{t-1} = 3) \right] \\ & + f(y_t|s_t = 2) \left[p_{12}\mathbb{P}(s_{t-1} = 1) + p_{22}\mathbb{P}(s_{t-1} = 2) + p_{32}\mathbb{P}(s_{t-1} = 3) \right] \\ & + f(y_t|s_t = 3) \left[p_{13}\mathbb{P}(s_{t-1} = 1) + p_{23}\mathbb{P}(s_{t-1} = 2) + p_{33}\mathbb{P}(s_{t-1} = 3) \right] \Big]. \end{aligned} \tag{2.12}$$

In order to make Equations 2.11 and 2.12 equivalent, we need to remove the Markov property from Equation 2.12. This can be done by making the following constraints for

the transition probabilities

$$\begin{aligned} p_{11} &= p_{21} = p_{31}, \\ p_{12} &= p_{22} = p_{32}, \\ p_{13} &= p_{23} = p_{33}. \end{aligned} \tag{2.13}$$

These constraints enforce that the probability of going from regime i at time $t-1$ to regime j at time t is now independent of regime at time $t-1$. Therefore $\mathbb{P}(s_t = j|s_{t-1} = i) = \mathbb{P}(s_t = j)$ and $p_{ij} = p_j$ for $i, j = 1, 2, 3$.

Applying these restrictions to the likelihood function in Equation 2.12, we get

$$\begin{aligned} L = \prod_{t=1}^{T} \Big[& f(y_t|s_t = 1)\left[p_1\mathbb{P}(s_{t-1} = 1) + p_1\mathbb{P}(s_{t-1} = 2) + p_1\mathbb{P}(s_{t-1} = 3)\right] \\ & + f(y_t|s_t = 2)\left[p_2\mathbb{P}(s_{t-1} = 1) + p_2\mathbb{P}(s_{t-1} = 2) + p_2\mathbb{P}(s_{t-1} = 3)\right] \\ & + f(y_t|s_t = 3)\left[p_3\mathbb{P}(s_{t-1} = 1) + p_3\mathbb{P}(s_{t-1} = 2) + p_3\mathbb{P}(s_{t-1} = 3)\right] \Big]. \end{aligned}$$

Since $\mathbb{P}(s_{t-1} = 1) + \mathbb{P}(s_{t-1} = 2) + \mathbb{P}(s_{t-1} = 3) = 1$, we get

$$L = \prod_{t=1}^{T} \Big[f(y_t|s_t = 1)p_1 + f(y_t|s_t = 2)p_2 + f(y_t|s_t = 3)p_3 \Big], \tag{2.14}$$

where $p_i = \mathbb{P}(s_t = i)$, for $i = 1, 2,$ or 3. Equation 2.14 is equivalent to Equation 2.11, hence, showing that the EVT-PoT and 3-regime MRS models are equivalent if the constraints in Equation 2.13 are imposed. Therefore, the EVT-PoT model is a special case of the 3-regime MRS model.

2.4 Price Simulation and Forecasting Results

In this section, we provide simulation and forecasting results for the MRS, TVRS and the EVT-PoT models that we have discussed in Section 4. We compare summary statistics and density plots of empirical prices with simulated prices in order to determine which models are capable of capturing the distribution of empirical prices and replicating their characteristics. Furthermore, we perform out-of-sample forecasts for one-month time spans under different market conditions. We perform the Kolmogorov-Smirnoff test on our price forecasts and compare their densities with empirical prices.

2.4.1 Summary of In-sample Fits

	n	Log-Likelihood	AIC	BIC	Time (seconds)
2-Regime MRS	20	-280.85	601.70	727.67	49.81
3-Regime MRS	33	-178.42	442.84	630.69	114.66
4-Regime MRS	48	-146.15	388.30	690.63	192.16
5-Regime MRS	65	-92.61	315.22	724.63	311.52
2-Regime TVRS	30	-240.35	540.70	729.66	171.07
3-Regime TVRS	63	-168.82	463.64	860.45	601.14
4-Regime TVRS	108	-21.98	259.96	940.20	4655.84
5-Regime TVRS	165	-71.34	472.68	1511.90	17724.09
EVT PoT	8	-2243.50	4503	4553.39	4.37

Table 2.3: The log-likelihood, AIC and BIC values of model fits, as well as the parameter estimation computational time.

Table 2.3 provides the results from our model estimation procedure. We observe the log-likelihoods and compare the various model Akaike information criterion (AIC) and Bayesian information criterion (BIC), which are penalized-likelihood value, where n denotes the number of parameters in the model. We also observe the computational time (in seconds) of the estimation algorithm, which is a trade-off that often needs consideration.

We can see from Table 2.3 that the log-likelihood generally improves with the number of regimes in the model, as this can be expected due to the increase in parameters. The TVRS models outperform their MRS counterparts with the same number of regimes in terms of the log-likelihood value. While the models generally perform better in terms of AIC as the regimes increase, the BIC values become worse due to the quadratic increase in model parameters.

The EVT-PoT model is the least complex model in the table with only 8 parameters and a relatively simple estimation procedure. As a result, it has the lowest log-likelihood value across the models and the lower number of parameters are not able to compensate for the likelihood value as the AIC and BIC are also the highest values in the table. This is not to say that the EVT-PoT model performs worse in general, as we have chosen to study a relatively simple case of this framework. However, we are concerned with how the EVT-PoT model replicates the density of empirical prices, particularly in the upper and lower tails, so this result is not of much concern.

The algorithm computational time increases exponentially with the number of regimes, with time-varying models having significantly longer computational times than their standard counterparts, due to the larger number of parameters. Consideration of computational time largely depends on the practical application of the algorithm. For example, an electricity derivatives trader on the floor seeking forecasts would likely prefer a speedier estimation algorithm as new real-time information becomes available. On the other hand, long horizon forecasts will not rely as much on speed.

Table 2.4 provides summary statistics of price simulations from our models. The results are obtained from 1000 simulations for each model. While most MRS models are capable of replicating the actual mean reliably, the higher order moments do not provide good results. On the other hand, the EVT-PoT model is able to capture the higher moments reasonably well, with an almost identical skew level but slightly lower kurtosis than actual data. As we will observe from simulation figures, the MRS models

	Mean	Std	Skewness	Kurtosis
Actual Prices	76.73	54.87	4.70	55.15
2-Regime MRS	79.68	53.49	3.27	32.40
3-Regime MRS	80.99	58.29	5.24	83.40
4-Regime MRS	78.94	69.54	5.35	105.69
5-Regime MRS	107.36	33.39	8.95	129.61
2-Regime TVRS	72.62	46.40	2.22	15.31
3-Regime TVRS	78.25	53.02	2.51	16.10
4-Regime TVRS	76.24	53.42	4.52	65.38
5-Regime TVRS	78.24	69.67	6.45	131.27
EVT-PoT	77.44	59.09	4.71	46.03

Table 2.4: Summary statistics of first four moments of real price data and simulated price data

find it difficult to replicate the spike characteristics of actual prices correctly. Generally, spikes are too frequent, resulting in higher skew and kurtosis levels for models with more than three regimes. Meanwhile, when prices are hovering around the mean levels, the standard deviation is incredibly high.

2.4.2 Simulations and Density Analysis

In this section, we provide simulations of forecasted density plots with comparison to empirical density plots. We focus on the 3-regime standard and TVRS models and the EVT-PoT model for further analysis as the higher regime models do not perform significantly better for the complexity they introduce.

Figure 2.7 shows the probability density function of actual and simulated electricity prices for the 3-regime MRS and TVRS models and the EVT-PoT model. The time-varying model provides a relatively good fit to the empirical density function, however, both regime-switching models underestimate the lower tail of prices. In other words, the models are unable to accommodate scenarios where prices are consistently low with low

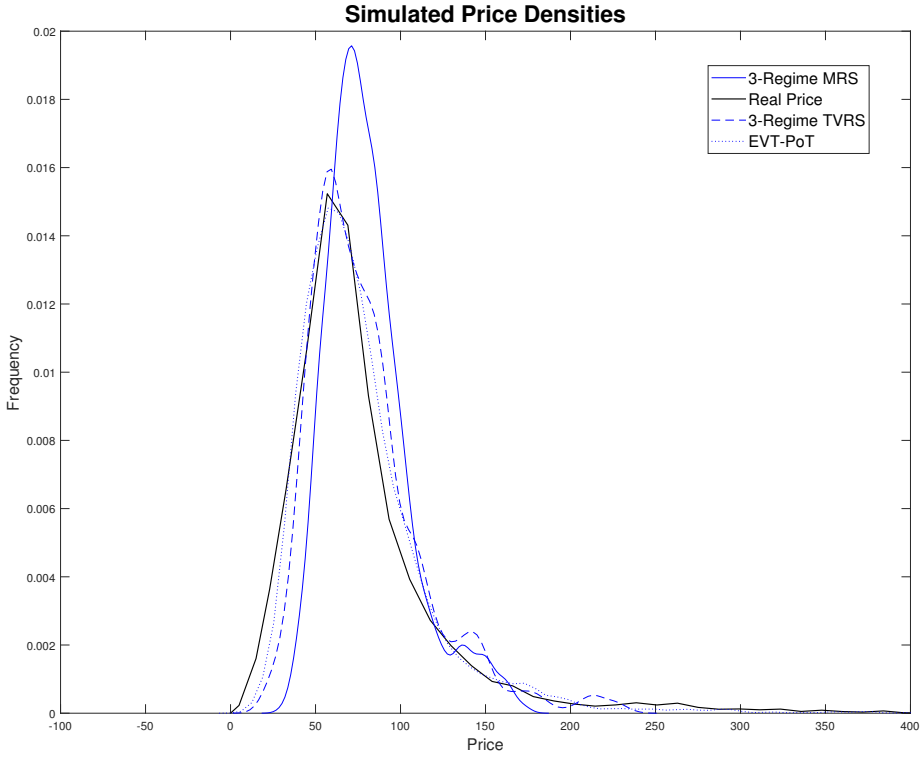


Figure 2.7: Probability density function of actual prices, 3-regime MRS model, TVRS model and EVT-PoT model.

volatility. The models apply too much emphasis on spikes resulting in underestimation of the lower tail and a slightly positive skewness. Nevertheless, both models capture the upper tail reasonably well. The density of the EVT-PoT simulated prices is much closer to the empirical density than the regime-switching models. The peak of the density function for EVT-PoT matches closely to the empirical density. Furthermore, the curvature to the right of the peak is almost identical, indicating that the upper tail of prices are replicated well. This leads us to believe that using GPD innovations within a regime-switching framework will likely improve their estimates and price simulations.

2.4.3 Out-Of-Sample Forecasts

In this section, we perform out-of-sample forecasts for the 3-regime MRS and TVRS models as well as the EVT-PoT model. We study the forecasting capabilities of each model under various market conditions in New Zealand by observing several different time periods. In particular, we want to observe forecasting results during several recurring market cycles. The first is the month of February, which is typically the period of the hydro accumulation phase for the coming winter months. January to March usually exhibit high prices and volatility if there is a lack of rainfall, despite being the warmest months of the year. As a result, we are interested in observing how the models are able to forecast this unique scenario. We would also like to study the peak winter periods around July, as once again prices are volatile during this period due to temperature conditions affecting demand. We also study the forecasting results during October 2018, a significant recent time period where prices were consistently high with significant volatility due to lack of rainfall and gas shortages. We also examine forecasts in October from previous years as a comparison to October 2018.

We perform 1000 simulations to obtain our forecasts and conduct the Kolmogorov-Smirnoff (KS) test to provide a quantitative result for the forecasts. We also discuss the root mean squared error (RMSE) and mean absolute percentage error (MAPE) values as another evaluation of forecast accuracy.

Since we perform out-of-sample forecasts, we re-run our estimation algorithm to incorporate all information from January 1, 2008 to one day before the forecast period.

Table 2.5 shows forecasting results for the three models under different market conditions. The KS test returns a decision for the null hypothesis that the empirical and simulated prices follow the same distribution. The result is 1 if the null hypothesis is rejected at the 5% significance level, and 0 if the null hypothesis cannot be rejected. The results for KS test shown in this table are an average from 1000 simulations and

		KS Test	p-value	RMSE	MAPE
February 2017	**3-Regime MRS**	**0.76**	**0.05**	35.91	52.51
	3-Regime TVRS	1.00	0.00	**23.77**	52.44
	EVT-PoT	0.99	0.00	46.74	**52.38**
February 2018	**3-Regime MRS**	0.36	0.24	38.70	67.62
	3-Regime TVRS	0.21	0.25	**35.84**	67.56
	EVT-PoT	**0.08**	**0.45**	53.97	**67.30**
July 2017	**3-Regime MRS**	**0.98**	0.00	84.68	**134.65**
	3-Regime TVRS	**0.98**	0.00	**83.24**	134.72
	EVT-PoT	1.00	0.00	86.73	134.68
July 2018	**3-Regime MRS**	0.66	0.05	51.33	80.72
	3-Regime TVRS	**0.08**	**0.25**	**46.58**	**79.81**
	EVT-PoT	0.15	0.23	54.65	80.91
October 2017	**3-Regime MRS**	0.63	**0.10**	**27.20**	57.77
	3-Regime TVRS	0.92	0.01	44.16	57.83
	EVT-PoT	**0.39**	0.09	70.43	**57.43**
October 2018	**3-Regime MRS**	1.00	0.00	**238.49**	**299.17**
	3-Regime TVRS	1.00	0.00	270.56	299.32
	EVT-PoT	1.00	0.00	268.35	299.35

Note: Bold fonts indicate better results.

Table 2.5: The average KS test and p-values from 1,000 simulations for each model. A lower KS test value is preferred, whereas a higher p-value is preferred.

lower values are preferred while the p-value column shows the average of corresponding p-values of the KS tests from the same set of simulations. A higher p-value is preferred and a value above 0.05 indicates the null hypothesis cannot be rejected. Precisely, we use the KS test strictly as a comparison among the models. In particular, we assess their capability to replicate the true price distribution, on average. As such, the test performed and the p-values shown are averages of 1,000 simulated forecasts. While the averaged p-values hold little meaning, the averaged KS test scores inform us of the proportion of simulations that return the ideal result. We also provide RMSE and MAPE results to measure the average differences between empirical and simulated prices. Note that, in this study, we focus on the forecasting results for February 2017 and 2018.

For the February 2017 test period, based on the KS tests including p-values, the 3-Regime MRS is the best performing model for the forecast distribution. However, the 3-Regime TVRS returns the least biased forecasts in terms of RMSE while the EVT-PoT is the best in terms of MAPE. Observing the forecast plots in Figures 2.8, we see that simulated prices in February 2017 are slightly overestimated, however, they are able to replicate the weekly seasonal cycle accurately. Prices in February 2018 were more erratic than in 2017, but the simulations are able to average out these deviations, as observed in Figure 2.9. Using the February 2018 as the out-of-sample data, the EVT-PoT still performs well in terms of KS tests, p-values, and the MAPE, followed by the TVRS from the RMSE.

For other test periods, the forecasting results are mixed. The TVRS seems to be the best model for July 2018 and 2018, both in terms of forecast distributions and biasedness. The MRS and the EVT-PoS are equally well for October 2017 while the MRS is the best for October 2018. Overall, in terms of the forecast distributions, the 3-regime MRS and EVT-PoT models generally provide better results. However, the 3-regime TVRS model has the lowest RMSE values in all scenarios except for October 2017 and October 2018.

Note that the EVT-PoT model performs relatively well, given that it is the least complex model and does not incorporate any time-varying parameters or external information about the market. The results strongly suggest extending the regime-switching framework to incorporate a heavy-tailed distribution such as the GPD. Furthermore, extending the EVT-PoT model to include auto-regressive behaviour, e.g., using a GARCH framework, may significantly improve its forecasting capability where we leave it for future research.

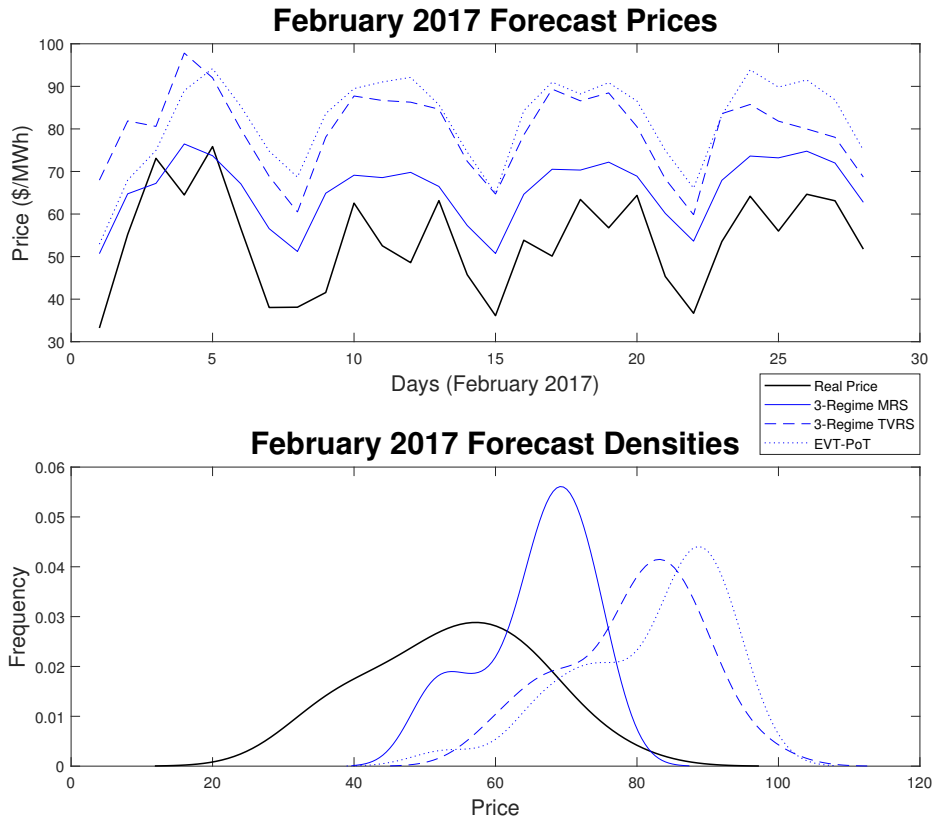


Figure 2.8: Comparison of actual price and forecast prices for February 2017.

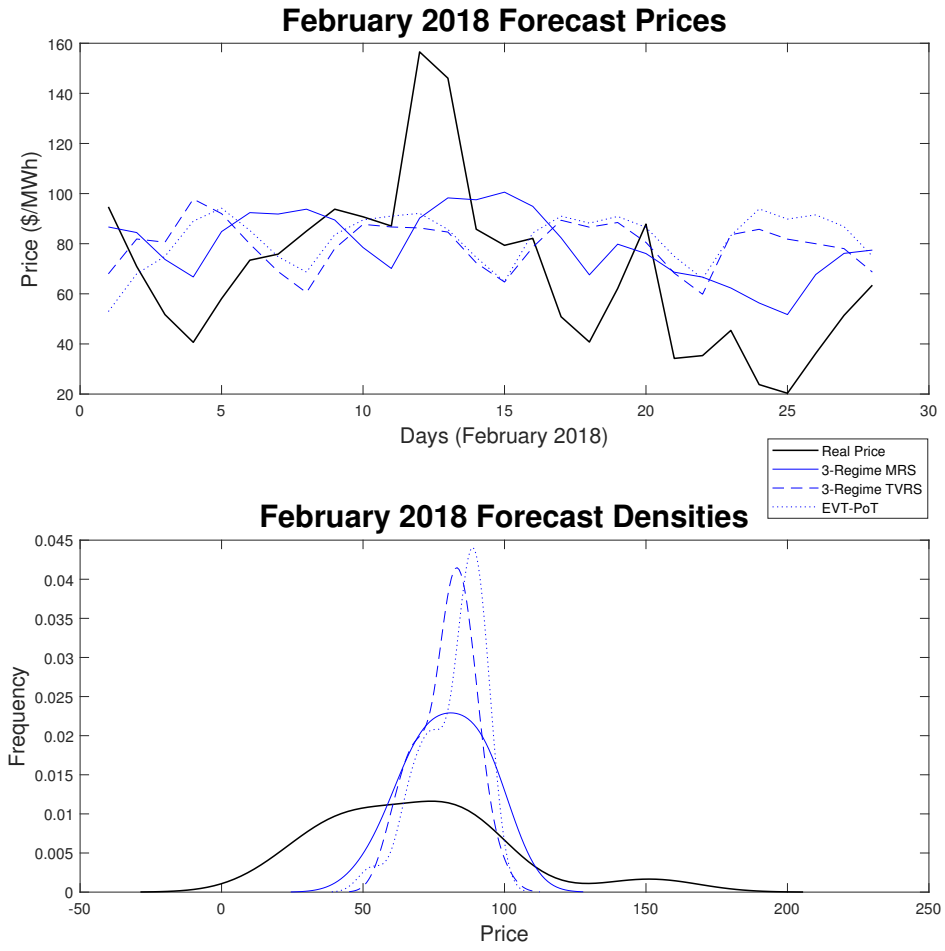


Figure 2.9: Comparison of actual price and forecast prices for February 2018.

2.5 Conclusion

In this study we study the forecasting capabilities of the Markov regime-switching (MRS) framework and its extensions, focusing on the New Zealand electricity market. We extend the framework to incorporate external market variables, add multiple regimes, and include time-varying transition probabilities (TVRS). We also discuss the peaks-over-threshold model (EVT-PoT) from the extreme-value-theory framework. We compare the EVT-PoT model to the standard 3-regime MRS model and show that the former is a special case of the MRS framework.

We apply the Hamilton's filter and use the maximum likelihood for parameter estimation. Our results show that the time-varying transition probabilities extensions generally perform better than their non-time-varying counterparts with the same number of regimes, in terms of the log-likelihood criterion. In-sample simulated prices depict several mini-spikes in the data leading to slightly higher variance in nearly every model compared to actual prices. The 3-regime MRS and TVRS models provide a reasonable in-sample fit to the density of empirical prices, however, they underestimate the lower tail. The EVT-PoT model outperforms the others in regards to empirical density in-sample fitting.

Finally, we study the out-of-sample forecasting capabilities of the 3-regime MRS and TVRS models and the EVT-PoT model under various market conditions. Particularly, we focus on the peak summer and winter months, and October 2018 where we obtain mixed results. The 3-regime MRS generally provides better density fits, however, the 3-regime TVRS and EVT-PoT usually have lower RMSE values. Extending the EVT-PoT model to incorporate auto-regressive behavior and external market information might be worth attempting as this may provide significantly better forecasting results but we leave it for future research.

There are several additional extensions as future work. The first is the incorporation

of time-varying parameters as we notice, in this study, that the models are not able to accurately capture the volatility persistence of prices. While we extend the MRS framework to include time-varying transition probabilities, using other time-varying parameters for the volatility, mean price levels and mean-reversion rate may produce better results. Furthermore, adding an auto-regressive component to the time-varying transition probabilities may be capable of improving forecasts by better capturing the persistence of these probabilities. The second extension is using non-Gaussian innovations in the MRS models. Several past studies use various heavy-tailed distributions for the innovations. Incorporating the generalized Pareto distribution into the MRS framework may provide better results.

Another extension is the use of better-suited external variables within the context of the models. In this study, we forecasted daily electricity prices and used information about quarterly generation mix. This was not an ideal choice as the generation mix fluctuates on a daily basis. However, this daily data was not easily accessible at the time. Furthermore, it has been shown in literature that implementing weather-related variables can significantly improve forecasting performance, particularly for highly-renewable markets. With this in mind, the study presented in the next chapter presents a mixed-frequency forecasting framework that incorporates hourly weather-related variables to forecasting half-hourly prices.

Chapter 3

Forecasting Half-Hourly Electricity Prices using a Mixed-Frequency Structural VAR framework

3.1 Introduction

Electricity in New Zealand is generated primarily through renewable sources, such as hydro, wind, and geothermal energy. Furthermore, New Zealand is committed to its path to a 100% renewable electricity market by 2030, with large wind-farm productions already in place and plans to build more, as well as a ban on offshore oil and gas exploration. As a result of these fundamental changes and the highly competitive nature of the market, electricity price volatility in New Zealand is likely to increase. In this study, we build electricity spot price models for half-hourly prices in the New Zealand electricity market. We focus our study on half-hourly prices since it reflects the complex price formation process during intra-day trading. We want to analyze the behaviour of this price formation process with respect to current and past information about several weather variables. Furthermore, since half-hourly prices are the building

blocks for lower-frequency prices, an ideal half-hourly price model can be implemented to characterize daily prices as well.

There have been several studies relating to the impact of weather variables on electricity markets worldwide. Huurman et al. (2012) analyze the predictive power of temperature, rainfall, and wind power in the Scandinavian electricity market. Their study is limited to forecasting day-ahead prices only. In our study, we observe the forecasting capabilities up to nine months ahead to capture multiple seasonal changes, with cross-validation to assess the robustness of our models. The explanatory variables we consider are electricity load (demand), temperature, rainfall, solar radiation, air pressure, humidity, wind speed, and wind direction. Suomalainen et al. (2015) analyzes the correlation of wind and hydro resources with electricity price and demand in New Zealand, with a focus on analyzing optimal wind power generation in different parts of the country. See Son & Kim (2017) and Mosquera-López, Uribe & Manotas-Duque (2017) for more applications of weather variables in electricity markets. Weron (2014) provides an excellent review of electricity price forecasting literature.

While our price and demand observations have half-hourly frequency, our weather variables are observed every hour. As a result of this frequency mismatch, we employ a mixed-frequency VAR (MF-VAR) framework. Our aim in this study is to explore the short and mid-term forecasting capability of the MF-VAR with incorporated weather data. For this reason, we produce out-of-sample forecasts for a nine-month period. We also implement a 5-fold cross-validation scheme that sequentially produces one-week ahead forecasts. Each fold in the cross-validation scheme is separated into one year of training data, and one week of test data to evaluate the forecasts. The training set dates are strictly before the test set dates to preserve the sequential nature of our time-series problem. This study is primarily focused on point forecasts rather than density forecasts. Furthermore, using only weather data, we attempt to replicate the key characteristics of electricity prices organically, i.e., persistent volatility, mean-reversion, seasonality,

and spikes. In particular, we are interested in replicating the heavy-tailed distribution of half-hourly electricity prices using the weather variables. In this way, we aim to provide a comprehensive study of the impact of weather on electricity price formation in New Zealand.

