UNIVERSITÀ DEGLI STUDI DI BERGAMO

Dipartimento di Scienze Aziendali

Corso di Laurea Magistrale in

International Management, Entrepreneurship and Finance

Classe n. LM 77

FORECASTING ELECTRICITY PRICES USING ARIMA MODELS

Relatore:

Chiar.ma Prof.ssa Michela Cameletti

Tesi di Laurea Magistrale

Linh Dan TRINH

Matricola n. 1062267

ANNO ACCADEMICO 2022 / 2023

ABSTRACT

Power generation soon became a hot topic in contribution to the fight against climate change. Energy sector has changed significantly after the 2020 pandemic. The importance of electricity price forecast is, therefore, undeniable for both electricity consumers and producers. With the liberalisation and complexity of electricity markets, accurate and timely forecasting of electricity prices has become an essential tool for market participants to make informed decisions about buying and selling electricity.

As renewable energy sources such as solar and wind power have become more prevalent in the electricity generation mix, electricity price prediction has become even more important in recent years. These energy sources are intermittent and difficult to predict, which can cause price volatility in electricity markets. This volatility has the potential to significantly impact the profitability of electricity producers as well as the cost of electricity to consumers.

Lots of work has been done trying to find a suitable model to best forecast this commodity. Accurate electricity price forecasting can also help with the integration of energy storage technologies, allowing for better management of electricity supply and demand. This can lead to lower electricity costs and increased reliability.

Overall, forecasting plays an important role in the development and operation of modern electricity markets. Its importance will only grow as the energy sector shifts to a more sustainable and diverse mix of energy sources. This thesis works with the ARIMA model and its variation to predict next – day electricity prices, with the empirical study from Japan energy exchange to help enterprises get better business decisions.

ACKNOWLEDGEMENT

Foremost, I would like to express my sincere gratitude to my supervisor Prof. Michela Cameletti for the continuous support of my thesis, for her patience and dedication within her busy schedule. Her guidance and expertise have been invaluable to me and have played a crucial role in the success of this thesis.

I am grateful to the International Renewable Energy Agency (IRENA) for providing me with the leadership training and being a huge inspiration and research resources for me to start the thesis.

I am deeply thankful to my parents for keeping their belief in me and their continuous support during this process. Without their encouragement and motivation, I would not have been able to complete this journey.

My sincere thanks also go to all my friends throughout my master journey. I am appreciative to have my supportive gangs from Via Berizzi 19, Via Berizzi 37, Via Paglia 22, all friends and professors who have been by my side suggesting me with different thesis topics when I was being Erasmus student at the University of Trier, everyone in Casa Vietnam for making it convenient for me to complete this work. To name particularly, I would like to mention Roberta Laratta for being there for me when I had nowhere to go to; Phanh Ngo for taking your time helping me to fix the citation; Nguyen Tien Thanh for your technical support and presence all this time; Nga, Ayca, Selda and Alessia, from the Sinh To Bo to the energy drink, advice, comfort and motivational sayings from you all. Without your tremendous mental support, the process would have taken a great deal longer to complete.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Electricity price, which is considered a tradable commodity in economic terms, is always a major concern for businesses, especially the ones participating in the energy market. One aspect of electricity is that prices are based solely on demand and seasons. The characteristics of non – storable and the dependency on weather of renewable energy still make it difficult for businesses to predict prices, thus, control their profitability. Meanwhile, other risks including counterparty risk, interest risk and liquidity risk also exist in parallel with all the companies' production and trading activities. Companies in the market ought to consider managing these risks if they want to stay in the market in the long - term.

1.1. Rationale of research

Electricity is a volatile asset because of its non – storability, uncertainty and inelasticity in demand (Deng and Oren, 2006). Better forecasts on system price fluctuation will not only reduce the need for provision of reserves and costs, but also improve the process of renewable energy (RE) integration.

Previous studies have shown that besides funding problems, price fluctuation is the most serious problem in financial risk to power companies (Global Sustainable Electricity Partnership (2011), Falbo et al. (2014), Gatzert et al. (2015)). Renewable energy's growing use may put strain on electricity systems' ability to respond to changes, since electricity production of RE source is fluctuated and most electricity systems have been established based on a system architecture that is primarily defined by programmable and non-embedded generation, a well-balanced mix of base - load and peak load. The spot market for electricity, apparently, is highly volatile as its nature, to the point that the price, which normally ranges between $0 and $100 per megawatt hour, could even suddenly increase to more than $10 000 or decrease to as much as -$1000 per megawatt hour. Such price volatility is a dangerous risk exposed to both retailers and generators. An electricity retailer has to negotiate and make contracts with independent customers, such as residents and enterprises, to deliver power at an agreed price, no matter how fluctuated the spot market price is. Whereas, generators, on the other side, face the problem of power quantity to produce to meet the demand. If a generator tried to deal with the spot market directly and any of these circumstances, whether over demand or less than demand power produced, the generator might not get a good return on its investment.

According to Kaminski (2013), since electricity is not storable in a system, its production and consumption must be balanced at all times. This is crucial not only for electricity traders but also investors in physical assets and electricity derivative traders. Eydeland and Wolyniec (2003) used electricity price as one of the products to trade and hedge as it has a huge impact on energy companies' business management. Weron (2006) modelled and forecasted electricity load price for the specific usage of electricity utilities, retail consumer, power generators, traders, hedge funds and other financial institutions.

Crises of unbalanced power generation and demand has led to a vast need of a common place to seek and sell energy. Since utilities cannot pass on their costs to retail customers, they are the most vulnerable in the market, as demonstrated by the California power crisis of 2000-2001 (Joskow, 2001). The Great East Japan earthquake in March 2011 caused the nuclear disaster in Fukushima, which led to the shutdown of nuclear power plants and thus, the electricity usage reduction plan of the Japan's government. A structural decrease from demand had been observed, and price forecasting became crucial for changes of the whole system (Honjo et al., 2018). The 2008 - 2009 financial crisis started from the US, which led to the global huge concern about Co2 emission and the green energy revolution. Adaptation of renewable energy causes a mass fluctuation in the supply chain and the price of the whole system's electricity (Varro et al., 2020). Recent political issues in 2022 triggered a sharp cut down in electricity supply for Europe and a notably price volatile had been seen. The average European price level (annual load-weighted base price of the evaluated countries) has risen sevenfold since 2020, to approximately 235 €/MWh (Kern et al., 2023).

The factors which impact electricity price forecasting have been addressed by a lot of researchers but still, a perfect methodology cannot be found. Bunn and Karakatsani (2003) pointed out some major challenges in forecasting models that require the inclusion of political, social, environmental aspects. Amjady and Hemmati (2006) stated the nature of electricity that caused difficulties from price forecasting, also claimed that the urge to predict this volatile asset is inevitable for companies to minimise their costs and maximise their profits. Liu et al. (2005) tried to use FMSE (Forecast Mean Square Error) and MWE (Mean Weekly Error) to compare the prediction accuracy of different models building on distinctive techniques like equilibrium and volatility analysis, time series, intelligent system method, etc. Contreras and Santos (2006) tried diverse methods of forecasting and concluded that time series models like the ARIMA model, in terms of precision and robustness, are the best among dynamic regression and transfer function models. A more general idea on where to start with predicting electricity prices, what are the market's characteristics and which model to be chosen can be found under the work from Niimura (2006).

1.2. Research questions

The research questions work within the financial risks that could affect any business. This study will produce a model on how the day - ahead electricity price can be predicted, as well as testing its validity by answering the following questions respectively:

- What are the data patterns and characteristics of electricity in the selected market?
- Which proposed model is best to describe the data?
- What will be the upcoming trend of the data?

1.3. Thesis structures

Part 1 – Introduction

This part aims to underscore the critical importance of electricity forecasting and provide a comprehensive context for the current research.

Part 2 – Literature review

This part thoroughly examines and assess previous studies conducted in the field of electricity forecasting, encompassing a comprehensive review of past research efforts.

Part 3 – Methodology

The methodology section provides a comprehensive overview of the data analysis and modelling approach, with a specific focus on time series analysis and modelling techniques.

Part 4 – Empirical study

This part describes the characteristics of the Japanese electricity market since and its process to final liberalisation, as well as detecting the types of patterns that Japanese electricity price gives. The empirical study utilizes data sourced from the Japan Electric

Power Exchange (JEPX) and employs an ARIMA modelling approach to forecast and predict day – ahead electricity prices.