Mixed-frequency models are commonly used by econometricians, stemming from the mixed-data sampling (MIDAS) model by Ghysels et al. (2004). Ghysels (2016) was the first to use the MF-VAR framework to forecast monthly data using quarterly variables in a macroeconomic setting. For other applications of the MF-VAR framework, see Eraker et al. (2014), Schorfheide & Song (2015) and Götz & Hauzenberger (2021).

Generally, applications of this model involve forecasting quarterly data using monthly variables. Using lower frequency variables to predict higher frequency data is known as a reverse mixed-frequency approach. Foroni et al. (2023) introduce the reverse unrestricted MIDAS (RU-MIDAS) as an alternative to the MF-VAR. They use macroeconomic variables to estimated models for electricity traded in the German and Italian electricity markets. In this study, we compare the forecasting capabilities of the MF-VAR and RU-MIDAS models using our weather variables. To our knowledge, we are the first to employ a reverse mixed-frequency framework for forecasting half-hourly electricity prices using hourly weather variables.

For parameter estimation of large datasets and parameter sets, the Gibbs sampler computational times are inefficient. As a result we employ the mean-field variational Bayes (MFVB) technique, see Bishop & Nasrabadi (2006), Ormerod & Wand (2010), Wand et al. (2011) and Koop & Korobilis (2018). The MFVB approximates the joint posterior distribution of parameters and is comparable in computational time to a least-squares approach. In our study, the MFVB provides an similar estimates to the Gibbs sampler. See Gefang et al. (2022) and Gefang et al. (2020) for MF-VAR estimation using the VB technique.

The incorporation of several weather variables may result in multicollinearity within

the model framework. To alleviate this issue, we employ the least absolute shrinkage and selection operator (LASSO). Introduced by Tibshirani (1996), LASSO is a variable selection method that introduces bias into the estimates to increase accuracy. This method introduces a regularization parameter λ into the least-squares estimation procedure, which eliminates less-significant regressors as λ increases. Uniejewski & Weron (2018a) employ the LASSO to identify significant lags of electricity prices in the Nord Pool and PJM electricity markets. Uniejewski et al. (2019) conduct a similar study in the German EPEX market. See Ludwig et al. (2015), Ziel (2016), Marcjasz et al. (2020) and Jedrzejewski et al. (2021) for more applications of the LASSO in electricity markets. For applications of LASSO for VAR models, see Hsu et al. (2008), Gefang (2014), Cavalcante et al. (2017) and Messner & Pinson (2019). Furthermore, we also incorporate the adaptive LASSO of Zou (2006) in our study. The adaptive LASSO is an oracle estimator that can reduce the bias in LASSO estimates and may improve forecasts. We also employ the random subspace methods introduced by Boot & Nibbering (2019), which select a subset of predictors at random using a specified probability distribution. This method reduces the complexity of the model, however, it does not eliminate multicollinearity. Other dimensionality reduction techniques such as factor analysis and principal component analysis (PCA) have been very popular in the literature, particularly in the machine learning context. However, with the success of LASSO and adaptive LASSO in our findings, a potential future study can include the comparison of LASSO dimensionality reduction with factor analysis and PCA.

The contributions of this study are as follows. Firstly, this study is the first to apply an MF-VAR approach to modeling and mid-term forecasting of half-hourly electricity prices. While mixed-frequency models are generally applied to larger time frames such as daily and monthly observations, we apply the MF-VAR framework to a shorter time frame of half-hourly and hourly data. Regarding the MF-VAR parameter estimation methodology using MFVB, we show that the MFVB is capable of producing close

approximations for the joint posterior distribution of parameters for electricity price datasets. Secondly, we provide a detailed comparison of the MF-VAR and RU-MIDAS models. We also show that the RU-MIDAS can be specified in a VAR framework under certain assumptions, which results in the RU-MIDAS-VAR model. Lastly, this study is the first to employ the LASSO technique for the RU-MIDAS framework. Our results show that utilizing the LASSO and adaptive LASSO techniques significantly improves forecasts for both models.

Our findings suggest that including weather variables and their time lags are beneficial for predicting electricity prices, as they are capable of replicating empirical price volatility, mean-reversion, spikes, and seasonality characteristics. We were able to achieve better forecasts using lags of 24 rather than lags of 240, suggesting that incorporating more past information may not be necessary. With that said, the LASSO method allows us to identify key explanatory variables in a vast parameter set, and significantly improves forecasting capabilities, as compared to standard estimation methods. The implementation of LASSO, in a way, allows us to use any lag value we prefer, as the optimal LASSO results will generally remove redundant parameters. We can see this in the forecasting results since the forecast metrics for LASSO with 24 lags and 240 lags have similar results. Regarding the comparison between the MF-VAR and RU-MIDAS-VAR models, our findings suggest that there is little difference in forecasting accuracy between the models, despite the RU-MIDAS-VAR using more up-to-date data for prediction. Furthermore, the LASSO provides similar forecasts for both models, suggesting that the reduced parameter space is similar. Finally, we find the MFVB parameter estimates to be similar to the Gibbs sampler estimates. The forecast results for these two methods are nearly identical, with the VB performing better in some scenarios. The application of VB to large electricity datasets is extremely beneficial due to the reduced computational burden. In particular, while Gibbs sampling with 100,000 iterations took approximately 45 minutes, the VB ELBO converged in nearly

30 seconds, with a convergence threshold of 10^{-7}. This was a significant improvement in computational time, with little different in the forecast results.

This chapter is organized as follows. Section 3.2 introduces the models and estimation methods. Section 3.3 presents our data, shows the forecasting techniques employed, and discusses the results. Section 3.4 provides a conclusion to this study.

3.2 Models and Methods

In this section, we present the mixed-frequency vector auto-regression (MF-VAR) framework. We discuss several estimation methods including Gibbs sampling and variational Bayes (VB). We also discuss the RU-MIDAS model of Foroni et al. (2023) and make certain assumptions in their model so that it can be specified in a VAR framework. In this setting, the RU-MIDAS-VAR is very similar to the MF-VAR model, with a minor distinction. Finally, we introduce variable-selection techniques including the LASSO and random subspace methods.

3.2.1 Mixed-Frequency VAR

Since its introduction by Sims (1980), vector autoregression models have become an essential tool in macroeconomic and financial analysis. A standard VAR(p) model has the form

$$\mathbf{y}_t = \mathbf{c} + \mathbf{a}_1\mathbf{y}_{t-1} + ... + \mathbf{a}_p\mathbf{y}_{t-p} + \epsilon_t, \tag{3.1}$$

where $\mathbf{y_t}$ is an $k \times 1$ vector of dependent variables, $\mathbf{c}$ is an $k \times 1$ vector of constants, $\mathbf{a}_i$ for $i = 1, 2, ..., p$ is a $k \times k$ matrix of coefficients, and $\epsilon_t \sim \text{MVN}(0_{k\times 1}, \Sigma_{k\times k})$.

Since our price and demand data are observed half-hourly, and weather variables are observed hourly, we employ a mixed-frequency approach. To build a mixed-frequency framework, we use the approach of Ghysels (2016) and separate our price dataset, denoted P_t into two time-series, where the first time-series consists of all half-hour prices from the first half-hour of each hour, denoted P_{1t}, and the second time-series contains all second half-hour prices for each hour, denoted P_{2t}. We adopt a similar approach for half-hourly demand, denoted D_{1t} and D_{2t}. In our framework, P_{1t}, P_{2t}, D_{1t} and D_{2t} are column vectors. Finally, we denote $\mathbf{x}_t$ as a row vector of the hourly

observations for our external variables at time t, i.e.

$$\mathbf{x}_t = [\text{temperature}_t, \text{rain}_t, \text{pressure}_t, \text{humidity}_t, \text{solar}_t, \text{wind\_speed}_t, \text{wind\_direction}_t],$$

where each variable is a column vector.

As a result, the vector $\mathbf{y}_t$ in Equation 3.1 has the form $\mathbf{y}_t = [P_{1t}, P_{2t}, D_{1t}, D_{2t}, \mathbf{x}_t]^\top$. Since $\mathbf{y}_t$ contains 11 dependent variables (2 price variables, 2 demand variables, and 7 weather variables), we have $k = 11$. For estimation and forecasting, we consider time lags of $p = 24$ and $p = 240$ which correspond to lags of up to 1 day and 10 days, respectively.

3.2.2 Parameter Estimation

It is convenient for parameter estimation to use a concise notation for the VAR

$$\mathbf{Y} = \mathbf{AZ} + \mathbf{U}, \tag{3.2}$$

where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T]$ is a $k \times T$ matrix of our data-set, $\mathbf{A}$ is a $k \times (kp + 1)$ matrix of parameters, $\mathbf{U} = [\epsilon_1, \epsilon_2, ..., \epsilon_T]$ is a $k \times T$ matrix of error terms, and T is the total number of observations for each variable, in our case $T = 17280$. The matrix $\mathbf{Z}$ has the form

$$\mathbf{Z} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \mathbf{y}_0 & \mathbf{y}_1 & \cdots & \mathbf{y}_{T-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{1-p} & \mathbf{y}_{2-p} & \cdots & \mathbf{y}_{T-p} \end{bmatrix},$$

such that $\mathbf{Z}$ is a $(kp + 1) \times T$ matrix of observations.

From Equation 3.2, we can estimate $\mathbf{A}$ using the multivariate least squares technique,

which has the analytical expression

$$\hat{\mathbf{A}} = \mathbf{Y}\mathbf{Z}^\top(\mathbf{Z}\mathbf{Z}^\top)^{-1},$$

and the covariance matrix Σ is estimated as

$$\hat{\Sigma} = \frac{1}{T - kp - 1}(\mathbf{Y} - \hat{\mathbf{A}}\mathbf{Z})(\mathbf{Y} - \hat{\mathbf{A}}\mathbf{Z})^\top.$$

Gibbs Sampler for VAR

Due to the popularity of the VAR framework, the application of Gibbs sampling for VAR estimation has been discussed in several works. It can be shown that the posterior distribution of $\mathbf{A}$ in Equation 3.2 conditional on Σ is the multivariate normal distribution. Mathematically, $\mathbf{A}|\Sigma, \mathbf{Y} \sim \text{MVN}(M, V)$, where

$$\begin{aligned} M &= (\Sigma_0^{-1}A_0 + T\Sigma^{-1}\bar{Y})V, \\ V &= (\Sigma_0^{-1} + T\Sigma^{-1})^{-1}, \end{aligned} \tag{3.3}$$

where M and V are the mean and variance of the multivariate normal distribution, respectively, A_0 and Σ_0 are prior hyperparameters, and $\bar{Y}$ is the sample mean.

The conjugate prior for the covariance matrix Σ is the inverse Wishart distribution i.e. $\Sigma|\mathbf{A}, \mathbf{Y} \sim \text{IW}(S, \nu)$, where

$$\begin{aligned} S &= S_0 + (\mathbf{Y} - \mathbf{Z}M^\top)^\top(\mathbf{Y} - \mathbf{Z}M^\top), \\ \nu &= T + \nu_0, \end{aligned} \tag{3.4}$$

where S is a positive-definite scale matrix, ν is the degrees of freedom parameter, and S_0 and ν_0 are prior hyperparameters.

A brief explanation of the Gibbs sampling algorithm is as follows.

Step 1: Set initial values for the prior hyperparameters and Σ. The initial value for σ can be chosen as the least squares estimate.

Step 2: Draw VAR coefficients in matrix $\mathbf{A}$ from the multivariate normal distribution and using Equation 3.3.

Step 3: Use the result from Step 2 to draw values for the covariance matrix Σ from the inverse Wishart distribution and using Equation 3.4.

Repeat Steps 2 and 3 for an arbitrarily large number of simulations to ensure convergence.

In this study, we adopt a hierarchical shrinkage prior in the form of the Horseshoe prior, see Zou (2006) and Cross et al. (2020). There is much discussion on the usage of informative priors for Bayesian estimation. While it is difficult to identify a single best prior selection, several studies suggest that the usage of flat priors may lead to inaccurate estimates, which further leads to inadequate predictions, see Litterman (1986) and Bańbura et al. (2010). For further discussion of informative priors, see Koop et al. (2010).

Mean Field Variational Bayes for VAR

Variational Bayesian (VB) inference methods have seen a rise in popularity in recent years as an estimation technique when MCMC methods are computationally challenging. In this study, we provide a brief explanation of the VB technique. For a more comprehensive, theoretical understanding, refer to Ormerod & Wand (2010) and Blei et al. (2017).

Consider the context where we would like to obtain the posterior distribution $p(\theta|y)$,

where θ denotes the parameter set and y denotes the data. The VB methods approximates $p(\theta|y)$ with another, simpler density $q(\theta)$. The new density is found by minimizing the Kullback-Leibler divergence between the two densities, or by maximizing the Evidence Lower Bound (ELBO)

$$\text{ELBO} = \mathbb{E}(\log p(\theta, y)) - \mathbb{E}(\log q(\theta)).$$

where $\mathbb{E}(\cdot)$ indicates the expected value of the densities.

For mean field variational Bayes (MFVB), the choice of an approximating density $q(\theta)$ is restricted such that

$$q(\theta) = \prod_{i=1}^{N} q_i(\theta_i),$$

where θ_i for $i = 1, ..., N$ are blocks of the complete parameter set θ.

If the parameter blocks are assumed to be independent, the ELBO is written as

$$\text{ELBO} = \mathbb{E}(\log p(y|\theta)) + \prod_{i=1}^{N} \mathbb{E}(\log p(\theta_i)) - \prod_{i=1}^{N} \mathbb{E}(\log q_i(\theta_i)).$$

The optimal density $q_i(\theta_i)$ is then determined as

$$q_i(\theta_i) = \exp\big[\mathbb{E}_{q(\theta_{-i})}(\log p(\theta_i|y, \theta_{-i}))\big],$$

where θ_{-i} denotes all parameter blocks apart from θ_i.

We follow the approach of Gefang et al. (2022) for VB inference in a VAR framework. Referring back to Equation (3.1), we modify it to take the form

$$\mathbf{y}_t = \mathbf{X}_t\alpha + \epsilon_t, \tag{3.5}$$

such that $\mathbf{X}_t = \mathbf{I}_k \otimes [1, \mathbf{y}_{t-1}^\top, ..., \mathbf{y}_{t-p}^\top]$ is a $k \times K^*$ matrix of explanatory variables and $\alpha = vec([\mathbf{c}, \mathbf{a}_1, ..., \mathbf{a}_p]^\top)$ is a $K^* \times 1$ vector of coefficients, where $K^* = k(kp + 1)$. Having the VAR in the form of Equation (3.5) is useful because the model can then be expressed as k independent equations with the ith equation having the form

$$y_{i,t} = \mathbf{w}_{i,t}\theta_i + \epsilon_{i,t},$$

where the ith equation corresponds to the evolution of the ith dependent variable on the left-hand side of the original framework in Equation 3.1. From here onward, the notation $\mathbf{W}_i = (\mathbf{w}_{i,1}, ...\mathbf{w}_{i,T})^\top$, $\mathbf{Y}_i = (y_{i,1}, ...y_{i,T})^\top$ and $\epsilon_i = (\epsilon_{i,1}, ...\epsilon_{i,T})^\top$ are used. This specification of the VAR allows equation-by-equation estimation, which can greatly reduce computational times, particularly for larger data sets. In this context, the equation-by-equation VAR specification is essentially an ARX representation where each dependent variables is estimated as a separate ARX model. Furthermore, this specification eases the assumptions of the VAR, i.e., the dependent variables are assumed to be independent of each other. By breaking this assumption, we can separate the VAR structure into individual equation with uncorrelated errors $\epsilon_{i,t}$. This assumption may not always hold, but it eases our computational burden for Bayesian estimation.

The priors for the parameters in the ith equation are

$$\begin{aligned} \theta_i &\sim N(0, \bar{V}_i), \\ \sigma_i^{-2} &\sim \text{Gamma}(\bar{\nu}, \bar{s}) \end{aligned} \tag{3.6}$$

where $\bar{V}_i$, $\bar{\nu}$, and $\bar{s}$ are prior hyperparameters.

Using the derivation of full conditional posteriors by You et al. (2014), the MFVB

approximation densities for the ith equation are

$$q(\theta_i) \sim N(\underline{\theta}_i, \underline{V}_i)$$
$$q(\sigma_i^{-2}) \sim \text{Gamma}\left(\bar{\nu} + \frac{T}{2}, \underline{s}_i\right),$$

where $\underline{V}_i$, $\underline{\theta}_i$, and $\underline{s}_i$ are parameters of the density functions dependent on the observations and hyperparameters such that

$$\underline{V}_i = \left[\left(\frac{\bar{\nu} + \frac{T}{2}}{\underline{s}_i}\right)\mathbf{W}_i^\top \mathbf{W}_i + \bar{V}_i^{-1}\right]^{-1}$$
$$\underline{\theta}_i = \left(\frac{\bar{\nu} + \frac{T}{2}}{\underline{s}_i}\right)\underline{V}_i \mathbf{W}_i^\top \mathbf{Y}_i$$
$$\underline{s}_i = \bar{s} + \frac{1}{2}\|\mathbf{Y}_i - \mathbf{W}_i\underline{\theta}_i\|^2 + \frac{1}{2}tr(\mathbf{W}_i^\top \mathbf{W}_i \underline{V}_i).$$

In this study, we employ a hierarchical shrinkage prior in the form of a Horseshoe prior, see Carvalho et al. (2010) and Cross et al. (2020). The horseshoe prior adds the assumption that

$$\bar{V}_i = \text{diag}(\lambda_{i,1}\tau_i, ..., \lambda_{i,K^*}\tau_i),$$

where the priors for the parameters are

$$\lambda_{i,j}^{-1}|\eta_{i,j} \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{\eta_{i,j}}\right),$$
$$\tau_i^{-1}|\phi_i \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{\phi_i}\right),$$
$$\eta_{i,1}^{-1}, ..., \eta_{i,K^*}^{-1}, \phi_i^{-1} \sim \text{Gamma}\left(\frac{1}{2}, 1\right),$$

where i denotes the VAR equation, j denotes the coefficients in that equation, and K^* is the total number of coefficients. The full derivation of the conditional posteriors can

be found in Gefang et al. (2023).

The iterative procedure begins by specifying the prior hyperparameters and providing initial values for $\underline{V}_i$ and $\underline{\theta}_i$. The $ELBO$ is calculated after each iteration, and the procedure runs until the change in $ELBO$ meets the specified convergence criterion. Finally, we use the $ELBO$ specified by Gefang et al. (2023)

$$\begin{aligned} ELBO_i = \frac{1}{2}\log\left(|\underline{V}_i|\right) - \frac{1}{2}\log\left(|\bar{V}_i|\right) - \frac{1}{2}\left[\underline{\theta}_i^{\top}\bar{V}_i^{-1}\underline{\theta}_i + tr(\bar{V}_i^{-1}\underline{V}_i)\right] \\ - \left(\bar{\nu} + \frac{T}{2}\right)\log\left(\underline{s}_i\right) + Const., \end{aligned}$$

which is derived from the assumption that $q(\theta) = q(\theta_i)q(\sigma_i^{-2})$.

For the prior hyperparameters, we set $\bar{\nu} = 5, \bar{s} = 0.05$. These values are chosen using a grid-search, although the parameter estimates were not sensitive to the choice of these hyperparameters. We chose values for $\bar{V}_i, \underline{V}_i$ and $\underline{\theta}_i$ according to least squares estimates. Our stopping criteria for the change in $ELBO$ is set to a tolerance of (10^{-7}). The iteration procedure generally converged relatively fast, with total computational time comparable to a least-squares approach. Furthermore, the parameter estimates were similar to Gibbs sampler estimates, which signifies a good approximation.

3.2.3 RU-MIDAS Model

The RU-MIDAS model proposed by Foroni et al. (2023) is an extension of the popular MIDAS model by Ghysels et al. (2004), which allows the use of low frequency variables to predict high frequency data. The RU-MIDAS in single-equation structure for half-hourly and hourly frequencies, and disregarding any external variables, has the form

$$P_t = \alpha_{11}(1 - D_2)P_{t-\frac{1}{2}} + \alpha_{12}D_2P_{t-\frac{1}{2}} + \alpha_{21}(1 - D_2)P_{t-1} + \alpha_{22}D_2P_{t-1} + \epsilon_t, \qquad (3.7)$$

where P_t denotes price at time t for $t = 0, \frac{1}{2}, 1, \frac{3}{2}, ...$, and $D_2 = 1$ for the first half hour of every hour, and $D_2 = 0$ otherwise.

Alternatively, we can separate Equation (3.7) into two cases

$$P_t = \alpha_{12} P_{t-\frac{1}{2}} + \alpha_{22} P_{t-1} + \epsilon_t, \quad \text{when } D_2 = 1,$$
$$P_t = \alpha_{11} P_{t-\frac{1}{2}} + \alpha_{21} P_{t-1} + \epsilon_t, \quad \text{when } D_2 = 0.$$

The RU-MIDAS makes an underlying assumption that the error term ϵ_t for each half-hourly observation is independent. If we relax this assumption, we can re-structure the RU-MIDAS into a VAR framework of the form

$$\begin{bmatrix} P_{1,t} \\ P_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & \alpha_{12} & \alpha_{22} \\ \alpha_{11} & \alpha_{21} & 0 \end{bmatrix} \begin{bmatrix} P_{1,t} \\ P_{2,t-1} \\ P_{1,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}, \tag{3.8}$$

where $[\epsilon_{1,t}, \epsilon_{2,t}]^\top \sim \text{MVN}(0, \Sigma)$. We call this the RU-MIDAS-VAR model in this study. Equation (3.8) can easily be extended to incorporate more time lags and external variables. However, there is a key distinction to be made between this model and the standard VAR framework. In the standard VAR model, the second half hour of each hour i.e $P_{2,t}$ only uses information up to $P_{2,t-1}$. In other words, it is not affected by $P_{1,t}$. In contrast, we can observe in Equation (3.8) that $P_{2,t}$ is influenced by $P_{1,t}$ in the RU-MIDAS-VAR framework. The RU-MIDAS-VAR incorporates more updated information in the prediction process. In Section 3, we compare the forecasting capabilities of both models to assess the impact of more frequent updating.

The RU-MIDAS-VAR model we consider in this study includes external variables, denoted by the vector $\mathbf{x_t}$, and so the model can be rewritten as

$$\begin{bmatrix} P_{1,t} \\ P_{2,t} \\ \mathbf{x_t} \end{bmatrix} = \begin{bmatrix} 0 & \alpha_{1,2} & \alpha_{2,2} & \dots \\ \alpha_{1,1} & \alpha_{2,1} & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} P_{1,t} \\ P_{2,t-1} \\ P_{1,t-1} \\ \vdots \\ P_{2,t-p} \\ P_{1,t-p} \\ \mathbf{x}_{t-1} \\ \vdots \\ \mathbf{x}_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \vdots \\ \epsilon_{k+2,t} \end{bmatrix}, \tag{3.9}$$

where $\mathbf{x}_t$ is a column vector of all external variables, p is the total number of lags, and k is the number of external variables.

3.2.4 Variable Selection Techniques

LASSO

The least absolute shrinkage and selection operator (LASSO) of Tibshirani (1996) has seen popularity in scientific fields as a variable selection tool. It can be thought of as a generalized version of a linear regression, where instead of minimizing the sum of squared residuals, we introduce a penalty function to the optimization problem. The LASSO, while initially introduced for linear regression models, is easily applied to multivariate regression problems. Recall from Equation 3.2 the concise notation for the VAR. The LASSO estimator $\tilde{A}$ is obtained by solving the optimization problem

$$\tilde{A} = \min_{\mathbf{A}} \left\{ (\mathbf{Y} - \mathbf{AZ})^{\top} (\mathbf{Y} - \mathbf{AZ}) + \lambda \sum_{j=1}^{k(kp+1)} |A_j| \right\}, \tag{3.10}$$

where A_j for $j = 1, ..., k(kp+1)$ is the jth parameter in the full parameter set, and $\lambda \geq 0$ is the LASSO tuning parameter. Note that for $\lambda = 0$ we obtain the least squares estimate. As λ increases, less significant parameters move to zero until all parameters are zero. In this way, we can determine an optimal λ value to suit our variable selection needs. In this study, we select several λ values according to the forecast metrics RMSE, MAPE and MAD. We perform the LASSO for the MF-VAR and RU-MIDAS-VAR models. The results are discussed in Section 3.

Adaptive LASSO

The adaptive LASSO of Zou (2006) is an evolution of the LASSO and is considered an oracle estimator. By adding an additional weight to the LASSO estimates, the adaptive LASSO is able to reduce the bias of the LASSO estimator. The adaptive LASSO estimator $\tilde{A}$ is obtained by solving the optimization problem

$$\hat{A} = \min_{\mathbf{A}} \left\{ (\mathbf{Y} - \mathbf{AZ})^{\top}(\mathbf{Y} - \mathbf{AZ}) + \lambda \sum_{j=1}^{k(kp+1)} \hat{w_j}|A_j| \right\},$$

where A_j for $j = 1, ..., k(kp+1)$ is the jth parameter in the full parameter set selected by the LASSO estimator, and $\hat{w_j}$ is a weight assigned to that parameter. In this study, we choose the weights to be the the inverse of the absolute sum of the LASSO-estimated parameters, i.e. $\hat{w_j} = \left\|\tilde{A}\right\|^{-1}$, where $\tilde{A}$ is the LASSO estimator described in Equation 3.10.

Random Subspace Methods

Random subspace methods were introduced by Boot & Nibbering (2019) with an application to high-dimensional linear regression models. The idea of random subspace methods is to apply random weights drawn from some probability distribution, on the observation set. The resulting dataset is used to estimate parameters and provide

predictions. This process is repeated with new weights each time, and the results are averaged to reduce the forecast variance while preserving most of the signal. Consider the model

$$y_{t+1} = w_t^\top \beta_w + x_t^\top R \beta_{x,r} + \epsilon_t, \tag{3.11}$$

where w_t is considered the essential observation set, and will not be filtered, β_w are essential regression coefficients, x_t is the remaining observation set, and R is a $p_x \times n$ matrix, where $n < p_x$. Here p_x denotes the number of regression coefficients in the 'non-essential' parameter set and n denotes the number of 'non-essential' parameters that will remain after dimension-reduction. The choice of R determines the type of random subspace method, and Boot & Nibbering (2019) suggests two methods for choosing R, the random subset regression and random projection regression.