2. LITERATURE REVIEW

2.1. Electricity market characteristics

Zweifel, Praktiknjo and Erdmann (2017) stated that there are three main features of the electricity market. The first one is the consumer surplus of electricity, which creates a distinction to the other types of commodities in the demand function. The second characteristic to be mentioned is the non – storability of electricity. Because of this feature, all market participants must close their open positions (typically every quarter of an hour) with reference to a specific execution period. A position is open if actual demand or supply differs from the contracted (or forecast) quantities. These forecasts depend on many parameters and must regularly be revised whenever new information becomes available. In particular, such a revision becomes necessary when retailers attract or lose new customers. The final special feature of the market is about its different market design options. There are four typical types of market designed from monopoly to fully liberalized market. With a monopoly, all electricity must be offered to a single buyer (usually government – managed). Another with more freedom one is mandatory power pool, in which all generated electricity must be offered to a pool that serves as an exchange for retailers and other affected customers, who are permitted to submit individual bids to satisfy their power needs. Next level of liberalisation is free wholesale competition, in which power can be traded either through bilateral long – term contracts (forward and future contracts with physical settlement), or anonymously through an exchange. The final level of openness of the market is a fully liberalised market with retail competition, which is also the most sophisticated design option because it also gives small customers the right to freely choose their supplies.

2.2. Liberalisation of electricity market

Power trading platforms can only be set up and operate well within fully deregulated power markets. A deregulated market consists of spot market for immediate selling and buying, and forward market for future hedgers. In the spot market there are day – ahead market (the market where price is announced ahead for traders to conduct transaction), intra – day market (a place that enables market participants to adjust the predefined schedules by sending supplementary supply offers or demand bids), daily products market (serves as the platform for trading daily products where participants are obligated to deliver energy).

On the day – ahead markets, both electricity generators and retailers are exposed significantly to price volatility. They can mitigate the risk by entering into long-term derivatives contracts, in which prices are normally agreed upon at the time of contract signing and payment is made later, at the time of delivery.

In a liberalised market, generators have two options to supply electricity. Either they can distribute generated power directly to a final customer through the P2P network, or in the wholesale market, which afterwards, is provided to all the retailers by the network grid's distributors and transmission.

2.3. Predicting day-ahead electricity price

According to Aggarwal, Saini and Kumar (2007), electricity price forecasting models are divided into three main methods, which are Game Theory Models, Time Series Models and Simulation Models. Within the Time Series Model, we also have three particular approaches, including Parsimonious Stochastic Models, Artificial Intelligence based models, and Regression or Causal Models. In the Artificial Intelligence based approach, often researchers – built models by Neural Network based method or Data – mining method.

Autoregressive Iterated Moving Average (ARIMA) model was the favourite model for a lot of researchers in the work to forecast commodities price. In two researches from Hagan (1987) and from Gross (1987), ARIMA was the methodology used to create the model for short term load forecasting and gained extremely well – suited results. A monthly model has been developed by Abdel – Aal and Al – Ghani (1997) to predict Saudi Arabia's electricity consumption price of one year ahead using univariate Box – Jenkins time – series analysis with 5 – year data, resulted best fitted by ARIMA with an average percentage error of 3.8% computed by MPE (Mean percentage error). The same method was used by Saab, Badr and Nasr (2001) to predict Lebanon's price, but the Autoregressive AR (1) model was the best fit within data from this country. Lee, et al. (2006) provided a short – term day – ahead system marginal price forecasting model using ARIMA model for the Korean Power Exchange.

ARIMAX model, which is an extension of ARIMA model with explanatory variables, was adopted by some researchers. Chinn (2001) applied ARIMAX models to examine the relationship between spot and futures price of electricity, as well as other commodities including crude oil, gas, heating oil and natural gas. Yunta et al. (1999) also adopted the univariate ARIMAX technique with meteorology as an explanatory variable for the Spanish power exchange.

Regarding to the techniques to remove seasonality, the use of general exponential smoothing to create an adaptive forecasting system based on observed integrated hourly demand values is investigated before the deliberation of the electricity market in a paper by Christiaanse (1971). Weron and Misiorek (2005) used two approaches to create two models to predict and compare electricity load. One model used the differencing technique, while the other one used moving average smoothing and logarithm to remove seasonality.