Random Subset Regression In random subset regression, a new subset of n parameters is chosen at random. Define an index $I = 1, ..., n$ and a scalar $c(I)$ such that $1 < c(I) < p_x$. Denote $E_{c(I)}$ a $p_x \times 1$ vector with its $c(I)$th entry equal to one, the the random matrix R is chosen such that

$$R = \left[E_{c(1)}, ..., E_{c(n)}\right], \quad E_{c(i)} \neq E_{c(j)} \text{ if } i \neq j. \tag{3.12}$$

Random Projection Regression In random projection regression, a new set of predictors is created by taking weighted averages drawn from the Gaussian distribution. In this case, each entry of R is independent and identically distributed as

$$[R]_{ij} \sim N(0, 1), \quad 1 \leq i \leq p_x, \quad 1 \leq, j, \leq n. \tag{3.13}$$

Using the VAR form in Equation (3.2), the R matrix is extended to fit a multivariate

regression setting, and it is easily applied to the MF-VAR and RU-MIDAS-VAR models.

We present the results of random subspace methods in Section 3.

3.3 Data and Results

3.3.1 Description and Key Features

Our electricity price dataset consists of half-hourly electricity prices from the Otahuhu node in New Zealand from January 1, 2018 to September 30, 2020. The New Zealand electricity market is a real-time market, where electricity prices are determined by matching demand and supply every five minutes. The prices are determined as the highest bids made by electricity generators that is capable of clearing all demand. The five-minute prices are averaged to produce the half-hourly prices.

Figure 3.1 displays the electricity price time series. It is clear from this figure that half-hourly prices exhibit several key characteristics; extreme volatility, upward spikes, and mean-reversion. The figure also shows the impact of several key events. A consistent hydro shortage, due to a lack of rainfall in nearby regions, can significantly increase electricity prices until lake levels return to normal. In October 2018, a hydro shortage, with an unexpected outage in the largest gas field in New Zealand, saw prices consistently around $500/MWh.

Figure 3.2(a) displays the autocorrelation of our price time series. The cyclical peaks that occur at lags of multiples of 48 indicate the intra-day seasonality behaviour. Since prices are largely influenced by demand, lower demand periods at night-time exhibit lower prices than during day-time. The intra-day seasonality can be further observed in Figure 3.2(b), which displays the mean prices for every half-hour period in our dataset. This figure is a good depiction of the intra-day seasonality behaviour of electricity prices. There are two peaks during the day, the first at approximately 9:00am as businesses open, and the second around 7:00pm as residential demand increases. There is a large drop-off thereafter, as demand subsides during late hours. For forecasting purposes, we separate our price time series into four distinct periods; the morning peak from

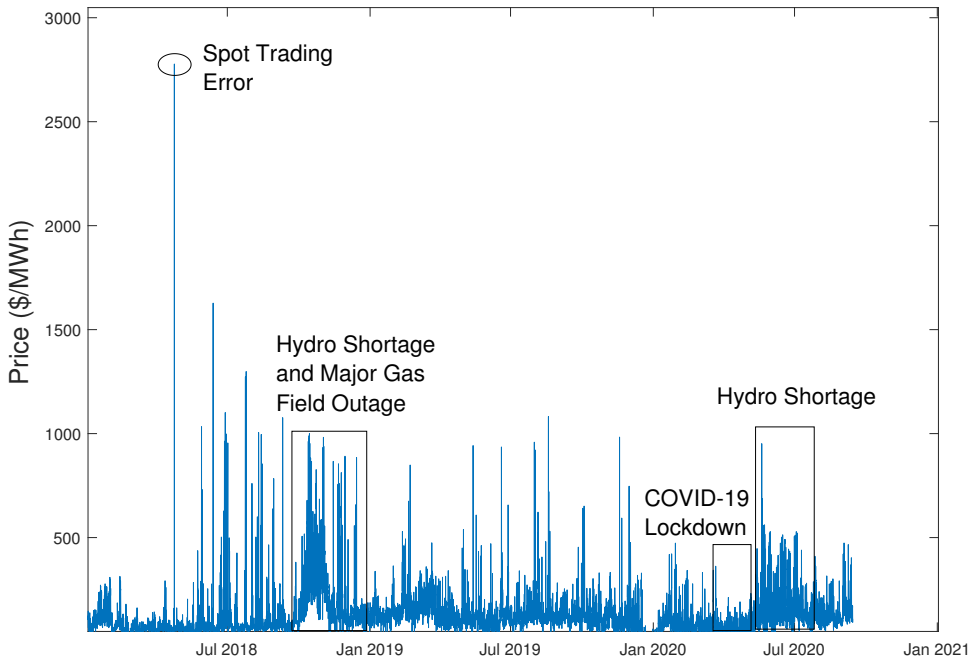
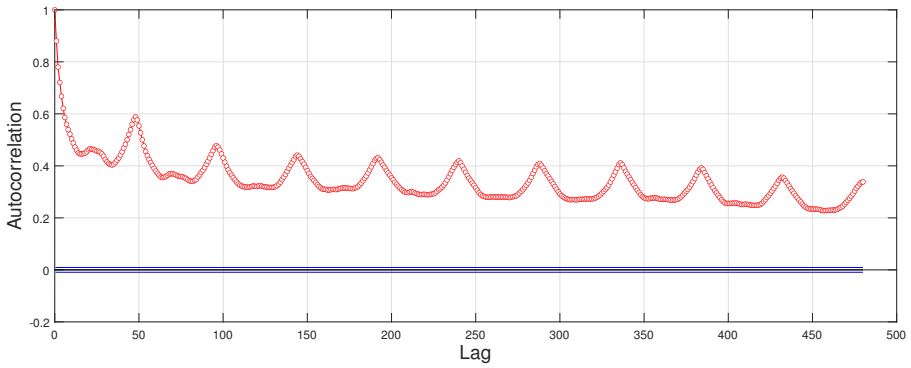


Figure 3.1: Half-hour electricity prices from Otahuhu node in New Zealand

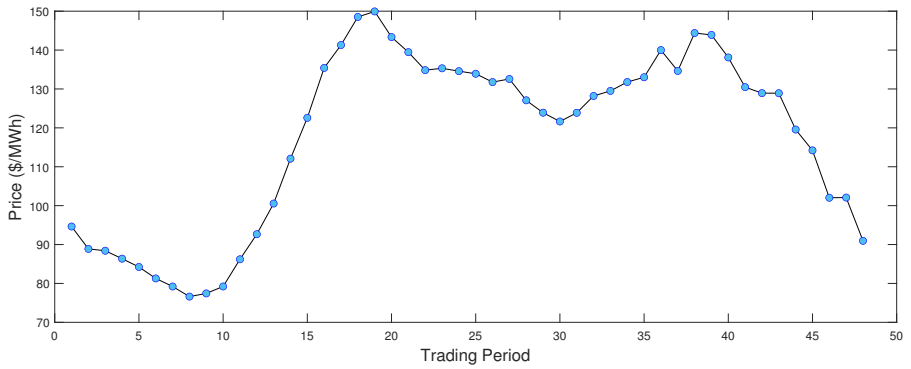
7:00am to 11:00am, the evening peak from 6:00pm to 10:00pm, the midday off-peak from 11:00am to 6:00pm, and the night off-peak from 10:00pm to 7:00 am. Figure 3.3 below provides our reasoning to make this distinction.

Figure 3.3 displays a box plot of half-hourly electricity prices for our sample period. The bottom and top edges of the box represents the 25th and 75th percentiles of data, while the red line inside the box represents the median price. The red crosses outside the box represents any outliers. Electricity prices exhibit a large quantity of outliers in the form of spikes that are observed in Figure 3.1. Furthermore, the size and intensity of the outliers are increased during peak price hours. As a result, the peak hours forecasts will have a higher priority than off-peak forecasts.

Our external market data consists of half-hourly demand for electricity in the North Island of New Zealand and several hourly weather variables including temperature, rainfall, air pressure, humidity, solar radiation, wind speed, and wind direction. Weather



(a) Auto-correlation of half-hourly prices



(b) Average price for each half-hour trading period

Figure 3.2: Price auto-correlation and average price per trading period

data are observed from the Auckland International Airport weather station by MetService. Table 3.1 presents descriptive statistics of price, demand, and weather variables.

In this study, all variables are log-transformed to reduce the impact of outliers and to improve the normality of the data. Furthermore, the wind direction variables is first transformed using a sine transformation to emphasize its cyclical nature.

Table 3.2 presents the correlation coefficients between price and external variable time series. For the purpose of computing these correlations, we have taken the mean of half-hourly prices and the sum of half-hourly demand to convert them to hourly variables. We can observe instances where some weather variables have a strong correlation among each other, such as temperature and solar radiation, or wind speed and humidity. Regardless, we accept them as explanatory variables and later perform

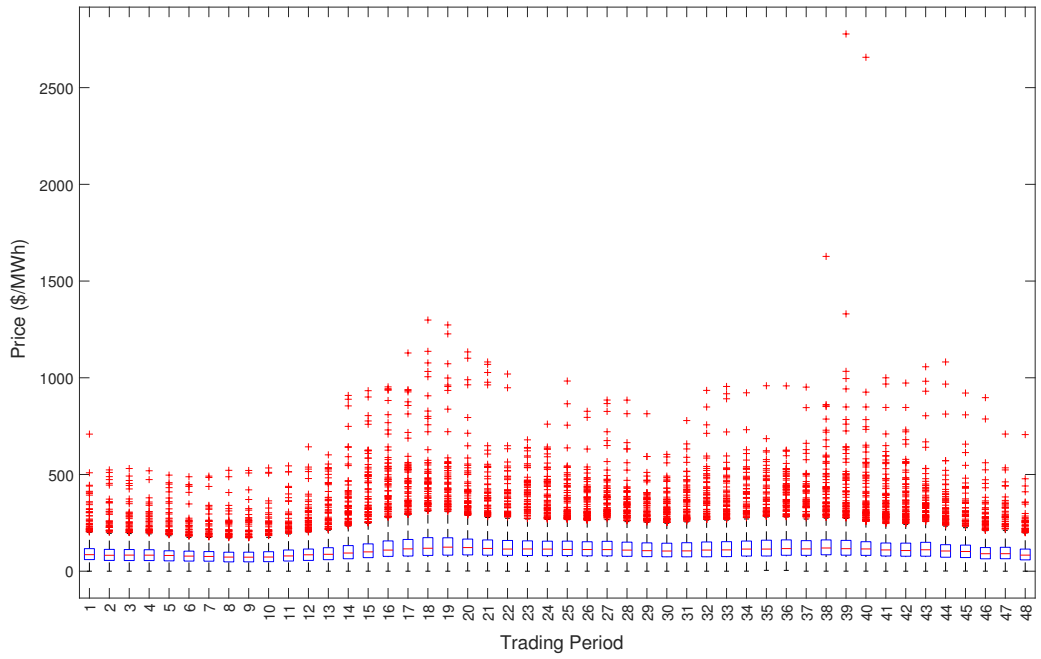


Figure 3.3: Box plot of prices for each trading period

variable-selection techniques using the LASSO and random subspace methods in an effort to remove less significant explanatory variables and improve price forecasts.

3.3.2 Forecast Results

In this section, we present the out-of-sample forecasting results for the various models and methods we have employed in this study. We also separate the results

	Mean	Median	Std	Skewness	Kurtosis
Price (NZ$/MWh)	117.67	98.90	91.28	4.32	52.20
Demand (MWh)	1437.73	1463.03	303.78	0.03	2.18
Temperature (Celsius)	15.71	15.50	4.32	0.11	2.63
Rainfall (mm)	0.13	0.00	0.74	17.97	701.54
Solar Radiation (mJ/m^2)	172.45	1.00	83.30	1.58	4.60
Air Pressure (hPa)	1015.80	1016.50	7.91	-0.37	3.29
Humidity (%)	80.47	82.00	12.10	-0.45	2.45
Wind Speed (km/h)	15.88	15.00	9.78	0.77	3.38
Wind Direction (degrees)	180.59	210	100.11	-0.22	2.01

Table 3.1: Descriptive statistics of variables

	Price	**Demand**	**Temp**	**Rain**	**Solar**	**Pressure**	**Humidity**	**Wind Speed**	**Wind Direction**
Price	1	0.344	-0.0479	0.022	0.108	0.0986	-0.084	0.033	-0.056
Demand	0.344	1	-0.076	0.026	0.246	-0.022	-0.212	0.154	0.061
Temp	-0.0479	-0.076	1	-0.029	0.493	-0.083	-0.464	0.254	-0.021
Rain	0.022	0.026	-0.029	1	-0.073	-0.177	0.166	0.080	0.027
Solar	0.108	0.246	0.493	-0.073	1	0.033	-0.627	0.273	0.093
Pressure	0.0986	-0.022	-0.083	-0.177	0.033	1	-0.084	-0.378	-0.197
Humidity	-0.084	-0.212	-0.464	0.166	-0.627	-0.084	1	-0.412	-0.150
Wind Speed	0.033	0.154	0.254	0.080	0.273	-0.378	-0.412	1	0.166
Wind Direction	-0.056	0.061	-0.021	0.027	0.093	-0.197	-0.150	0.166	1

Table 3.2: Correlation coefficient table

for peak and off-peak periods. Since electricity prices behave much more erratically during peak periods, the peak forecast metrics are more relevant to our study than off-peak forecasts. We consider the mean absolute deviation (MAD), root mean squared errors (RMSE), and mean absolute percentage errors (MAPE) as forecast metrics. Our estimation data set consists of price, demand, and weather data from January 01, 2018, to December 31, 2019. We only consider data starting from January 01, 2018, because there has been a significant fundamental change in the behavior of the New Zealand electricity market beyond this time period, due to the ban of offshore oil and gas exploration, as well as major outages in significant gas fields in New Zealand. The year 2018 was a transitional period into a more renewable environment for the New Zealand electricity market, and a major step towards a fully renewable scheme. Our forecasting period is from January 1, 2020, to September 30, 2020. During the course of our work, global markets were impacted by the COVID-19 pandemic. However, the impact of COVID-19 on New Zealand electricity prices was minimal. We notice a period in April 2020, during the lockdown in New Zealand, where prices are lower than usual. The lockdown reduced nationwide industrial demand for electricity, as the majority of businesses were closed for this period. Other than this, there are no significant effects of the pandemic on electricity price itself, regardless of the impact on other aspects of the electricity market.

In particular, we conduct two separate studies. The first is a simple, single nine-month out-of sample forecast for the period from January 1, 2020 to September 30, 2020. This forecast is used to assess the long-term capabilities of the models. A second study incorporates a cross-validation scheme of several one-week forecasts, for the same time period. The purpose of this is to compare the short-term capabilities of the models. For this study, we perform a rolling estimation technique, so parameters are updated after every week. In the single out-of-sample nine-month forecast, we use forecasts of external variables as inputs, whereas in the cross-validation scheme, we use the realized values. In this way, we assure that no future information is used to make price forecasts.

Figure 3.4 displays the actual electricity price along with simulations generated from the MF-VAR and RU-MIDAS-VAR models. When running multiple simulations, we found that some spikes persistently appear in specific periods in our forecast horizon, suggesting that the weather variables have some impact on price spiking. However, we notice, in several periods, a large mismatch between simulated prices and actual prices. As we expected, there is a vast amount of information in the price data that cannot be explained by weather variables alone. The VAR framework with weather data is capable of capturing key characteristics, however, more data on market dynamics as well as macroeconomic variables, such as commodity prices, may improve our forecasts. Furthermore, we may be able to improve the simulated price density using heavy-tailed distribution within the VAR framework.

To study the capability of our forecasting models to replicate the mean-reversion and seasonality characteristics of electricity prices, we observe the half-life, and spectral density plots, respectively. The half-life values are derived from the lag 1 autocorrelation coefficient, and they denote the decaying rate of deviations from the long-term mean. Similar half-life values between different time-series' indicate similar mean-reversion

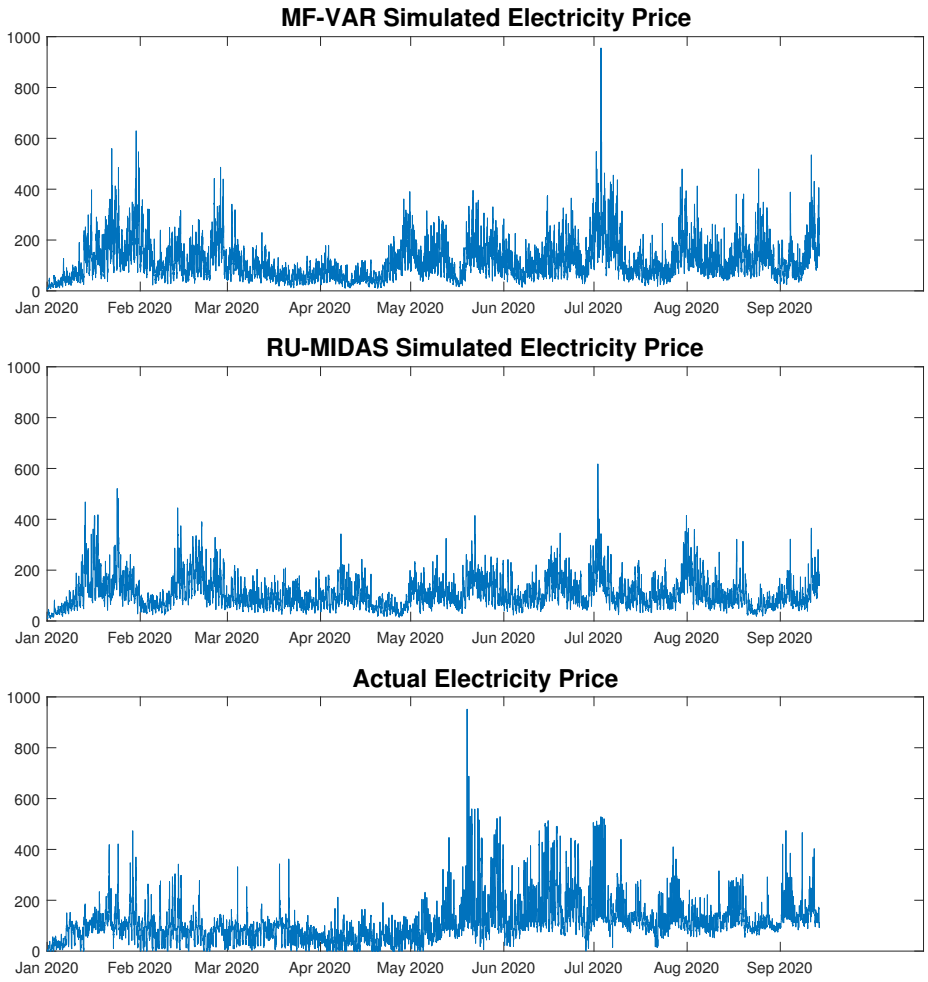


Figure 3.4: Simulations of actual and model prices

characteristics. Table 3.3 displays the AR(1) coefficient and corresponding half-life values for the actual prices and prices simulated from the MF-VAR and RU-MIDAS-VAR models with OLS estimates. The half-life is approximately 0.7 for all three time-series, indicating similar mean-reversion rates. The spectral density will show the distribution of variance for each frequency component in the time series. Peaks in the spectral density correspond to dominant frequencies, which are related to seasonality. Visualizing the spectral density plots and observing the correlation of the spectral densities between actual and simulated prices can help us identify the captured seasonality of the models. This visualization is provided in Figure 3.5. In this figure, we observe the logarithm of

power spectral density to improve interpretability. We notice that the spectral density plots of actual and simulated prices are similar, indicating that the models are able to capture the seasonality of electricity prices. Furthermore, the correlation between the spectral densities of actual prices and the MF-VAR simulated price is 0.999, and the correlation between the spectral densities of actual prices and the RU-MIDAS-VAR simulated price is 0.997.

Model	**AR(1) Coef.**	**Half-Life**
Actual	-0.9707	0.7140
MF-VAR	-0.9745	0.7113
RU-MIDAS-VAR	-0.9899	0.7002

Table 3.3: Half-life values

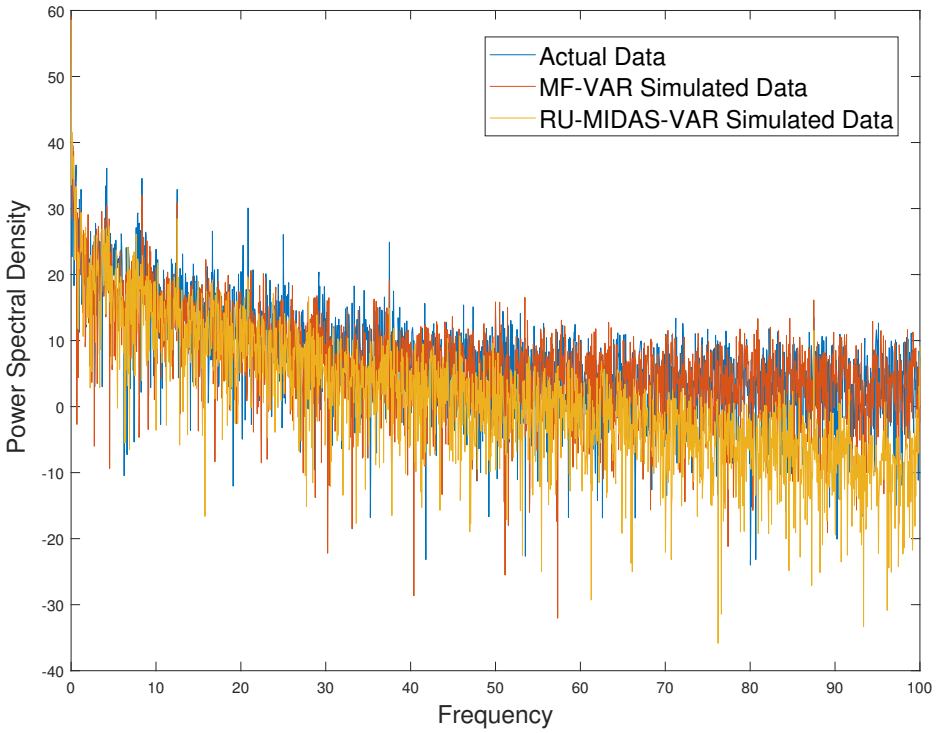


Figure 3.5: Spectral density plots of actual and simulated prices.

In Tables 3-7, MLS, MFVB, and GS refer to the multivariate least-squares, mean-field variational Bayes, and Gibbs sampling techniques, respectively. Table 3.4 shows the out-of-sample forecast results for the models and methods employed in this study. For Bayesian estimation, our experiment uses 100,000 iterations to obtain parameter estimates, where the first 10,000 samples are burned. We also perform thinning by selecting every fifth sample. From the remaining samples, we use the arithmetic mean to obtain parameter estimates, which are then used for forecasting purposes. The left table considers a model with lags of $p = 24$ for each variable, whereas the right table shows results for lags of $p = 240$. For most estimation methods, the lower lag model performs better. The LASSO method performs significantly better for $p = 240$. Comparing the least-squares estimated for the MF-VAR and RU-MIDAS-VAR models, we observe that the MF-VAR model performs marginally better for both lag values. The MFVB and Gibbs sampler estimates are also similar to the least-squares estimates, with the Gibbs sampler performing better in terms of MAPE. The subset regression and projection regression methods provide better forecasts than least-squares for $p = 240$, but significantly underperform for $p = 24$.

We provide two LASSO estimates with different λ values so we may study their optimal performance for all forecast metrics. Generally, the LASSO results are similar for both $p = 24$ and $p = 240$, since, for the latter case, the LASSO likely removes any insignificant, higher lag data. We also notice that, for $p = 240$, the LASSO identifies a parameter set that significantly reduces the MAPE compared to other methods. In particular, we find that the LASSO chosen coefficients include recent price, demand, and solar energy values. Wind speed and wind direction of beyond lag 10 are also relevant, indicating a delayed effect of wind power. Less prominent, but still included features are recent rainfall and humidity values. We find air pressure to hold no important information and it is mostly excluded in the parameter set. Furthermore, for the study of 240 lags, we find that price, demand, wind metrics, and rainfall can have hold

information about current prices even beyond a week. However, solar, humidity, and air pressure do not hold significant information beyond the previous day.

	Lag p = 24				Lag p = 240		
Model	**MAD**	**RMSE**	**MAPE**	**Model**	**MAD**	**RMSE**	**MAPE**
MF-VAR MLS	45.9472	72.2045	34.6090	**MF-VAR MLS**	58.9771	87.7004	38.2073
MF-VAR MFVB	46.9001	73.4641	35.2611	**MF-VAR MFVB**	58.3423	86.4735	37.7803
MF-VAR GS	47.7419	75.9605	**30.4006**	**MF-VAR GS**	58.4798	87.5852	35.5844
MF-VAR LASSO				**MF-VAR LASSO**			
$\lambda = 0.017200$	42.7937	72.6223	38.3311	$\lambda = 0.033160$	**44.9211**	**73.1058**	39.1370
$\lambda = 0.000180$	45.9872	72.4249	36.1998	$\lambda = 0.008214$	57.3355	88.9184	18.5820
Adaptive LASSO	43.1545	**70.4230**	41.2096	**Adaptive LASSO**	46.0301	76.0408	29.9470
Subset Regression	54.6850	83.2752	56.9606	**Subset Regression**	57.2314	86.9146	40.8667
Projection Regression	51.4763	80.8129	53.9186	**Projection Regression**	55.5147	85.4771	38.6816
RU-MIDAS-VAR MLS	46.0136	72.2729	34.5711	**RU-MIDAS-VAR MLS**	59.0786	87.8343	38.3589
RU-MIDAS-VAR LASSO				**RU-MIDAS-VAR LASSO**			
$\lambda = 0.009849$	**42.6590**	72.3683	37.9954	$\lambda = 0.018975$	45.3994	75.5241	36.6278
$\lambda = 0.000078$	46.0803	72.5578	35.9015	$\lambda = 0.006214$	58.7990	90.1122	**17.6100**
Adaptive LASSO	43.5719	71.8944	38.4876	**Adaptive LASSO**	45.5970	75.8669	38.7187
Subset Regression	52.4915	81.7978	56.3428	**Subset Regression**	57.7477	87.2958	42.0268
Projection Regression	51.2251	80.8534	53.9446	**Projection Regression**	55.9291	85.9688	39.3641

Table 3.4: Out-of-sample nine-month forecast results. *Note*: Bold font indicates best result

Tables 3.5 and 3.6 separate the results of Table 3.4 into peak and off-peak forecasts for $p = 24$ and $p = 240$, respectively. From prior analysis, it is clear that electricity price exhibit larger volatility and average prices during peak periods. This is a direct result of higher demand and significant generation costs during peak periods. This also necessitates a larger focus on forecasting capabilities for peak periods, since there is much more risk involved for generators and retailers of electricity during these periods. Therefore, we would like to examine the peak period forecasting capabilities of our models. We notice that the forecasts are much better for the off-peak periods than the peak periods. We expected this since the peak periods exhibit much higher volatility and frequency of spikes. In general, the LASSO and adaptive LASSO models perform significantly better than the other methods, especially so in Table 3.6 for $p = 240$, where the MF-VAR LASSO significantly outperforms even the RU-MIDAS-VAR LASSO model. Again, the MF-VAR and RU-MIDAS-VAR least-squares estimates provide similar results, with the MF-VAR performing slightly better. An interesting result is

that the MAPE for evening peaks is significantly lower than other periods of the day, despite the MAD and RMSE being higher. The reason for this is the MAPE has a more severe penalty for negative errors and the models are consistently overestimating evening prices. Finally, Table 3.8 displays the out-of-sample cross-validation results for rolling 1-week periods. We further separate cross-validation results into peak and off-peak periods. Note that, in some cases, only a single LASSO value is shown, since the LASSO value obtains the best result for every forecast metric. The forecast performance of the models relative to each other in the cross-validation context is similar to the fixed-window nine-month forecast, suggesting that our results are robust to changes in the forecast horizon. The raw values in cross-validation are lower than their nine-month forecast counterparts in Tables 3.4, 3.5 and 3.6, since the former produces one-week forecasts while the latter produces a nine-month forecast.

Table 3.7 displays the results of the two-sided Diebold-Mariano (DM) test to compare the predictive accuracy of forecasts from two models for lag $p = 240$, see Diebold (2015) for details on the DM test. The null hypothesis of the DM test is that the two forecasts have similar predictive accuracy. Since this is a two-tailed test, a significance level $\alpha = 0.05$ suggests that the null hypothesis of no difference in forecasts will be rejected if the computed DM statistics fall outside the range of -1.96 to 1.96. In cases where the forecasts are obtained multiple times, such as in Bayesian inference methods and random subspace methods, the test statistics are computed several times and averaged. We observe from the table that the MFVB, Gibbs sampler, MF-VAR LASSO, and RU-MIDAS-VAR LASSO obtain similar forecasts. This reflects the similarity between the two procedures in that a prior selection in Bayesian inference is a form of regularization. We also observe that the MLS, subset regression, and projection regression forecasts have similar predictions between their MF-VAR and RU-MIDAS-VAR variants, as do the LASSO and adaptive LASSO methods.

We also present the result of the Model Confidence Set (MCS) developed by Hansen

	Morning Peak				Midday Off-Peak		
Model	**MAD**	**RMSE**	**MAPE**	**Model**	**MAD**	**RMSE**	**MAPE**
MF-VAR MLS	54.6073	**82.1370**	14.8220	**MF-VAR MLS**	47.5784	73.1042	3.7002
MF-VAR MFVB	55.3793	83.5628	15.7624	**MF-VAR MFVB**	48.8180	74.6297	3.9611
MF-VAR GS	55.2612	86.3424	13.2280	**MF-VAR GS**	48.3737	76.4844	**3.2597**
MF-VAR LASSO				**MF-VAR LASSO**			
$\lambda = 0.002223$	54.6468	82.5341	15.6724	$\lambda = 0.018890$	**44.0121**	73.1371	3.9950
$\lambda = 0.013020$	55.0219	89.1511	**13.1186**	$\lambda = 0.013020$	44.2649	**72.8327**	3.7915
Adaptive LASSO	**53.8210**	83.2561	15.5626	**Adaptive LASSO**	45.3429	72.9665	3.9627
Subset Regression	68.3809	101.3910	20.0747	**Subset Regression**	57.0204	85.4541	5.8694
Projection Regression	65.5759	98.7949	19.5033	**Projection Regression**	53.1245	82.0871	5.6827
RU-MIDAS-VAR MLS	54.7064	82.2539	14.7734	**RU-MIDAS-VAR MLS**	47.6310	73.1516	3.6902
RU-MIDAS-VAR LASSO				**RU-MIDAS-VAR LASSO**			
$\lambda = 0.001056$	55.0073	82.9469	15.7267	$\lambda = 0.013020$	47.6601	73.6399	3.9032
$\lambda = 0.000078$	55.0474	82.9350	15.5213	$\lambda = 0.017212$	47.6984	75.8949	3.9001
Adaptive LASSO	53.5300	83.7917	14.9882	**Adaptive LASSO**	44.1214	72.9532	3.8927
Subset Regression	66.7269	100.1605	19.4789	**Subset Regression**	54.4827	83.5487	5.7332
Projection Regression	65.5392	99.1871	19.1159	**Projection Regression**	52.6773	81.7077	5.7334
	Evening Peak				**Night Off-Peak**		
Model	**MAD**	**RMSE**	**MAPE**	**Model**	**MAD**	**RMSE**	**MAPE**
MF-VAR MLS	61.1290	104.7811	0.8279	**MF-VAR MLS**	39.6659	59.4733	66.3977
MF-VAR MFVB	62.2437	106.5511	0.8689	**MF-VAR MFVB**	40.4225	60.3357	67.3719
MF-VAR GS	64.7439	110.7438	0.7590	**MF-VAR GS**	41.9589	62.7301	**58.2653**
MF-VAR LASSO				**MF-VAR LASSO**			
$\lambda = 0.014290$	60.7020	104.8512	0.7521	$\lambda = 0.020732$	**35.2874**	**57.2525**	76.4216
$\lambda = 0.013020$	60.7203	107.1022	**0.7467**	$\lambda = 0.000078$	39.5438	59.4261	68.3569
Adaptive LASSO	56.8763	**98.7281**	0.8817	**Adaptive LASSO**	35.4375	59.0505	79.9593
Subset Regression	72.7502	116.9132	1.1265	**Subset Regression**	46.7055	68.4308	115.0509
Projection Regression	71.8114	116.6687	1.0519	**Projection Regression**	44.1571	65.8535	105.5108
RU-MIDAS-VAR MLS	61.1883	104.8416	0.8257	**RU-MIDAS-VAR MLS**	39.7358	59.5502	66.3361
RU-MIDAS-VAR LASSO				**RU-MIDAS-VAR LASSO**			
$\lambda = 0.000417$	61.1077	105.2518	0.8456	$\lambda = 0.009849$	39.2643	59.4555	74.0237
$\lambda = 0.006789$	61.1593	105.4274	0.8437	$\lambda = 0.000078$	39.5477	59.4938	68.7988
Adaptive LASSO	**56.4591**	99.9185	0.8355	**Adaptive LASSO**	35.2920	59.9381	74.6538
Subset Regression	71.3481	115.9424	1.0911	**Subset Regression**	43.9589	65.5341	109.3512
Projection Regression	70.8592	116.2536	1.0469	**Projection Regression**	42.8587	64.8399	104.4356

Table 3.5: Out-of-sample forecasts for peak and off-peak time periods for lag $p = 24$. *Note*: Bold font indicates best result

	Morning Peak				Midday Off-Peak		
Model	**MAD**	**RMSE**	**MAPE**	**Model**	**MAD**	**RMSE**	**MAPE**
MF-VAR MLS	67.9381	97.6032	23.2370	**MF-VAR MLS**	66.3473	95.4826	8.7707
MF-VAR MFVB	67.2869	96.2480	22.8303	**MF-VAR MFVB**	65.6230	94.0924	8.5966
MF-VAR GS	67.6195	97.9506	20.3349	**MF-VAR GS**	65.0256	95.9446	8.1388
MF-VAR LASSO				**MF-VAR LASSO**			
$\lambda = 0.002690$	**55.6432**	**80.1613**	20.3435	$\lambda = 0.033160$	**47.2202**	**77.4684**	4.0638
$\lambda = 0.009894$	68.8604	105.0660	**8.8490**	$\lambda = 0.009015$	61.6604	93.4404	**2.6254**
Adaptive LASSO	57.1686	86.5855	14.0758	**Adaptive LASSO**	49.0756	79.2626	3.4859
Subset Regression	73.6302	106.0827	15.5964	**Subset Regression**	63.9974	93.0562	5.9785
Projection Regression	72.9490	105.5436	16.2854	**Projection Regression**	59.8845	89.7752	5.1609
RU-MIDAS-VAR MLS	68.0862	97.7560	23.3134	**RU-MIDAS-VAR MLS**	66.5073	95.6890	8.7869
RU-MIDAS-VAR LASSO				**RU-MIDAS-VAR LASSO**			
$\lambda = 0.039941$	63.9649	94.3260	27.8201	$\lambda = 0.036393$	65.9945	97.8141	9.3677
$\lambda = 0.001689$	66.5887	95.8687	25.8593	$\lambda = 0.000459$	67.1229	98.2629	9.3552
Adaptive LASSO	60.7567	96.0937	13.0557	**Adaptive LASSO**	48.1070	78.5502	4.0298
Subset Regression	75.2452	107.3810	17.4752	**Subset Regression**	63.0006	92.2766	5.5519
Projection Regression	72.7223	105.2908	16.3465	**Projection Regression**	59.7987	89.6835	5.2760
	Evening Peak				**Night Off-Peak**		
Model	**MAD**	**RMSE**	**MAPE**	**Model**	**MAD**	**RMSE**	**MAPE**
MF-VAR MLS	78.9431	118.5856	2.0710	**MF-VAR MLS**	47.8377	70.5075	68.4979
MF-VAR MFVB	78.0582	116.9086	2.0301	**MF-VAR MFVB**	47.3563	69.5981	67.8214
MF-VAR GS	79.1230	117.0852	1.9854	**MF-VAR GS**	47.7414	70.3040	63.3899
MF-VAR LASSO				**MF-VAR LASSO**			
$\lambda = 0.036393$	61.3004	98.0360	0.8575	$\lambda = 0.033160$	**36.5449**	**57.2421**	75.9995
$\lambda = 0.013079$	70.9618	114.9250	**0.7219**	$\lambda = 0.007484$	46.5724	69.6286	**34.7978**
Adaptive LASSO	**59.4086**	**97.0294**	0.9291	**Adaptive LASSO**	37.2777	58.7348	57.6482
Subset Regression	80.9618	122.4441	1.1081	**Subset Regression**	45.2735	67.4052	78.2747
Projection Regression	80.9158	123.0694	1.0370	**Projection Regression**	44.2874	66.7439	75.8000
RU-MIDAS-VAR MLS	79.0620	118.7199	2.0788	**RU-MIDAS-VAR MLS**	47.8871	70.5774	68.7853
RU-MIDAS-VAR LASSO				**RU-MIDAS-VAR LASSO**			
$\lambda = 0.036393$	75.1225	114.8347	2.1224	$\lambda = 0.015754$	43.1750	65.0516	68.3824
$\lambda = 0.030214$	75.2718	114.9091	2.1174	$\lambda = 0.006214$	45.0147	67.2977	39.9329
Adaptive LASSO	62.8122	108.2875	0.8411	**Adaptive LASSO**	36.8866	58.5351	75.1392
Subset Regression	80.9481	122.6194	1.1255	**Subset Regression**	45.4298	67.5162	80.0261
Projection Regression	80.9835	123.1973	1.0391	**Projection Regression**	44.2645	66.7739	74.9143

Table 3.6: Out-of-sample forecasts for peak and off-peaks time periods for lag $p = 240$. *Note*: Bold font indicates best result

et al. (2011). The MCS procedure contains a series of tests to construct a set of superior models, which contains the best model with a given confidence level. The test null hypothesis is that all models have equal predictive ability. The test is calculated for a specific loss function, which can be chosen depending on the specific forecasting task. In our study, we choose the loss function to simply be the absolute errors. Furthermore, we conduct the test for a confidence level of $\alpha = 0.05$. We have implemented the test using the `MFE Toolbox` in `MATLAB`, see Sheppard (2009). The MCS procedure selects the MF-VAR LASSO as the only model in the superior set of models for both

		MF-VAR							RU-MIDAS-VAR				
		MLS	**MFVB**	**GS**	**LASSO**	**Adaptive LASSO**	**Subset Regression**	**Projection Regression**	**MLS**	**LASSO**	**Adaptive LASSO**	**Subset Regression**	**Projection Regression**
MF-VAR	**MLS**	1	2.906	3.053	2.947	3.012	2.244	2.060	**0.746**	2.894	2.964	2.448	2.289
	MFVB	2.906	1	**-1.960**	**-0.401**	**-0.433**	3.134	3.959	-3.400	**0.142**	**0.358**	-2.506	-2.964
	GS	3.053	**-1.960**	1	**0.413**	**0.561**	3.299	3.969	-3.546	**-1.808**	-2.546	-2.851	-3.110
	LASSO	2.947	**-0.401**	**0.413**	1	**0.138**	4.039	3.776	-3.326	**-0.381**	**-0.391**	-4.081	-3.612
	Adaptive LASSO	3.012	**-0.433**	**0.561**	**0.138**	1	4.039	3.665	-3.566	**-0.543**	**-0.278**	-3.595	-3.426
	Subset Regression	2.244	3.134	3.299	4.039	4.039	1	**-1.893**	-2.341	3.277	3.619	**0.767**	3.565
	Projection Regression	2.060	3.959	3.969	3.776	3.665	**-1.893**	1	-2.060	3.957	4.287	2.755	2.229
RU-MIDAS-VAR	**MLS**	**0.746**	-3.400	-3.546	-3.326	-3.566	-2.341	-2.060	1	3.396	3.762	2.627	2.403
	LASSO	2.894	**0.142**	**-1.808**	**-0.381**	**-0.543**	3.277	3.957	3.396	1	**-0.394**	2.743	2.990
	Adaptive LASSO	2.964	**0.358**	-2.546	**-0.391**	**-0.278**	3.619	4.287	3.762	**-0.394**	1	2.493	3.118
	Subset Regression	2.448	-2.506	-2.851	-4.081	-3.595	**0.767**	2.755	2.627	2.743	2.493	1	**-1.912**
	Projection Regression	2.289	-2.964	-3.110	-3.612	-3.426	3.565	2.229	2.403	2.990	3.118	**-1.912**	1

Table 3.7: Diebold-Mariano test for comparing similarity in forecasts with different models. *Note*: Bold font indicates failure to reject the null hypothesis of no difference between forecasts with significance level $\alpha = 0.05$.

lags of $p = 24$ and $p = 240$, and the corresponding p-value of this test is 2e-4, indicating that the null hypothesis of equal predictive ability amongst all models is rejected at the 5% level.

Lag p = 24				Lag p = 240			
Overall Cross-Validation							
Model	**MAD**	**RMSE**	**MAPE**	**Model**	**MAD**	**RMSE**	**MAPE**
MF-VAR MLS	**39.0213**	**54.2849**	25.9527	**MF-VAR MLS**	43.9939	59.8975	20.7390
MF-VAR MFVB	39.0588	54.4013	26.0758	**MF-VAR MFVB**	43.9310	59.9331	20.9073
MF-VAR GS	39.0699	54.3874	**25.7957**	**MF-VAR GS**	44.4511	60.4291	20.4364
MF-VAR LASSO				**MF-VAR LASSO**			
$\lambda = 0.000213$	39.3532	54.6452	26.5770	$\lambda = 0.003174$	**39.1645**	**55.5124**	19.0060
$\lambda = 0.000092$	39.4099	54.6039	26.0981	$\lambda = 0.001374$	40.1439	55.9094	**18.3995**
Adaptive LASSO	39.5322	55.4278	26.5786	**Adaptive LASSO**	39.5543	55.9071	19.6482
Subset Regression	45.1129	64.4887	48.1684	**Subset Regression**	43.3860	61.8194	20.8003
Projection Regression	43.9946	63.5313	46.1500	**Projection Regression**	43.1366	61.7508	20.5415
RU-MIDAS-VAR MLS	39.0448	54.2908	25.9188	**RU-MIDAS-VAR MLS**	43.9905	59.8949	20.6971
RU-MIDAS-VAR LASSO				**RU-MIDAS-VAR LASSO**			
$\lambda = 0.000092$	39.0548	54.4545	27.0251	$\lambda = 0.001817$	39.7438	56.0688	19.8824
-	-	-	-	$\lambda = 0.000653$	40.8673	56.5273	19.2821
Adaptive LASSO	39.5127	55.5343	26.6036	**Adaptive LASSO**	40.0017	55.9823	19.6184
Subset Regression	44.5340	63.9418	47.6400	**Subset Regression**	43.2785	61.8100	20.7348
Projection Regression	44.1214	63.5890	46.2788	**Projection Regression**	43.2121	61.7507	20.8290
Morning Peak Cross-Validation							
Model	**MAD**	**RMSE**	**MAPE**	**Model**	**MAD**	**RMSE**	**MAPE**
MF-VAR MLS	43.8639	**58.2397**	10.0080	**MF-VAR MLS**	51.0876	65.7240	8.1383
MF-VAR MFVB	43.9907	58.2995	10.3847	**MF-VAR MFVB**	50.7665	65.5618	8.4403
MF-VAR GS	44.0088	58.5488	**9.8737**	**MF-VAR GS**	51.5391	66.3977	7.2382
MF-VAR LASSO				**MF-VAR LASSO**			
$\lambda = 0.000092$	44.7242	59.1166	10.4724	$\lambda = 0.002188$	**46.2319**	**60.9060**	7.5837
-	-	-	-	$\lambda = 0.014065$	52.3180	68.4850	5.4996
Adaptive LASSO	**43.8443**	59.0427	10.0107	**Adaptive LASSO**	47.0004	60.5723	6.2794
Subset Regression	57.2532	73.8674	13.1330	**Subset Regression**	55.0560	71.3817	11.0370
Projection Regression	57.8610	74.6463	12.5676	**Projection Regression**	55.4509	71.7095	11.2152
RU-MIDAS-VAR MLS	43.8844	58.2552	9.9796	**RU-MIDAS-VAR MLS**	51.0863	65.7137	8.0282
RU-MIDAS-VAR LASSO				**RU-MIDAS-VAR LASSO**			
$\lambda = 0.000092$	44.3937	58.9833	10.3828	$\lambda = 0.000947$	46.9679	61.5296	8.2155
-	-	-	-	$\lambda = 0.009694$	53.0959	69.4049	**5.1654**
Adaptive LASSO	44.0027	59.4278	11.8893	**Adaptive LASSO**	47.4248	61.1734	6.2672
Subset Regression	56.6396	73.3442	12.7472	**Subset Regression**	55.5680	71.8681	11.5234
Projection Regression	56.1396	72.9728	13.0419	**Projection Regression**	55.4070	71.6640	11.7774
Evening Peak Cross-Validation							
Model	**MAD**	**RMSE**	**MAPE**	**Model**	**MAD**	**RMSE**	**MAPE**
MF-VAR MLS	52.7036	69.5604	0.6461	**MF-VAR MLS**	55.7896	73.6556	0.8524
MF-VAR MFVB	52.6321	**69.4464**	0.6617	**MF-VAR MFVB**	55.9292	73.8143	0.8721
MF-VAR GS	52.7788	69.7454	0.6286	**MF-VAR GS**	56.5841	74.1283	0.8266
MF-VAR LASSO				**MF-VAR LASSO**			
$\lambda = 0.000194$	52.5444	69.6440	0.6617	$\lambda = 0.004196$	49.8996	68.4681	0.6113
$\lambda = 0.000092$	52.6011	69.6018	0.6581	$\lambda = 0.015436$	53.1975	72.7117	**0.5464**
Adaptive LASSO	52.6772	69.7427	0.6613	**Adaptive LASSO**	**49.8631**	**68.3418**	0.5994
Subset Regression	61.8930	81.6559	0.8904	**Subset Regression**	53.5917	73.3821	0.9414
Projection Regression	61.0245	80.6504	0.8715	**Projection Regression**	54.4470	74.4293	0.9545
RU-MIDAS-VAR MLS	52.7315	69.5781	**0.6460**	**RU-MIDAS-VAR MLS**	55.7557	73.6337	0.8530
RU-MIDAS-VAR LASSO				**RU-MIDAS-VAR LASSO**			
$\lambda = 0.000234$	52.5322	69.7340	0.6729	$\lambda = 0.001817$	51.0066	68.5294	0.7254
$\lambda = 0.000101$	52.6387	69.7206	0.6629	$\lambda = 0.010640$	53.2158	72.8385	0.5546
Adaptive LASSO	**52.5016**	69.5445	0.6602	**Adaptive LASSO**	51.0101	68.5017	0.6478
Subset Regression	61.1652	81.0363	0.8681	**Subset Regression**	54.1344	73.7514	0.9815
Projection Regression	61.6345	81.4736	0.8602	**Projection Regression**	54.1406	73.7965	0.9665
Afternoon Off-Peak Cross-Validation							
Model	**MAD**	**RMSE**	**MAPE**	**Model**	**MAD**	**RMSE**	**MAPE**
MF-VAR MLS	39.8537	54.4063	2.9195	**MF-VAR MLS**	47.5721	62.9285	2.9189
MF-VAR MFVB	**39.7880**	54.4804	3.0612	**MF-VAR MFVB**	47.4596	63.0013	3.0012
MF-VAR GS	39.9843	54.6214	**2.8014**	**MF-VAR GS**	48.3398	64.0571	2.7732
MF-VAR LASSO				**MF-VAR LASSO**			
$\lambda = 0.000092$	40.7583	54.9285	3.0502	$\lambda = 0.039136$	**42.2081**	60.2456	2.4681
-	-	-	-	$\lambda = 0.011677$	43.9839	**60.0132**	**1.9250**
Adaptive LASSO	40.8676	56.0234	2.9358	**Adaptive LASSO**	43.0138	60.3792	2.3438
Subset Regression	45.7822	63.7576	4.2465	**Subset Regression**	45.8491	63.8941	3.5062
Projection Regression	44.6630	62.6429	4.3299	**Projection Regression**	45.7327	63.5678	3.5471
RU-MIDAS-VAR MLS	39.8304	**54.3609**	2.9204	**RU-MIDAS-VAR MLS**	47.5807	62.9313	2.9191
RU-MIDAS-VAR LASSO				**RU-MIDAS-VAR LASSO**			
$\lambda = 0.000092$	40.2548	54.7857	3.0817	$\lambda = 0.026975$	42.6140	61.0330	2.5166
-	-	-	-	$\lambda = 0.004196$	42.7366	58.9088	2.0225
Adaptive LASSO	40.5783	54.3715	3.0957	**Adaptive LASSO**	42.5791	61.5725	2.4816
Subset Regression	45.0336	63.1032	4.2150	**Subset Regression**	45.7977	63.8068	3.5176
Projection Regression	44.3662	62.3619	4.3749	**Projection Regression**	45.6615	63.6501	3.4872
Night Off-Peak Cross-Validation							
Model	**MAD**	**RMSE**	**MAPE**	**Model**	**MAD**	**RMSE**	**MAPE**
MF-VAR MLS	33.3966	44.9116	49.5500	**MF-VAR MLS**	36.3687	48.4260	39.1064
MF-VAR MFVB	33.5049	45.0345	49.6254	**MF-VAR MFVB**	36.3521	48.4165	39.3302
MF-VAR GS	33.3630	44.8812	**49.3355**	**MF-VAR GS**	36.5378	48.5702	38.7799
MF-VAR LASSO				**MF-VAR LASSO**			
$\lambda = 0.018521$	32.3393	46.0774	68.6164	$\lambda = 0.003174$	**32.0412**	**44.2983**	36.0362
$\lambda = 0.000372$	33.1203	44.8136	51.7147	$\lambda = 0.001374$	32.5423	44.5084	**34.5579**
Adaptive LASSO	33.1236	45.9539	52.0444	**Adaptive LASSO**	32.2955	44.6052	35.7815
Subset Regression	37.1682	51.9024	94.5028	**Subset Regression**	35.6758	49.3582	36.3426
Projection Regression	37.4977	52.3368	95.8479	**Projection Regression**	35.5013	49.1460	38.0939
RU-MIDAS-VAR MLS	33.4498	44.9400	49.4855	**RU-MIDAS-VAR MLS**	36.3638	48.4273	39.0438
RU-MIDAS-VAR LASSO				**RU-MIDAS-VAR LASSO**			
$\lambda = 0.012766$	**32.2032**	45.9315	69.5419	$\lambda = 0.024579$	32.1665	45.8040	55.9070
$\lambda = 0.000213$	32.9623	**44.6521**	52.6003	$\lambda = 0.000542$	33.5490	45.3027	36.1274
Subset Regression	36.7217	51.4238	93.0253	**Subset Regression**	35.4182	49.1436	37.9251
Projection Regression	36.3093	51.1871	90.0338	**Projection Regression**	35.4040	49.1487	38.0873

Table 3.8: One-week cross-validation results for lags $p = 24$ (left) and $p = 240$ (right). *Note*: Bold font indicates best result

3.4 Conclusion

The aim of this study was to study the half-hourly electricity price forecasting capability of an MF-VAR model which incorporated weather data and demand. To our knowledge, this study is the first to study the dynamics of intra-day electricity prices using the MF-VAR framework. Our forecasting results show that incorporating weather data and demand is capable of replicating several key characteristics observed in electricity prices; persistent volatility, spikes, and mean-reversion. Our out-of-sample density plots show that the MF-VAR model captured the lower and upper tails of actual price density well, however, there was a mismatch around median prices. The VAR is a capable framework for electricity price forecasting but additional data relating to market dynamics and macroeconomic factors may improve our forecasting capabilities.