Artificial Neural Network (ANN) techniques, which were previously used primarily for load forecasting, are now also being used for price prediction. Wang and Ramsay (1998) used this method to predict the system marginal price with references to weekends and public holidays errors. Szkuta, Sanabria and Dillon (1999) describes the use of ANN computing techniques to implement system marginal price short – term forecasting using real – world data from Victoria (Australia) deregulated power system. Richter et al. (2000) used some techniques within ANN methodology and compared them to identify patterns of the time series data, which would be the leading indicators for future electricity price forecast. Hippert, Pedreira and Souza (2001) collected a range of papers on load forecasting based on ANN techniques to examine the usefulness of this approach.

Dynamic regression and transfer function models from the Time Series based approach are used by Nogales and Contreras (2002) to predict next – day electricity price for the Spanish and Californian market. A more complicated way, the support vector regression (SVR) model, was also used to predict Northern China's electricity price. The model combines the seasonal adjustment mechanism and a chaotic immune algorithm (Hong et al., 2011).

3. METHODOLOGY

The quantitative method is used in the thesis to reveal patterns and correlations by measuring variables and testing relationships between variables. This thesis, in particular, follow the approach by Contreras, et al. (2003) in order to apply it on another empirical case study. A quantitative instrument was used in a time series experiment with historical data from a publicly accessible web.

3.1. Time series analysis and application

3.1.1. Data types and patterns

In time series data, it is crucial to identify if a set of data is stationary or nonstationary. What are known as stationary models are a class of stochastic models, which assumes that the process is in statistical equilibrium, with probabilistic properties that do not alter over time. It varies around a fixed constant mean level with constant variance. However, forecasting, on the other hand, is often worked with time series that are better represented as nonstationary. Nonstationary time series, in contrast, have non – constant mean over time and/or non – constant variance. Researchers since previous decades have soon realised this problem, thus, methods like differences and exponentially weighted moving averages were proposed to smooth and transform data into stationary series. Even though these methods cannot be applied to complicated data efficiently, they still produce a suitable forecast function that could suggest a nonstationary model best describe original data. Example of stationary and non – stationary data is shown in Figure 1 and Figure 2.

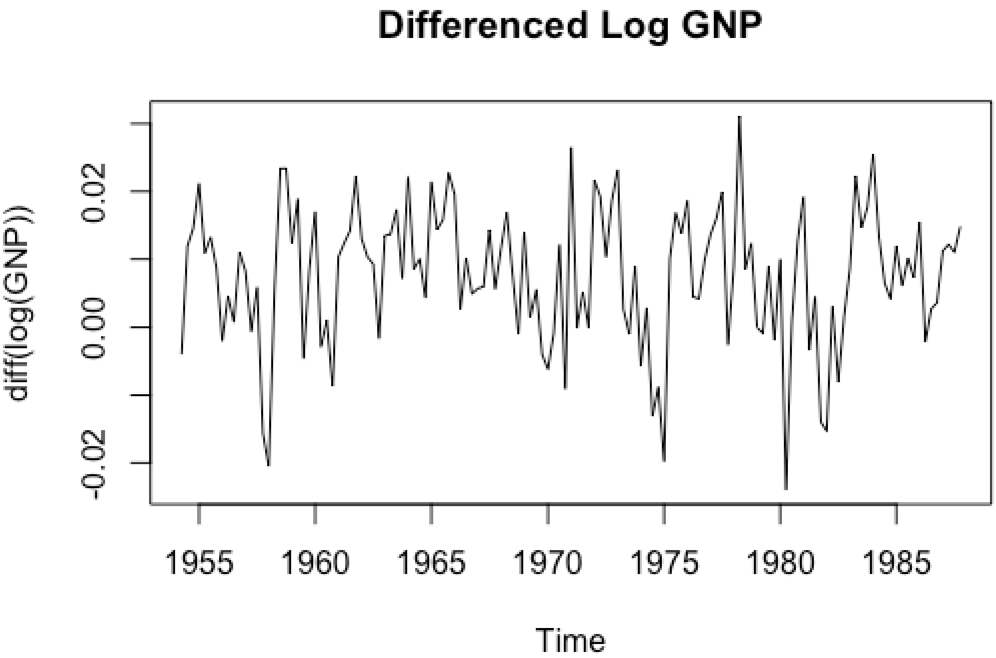


Figure 1. Differenced US log gross national product (GNP) 1955 – 1985

(Source: Alex (2020))