Furthermore, we applied several techniques for estimating an MF-VAR framework. In particular, we used the mean-field variational Bayes as an approximation for Gibbs sampling estimates. Our results were promising and suggested that the MFVB provides impressive approximations to the Gibbs sampler estimates when applied to electricity prices in New Zealand, and within a VAR framework. The MFVB, therefore, proves to be a useful alternative Bayesian tool, particularly for large sets of data and parameters, due to the rapid computational time. In particular, while Gibbs sampling with 100,000 iterations took approximately 45 minutes, the VB ELBO converged in nearly 30 seconds, with a convergence threshold of 10^{-7}. This was a significant improvement in computational time, with little different in the forecast results.

We also discussed an alternative mixed-frequency model, the RU-MIDAS model. We showed that easing the error restriction in this model allows us to specify the RU-MIDAS within a VAR framework, and we refer to it as the RU-MIDAS-VAR model. This study is the first to study this relationship. Furthermore, we show that the RU-MIDAS-VAR closely resembles the MF-VAR, however, it uses more up-to-date

information for its lags. Despite this, our forecast results show that the MF-VAR and the RU-MIDAS-VAR provide similar price predictions, with the MF-VAR performing marginally better.

We also incorporate the LASSO, adaptive LASSO, and random subspace methods as variable selection tools for our models. Our forecast results show that the LASSO performs extremely well for both the MF-VAR and RU-MIDAS-VAR models within the New Zealand electricity market. The LASSO and adaptive LASSO outperform all other methods in nearly all metrics, particularly so when a larger parameter set is used. The LASSO also allows us to identify the optimal parameter set for a specific criterion. The random subspace methods as dimension reduction techniques performed better than least-squares estimates for a larger parameter set, however, they did not perform well otherwise. They may still have useful practical applications for studies with large datasets and parameter sets.

Finally, we extended our forecasting studies by observing forecasts for specific periods in a day, i.e., peak and off-peak periods. The models generally perform better for off-peak periods, as one would expect, since peak periods exhibit much higher volatility and spike frequency. However, to adjust for this, it may be beneficial to test heavy-tailed distribution such as the generalized Pareto distribution or to incorporate stochastic volatility within a VAR framework.

A key finding in this study was the improvement in forecasting accuracy with the implementation of LASSO as a feature selection tool. Interestingly, a recent successful class of models for forecasting electricity prices is a simple linear regression model with an extensive feature set and a rigorous feature selection methodology. One of these models is the LASSO-estimated auto-regressive (LEAR) model, see Lago et al. (2021). We found extensive literature in the electricity price forecasting space focusing on the use of machine learning models with large feature sets and feature selections methods. However, apart from the LEAR model, there is lacking literature on the use

of large feature sets with heteroskedastic models. This was our motivation for the study presented in the next chapter. This study will focus on the daily electricity price forecasting problem and implements GARCH and stochastic volatility models, with a large feature set consisting of weather data, and market-related variables. Furthermore, several feature selection tools are analyzed for these models. The idea is to build a parsimonious set of features for GARCH and stochastic volatility models, and compare forecasting performance with the LEAR model, and other machine learning models.

Chapter 4

Electricity Price Forecasting in New Zealand: A Comparative Analysis of Statistical and Machine Learning Models with Feature Selection

4.1 Introduction

4.1.1 New Zealand Electricity Market and Price Forecasting

The New Zealand electricity market (NZEM) is a wholesale market operated by the Electricity Authority (EA), which is the independent regulator of the electricity industry in New Zealand. The EA is responsible for ensuring that the market is fair and efficient and that the market rules are followed. Unlike most European markets, the NZEM follows a real-time market design, where prices are determined every 30 minutes based on the supply and demand of electricity in that interval. Generators submit offers for the next half-hour period, through a Wholesale Information and Trading System (WITS).

The offer consists of a specified quantity of electricity generation at a nominated price. The system operator, Transpower, uses a scheduling, pricing, and dispatch (SPD) method to rank the generation offers in order of price. The SPD method is a merit-order dispatch, where the cheapest generation offers are dispatched first, until all demand for the period is satisfied. The highest-priced bid, offered by a generator, required to meet demand for a given half-hour period is set as the spot price for that trading period.

The NZEM is characterized by a high penetration of renewable energy sources, such as hydro and wind power. In 2021, 82.1% of electricity was generated using renewable sources, see Ministry of Business, Innovation and Employment (2021). While this is the right step to reduce the carbon footprint, it also results in a significant increase in the variability of electricity prices, due to the high dependence on suitable climate conditions for hydro and wind power generation. Furthermore, the New Zealand government has a strong focus on achieving 100% renewable electricity generation by 2030. As a result, it becomes increasingly crucial to develop accurate and reliable electricity price forecasting models to help the electricity market participants make informed decisions.

Electricity price forecasting (EPF) has been a topic of interest for more than two decades now, since the deregulation trend of electricity markets around the world. There have been numerous models and methodologies implemented in EPF literature. Literature on EPF models generally falls under one of five categories; game theory models, fundamental models, reduced-form models, statistical models, and machine learning models. In this study, we focus on the comparison between statistical models and machine learning models as they have seen the most effective results. See Weron (2014); Hong et al. (2020); Lago et al. (2021) for comprehensive reviews of advances in EPF in the last two decades.

In this study, we focus on the daily EPF problem for the New Zealand electricity market, which is a real-time market. We focus on daily prices since many of these

explanatory features are not available on a lower frequency, and we do not wish to conduct a mixed-frequency study, since the primary focus of this research is the comparison between GARCH, SV, LEAR, and machine learning benchmarks. Daily prices are calculated as the average of half-hourly prices on a given day. We forecast prices for the five major regions in New Zealand, the upper North Island (UNI), the central North Island (CNI), the lower North Island (LNI), the upper South Island (USI), and the lower South Island (LSI). These regions encompass the whole of New Zealand and several metrics in the New Zealand electricity market are represented based on these regions. The full dataset consists of prices from 01/01/2014 to 31/10/2022. We apply a rigorous time-series cross-validation scheme when producing forecasts, to obtain robust estimates. A key consideration in our work is the incorporation of numerous explanatory features relating to the New Zealand electricity market. This includes load, generation by fuel type, forward prices from the derivatives market, reserve prices, transmission information, and weather features. Furthermore, the weather features include temperature, precipitation information, humidity, wind information, dew, cloud coverage, and solar energy. We construct a unique set of features for each region mentioned above. To our knowledge, there has not been such a comprehensive study of EPF in the New Zealand real-time market, which is characterized by a high concentration of renewable energy sources.

With a large set of explanatory features, we identify several groups of highly correlated features. To remedy this, we incorporate several feature selection techniques. In particular, we apply the LASSO, mutual information (MI), and recursive feature elimination (RFE) feature selection methods individually on all models described above apart from the LEAR model. We test the performance of each model with each of these methods, as a consistent measure of their performance. One of our key findings suggests that incorporating LASSO with GARCH-t (LE-GARCH-t) improved the GARCH model's predictions by up to 40% over GARCH-t with all features. We see

similar results with LE-GARCH, LE-SV-*t*, LE-SV, and also variants of these models using MI. We also conduct Diebold-Mariano tests, see Diebold & Mariano (2002), on the model forecasts. We find the models mentioned to be consistently in the group of top-performers, along with the LEAR benchmark.

4.1.2 Electricity Price Forecasting using GARCH and Stochastic Volatility Models

The well-known GARCH model of Bollerslev (1986) is well-suited for heteroskedastic time series data. Its ability to model the conditional variance of a time series has made it a popular choice in financial econometrics. However, to no surprise, it has seen many application in EPF literature as well, due to the presence of heteroskedasticity in electricity prices, Tan et al. (2010).

Knittel & Roberts (2005) provided one of the first applications of GARCH models to EPF. In particular, they successfully implemented an AR-EGARCH specification to model California energy prices during a crisis period in 2000. Generally, the earlier implementations of GARCH models, from approximately mid-2000s to early 2010s, were shown to be effective when compared with other models, for example, see Diongue et al. (2009); Wu & Shahidehpour (2010).

Most implementations of the GARCH specification in EPF involve some form of hybrid setting with other statistical models, most commonly, an AR, ARMA, or ARIMA specification. For example, Tan et al. (2010) use a wavelet transform to decompose electricity prices into subseries, and predict them using an ARIMA-GARCH model. Wu & Shahidehpour (2010) present an adaptive wavelet neural network with an ARMAX-GARCH specification for the PJM market. Most recently, Lago et al. (2018) compare the performance of an ARIMA-GARCH models along with 26 other models, for predictions

in several EPEX markets. Their results suggest that the performance of the ARIMA-GARCH and other traditional statistical models is worse than that of the machine learning models they used in their study. However, importantly, they considered an ARIMA-GARCH model without the implementation of exogenous variables, whereas most of the machine learning models had access to these variables.

While there is vast literature on the use of GARCH models for EPF, only a small percentage of those implement GARCH with exogeneous variables. Wu & Shahidehpour (2010) consider an ARMAX-GARCH model, however they only have a single explanatory variable, which is electricity load. Gianfreda & Grossi (2012) employ a Reg-ARFIMA-GARCH model with several fundamental explanatory variables in the Italian electricity market. Huurman et al. (2012) study the prediction capabilities of ARIMAX-GARCH and ARIMAX-GARCHX frameworks using weather variables. All of these studies suggest that incorporating exogoneous variables improve model forecasting capabilities.

To our knowledge, newer studies, which provide GARCH specifications as a benchmark for statistical models, have rarely considered GARCH with exogeneous variables or thorough feature selection. This is likely due to the primary focus on machine-learning methods, as well as the notion that introducing exogeneous variables with GARCH is unnecessary since the variables will be able to capture the heteroskedastic behavior of electricity prices, see Karakatsani & Bunn (2010). However, in this study, we find that the inclusion of exogoneous variables in GARCH models can improve forecasting performance, and that feature selection is important for GARCH models as well. In this study, we implement the GARCH(1,1) model and the GARCH(1,1)-*t* model. Both variants are estimated using the `ARCH` package in `Python`, see Sheppard et al. (2022), which implements a maximum likelihood estimation technique.

As a comparison for GARCH, we consider a stochastic volatility (SV) framework, where the volatility is specified as a latent stochastic process, as proposed by Taylor

(1994). They are unlike the GARCH specification, where the evolution of volatility is deterministic. Despite the early evidence in favor of SV models, see Ghysels et al. (1996); S. Kim et al. (1998), these models have not found comparable success in the field of time-varying volatility modeling. This is largely due to the estimation difficulty and incompatibility of estimation methods with different SV frameworks, see Kastner & Frühwirth-Schnatter (2014). As a comparison, the GARCH family has numerous variants, and requires only a few tweaks in the estimation procedure.

SV models have not seen significant implementation or success in EPF, however, they have been applied in financial econometrics settings and commodities markets, see Chan (2013). Furthermore, there are several studies comparing the forecasting capabilities of GARCH and SV models. For example, Chan & Grant (2016) provide a comprehensive study of forecasting oil, petroleum products and natural gas prices using several GARCH and SV variants. They conclude that the SV models almost always outperform their GARCH counterparts. In another study, Tiwari et al. (2019) model the dynamics of Bitcoin and Litecoin using GARCH and SV models. They also observe that the SV models consistently outperform GARCH models. They also suggest that the *t*-distribution variants of GARCH and SV models show better results. For these reasons, we include SV models in this study as a comparison for GARCH models. We consider the standard SV model as well as the SV-*t* variant using the `stochvol` package in `R`, see Kastner (2019), which implements a Bayesian Markov chain Monte Carlo (MCMC) estimation technique.

4.1.3 Popular Models in Electricity Price Forecasting Literature

In recent years, there have been several successful implementations of linear regression models with numerous input features for EPF. It has been observed that performing regularization techniques, such as the least absolute shrinkage and selection operator

(LASSO) of Tibshirani (1996), to reduce the feature set, can significantly improve forecasting performance, see Uniejewski et al. (2016); Lago et al. (2018); Ziel & Weron (2018); Uniejewski & Weron (2018b). The LASSO adds a penalty term to the objective function of the regression model, which is proportional to the sum of the absolute values of the coefficients. This encourages the coefficients to be small and some coefficients to be zero, and thus, reduces the number of features used in the model. Such an approach is referred to as the LASSO-estimated autoregressive (LEAR) model in Lago et al. (2021). Due to the success of the LEAR model, and its simplicity in theory, we consider it a benchmark in this study, as suggested by Lago et al. (2021).

Our second benchmark model is a four-layer deep neural network (DNN) with two-hidden layers, a natural extension of the traditional multilayer perceptron. DNNs are simple but powerful models that form the basis for other advanced machine learning models. They have seen much use in EPF, and often outperform more advanced machine learning models. For example, Lago et al. (2018) provide a thorough study of 27 models, of which 15 are statistical models and 7 are machine learning models. They find the DNN, the long short-term memory (LSTM) model, and the gated recurrent unit (GRU) model, to obtain a predictive accuracy that is better than all other models. Furthermore, of these three, they find that DNN is the best performing. For other application of DNN in EPF, see Luo & Weng (2019); Atef & Eltawil (2019).

We also consider two recurrent neural network (RNN) models in this study, particularly the long short-term memory (LSTM) model Hochreiter & Schmidhuber (1997), and the gated recurrent unit (GRU) Cho et al. (2014). They are both capable of learning long-term dependencies in time series data, and are able to capture the temporal structure of the data. Both models have seen successful applications in energy-related literature, see Coelho et al. (2017); Fan et al. (2017); H.-z. Wang et al. (2017). As mentioned previously, Lago et al. (2018) find that LSTM, GRU, and DNN are the best performing models. In their study, the LSTM and GRU models are hybrid models

combined with a DNN. For applications of LSTM and GRU in EPF, see Chang et al. (2019); C.-J. Huang et al. (2021).

Finally, we also consider the extreme gradient boosting (XGBoost) model Chen & Guestrin (2016). The XGBoost model is a popular ensemble of regression trees Breiman (2017), based on the principle of boosting, which is a sequential technique for constructing an optimal combination of weak learners.

In this study, we perform hyperparameter optimization for the DNN, LSTM, GRU, and XGBoost models using the tree-structured Parzen estimator (TPE) algorithm Bergstra et al. (2011). TPE is a Bayesian optimization algorithm that is able to efficiently search the hyperparameter space. It is a sequential model-based optimization technique that uses a Gaussian process to model the objective function. It is able to efficiently search the hyperparameter space in order to find the optimal hyperparameters in few iterations. We implement the TPE algorithm using the `hyperopt` package in `Python`, see Bergstra et al. (2015).

4.1.4 Feature Selection Methods

In this study, we consider three feature selection methods, the LASSO, mutual information (MI), see Kozachenko & Leonenko (1987), and recursive feature elimination (RFE), see M.-L. Huang et al. (2014). Each of these techniques is individually applied to each of the models described above, apart from the LEAR model, since it has already undergone LASSO regularization.

MI is a popular choice for feature selection in EPF. For example, Gholipour Khajeh et al. (2018) implement an iterative two-stage feature selection technique using MI and correlation analysis for optimal selection of lagged prices. They implement this technique within a DNN to make short-term price prediction in the PJM electricity market. As another example, Ebrahimian et al. (2018) apply MI to select optimal

features from their set of lagged prices, load, and available generation.

Relating to RFE, Naz et al. (2019) implement the method on a set of 16 features applicable to retail electricity usage for price and load forecasting. In another study, Shao et al. (2017) take a hybrid approach to combining the RFE using a support vector machine (SVM) estimator. They also use MI as a pre-processing technique.

All of these methods have seen some success in EPF, particularly the LASSO. As a result, it will be an interesting comparison to see how these methods perform in conjunction with the models.

4.1.5 Motivation and Contributions

Machine learning models have seen numerous successful implementations in EPF. They are very capable tools for extracting patterns from non-linear time-series. H. Wang et al. (2019) provides a thorough review of machine learning models and methodologies for EPF. However, as Lago et al. (2021) points out, comparisons between statistical models and machine learning models have been limited. To be specific, advanced studies have not provided fair comparisons between these categories of models. Advanced machine learning models are often compared with simple statistical models, and without thorough feature engineering, feature selection, or cross-validation.

We consider this to be a crucial gap in literature. In particular, we find that the application of GARCH and SV models as comparisons for novel machine learning techniques in EPF has been limited to their basic forms, or at most, in a hybrid form with ARIMA, or other statistical models. We have found a lack of research works pertaining to GARCH and SV models with exogenous features, which is a key component of our study. Furthermore, when exogenous features are included, they are often not selected or engineered in a rigorous manner. This is in contrast to the extensive feature engineering and selection that is applied to machine learning models. This lack of rigor

is abundant in the literature, and is a key motivation for this study.

To resolve this gap in literature, this study provides a fair comparison of the auto-regressive mean with generalized autoregressive conditional heteroskedasticity volatility (GARCH) and auto-regressive mean with stochastic volatility (SV) models and their *t*-distribution variants with a variety of machine learning models, including the long short-term memory (LSTM) model, the gated recurrent unit (GRU), extreme gradient boosting (XGBoost), the LASSO-estimated auto-regressive (LEAR) model, and a four-layer deep neural-network (DNN).

Following the guidelines of Lago et al. (2021), we consider the LEAR and DNN models as benchmarks in this study. The least absolute shrinkage and selection operator (LASSO) as a feature-selection tool has been effective in many applications with a considerable number of explanatory variables, including EPF. Similarly, simple DNNs also provide an effective basis to compare with traditional and more advanced machine learning models. Whether LEAR is a machine learning model or statistical model is a topic for debate, since auto-regressive models are generally statistical models, however regularization is seen as a machine learning technique by some. Regardless of its classification, the model is a consistent benchmark for both categories of models.

We employ the GARCH and SV frameworks, in particular, due to their ability to capture time-varying volatility dynamics, which is abundantly present in electricity prices, see Tan et al. (2010). GARCH models have been applied in several forecasting studies, however, their results suggest they perform no better than simpler AR models, for example, see Lago et al. (2018). In another study, Karakatsani & Bunn (2010) suggest that the effectiveness of GARCH diminishes when the fundamental drivers of electricity price volatility are accounted for using explanatory variables. However, our results contradict this statement. It is true that the GARCH and SV models severely underperform the benchmarks when they are overburdened with explanatory features. However, with a parsimonious set of features, we find that the GARCH and SV models

outperform the benchmarks (LEAR and DNN models), on average, by 2% and 3% in terms of the symmetric mean absolute percentage error (sMAPE) and mean absolute scaled error (MASE), respectively.

It is important to discuss a potential shortcoming of our study. In particular, we focus on simpler machine learning models. We primarily do this because past literature, such as Lago et al. (2018, 2021), suggest that simpler base models such as LEAR, DNN, and LSTM, can outperform newer, complex models. However, recent research also speaks of the usefulness of transformers for time-series forecasting, which we have excluded from our study. On the other hand, we also employ fairly simple base statistical models in the form of GARCH and SV. We believe the level of complexity between the statistical and machine learning models in our study is comparable, and thus, provides a fair comparison.

With these contributions, we hope to provide a fair and comprehensive comparison between statistical and machine learning models in EPF.

4.1.6 Structure

The remainder of this chapter is structured as follows. Section 4.2 introduces the data and discusses the pre-processing and feature selection techniques. Section 4.3 describes the models used in this study. Section 4.4 presents the forecasting results of our study and provides a discussion. Finally, Section 4.5 concludes the study.

4.2 Data and Preprocessing Techniques

4.2.1 Price Data

The data consists of daily electricity prices across New Zealand from 01/01/2014 to 31/10/2022. In particular, the price data, as well as some of the features, are separated into five regions across New Zealand, the upper North Island (UNI), the central North Island (CNI), the lower North Island (LNI), the upper South Island (USI), and the lower South Island (LSI). We attempt predictions for all five regions in separate univariate frameworks. These regions together encompass the entire New Zealand electricity grid. We have chosen to study these regions, rather than specific nodes, since they provide broader insights into the behaviour of electricity prices within each region. Furthermore, several metrics relating to the electricity market, provided officially by the NZ electricity authorities, are separated by these regions. Electricity price and demand vary in each region based on population density and industrial activity. For example, much of the industrial demand for electricity is present in the CNI, whereas the LSI hosts a large amount of generation capacity.

	UNI	CNI	LNI	USI	LSI
mean	107.45	102.85	101.29	103.59	95.37
std	73.81	73.51	89.54	76.82	69.85
skew	3.05	3.32	15.36	2.86	2.60
kurtosis	26.92	27.61	505.24	20.10	16.60
min	0.02	0.02	0.02	0.02	0.02
25%	60.91	57.86	56.54	55.05	50.63
50%	84.10	80.27	79.39	80.85	74.11
75%	131.37	124.87	122.10	127.64	117.81
max	1330.57	1237.19	3288.31	1250.05	1092.47

Table 4.1: Descriptive statistics of the electricity price data. All values are in NZD/MWh.

The descriptive statistics of the electricity price data are presented in Table 4.1.

All regions typically behave similarly, however, we observe that LNI has much larger skewness and kurtosis, indicating that this region is much more prone to positive jumps in prices. The LNI region hosts several lakes and is attributed with the highest generation using hydro fuel, so it is not surprising that the region has a higher skewness and kurtosis.

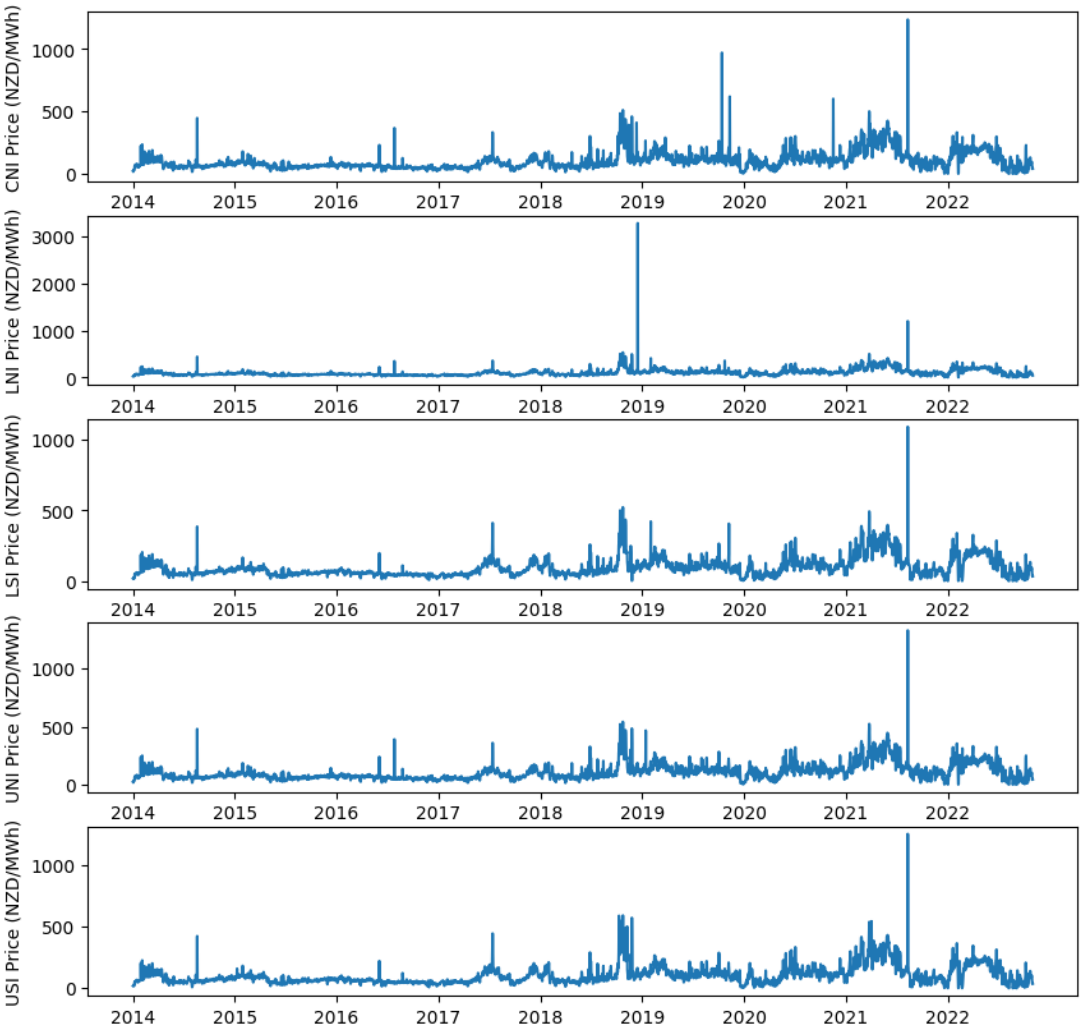


Figure 4.1: Time-series of aggregated daily prices for each region.

Figure 4.1 displays the time-series of aggregated daily prices for each region. We observe that the prices exhibit a high degree of volatility, with several spikes in the data. The prices are also highly correlated with each other. Some of the key characteristics of

the price data which are evident from the time-series plots are extremely high volatility, clustering of high prices, and tendency for price jumps. These characteristics are typical of electricity price data, and are a result of the complex interactions between supply and demand in the electricity market. Another characteristic, not visible from the time-series plots, is the presence of seasonality on the daily, weekly, and monthly scales. The motivation to use GARCH and SV models as the benchmark for statistical models is to capture the volatility clustering and persistence in the data. However, more important than the model architecture, is the feature selection process, which is crucial for capturing the complex interactions between the price data and the features. With a parsimonious set of features, the forecasts from our statistical models can be significantly improved. Also, the pattern recognition capabilities of machine learning models can be leveraged to capture the non-linear relationships between the price data and the features.

4.2.2 Features

In this study, we include a variety of features to improve the performance of the models. The time period for all features is the same as that for price data, i.e., from 01/01/2014 to 31/10/2022. Additionally, all features are available at a daily interval. Some of the features are segmented into five individual series, corresponding to the five regions mentioned above. The features under consideration are:

- **Electricity load (MWh)**: Average daily electricity load (demand) for each region under consideration.
- **Generation by fuel type (MWh)**: Average daily generation for each of the following fuels: hydro, wind, coal, gas, geothermal, diesel, and wood. This data is not segmented for each region, it is an average across New Zealand for each fuel type.

- **Reserve prices (NZD/MWh)**: Reserves are generation capacity that is made available to be used in the event of a sudden failure of a generating or transmission facility in order to maintain system frequency at 50 Hz. Fast instantaneous reserve (FIR) is available within six seconds and must be able to operate for one minute. Sustained instantaneous reserve (SIR) is available within 60 seconds and must be available for 15 minutes. FIR and SIR prices are segmented into two regions: North Island and South Island.

- **Forward prices (NZD/MWh)**: Forward prices are taken from the New Zealand electricity derivatives market, which is listed under the Australian Securities Exchange (ASX). Forward prices indicate the average price of electricity for the time period under consideration. We consider quarterly contracts, meaning that forward prices are estimates of average daily prices for specific calendar quarters. We also consider base contracts rather than peak contracts, meaning that all 24 hours in a day are considered when averaging. The following contract schemes are used as features:

  - **All maturities**: Average settlement price of all contracts currently being traded.
  - **Short-dated maturities**: Average settlement price of contracts maturing within the next 12 months of current date.
  - **Long-dated maturities**: Average settlement price of contracts maturing more than 12 months from current date.
  - **Year+1 maturities**: Average settlement price of contracts maturing in the year preceding current year.
  - **Year+2 maturities**: Average settlement price of contracts maturing in the second year preceding current year.

  - **Year+3 maturities**: Average settlement price of contracts maturing in the third year preceding current year.

  The forward prices are available for two specific nodes in the New Zealand national grid: Otahuhu in the North Island, and Benmore in the South Island. Therefore, each of the features described above are available for each of the two nodes.

- **HVDC transfer (GWh)**: The daily net energy transferred across the HVDC link that connects Benmore in the South Island to Haywards in the North Island. Positive values indicate net northward flow to Haywards, negative values indicate net southward flow to Benmore.

- **Weather data**: We have acquired daily weather data from several weather stations across New Zealand. Each of the variables described below are segmented into the five regions mentioned above. We collected all weather data from several weather stations across New Zealand, and aggregated the data based on their location for each of the five regions mentioned above. The variables are:

  - **Wind Speed (kph)**: Average daily wind speed.
  - **Wind Direction (degrees)**: Average daily wind direction.
  - **Precipitation (mm)**: Average daily precipitation.
  - **Precipitation Coverage (%)**: Average daily precipitation coverage. Precipitation coverage is the proportion of the day that measurable precipitation occurs.
  - **Temperature (C)**: Average daily temperature.
  - **Dew (C)**: Average daily dew.
  - **Humidity (%)**: Average daily humidity.

- **Cloud Cover (%)**: Average daily cloud cover. Cloud cover is the proportion of sky near the region that is covered by clouds.
- **Solar Energy (MJ/m^2)**: Average daily solar energy.

Again, each of these weather variables is available for each of the five regions under consideration. Additionally, we include average wind speed and wind direction specifically around the wind farms across New Zealand and precipitation and precipitation coverage across hydro dams and lakes across New Zealand as four additional features.

A correlation heatmap of the price data and features is presented in Figure 4.2. For clarity purposes, we provide broad labels for all features, with smaller labels indicating the segmented data. We observe that several groups of features are highly correlated with each other. Particularly, the forward prices exhibit multicollinearity, as do some of the weather variables.

4.2.3 Data Preprocessing

The price data and features are all standardized using one of three techniques, depending on original probability distribution. Series, which exhibit heavy-tailed distributions, are first standardized using either the Box-Cox transformation or the Yeo-Johnson transformation, depending on whether the data is strictly positive or not, and are then scaled using min-max scaling. Series, which exhibit close to normal distribution properties, are simply scaled using min-max scaling. As a special case, wind direction, which is originally measured in degrees, is first transformed using a sine transformation, and then scaled using min-max scaling. To further elaborate on the transformations, we find that electricity prices, coal generation, diesel generation, reserve prices, forward prices, and precipitation series tend to display a skewed distribution, with median values

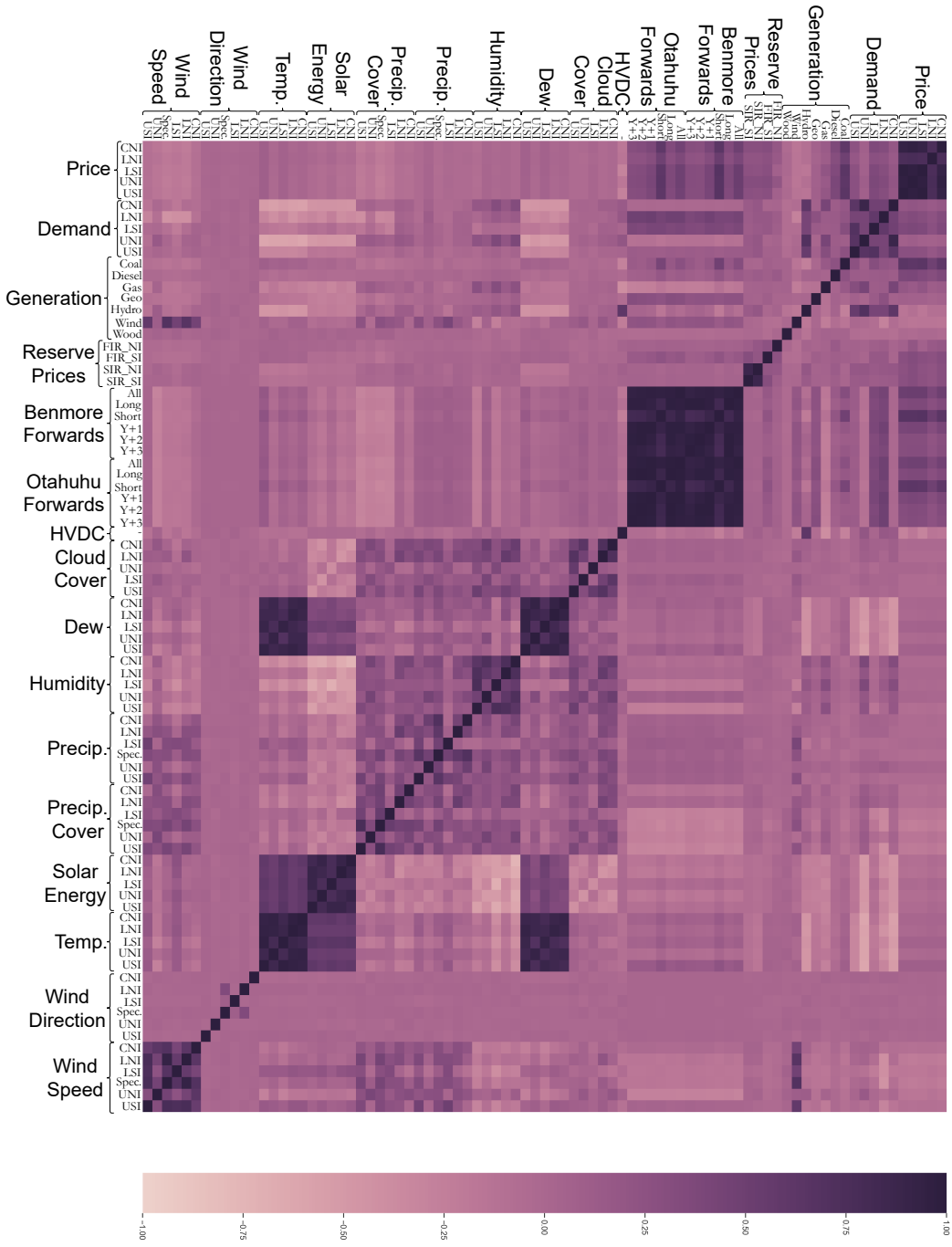


Figure 4.2: Correlation heatmap of price and features.

being far from the mean. This suggests a higher tendency for extreme values, due to which, we choose to transform them using the Box-Cox or Yeo-Johnson logarithmic transformations. The remaining features typically have symmetric distributions, and it

suffices to simply scale them using min-max scaling. The transformations and min-max scaling are defined as follows:

$$
\begin{aligned}
\text{Box-Cox: } x' &= \begin{cases} \frac{x^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln(x) & \text{if } \lambda = 0 \end{cases} \\
\text{Yeo-Johnson: } x' &= \begin{cases} \frac{(x+1)^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0, x \geq 0 \\ \ln(x+1) & \text{if } \lambda = 0, x \geq 0 \\ -\frac{(-x+1)^{2-\lambda}-1}{2-\lambda} & \text{if } \lambda \neq 2, x < 0 \\ ln(-x+1) & \text{if } \lambda = 2, x < 0 \end{cases} \\
\text{Min-Max: } x' &= \frac{x - \min(x)}{\max(x) - \min(x)},
\end{aligned}
\tag{4.1}
$$

where x is the original data, x' is the transformed data, and λ is the transformation parameter. The transformation parameter is determined using the Box-Cox and Yeo-Johnson transformations, and the min-max scaling range is $[0, 1]$. The transformation methods for each series are summarized in Table 4.2. Keep in mind that features which are segmented into different regions have different scaling parameter λ for each region. The relevant scaling parameters are estimated through maximum likelihood estimation using the `Scikit-Learn` package in `Python`.

After transforming the data, we add several features to the dataset. We introduce lags of 1, 2, 3, 7, and 14 days for each of the price series, as well as the features. Additionally, we add one-hot encoded features representing the day of the week, as well as a binary dummy variable representing whether the day is a holiday or not. To summarize, the full list of features is as follows:

- Price series for each region, lagged by 1, 2, 3, 7, and 14 days: 25 features.
- Demand series for each region, lagged by 1, 2, 3, 7, and 14 days: 25 features.

Series	Transformation
Price	Box-Cox + Min-Max
Demand	Min-Max
Hydro Generation	Min-Max
Wind Generation	Min-Max
Coal Generation	Yeo-Johnson + Min-Max
Gas Generation	Min-Max
Geothermal Generation	Min-Max
Diesel Generation	Yeo-Johnson + Min-Max
Wood Generation	Min-Max
Reserve Prices	Yeo-Johnson + Min-Max
Forward Prices	Box-Cox + Min-Max
HVDC Transfer	Min-Max
Wind Speed	Min-Max
Wind Direction	Sine + Min-Max
Precipitation	Yeo-Johnson + Min-Max
Precipitation Coverage	Min-Max
Temperature	Min-Max
Dew	Min-Max
Humidity	Min-Max
Cloud Coverage	Min-Max
Solar Energy	Min-Max

Table 4.2: Transformation methods for price and features.

- Generation series for each fuel type mentioned above, lagged by 1, 2, 3, 7, and 14 days: 35 features.
- Fast and instantaneous reserve prices for both islands, lagged by 1, 2, 3, 7, and 14 days: 20 features.
- All forward contracts mentioned above for the Benmore and Otahuhu nodes, lagged by 1, 2, 3, 7, and 14 days: 60 features.
- Daily HVDC transfer, lagged by 1, 2, 3, 7, and 14 days: 5 features.
- All weather variables mentioned above for each region, plus an additional region for wind speed, wind direction, precipitation, and precipitation coverage, all lagged by 1, 2, 3, 7, and 14 days: 245 features.

- One-hot encoded features representing the day of the week: 7 features.
- Binary dummy variable representing whether the day is a holiday or not: 1 feature.

In total, the full feature set contains 423 features. The full list of features is used for the feature selection process described later in this section. To make the point clear, we strictly use lagged variables for prediction at time t, and we do not use any information from the future.

4.2.4 Cross-Validation Scheme

Prior to feature selection, we perform a 5-fold time-series cross-validation scheme to the split the data into 5 subsets of training and test sets. Since we are working with time-series data, it is crucial to ensure that the sequential nature of the data is preserved during the cross-validation process. Generally, in time-series cross-validation, the training set contains all available information prior to the test set. However, in this study, we have fixed the length of the training set for each fold to be the same. To be precise, each training set is 1456 days long, and each test set consists of the 364 days immediately succeeding its training set. These values are chosen so that the complete dataset is split into 5 folds of equal length. Figure 4.3 illustrates the cross-validation scheme.

The training set for each fold is used to train the model, and the test set is used to evaluate the model. However, in certain cases, we require a validation set to tune hyperparameters. In this case, we split the training set into a training set and a validation set. To be specific, we leave the first 1092 days in the training set, and use the remaining 364 days as the validation set. We utilize a validation set for tuning hyperparameters during feature selection and training with machine learning models.

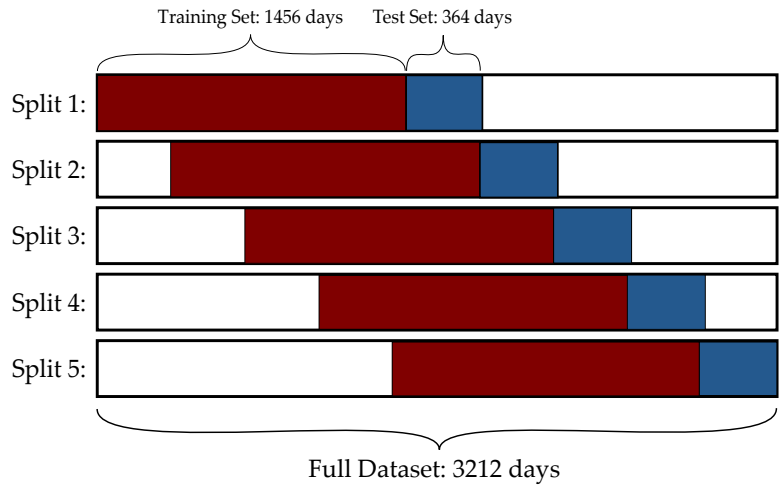


Figure 4.3: Cross-validation scheme for model training and testing.

4.2.5 Feature Selection

We perform feature selection on each cross-validation fold separately. This is because we want to ensure that the features selected for each fold are not dependent on any data from the test sets. We use validation sets to select optimal hyperparameters for each feature selection method.

LASSO Regularization

LASSO regression is a type of linear regression that is popular for feature selection and regularization. The LASSO adds a penalty term to the ordinary least squares cost function in order to reduce the number of features used in the model. Consider a linear regression model with p features, $x_1, x_2, \ldots, x_p$, and a target variable y

$$y_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \ldots + \beta_p x_{p,t} + \epsilon_t, \tag{4.2}$$

the LASSO estimator for the regression coefficients is given by

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{t=1}^{T} (y_t - \beta_1 x_{1,t} - \beta_2 x_{2,t} - \ldots - \beta_p x_{p,t})^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}. \tag{4.3}$$

This is essentially the LEAR model when we consider the features to have autoregressive components, as is the case with our features described in Section 4.2.3.

The regularization parameter λ controls the degree to which the LASSO estimator shrinks the regression coefficients towards zero. The larger the value of λ, the more coefficients are shrunk towards zero. In this study, we use a validation set to select the optimal value of λ. For each cross-validation set, we perform a grid search over the values of λ in the range $[10^{-4}, 10^{-1}]$ with a step size of 10^{-3}. For each λ value, we assess the performance of the model on the validation set using the mean squared error (MSE) metric. The value of λ that minimizes the MSE is selected as the optimal value for the corresponding cross-validation fold.

Mutual Information

For two random variables, X and Y, the mutual information between X and Y is defined as

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}. \tag{4.4}$$

The mutual information (MI) between two random variables is a measure of the amount of information that X provides about Y. In this study, we use the MI between the target variable and each feature to select the most relevant features. For feature selection, MI requires us to provide the number of features to select, k. For each cross-validation set, we perform a grid search over the values of k in the range $[1, 423]$. For each k value, we estimate a linear regression model using the training set and assess the performance of the model on the validation set using the MSE metric. The value of k that minimizes the MSE is selected as the optimal value for the corresponding cross-validation fold.

Recursive Feature Elimination

Recursive feature elimination (RFE) is a feature selection method that recursively removes features from the feature set. The method starts with all features in the feature set, and iteratively removes the least important features. The importance of each feature is determined by a given estimator. To maintain consistency with the other feature selection methods, we use a linear regression estimator to determine the importance of each feature. As with MI, RFE requires us to provide the number of features to select, k. For each cross-validation set, we perform a grid search over the values of k in the range $[1, 423]$. For each k value, we estimate a linear regression model using the training set and assess the performance of the model on the validation set using the MSE metric. The value of k that minimizes the MSE is selected as the optimal value for the corresponding cross-validation fold.

4.3 Models

4.3.1 LEAR Model

The LASSO estimated auto-regressive (LEAR) model was first implemented in Uniejewski et al. (2016) under the name LassoX. It is essentially a linear regression model with a large number of features, where the coefficients are estimated using LASSO regularization. The model has the following specification

$$y_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \ldots + \beta_n x_{n,t} + \epsilon_t, \epsilon_t \quad \sim \mathcal{N}(0, \sigma^2). \tag{4.5}$$

where y_t is the electricity price at time t, $x_{i,t}$ is the ith feature at time t, and ϵ_t is the error term at time t. The parameters $\sigma^2, \beta_1, \ldots, \beta_n$ are to be estimated.

The LASSO estimator for the regression coefficients is given by

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{t=1}^{T} (y_t - \beta_1 x_{1,t} - \beta_2 x_{2,t} - \ldots - \beta_n x_{n,t})^2 + \lambda \sum_{i=1}^{n} |\beta_i| \right\}. \tag{4.6}$$

Following the guidelines of Lago et al. (2021), we consider the LEAR model as a benchmark in this study.

4.3.2 GARCH Models

We consider a GARCH(1,1) with exogeneous variables in the mean equation, which has the following specification

$$
\begin{aligned}
y_t &= \mu_t + \epsilon_t, \\
\mu_t &= \mu_0 + \sum_{i=1}^{n} \phi_i x_{i,t}, \\
\epsilon_t &\sim \mathcal{N}(0, \sigma_t^2), \\
\sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2,
\end{aligned}
\tag{4.7}
$$

where y_t is the electricity price at time t, μ_t is the mean of the electricity price at time t, and ϵ_t is the error term at time t. The mean equation is dependent on the variables $x_{i,t}$, $i = 1, \ldots, n$, which are the features we employ in this study, see Section 4.2.3. The variance equation is dependent on the previous error term ϵ_{t-1} and the lagged variance σ_{t-1}^2. The parameters ω, α, β, μ_0, and ϕ_i are to be estimated.

The features $x_{i,t}$, $i = 1, \ldots, n$ differ based on which feature selection algorithm is employed. Using LASSO feature selection results in the same feature set utilized in the LEAR model. We denote this the LASSO-estimated GARCH (LE-GARCH) model. Similarly when using the MI or RFE generated features, we denote the model as MI-GARCH and RFE-GARCH, respectively. When using all features, the model is simply denoted as the GARCH model.

In the case of the GARCH-*t* model, the error term ϵ_t is distributed as a *t*-distribution with ν degrees of freedom, $\epsilon_t \sim t_\nu(0, \sigma^2)$. Similar to the GARCH, the GARCH-*t* model is denoted as LE-GARCH-*t*, MI-GARCH-*t*, and RFE-GARCH-*t* when using the LASSO, MI, and RFE feature selection algorithms, respectively.

We use the `ARCH` package in `Python` to estimate the GARCH models, see Sheppard et al. (2022). The package uses maximum likelihood estimation (MLE) to estimate the parameters of the GARCH models.

4.3.3 SV Models

We consider a SV model with exogeneous variables in the mean equation, which has the following specification

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, \\ \mu_t &= \mu_0 + \sum_{i=1}^{n} \phi_i x_{i,t}, \\ \epsilon_t &\sim \mathcal{N}(0, e^{h_t}), \\ h_t &= \omega + \alpha(h_{t-1} - \omega) + \epsilon_t^h, \\ \epsilon_t^h &\sim \mathcal{N}(0, \sigma_h^2), \end{aligned} \tag{4.8}$$

where y_t is the electricity price at time t, μ_t is the mean of the electricity price at time t and ϵ_t is the error term at time t. The mean equation is dependent on the features $x_{i,t}$, $i = 1, \ldots, n$. The log-volatility h_t is dependent on the previous log-volatility h_{t-1} and the error term ϵ_t^h. The parameters ω, α, μ_0, ϕ_i, and σ_h^2 are to be estimated.

Similar to the GARCH models, the features $x_{i,t}$, $i = 1, \ldots, n$ differ based on which feature selection algorithm is employed. When utilizing the LASSO-estimated feature set, we denote the model as the LASSO-estimated SV (LE-SV) model. Similarly when using the MI or RFE generated features, we denote the model as MI-SV and RFE-SV, respectively. When using all features, the model is simply denoted as the SV model.

In the case of the SV-*t* model, the error term ϵ_t is distributed as a *t*-distribution with ν degrees of freedom, $\epsilon_t \sim t_\nu(0, e^{h_t})$. Similar to the SV, the SV-*t* model is denoted as LE-SV-*t*, MI-SV-*t*, and RFE-SV-*t* when using the LASSO, MI, and RFE feature selection algorithms, respectively.

We use the `stochvol` package in `R` to estimate the SV models, see Kastner (2019). The package estimates SV parameters via Bayesian Markov chain Monte Carlo (MCMC) sampling.

4.3.4 DNN Model

We consider a deep neural network (DNN) with two hidden layers in this study. The network takes as input the features $x_{i,t}$, $i = 1, \ldots, n$, and outputs a price prediction y_t. The two hidden layers have n_1 and n_2 neurons, respectively. The activation function for the hidden layers is the rectified linear unit (ReLU) function. The network is trained using the Adam optimizer Kingma & Ba (2014). We train the model for 100 epochs using a batch size of 64. Furthermore, we utilize a validation set to implement early stopping. Figure 4.4 shows the network architecture. When a specific feature selection algorithm is used, we denote the model as the LE-DNN, MI-DNN, and RFE-DNN models, referring to the LASSO, mutual information, and recursive feature elimination algorithms, respectively. When all features are used, the model is simply denoted as the DNN model.

We use the tree-based Parzen estimator (TPE) algorithm to optimize several hyperparameters of the DNN model. The hyperparameters are the number of neurons in the first and second hidden layers, n_1 and n_2, respectively, the learning rate of the Adam optimizer, the dropout rate after each fully-connected layer, and whether or not to use batch normalization. The hyperparameters are optimized using a training set, and the model is evaluated on a validation set, which is independent of the test set used for predictions. The optimized hyperparameters are shown in Table 4.3.

4.3.5 LSTM and GRU Models

To extend the neural network architecture to account for the temporal nature of the electricity price, we consider recurrent neural networks (RNNs) in the form of the LSTM and GRU models. In our study, both model implementations are fairly simple and have only one hidden layer. The LSTM and GRU models have the same architecture as the DNN model, except that the hidden layer is recurrent. The activation function for

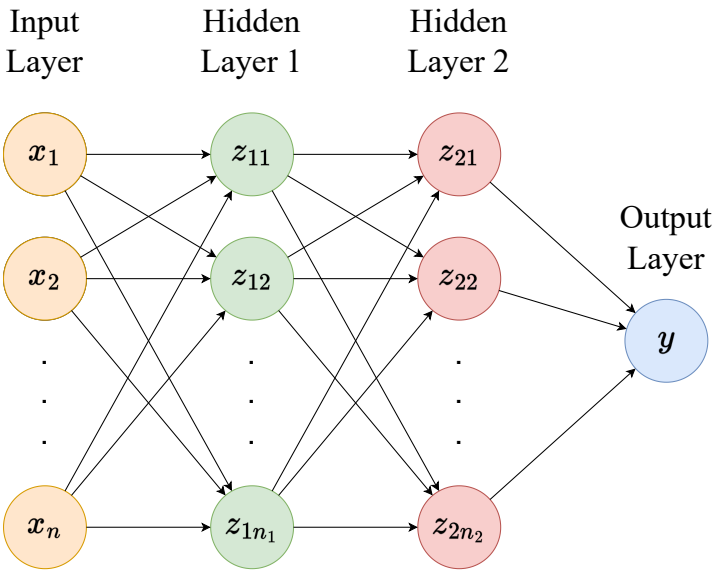


Figure 4.4: Network architecture of the DNN model.

these models is the Tanh function, and the recurrent activation is the sigmoid function. Both models utilize the Adam optimizer and early stopping using a validation set. Both models are trained for 100 epochs with a batch size of 64, same as the DNN model.

When a specific feature selection algorithm is used, we denote the LSTM model as the LE-LSTM, MI-LSTM, and RFE-LSTM models, referring to the LASSO, mutual information, and recursive feature elimination algorithms, respectively. When all features are used, the model is simply denoted as the LSTM model. Similarly for the GRU model, the LE-GRU, MI-GRU, and RFE-GRU models are denoted for the respective features, and when all features are used, the model is simply denoted as the GRU model.

The hyperparameters of the LSTM and GRU models are optimized using the TPE algorithm. The hyperparameters are the number of neurons in the hidden layer, n_{LSTM} and n_{GRU}, respectively, and the number of previous time steps used as input to the model. The optimized hyperparameters are shown in Table 4.3.

Model	Hyperparameter	Value
DNN	n_1	128
	n_2	64
	Learning rate	0.001
	Dropout rate	0.2
	Batch normalization	True
LSTM	n_{LSTM}	80
	Number of time steps	7
GRU	n_{GRU}	96
	Number of time steps	7

Table 4.3: Optimal hyperparameters of the DNN, LSTM, and GRU models.

4.3.6 XGBoost Model

The extreme gradient boosting (XGBoost) model is a tree-based model that uses gradient boosting to produce a prediction. We optimize the hyperparameters of the XGBoost model using the TPE algorithm. The hyperparameters to optimize are the number of boosting iterations, $n_{\text{estimators}}$, the learning rate, η, the maximum depth of the trees, $n_{\text{max\_depth}}$, the subsample ratio of the training instances, subsample, the minimum loss reduction required to make a further partition on a leaf node of the tree, γ, the minimum sum of instance weight (hessian) needed in a child, min_child_weight, and the L1 regularization term on weights, α. The optimized hyperparameters are shown in Table 4.4.

When a specific feature selection algorithm is used, we denote the model as the LE-XGBoost, MI-XGBoost, and RFE-XGBoost models, referring to the LASSO, mutual information, and recursive feature elimination algorithms, respectively. When all features are used, the model is simply denoted as the XGBoost model.

Hyperparameter	Value
$n_{\text{estimators}}$	1000
η	0.1
$n_{\text{max\_depth}}$	5
subsample	0.8
γ	0.1
min_child_weight	1
α	0.1

Table 4.4: Optimal hyperparameters of the XGBoost model.