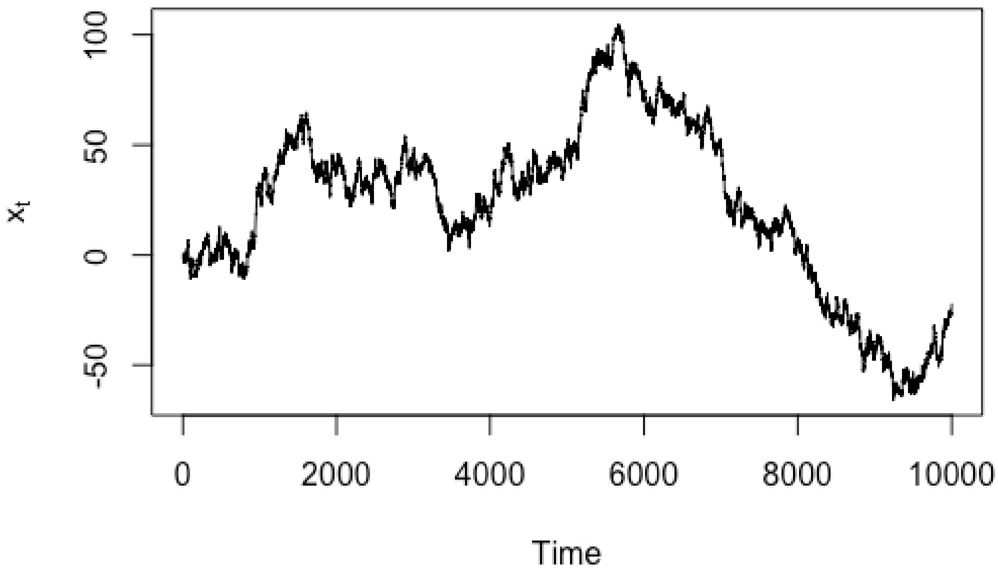


Figure 2. A random walk series (Source: Alex (2020))

Besides data type, one when working with a set of data also has to detect the typical patterns or trends of the data movement overtime to choose a suitable model. A trend is observed when there is a long – term increase or decrease in the data. The set of data does not have to be a perfect straight line, but a gradual shift in the direction of data overtime, either upward or downward, can be defined as an increasing or decreasing trend. Another pattern of data to be examined is seasonality. When a time series is affected by seasonal factors such as the time of year or the day of the week, a seasonal pattern is formed. Seasonality has a fixed and predictable frequency. A final characteristic to mind when seeing data is called cycle, which occurs when the data exhibits nonlinear rises and falls. These variations are frequently correlated with the "business cycle", which usually resulted from an economic cycle that often lasts for around 4 – 10 years (Tejvan P., 2018). Figure 3, 4, 5 and 6 show the examples of data with trends, patterns of seasonality and cycle.