4.4 Results

In this section, we present the forecast metrics and the Diebold-Mariano test results for the models under consideration. The forecast metrics include mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), symmetric mean absolute percentage error (sMAPE), and mean absolute scaled error (MASE). We also report the Diebold-Mariano (DM) test results to determine whether the improvement in forecast accuracy of one model over another is statistically significant.

4.4.1 Evaluation Metrics

We provide a brief discussion of the metrics used for evaluation of model performance. The three most popular metrics used in EPF research are the MAE, the RMSE, and the MAPE:

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|y_t - \hat{y}_t| \tag{4.9}$$

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2} \tag{4.10}$$

$$\text{MAPE} = \frac{100}{T}\sum_{t=1}^{T}\frac{|y_t - \hat{y}_t|}{y_t} \tag{4.11}$$

where y_t is the actual value and $\hat{y}_t$ is the forecasted value of the t^{th} observation. The MAE and RMSE are symmetric metrics, while the MAPE is asymmetric. The former two are more sensitive to large errors, while the MAPE is more sensitive to small errors and outliers, and is also not defined for observations with zero actual values. To alleviate some of these issues, we also consider the symmetric mean absolute percentage error

(sMAPE):

$$\text{sMAPE} = \frac{100}{T}\sum_{t=1}^{T}\frac{|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t|} \tag{4.12}$$

Due to its ease of interpretation, several studies have suggested the use of the mean absolute scaled error (MASE) as a metric for EPF, see Hyndman & Koehler (2006); Weron (2014); Hyndman & Athanasopoulos (2018). The MASE is defined as:

$$\text{MASE} = \frac{1}{T}\sum_{t=1}^{T}\frac{|y_t - \hat{y}_t|}{\frac{1}{T-1}\sum_{i=1}^{T}|y_i - y_{i-1}|}, \tag{4.13}$$

where the denominator inside the outer sum represents the in-sample MAE of a naive one-step ahead forecast. The MASE provides a simple interpretation of being better/-worse than a one-step ahead naive forecast if it is lower/higher than one. Furthermore, since the MASE uses the MAE of the naive one-step ahead forecast as a benchmark for accuracy, it is less sensitive to outliers than the other metrics, and provides a more stable baseline for comparison. This is especially important when comparing the performance of EPF models, since electricity prices are prone to large spikes and outliers.

The evaluation metrics for model forecasts are presented in Tables 4.5 and 4.6, for the North Island and South Island regions, respectively. The results shown in these tables are averaged results from the 5-fold cross-validation scheme.

We can make several observations from these tables. Firstly, the LE-GARCH-*t* has the best performance in terms of the MAE, and often in terms of other metrics as well, in all regions except for the LSI region. Other well-performing statistical models include the LE-GARCH, LE-SV-*t*, MI-GARCH-*t*, and MI-SV-t. These models tend to outperform the other machine learning models, and often outperform the DNN and LEAR benchmark models as well. Secondly, we notice that the LEAR model outperforms the GARCH and SV models in the LSI region. LSI prices are generally lower than other regions, and also exhibit lower volatility and price spiking tendencies.

Therefore, the outperformance may be due to the GARCH and SV models overfitting in this region. Thirdly, the GARCH, GARCH-*t*, SV, and SV-*t* models with all features included are generally the worst-performing models. This speaks to the importance of regularization or feature selection to construct a parsimonious set of features for these models. In some scenarios, we can observe performance improvement of up to 45% from the GARCH model with all features included to the LE-GARCH or MI-GARCH models. Finally, we observe that the LASSO feature selection generally obtains the best performance metrics, followed by MI. RFE tends to underperform in most cases, but still usually outperforms GARCH and SV models with all features. Table 4.7 summarizes the main results by displaying the best model according to each metric, for each region.

Table 4.8 shows the top three best performing models in each region according to the MASE metric. The LE-GARCH-*t* and LE-SV-*t* are in the top three in every region, with the former being the best model in terms of MASE for every region, except the LSI. The LEAR is the best model in the LSI region, and only appears in the top three in this region.

4.4.2 The Diebold-Mariano Test

We utilize the Diebold-Mariano (DM) test, see Diebold & Mariano (2002), to assess the statistical significant of our forecast results. We use absolute errors to form the loss differential series:

$$\Delta_t = |\epsilon_{t,A}| - |\epsilon_{t,B}|, \tag{4.14}$$

where $\epsilon_{t,A}$ and $\epsilon_{t,B}$ are the forecast errors of models A and B at time t, respectively. The DM test statistic is defined as:

$$\text{DM} = \sqrt{N}\frac{\hat{\mu}}{\hat{\sigma}} \sim \mathcal{N}(0,1), \tag{4.15}$$

Model	Central North Island					Upper North Island					Lower North Island				
	MAE	RMSE	MAPE	sMAPE	MASE	MAE	RMSE	MAPE	sMAPE	MASE	MAE	RMSE	MAPE	sMAPE	MASE
DNN	0.4540	0.4719	1.9867	0.7801	1.5015	0.4385	0.4	2.3694	0.7462	1.4734	0.4681	0.4695	2.3892	0.8849	1.5882
GARCH	0.6624	1.115	3.7286	0.8664	2.2058	0.6042	0.8862	4.5744	0.8882	2.0484	0.6296	1.0056	12.2467	0.8828	2.1516
GARCH-*t*	0.6068	0.9114	3.6764	0.8539	2.0248	0.6165	0.8738	3.5651	0.9	2.0692	0.5719	0.8504	7.7705	0.8862	1.964
GRU	0.4202	0.3824	2.5842	0.7263	1.3902	0.4068	0.3443	2.9169	0.7055	1.3708	0.4291	0.4119	3.6281	0.738	1.4538
LEAR	0.3909	0.3556	2.3845	0.686	1.3041	0.3832	**0.3217**	2.9951	0.6822	1.2954	0.3854	**0.3564**	3.1294	0.6965	1.3085
LSTM	0.4536	0.4124	3.2971	0.7492	1.5186	0.4141	0.3587	2.3933	0.7187	1.3966	0.4136	0.389	3.1208	0.7303	1.3996
SV	0.5431	0.7275	2.8784	0.8265	1.8226	0.5428	0.7044	3.4758	0.8299	1.8471	0.5429	0.7232	10.0983	0.8512	1.8549
SV-*t*	0.5431	0.7281	2.9107	0.8353	1.8237	0.5411	0.689	3.6033	0.831	1.8405	0.5416	0.7213	10.1176	0.8475	1.8502
XGBoost	0.4815	0.5239	2.4817	0.7408	1.6434	0.4653	0.4792	3.8907	0.7589	1.6018	0.5787	0.772	4.1283	0.8215	2.0038
LE-DNN	0.447	0.4437	2.0859	0.7852	1.4928	0.5015	0.5047	**2.0238**	0.9041	1.7293	0.4594	0.4828	2.787	0.7987	1.5553
LE-GARCH	0.3976	0.3609	2.2955	0.6943	1.3218	0.3879	0.3337	3.0548	0.6868	1.3086	0.3901	0.3775	2.8878	0.6937	1.3199
LE-GARCH-*t*	**0.3793**	**0.3508**	1.8658	**0.6704**	**1.2626**	**0.3796**	0.3228	2.508	0.6808	**1.2784**	**0.3817**	0.3711	2.6236	**0.688**	**1.2909**
LE-GRU	0.4212	0.386	2.5103	0.7312	1.3991	0.4132	0.3603	2.6946	0.7147	1.3919	0.4202	0.401	3.4032	0.7328	1.4173
LE-LSTM	0.4188	0.3931	2.3699	0.7249	1.3947	0.411	0.3652	2.2914	0.7274	1.3882	0.4114	0.3951	3.0611	0.7425	1.3845
LE-SV	0.3816	0.3537	1.9814	0.6728	1.2694	0.38	0.3228	2.6005	0.6768	1.2806	0.3842	0.3727	2.6887	0.6929	1.2997
LE-SV-*t*	0.3824	0.3548	1.9055	0.6745	1.2718	0.3797	0.3224	2.5207	**0.6754**	1.2787	0.3833	0.3705	2.6317	0.6945	1.2971
LE-XGBoost	0.4418	0.4759	2.1669	0.7273	1.48	0.4263	0.4104	3.1469	0.7287	1.4468	0.492	0.5574	3.2254	0.7812	1.6936
MI-DNN	0.5088	0.5463	**1.8311**	0.8821	1.7157	0.4865	0.4827	2.1955	0.8716	1.6582	0.4892	0.5904	**2.1365**	0.8321	1.6624
MI-GARCH	0.411	0.4035	1.9141	0.723	1.3658	0.3994	0.3534	2.1815	0.7245	1.3527	0.4367	0.4955	5.494	0.7411	1.4627
MI-GARCH-*t*	0.4221	0.4157	1.895	0.7385	1.4022	0.4111	0.3652	2.1956	0.7545	1.3918	0.4105	0.4154	3.7797	0.7424	1.382
MI-GRU	0.4194	0.3794	2.79	0.7268	1.4008	0.431	0.3808	3.1534	0.7209	1.4624	0.4125	0.3946	3.5536	0.7202	1.4042
MI-LSTM	0.416	0.3942	2.5193	0.7192	1.3954	0.411	0.3567	2.4853	0.7135	1.3896	0.4165	0.4053	3.0756	0.7436	1.4172
MI-SV	0.4154	0.4296	1.8261	0.726	1.3806	0.4037	0.3926	2.2591	0.7228	1.3667	0.4002	0.3923	3.2356	0.7256	1.3475
MI-SV-*t*	0.4157	0.4295	1.8191	0.7285	1.3802	0.4035	0.3919	2.2488	0.7231	1.3656	0.4012	0.3931	3.2443	0.7295	1.3506
MI-XGBoost	0.4669	0.499	2.1717	0.7631	1.5727	0.4831	0.4844	4.2121	0.7773	1.661	0.5174	0.6166	3.3341	0.8073	1.7835
RFE-DNN	0.532	0.5926	3.3314	0.8114	1.7617	0.626	0.8712	3.3333	0.9422	2.053	0.6156	1.4859	27.6257	0.8714	2.0741
RFE-GARCH	0.4688	0.5291	2.2	0.8277	1.5676	0.6261	1.1165	4.0893	0.8585	2.0507	0.6587	0.9239	5.2044	0.9508	2.2683
RFE-GARCH-*t*	0.5028	0.5359	2.9153	0.871	1.7081	0.5921	0.9435	3.504	0.8737	1.952	0.5944	0.7748	4.139	0.9972	2.0519
RFE-GRU	0.4395	0.42	3.143	0.7287	1.4724	0.6217	1.0008	3.6674	0.8622	2.0286	0.5152	0.5947	9.3575	0.9156	1.7536
RFE-LSTM	0.4363	0.4199	2.9704	0.7291	1.4646	0.6027	0.9343	3.6783	0.8559	1.9701	0.5069	0.5765	8.0118	0.9014	1.7346
RFE-SV	0.4541	0.4667	2.4743	0.759	1.5175	0.5815	0.9054	3.3983	0.8614	1.9117	0.5008	0.605	6.4682	0.8931	1.7043
RFE-SV-*t*	0.4527	0.4658	2.3219	0.755	1.5128	0.5828	0.9058	3.5744	0.861	1.9154	0.5043	0.6134	6.6761	0.8974	1.7159
RFE-XGBoost	0.5195	0.6091	3.8196	0.7921	1.7418	0.7043	1.2079	6.435	0.9105	2.3318	0.6786	0.9425	7.0961	0.8969	2.3484

Table 4.5: Forecast metrics for North Island regions. *Note*: Bold values indicate best performance for given metric and region.

where $\hat{\mu}$ is the sample mean and $\hat{\sigma}$ are the sameple mean and sample standard deviation of the loss differential series, respectively, and N is the number of observations in the loss differential series.

The DM test is a two-tailed test, meaning that we are testing for the possibility that either model has better predictive accuracy. The null hypothesis is that both models have the same predictive accuracy, while the alternative hypothesis is that one model has better predictive accuracy than the other. If the p-value of the DM test is less than the significance level, α = 0.05, we reject the null hypothesis and conclude that one model has better predictive accuracy than the other. On the other hand, if the p-value is greater than α, we fail to reject the null hypothesis and conclude that the two models have the same predictive accuracy.

Figures 4.5 and 4.6 show the DM test statistics for the Central North Island and Lower South Island regions, respectively. We select these regions to illustrate the DM

Model	Upper South Island					Lower South Island				
	MAE	RMSE	MAPE	sMAPE	MASE	MAE	RMSE	MAPE	sMAPE	MASE
DNN	0.4861	0.4942	2.3833	0.7975	1.8321	0.4669	0.447	2.4921	0.7819	1.7794
GARCH	0.531	0.6378	3.6149	0.7908	2.0412	0.4904	0.5111	4.139	0.7922	1.9266
GARCH-*t*	0.5451	0.6346	3.403	0.849	2.0639	0.4979	0.5684	3.3643	0.7954	1.9636
GRU	0.3923	0.3178	2.4352	0.6796	1.4842	0.3821	0.3089	2.8662	0.6655	1.4785
LEAR	0.3559	0.2839	2.3291	0.6454	1.3473	**0.3449**	**0.2692**	2.4572	**0.6283**	**1.3365**
LSTM	0.3999	0.3368	2.5603	0.6863	1.5137	0.3979	0.3326	3.269	0.688	1.5381
SV	0.4853	0.5241	2.801	0.7797	1.8558	0.4836	0.5049	3.8839	0.7727	1.8931
SV-*t*	0.4804	0.5175	2.8234	0.7682	1.835	0.4833	0.5072	3.8267	0.7759	1.8913
XGBoost	0.5119	0.5273	3.145	0.7611	2.0024	0.4132	0.3758	3.7167	0.7278	1.6303
LE-DNN	0.435	0.3917	2.3977	0.7528	1.6492	0.4636	0.4575	3.4647	0.7632	1.7794
LE-GARCH	0.3532	**0.2814**	2.3012	**0.6393**	1.3346	0.3525	0.2875	2.8456	0.6427	1.369
LE-GARCH-*t*	**0.352**	0.2858	2.1612	0.6396	**1.328**	0.3486	0.2761	2.7815	0.6402	1.3513
LE-GRU	0.3765	0.3052	2.5336	0.6689	1.4214	0.3984	0.3195	3.5424	0.7009	1.5303
LE-LSTM	0.3879	0.3254	2.4816	0.6927	1.4625	0.3816	0.3058	2.8212	0.6923	1.4685
LE-SV	0.3541	0.285	2.2142	0.6459	1.3357	0.3492	0.2746	2.8589	0.6401	1.3533
LE-SV-*t*	0.3538	0.2857	2.1933	0.645	1.3343	0.348	0.2735	2.7157	0.6346	1.3487
LE-XGBoost	0.4234	0.3979	2.9577	0.7078	1.626	0.4078	0.3578	3.7302	0.6956	1.6069
MI-DNN	0.4258	0.4197	2.1775	0.7261	1.6095	0.4475	0.4216	3.521	0.7873	1.7825
MI-GARCH	0.3598	0.2849	2.244	0.651	1.3649	0.3618	0.2916	2.7337	0.6582	1.4012
MI-GARCH-*t*	0.3645	0.2906	**2.1138**	0.6773	1.385	0.3619	0.2927	2.5224	0.6634	1.4051
MI-GRU	0.4002	0.327	2.7019	0.6998	1.5212	0.3871	0.3096	3.1054	0.6788	1.5027
MI-LSTM	0.3917	0.316	2.615	0.6965	1.4909	0.3744	0.3	2.9024	0.6696	1.4653
MI-SV	0.3637	0.29	2.1543	0.6646	1.3806	0.3603	0.2902	2.6503	0.6534	1.3959
MI-SV-*t*	0.363	0.2904	2.1255	0.6638	1.3779	0.3582	0.2888	**2.3892**	0.6478	1.3885
MI-XGBoost	0.421	0.3883	2.3878	0.7126	1.6104	0.4447	0.4325	2.659	0.7513	1.7422
RFE-DNN	0.5547	0.6397	2.9509	0.8356	2.0991	0.5629	0.6129	2.6783	0.995	2.2083
RFE-GARCH	0.4657	0.5197	2.3796	0.8076	1.7875	0.6296	1.2122	3.3459	0.8851	2.5001
RFE-GARCH-*t*	0.4616	0.5193	2.2769	0.7959	1.7723	0.6234	1.1379	3.1634	0.8895	2.4694
RFE-GRU	0.4472	0.4775	2.2395	0.7667	1.7082	0.5876	0.9465	3.4272	0.8627	2.3161
RFE-LSTM	0.4499	0.4694	2.4281	0.7489	1.7186	0.5935	0.9398	3.2453	0.8674	2.3429
RFE-SV	0.4633	0.5411	2.2794	0.7764	1.7744	0.4733	0.4925	3.0263	0.8407	1.8494
RFE-SV-*t*	0.4635	0.546	2.2716	0.7721	1.7759	0.4715	0.4911	2.7034	0.8354	1.8417
RFE-XGBoost	0.6009	0.7026	3.4217	0.8971	2.358	0.597	0.7286	3.9014	0.951	2.4022

Table 4.6: Forecast metrics for South Island regions. *Note*: Bold values indicate best performance for given metric and region.

	MAE	**RMSE**	**MAPE**	**sMAPE**	**MASE**
Central North Island	LE-GARCH-*t*	LE-GARCH-*t*	MI-DNN	LE-GARCH-*t*	LE-GARCH-*t*
Upper North Island	LE-GARCH-*t*	LEAR	LE-DNN	LE-SV-*t*	LE-GARCH-*t*
Lower North Island	LE-GARCH-*t*	LEAR	MI-DNN	LE-GARCH-*t*	LE-GARCH-*t*
Upper South Island	LE-GARCH-*t*	LE-GARCH	MI-GARCH-*t*	LE-GARCH	LE-GARCH-*t*
Lower South Island	LEAR	LEAR	MI-SV-*t*	LEAR	LEAR

Table 4.7: Best model for each metric in each region.

	Best Model	**Second Best Model**	**Third Best Model**
Central North Island	LE-GARCH-*t*	LE-SV-*t*	LE-SV
Upper North Island	LE-GARCH-*t*	LE-SV	LE-SV-*t*
Lower North Island	LE-GARCH-*t*	LE-SV-*t*	LE-SV
Upper South Island	LE-GARCH-*t*	LE-SV-*t*	LE-GARCH
Lower South Island	LEAR	LE-SV-*t*	LE-GARCH-*t*

Table 4.8: Top three models in each region according to their MASE.

test results because the LE-GARCH-*t* is the best-performing model in the CNI, whereas the LEAR model is the best-performing model in the LSI region. We observe the test statistics rather than p-values because they provide a better visualization of the relative performance of the models. A negative test statistic lower than -1.67 indicates the row-wise model is statistically better, and a positive test statistics higher than 1.67 indicates the column-wise model is statistically better. In the heatmap, blue cells represent a better row-wise model, and red cells indicate a better column-wise model. Lighter colors indicate stronger significance in predictive accuracy (higher absolute test statistic). Grey-shaded cells indicate that there is no statistical significance between the predictive accuracy of the two models. A darker shade of grey corresponds to a lower test statistic, indicating stronger insignificance in predictive accuracy. The models in the heatmap are also sorted by their performance, from worst to best, in the corresponding region.

Sorting the models from worst to best based on their overall DM test scores provides a nice visual representation of the comparison between models. We notice the GARCH and GARCH-t models with all features included are generally outperformed by other models. On the other hand, the LE-GARCH-t and LE-SV-t are in the top three models for both regions. There are several other GARCH and SV models which are in the top ten in both regions.

The top performing model for CNI prices is the LE-GARCH-*t* model. However, according to the DM test, its predictive accuracy is statistically insignificant from the

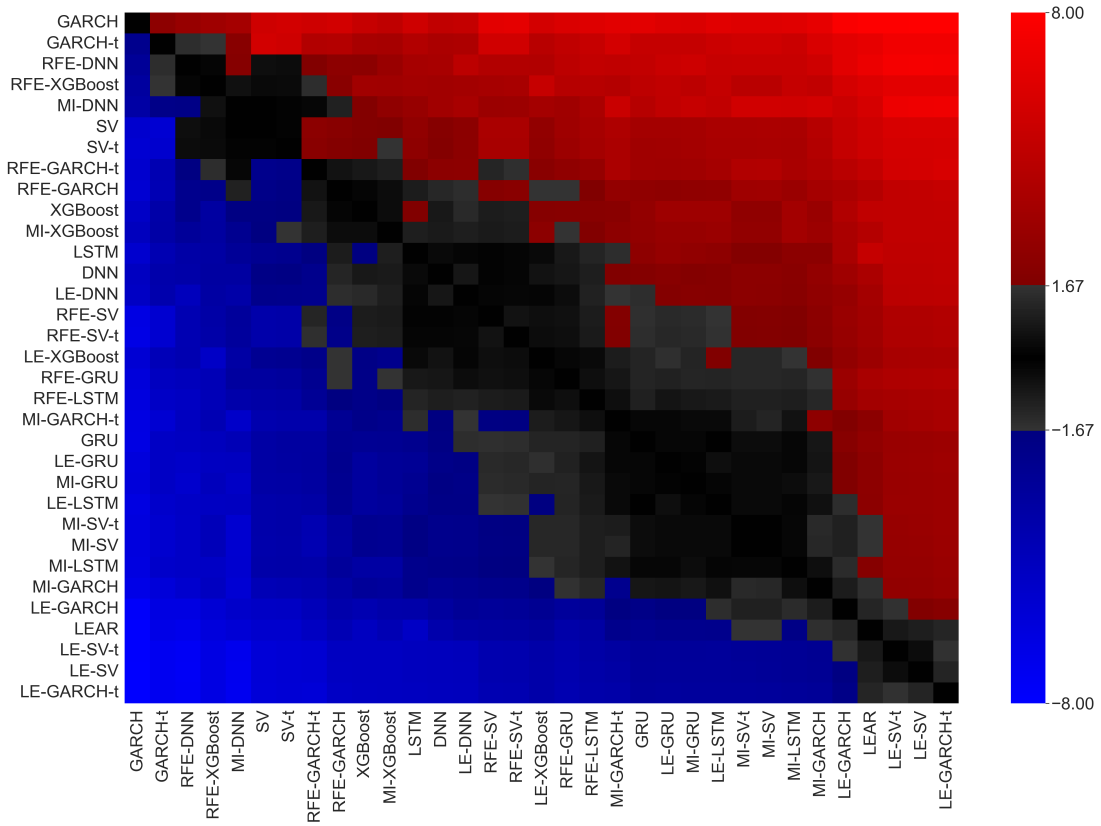


Figure 4.5: DM test statistics for Central North Island.

predictive accuracy of the LE-SV, LE-SV-*t*, and LEAR models. For LSI prices, the top performing model is the LEAR. However, the DM test suggests that its predictive accuracy is statistically insignificant from the predictive accuracy of the LE-SV-*t*, LE-GARCH-*t*, LE-SV, LE-GARCH, MI-SV-*t*, and the MI-SV models.

According to the DM test, we cannot conclude whether the GARCH-related or SV-related model is significantly better than the LEAR benchmark model. However, after observing the forecast metrics, we can confirm that these statistical models are on par with the LEAR, if not slightly better on the merit of forecast metrics. Additionally, the results show overwhelming evidence that the GARCH and SV models, and their *t* variants, provide significantly better forecasts when feature selection is conducted, as opposed to when all features are included.

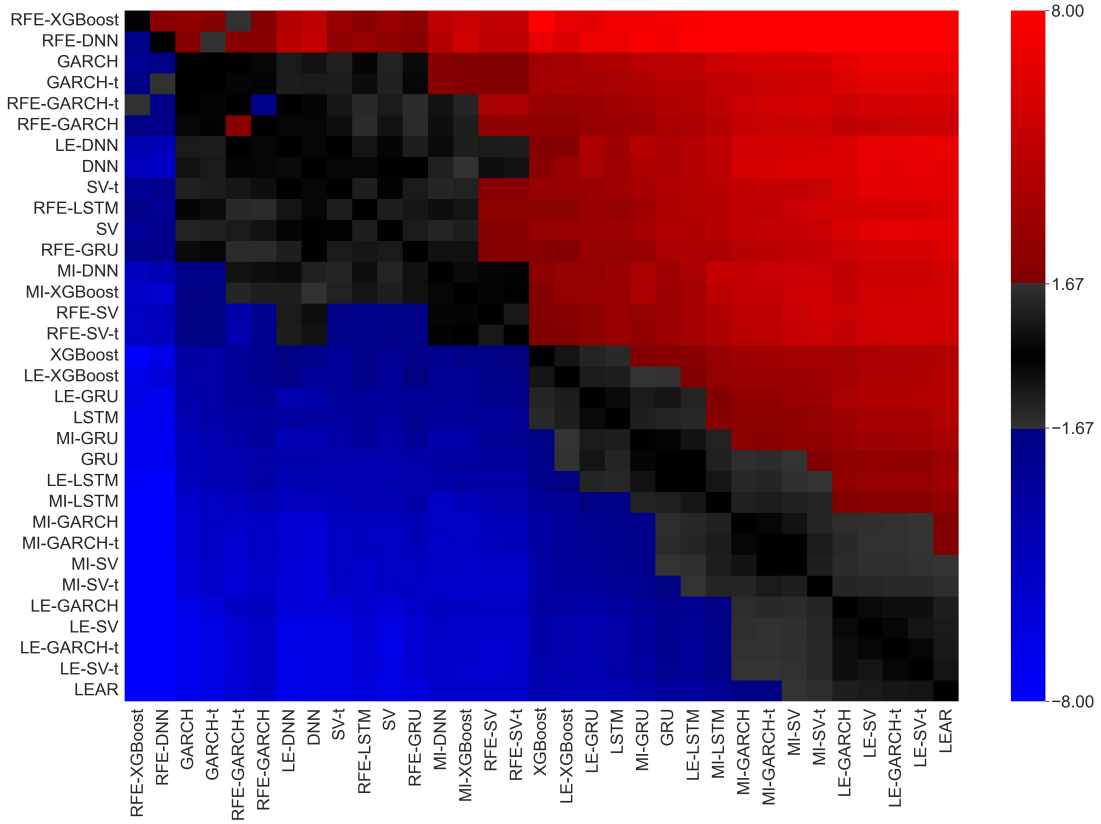


Figure 4.6: DM test statistics for Lower South Island.

4.5 Conclusion

This study presents a comparison of statistical models and machine learning models for daily electricity price forecasting in the New Zealand electricity market. We predict prices for the five encompassing regions in New Zealand. In particular, we compare the GARCH and SV models and their t-distribution variants with a variety of popular models in electricity price forecasting research, including LSTM, GRU, XGBoost, LEAR, and a four-layer DNN. The latter two models are considered benchmarks in this study. We use several exogenous variables for all models, including demand, generation fuel, forward prices and reserve prices, the HVDC transfer rate, and several weather-related variables. We also implement feature selection techniques, including LASSO, mutual information, and recursive feature elimination to create parsimonious feature sets for each region and each cross-validation set.

Our results suggest that the GARCH and SV statistical models, and their t-distributed variants, are very capable forecasting tools for EPF when paired with a variety of features, and feature selection methods to create an appropriate set of features. We find the LE-GARCH-t, the LE-SV-t, the LE-GARCH, the LE-SV to be consistently top-performing models, out-performing both the LEAR and DNN benchmark models, in terms of the forecast metrics considered. We also find their performance to increase by over 40% compared to the GARCH and SV models with all features included. This result speaks to the importance of implementing feature selection techniques.

We also investigated the Diebold-Mariano (DM) test results for all models considered. We found the difference in predictive accuracy of the top-performing GARCH and SV models to be statistically insignificant from the predictive accuracy of the LEAR model. This result suggests that the GARCH and SV models are not significantly better than the LEAR model. However, taking into account the forecast metrics, we can conclude that the GARCH and SV models are capable of out-performing the DNN,

LSTM, GRU, and XGBoost models, in their simple forms. They are also capable of out-performing the LEAR model with respect to traditional forecast metrics.

A natural extension of our work would be to study more advanced statistical and machine-learning models. This includes the plethora of GARCH variants, such as the exponential GARCH (EGARCH), and the wide variety of novel machine-learning models that have been tested in recent literature. This could also include different classes of statistical models such as regime-switching and functional data models, see Shah et al. (2021); Jan et al. (2022). Another extension would be to study the performance of the models considered in this study on a more granular level, such as the hourly level. This would allow for a more detailed analysis of the performance of the models, and would also allow for a more detailed analysis of the performance of the models on different days of the week, and at different times of the day. Finally, an interesting study would involve the ensemble of these models to create hybrid models that are potentially capable of out-performing the individual models. A hybrid of statistical and machine learning models would also be an interesting study.

Chapter 5

Conclusion

This thesis demonstrated three forecasting studies for New Zealand electricity prices using various methods and models in electricity price forecasting literature. The underlying motivation for this thesis, and the three studies, was to provide an electricity price forecasting framework tailored specifically to the New Zealand electricity market, which was previously lacking in literature. This is of critical importance for market participants, due to the highly renewable nature of the market. Furthermore, with plans to reach 100% renewable generation by 2035, the market conditions will become more uncertain. This thesis relies on the use of external variables, such as generation information, weather-related data, and data from the derivatives market, in order to understand and explain the specific characteristics of electricity prices in New Zealand. The three studies presented in this thesis were conducted in chronological order, with each study building on the previous one. The studies also incorporated newer methodologies and models that were present in the field of electricity price forecasting literature at the time of the study. As a result, the research conducted for this thesis will provide a solid framework for market participants to make informed decisions, and for future research to build upon.

The first study focused on the application of Markov regime-switching (MRS)

models to forecast daily electricity prices. There was a focus on the comparison between the MRS framework and extreme-value theory (EVT). It was shown that the EVT peaks-over-threshold framework is a special case of the MRS framework. Results showed that MRS with time-varying transition probabilities provided reasonable fits to out-of-sample price density. Furthermore, the EVT class, despite its simplicity compared to MRS models, outperformed the other models in regards to in-sample density fitting. This was attributed to the use of the generalized Pareto distribution to model extreme prices. While this study incorporated some quarterly generation data as external variables, we found the lack of more significant features to be a detriment to forecasting performance.

Much of electricity pricing literature has suggested the incorporation of weather-related variables for electricity price forecasting exercises. In the second study of this thesis, a mixed-frequency framework was used to forecast half-hourly electricity prices using hourly weather variables. A mixed-frequency vector-autoregressive (MF-VAR) model was used, and compared alongside another mixed-frequency model, the reverse unrestricted mixed-data sampling (RU-MIDAS) model. The study demonstrated that the RU-MIDAS can be represented in a VAR format, which explained the similarity in forecasting performance. A key finding from this study was the use of least absolute shrinkage and selection operator (LASSO) regularization to select a subset of the most important features. The LASSO estimated models consistently outperformed others.

Seeing the success of LASSO with electricity related features stemmed the idea for the third study presented in this thesis. The third study provided a comparison of statistical models, namely GARCH and stochastic volatility (SV), and machine learning models, the LASSO-estimated auto-regressive model, a three-layer deep neural network (DNN), a long short-term memory (LSTM) model, a gated recurrent unit (GRU), and an extreme gradient boosting (XGBoost) model. The key focus was on providing a fair comparison for the GARCH and SV models, which is often overlooked in literature. A

fair comparison was provided in the methodology, where the same large initial feature set was used for all models, and several feature-selection methods were used to select the best subset of features. Our results showed the LE-GARCH-t, the LE-SV-t, the LE-GARCH, the LE-SV were consistently in the group of top-performing models, out-performing both the LEAR and DNN benchmark models, in terms of the forecast metrics considered. We also found their performance to increase by over 40% compared to the GARCH and SV models with all features included. This result speaks to the importance of implementing feature selection techniques.

The studies presented in this thesis have gradually applied newer methodologies and models that are present in the field of electricity price forecasting literature. This field is rapidly growing and newer statistical and machine-learning ideologies are constantly being tested. The studies presented in this thesis are a small contribution to this field. In regards to future work, each of the studies presented in this thesis can be extended in several ways. For example, the first two studies can be extended to include better-suited external data, such as generation information and weather-related data that were present in the third study, and also the use of heavy-tailed distribution to better model the behaviour of extreme prices. The third study provided a suitable forecasting framework for daily electricity prices. However, the same methodology for half-hourly prices would be beneficial for many applications, such as derivatives pricing. Furthermore, a similar comparison of more complex statistical and machine learning models will be beneficial for the field of electricity price forecasting, particularly with the rise in popularity of transformers, and more sophisticated deep learning and reinforcement learning frameworks.

References

Atef, S. & Eltawil, A. B. (2019). A comparative study using deep learning and support vector regression for electricity price forecasting in smart grids. In *2019 ieee 6th international conference on industrial engineering and applications (iciea)* (pp. 603–607).

Bańbura, M., Giannone, D. & Reichlin, L. (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*(1), 71–92.

Bazzi, M., Blasques, F., Koopman, S. J. & Lucas, A. (2017). Time-varying transition probabilities for markov regime switching models. *Journal of Time Series Analysis*, *38*(3), 458–478.

Becker, R., Hurn, S. & Pavlov, V. (2007). Modelling spikes in electricity prices. *Economic Record*, *83*(263), 371–382.

Bennedsen, M. (2017). A rough multi-factor model of electricity spot prices. *Energy Economics*, *63*, 301–313.

Bergstra, J., Bardenet, R., Bengio, Y. & Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems*, *24*.

Bergstra, J., Komer, B., Eliasmith, C., Yamins, D. & Cox, D. D. (2015). Hyperopt: a python library for model selection and hyperparameter optimization. *Computational Science & Discovery*, *8*(1), 014008.

Bierbrauer, M., Menn, C., Rachev, S. T. & Trück, S. (2007). Spot and derivative pricing in the eex power market. *Journal of Banking & Finance*, *31*(11), 3462–3485.

Bierbrauer, M., Trück, S. & Weron, R. (2004). Modeling electricity prices with regime switching models. In *International conference on computational science* (pp. 859–867).

Bishop, C. M. & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4) (No. 4). Springer.

Blei, D. M., Kucukelbir, A. & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Boot, T. & Nibbering, D. (2019). Forecasting using random subspace methods. *Journal of Econometrics*, *209*(2), 391–406.

Borovkova, S. & Permana, F. J. (2006). Modelling electricity prices by the potential jump-diffusion. In *Stochastic finance* (pp. 239–263). Springer.

Borovkova, S. & Schmeck, M. D. (2017). Electricity price modeling with stochastic time change. *Energy Economics*, *63*, 51–65.

Breiman, L. (2017). *Classification and regression trees*. Routledge.

Byström, H. N. (2005). Extreme value theory and extremely large electricity price changes. *International Review of Economics & Finance*, *14*(1), 41–55.

Cartea, A. & Figueroa, M. G. (2005). Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, *12*(4), 313–335.

Carvalho, C. M., Polson, N. G. & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, *97*(2), 465–480.

Cavalcante, L., Bessa, R. J., Reis, M. & Browell, J. (2017). Lasso vector autoregression structures for very short-term wind power forecasting. *Wind Energy*, *20*(4), 657–675.

Chan, J. C. (2013). Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, *176*(2), 162–172.

Chan, J. C. & Grant, A. L. (2016). Modeling energy price dynamics: Garch versus stochastic volatility. *Energy Economics*, *54*, 182–189.

Chang, Z., Zhang, Y. & Chen, W. (2019). Electricity price prediction based on hybrid model of adam optimized lstm neural network and wavelet transform. *Energy*, *187*, 115804.

Chen, T. & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Cho, K., Van Merriënboer, B., Bahdanau, D. & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.

Coelho, I. M., Coelho, V. N., Luz, E. J. d. S., Ochi, L. S., Guimarães, F. G. & Rios, E. (2017). A gpu deep learning metaheuristic based model for time series forecasting. *Applied Energy*, *201*, 412–418.

Cross, J. L., Hou, C. & Poon, A. (2020). Macroeconomic forecasting with large bayesian vars: Global-local priors and the illusion of sparsity. *International Journal of Forecasting*, *36*(3), 899–915.

De Jong, C. (2006). The nature of power spikes: A regime-switch approach. *Studies in Nonlinear Dynamics & Econometrics*, *10*(3).

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business & Economic Statistics*, *33*(1), 1–1.

Diebold, F. X., Lee, J.-H. & Weinbach, G. C. (1994). Regime switching with time-varying transition probabilities. *Business Cycles: Durations, Dynamics, and Forecasting*, *1*, 144–165.

Diebold, F. X. & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*(1), 134–144.

Diongue, A. K., Guegan, D. & Vignal, B. (2009). Forecasting electricity spot market prices with a k-factor gigarch process. *Applied Energy*, *86*(4), 505–510.

Ebrahimian, H., Barmayoon, S., Mohammadi, M. & Ghadimi, N. (2018). The price prediction for the energy market based on a new method. *Economic research-Ekonomska istraživanja*, *31*(1), 313–337.

Eraker, B., Chiu, C. W., Foerster, A. T., Kim, T. B. & Seoane, H. D. (2014). Bayesian mixed frequency vars. *Journal of Financial Econometrics*, *13*(3), 698–721.

Fan, C., Xiao, F. & Zhao, Y. (2017). A short-term building cooling load prediction method using deep learning algorithms. *Applied Energy*, *195*, 222–233.

Foroni, C., Ravazzolo, F. & Rossini, L. (2023). Are low frequency macroeconomic variables important for high frequency electricity prices? *Economic Modelling*, *120*, 106160.

Gefang, D. (2014). Bayesian doubly adaptive elastic-net lasso for var shrinkage. *International Journal of Forecasting*, *30*(1), 1–11.

Gefang, D., Koop, G. & Poon, A. (2020). Computationally efficient inference in large bayesian mixed frequency vars. *Economics Letters*, *191*, 109120.

Gefang, D., Koop, G. & Poon, A. (2022). Forecasting using variational bayesian inference in large vector autoregressions with hierarchical shrinkage. *International Journal of Forecasting*, *https://doi.org/10.1016/j.ijforecast.2021.11.012*.

Gefang, D., Koop, G. & Poon, A. (2023). Forecasting using variational bayesian inference in large vector autoregressions with hierarchical shrinkage. *International Journal of Forecasting*, *39*(1), 346–363.

Geman, H. & Roncoroni, A. (2006). Understanding the fine structure of electricity prices. *The Journal of Business*, *79*(3), 1225–1261.

Gholipour Khajeh, M., Maleki, A., Rosen, M. A. & Ahmadi, M. H. (2018). Electricity price forecasting using neural networks with an improved iterative training algorithm. *International Journal of Ambient Energy*, *39*(2), 147–158.

Ghysels, E. (2016). Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics*, *193*(2), 294–314.

Ghysels, E., Harvey, A. C. & Renault, E. (1996). Stochastic volatility. *Handbook of Statistics*, *14*, 119–191.

Ghysels, E., Santa-Clara, P. & Valkanov, R. (2004). The midas touch: Mixed data sampling regression models. *UCLA: Finance*, *retrieved from https://escholarship.org/uc/item/9mf223rs*.

Gianfreda, A. & Grossi, L. (2012). Forecasting italian electricity zonal prices with exogenous variables. *Energy Economics*, *34*(6), 2228–2239.

Götz, T. B. & Hauzenberger, K. (2021). Large mixed-frequency vars with a parsimonious time-varying parameter structure. *The Econometrics Journal*, *24*(3), 442–461.

Hagfors, L. I., Bunn, D., Kristoffersen, E., Staver, T. T. & Westgaard, S. (2016). Modeling the uk electricity price distributions using quantile regression. *Energy*, *102*, 231–243.

Haldrup, N. & Nielsen, M. Ø. (2006). A regime switching long memory model for electricity prices. *Journal of Econometrics*, *135*(1-2), 349–376.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357–384.

Hansen, P. R., Lunde, A. & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Higgs, H. & Worthington, A. (2008). Stochastic price modeling of high volatility, mean-reverting, spike-prone commodities: The australian wholesale spot electricity market. *Energy Economics*, *30*(6), 3172–3185.

Hochreiter, S. & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D. & Zareipour, H. (2020). Energy forecasting: A review and outlook. *IEEE Open Access Journal of Power and Energy*, *7*, 376–388.

Hsu, N.-J., Hung, H.-L. & Chang, Y.-M. (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics & Data Analysis*, *52*(7), 3645–3657.

Huang, C.-J., Shen, Y., Chen, Y.-H. & Chen, H.-C. (2021). A novel hybrid deep neural network model for short-term electricity price forecasting. *International Journal of Energy Research*, *45*(2), 2511–2532.

Huang, M.-L., Hung, Y.-H., Lee, W., Li, R.-K. & Jiang, B.-R. (2014). Svm-rfe based feature selection and taguchi parameters optimization for multiclass svm classifier. *The Scientific World Journal*, *2014*.

Huisman, R., Huurman, C. & Mahieu, R. (2007). Hourly electricity prices in day-ahead markets. *Energy Economics*, *29*(2), 240–248.

Huisman, R. & Mahieu, R. (2003). Regime jumps in electricity prices. *Energy Economics*, *25*(5), 425–434.

Huurman, C., Ravazzolo, F. & Zhou, C. (2012). The power of weather. *Computational Statistics & Data Analysis*, *56*(11), 3793–3807.

Hyndman, R. J. & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Hyndman, R. J. & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688.

Jan, F., Shah, I. & Ali, S. (2022). Short-term electricity prices forecasting using functional time series analysis. *Energies*, *15*(9), 3423.

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

Janczura, J. & Weron, R. (2010). An empirical comparison of alternate regime-switching models for electricity spot prices. *Energy Economics*, *32*(5), 1059–1073.

Janczura, J. & Weron, R. (2012). Efficient estimation of markov regime-switching models: An application to electricity spot prices. *AStA Advances in Statistical Analysis*, *96*(3), 385–407.

Jedrzejewski, A., Marcjasz, G. & Weron, R. (2021). Importance of the long-term seasonal component in day-ahead electricity price forecasting revisited: Parameter-rich models estimated via the lasso. *Energies*, *14*(11), 3249.

Kanamura, T. & Ōhashi, K. (2008). On transition probabilities of regime switching in electricity prices. *Energy Economics*, *30*(3), 1158–1172.

Karakatsani, N. V. & Bunn, D. W. (2010). Fundamental and behavioural drivers of electricity price volatility. *Studies in Nonlinear Dynamics & Econometrics*, *14*(4).

Kastner, G. (2019). Dealing with stochastic volatility in time series using the r package stochvol. *arXiv preprint arXiv:1906.12134*.

Kastner, G. & Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, *76*, 408–423.

Kim. (1994). Dynamic linear models with markov-switching. *Journal of Econometrics*, *60*(1-2), 1–22.

Kim, C.-J., Nelson, C. R. et al. (1999). State-space models with regime switching: classical and gibbs-sampling approaches with applications. *MIT Press Books*, *1*.

Kim, S., Shephard, N. & Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with arch models. *The Review of Economic Studies*, *65*(3), 361–393.

Kingma, D. P. & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Knittel, C. R. & Roberts, M. R. (2005). An empirical examination of restructured electricity prices. *Energy Economics*, *27*(5), 791–817.

Koop, G. & Korobilis, D. (2018). Variational bayes inference in high-dimensional time-varying parameter models. , *Retrieved from https://mpra.ub.uni-muenchen.de/id/eprint/87972*.

Koop, G., Korobilis, D. et al. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends® in Econometrics*, *3*(4), 267–358.

Kozachenko, L. F. & Leonenko, N. N. (1987). Sample estimate of the entropy of a random vector. *Problemy Peredachi Informatsii*, *23*(2), 9–16.

Lago, J., De Ridder, F. & De Schutter, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, *221*, 386–405.

Lago, J., Marcjasz, G., De Schutter, B. & Weron, R. (2021). Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, *293*, 116983.

Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions—five years of experience. *Journal of Business & Economic Statistics*, *4*(1), 25–38.

Lucia, J. J. & Schwartz, E. S. (2002). Electricity prices and power derivatives: Evidence from the nordic power exchange. *Review of Derivatives Research*, *5*(1), 5–50.

Ludwig, N., Feuerriegel, S. & Neumann, D. (2015). Putting big data analytics to work: Feature selection for forecasting electricity prices using the lasso and random forests. *Journal of Decision Systems*, *24*(1), 19–36.

Luo, S. & Weng, Y. (2019). A two-stage supervised learning approach for electricity price forecasting by leveraging different data sources. *Applied Energy*, *242*, 1497–1512.

Marcjasz, G., Uniejewski, B. & Weron, R. (2020). Beating the naïve—combining lasso with naïve intraday electricity price forecasts. *Energies*, *13*(7), 1667. doi: 10.3390/en13071667

Mari, C. (2008). Random movements of power prices in competitive markets: A hybrid model approach. *Journal of Energy Markets*, *1*(2), 87–103.

Mayer, K., Schmid, T. & Weber, F. (2015). Modeling electricity spot prices: combining mean reversion, spikes, and stochastic volatility. *The European Journal of Finance*, *21*(4), 292–315.

Meng, A., Wang, P., Zhai, G., Zeng, C., Chen, S., Yang, X. & Yin, H. (2022). Electricity price forecasting with high penetration of renewable energy using attention-based lstm network trained by crisscross optimization. *Energy*, *254*, 124212.

Messner, J. W. & Pinson, P. (2019). Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting. *International Journal of Forecasting*, *35*(4), 1485–1498. doi: 10.1016/j.ijforecast.2018.02.001

Ministry of Business, Innovation and Employment. (2015). *Chronology of new zealand electricity reform.* `https://www.mbie.govt.nz/dmsdocument/178-chronology-of-nz-electricity-reform-pdf`. Retrieved from `https://www.mbie.govt.nz/dmsdocument/178-chronology-of-nz-electricity-reform-pdf`

Ministry of Business, Innovation and Employment. (2021). *Energy in new zealand 2020.* Retrieved from `https://www.mbie.govt.nz/dmsdocument/11679-energy-in-new-zealand-2020`

Mosquera-López, S., Manotas-Duque, D. F. & Uribe, J. M. (2017). Risk asymmetries in hydrothermal power generation markets. *Electric Power Systems Research*, *147*, 154–164.

Mosquera-López, S., Uribe, J. M. & Manotas-Duque, D. F. (2017). Nonlinear empirical pricing in electricity markets using fundamental weather factors. *Energy*, *139*, 594–605.

Mount, T. D., Ning, Y. & Cai, X. (2006). Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters. *Energy Economics*, *28*(1), 62–80.

Naz, A., Javed, M. U., Javaid, N., Saba, T., Alhussein, M. & Aurangzeb, K. (2019). Short-term electric load and price forecasting using enhanced extreme learning machine optimization in smart grids. *Energies*, *12*(5), 866.

Ormerod, J. T. & Wand, M. P. (2010). Explaining variational approximations. *The American Statistician*, *64*(2), 140–153.

Paraschiv, F. (2013). Price dynamics in electricity markets. In *Handbook of risk management in energy production and trading* (pp. 47–69). Springer.

Paraschiv, F., Erni, D. & Pietsch, R. (2014). The impact of renewable energies on eex day-ahead electricity prices. *Energy Policy*, *73*, 196–210.

Paraschiv, F., Fleten, S.-E. & Schürle, M. (2015). A spot-forward model for electricity prices with regime shifts. *Energy Economics*, *47*, 142–153.

Paraschiv, F., Hadzi-Mishev, R. & Keles, D. (2016). Extreme value theory for heavy tails in electricity prices. *Journal of Energy Markets*, *9*(2).

Perlin, M. (2015). Ms_regress-the matlab package for markov regime switching models. *Available at SSRN 1714016*.

Schorfheide, F. & Song, D. (2015). Real-time forecasting with a mixed-frequency var. *Journal of Business & Economic Statistics*, *33*(3), 366–380.

Shah, I., Akbar, S., Saba, T., Ali, S. & Rehman, A. (2021). Short-term forecasting for the electricity spot prices with extreme values treatment. *IEEE Access*, *9*, 105451–105462.

Shao, Z., Yang, S., Gao, F., Zhou, K. & Lin, P. (2017). A new electricity price prediction strategy using mutual information-based svm-rfe classification. *Renewable and Sustainable Energy Reviews*, *70*, 330–341.

Sheppard, K. (2009). Mfe matlab function reference financial econometrics. *URL https://www. kevinsheppard. com/images/9/95/MFE_Toolbox_Documentation. pdf*.

Sheppard, K., Khrapov, S., Lipták, G., mikedeltalima, Capellini, R., alejandro cermeno, ... Çelik, B. (2022, June). *bashtage/arch: Release 5.3.1*. Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.6684078` doi: 10.5281/zenodo.6684078

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1–48.

Son, H. & Kim, C. (2017). Short-term forecasting of electricity demand for the residential sector using weather and social variables. *Resources, Conservation and Recycling*, *123*, 200–207.

Suomalainen, K., Pritchard, G., Sharp, B., Yuan, Z. & Zakeri, G. (2015). Correlation analysis on wind and hydro resources with electricity demand and prices in new zealand. *Applied Energy*, *137*, 445–462.

Tan, Z., Zhang, J., Wang, J. & Xu, J. (2010). Day-ahead electricity price forecasting using wavelet transform combined with arima and garch models. *Applied Energy*, *87*(11), 3606–3610.

Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*, *4*(2), 183–204.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Tiwari, A. K., Kumar, S. & Pathak, R. (2019). Modelling the dynamics of bitcoin and litecoin: Garch versus stochastic volatility models. *Applied Economics*, *51*(37), 4073–4082.

Tschora, L., Pierre, E., Plantevit, M. & Robardet, C. (2022). Electricity price forecasting on the day-ahead market using machine learning. *Applied Energy*, *313*, 118752.

Uniejewski, B., Marcjasz, G. & Weron, R. (2019). Understanding intraday electricity markets: Variable selection and very short-term price forecasting using lasso. *International Journal of Forecasting*, *35*(4), 1533–1547. doi: 10.1016/j.ijforecast.2019.02.001

Uniejewski, B., Nowotarski, J. & Weron, R. (2016). Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies*, *9*(8), 621.

Uniejewski, B. & Weron, R. (2018a). Efficient forecasting of electricity spot prices with expert and lasso models. *Energies*, *11*(8), 2039. doi: 10.3390/en11082039

Uniejewski, B. & Weron, R. (2018b). Efficient forecasting of electricity spot prices with expert and lasso models. *Energies*, *11*(8), 2039.

Wand, M. P., Ormerod, J. T., Padoan, S. A. & Frühwirth, R. (2011). Mean field variational bayes for elaborate distributions. *Bayesian Analysis*, *6*(4), 847–900.

Wang, H., Lei, Z., Zhang, X., Zhou, B. & Peng, J. (2019). A review of deep learning for renewable energy forecasting. *Energy Conversion and Management*, *198*, 111799.

Wang, H.-z., Li, G.-q., Wang, G.-b., Peng, J.-c., Jiang, H. & Liu, Y.-t. (2017). Deep learning based ensemble approach for probabilistic wind power forecasting. *Applied Energy*, *188*, 56–70.

Wang, L., Wu, J., Cao, Y. & Hong, Y. (2022). Forecasting renewable energy stock volatility using short and long-term markov switching garch-midas models: Either, neither or both? *Energy Economics*, *111*, 106056.

Weron, R. (2009). Heavy-tails and regime-switching in electricity prices. *Mathematical Methods of Operations Research*, *69*(3), 457–473.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, *30*(4), 1030–1081.

Weron, R., Bierbrauer, M. & Trück, S. (2004). Modeling electricity prices: jump diffusion and regime switching. *Physica A: Statistical Mechanics and its Applications*, *336*(1-2), 39–48.

Wu, L. & Shahidehpour, M. (2010). A hybrid model for day-ahead price forecasting. *IEEE Transactions on Power Systems*, *25*(3), 1519–1530.

You, C., Ormerod, J. T. & Mueller, S. (2014). On variational bayes estimation and variational information criteria for linear regression models. *Australian & New Zealand Journal of Statistics*, *56*(1), 73–87.

Ziel, F. (2016). Forecasting electricity spot prices using lasso: On capturing the autoregressive intraday structure. *IEEE Transactions on Power Systems*, *31*(6), 4977–4987.

Ziel, F. & Weron, R. (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics*, *70*, 396–420.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*(476), 1418–1429.