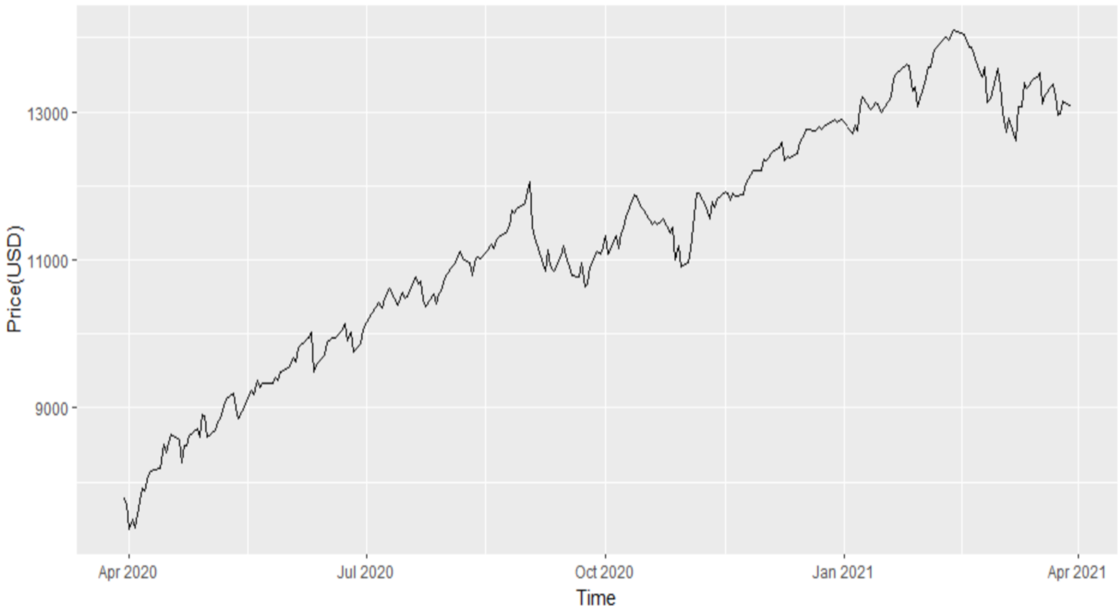


Figure 3. NASDAQ composite in USD for the period April 2020 – April 2021

(Source: Yahoo finance)

Figure 3 shows a strong increasing trend of the indices due to the social and economic events that occurred during the time, most of its was the sharp tighten monetary policy from Fed to cope with the pandemic (Thorbeck, 2021).

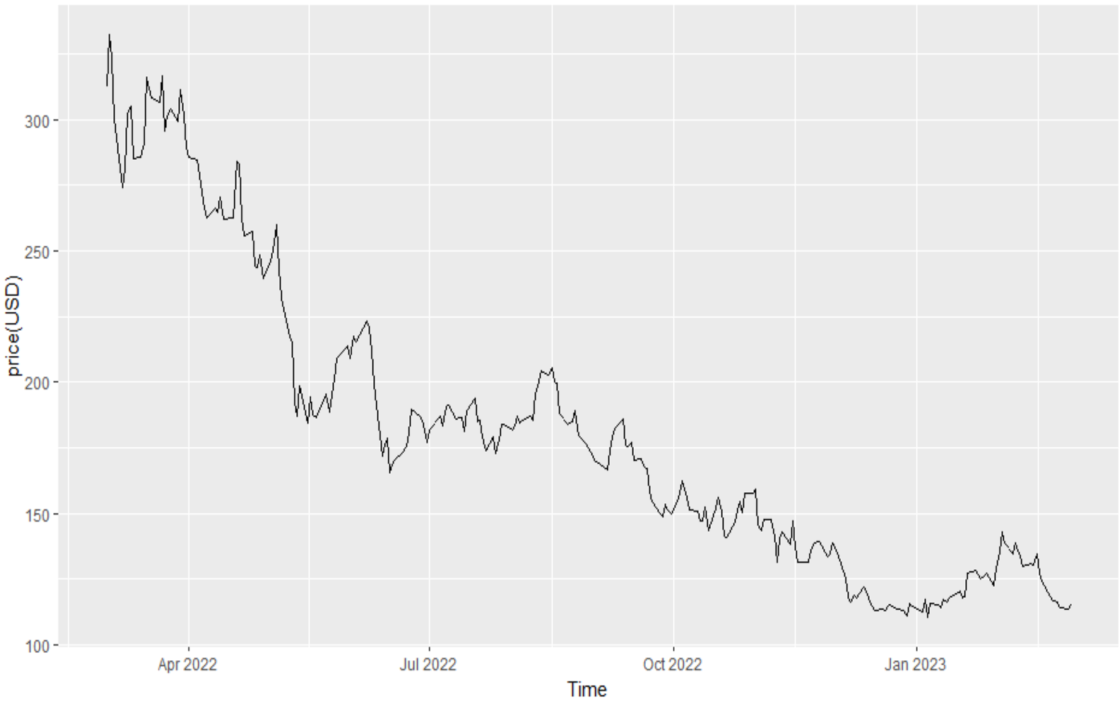


Figure 4. Signature Bank stock price (USD) March 2022 – March 2023

(Source: Yahoo finance)

Figure 4 shows an obvious downward trend as price constantly going down in 1 year.

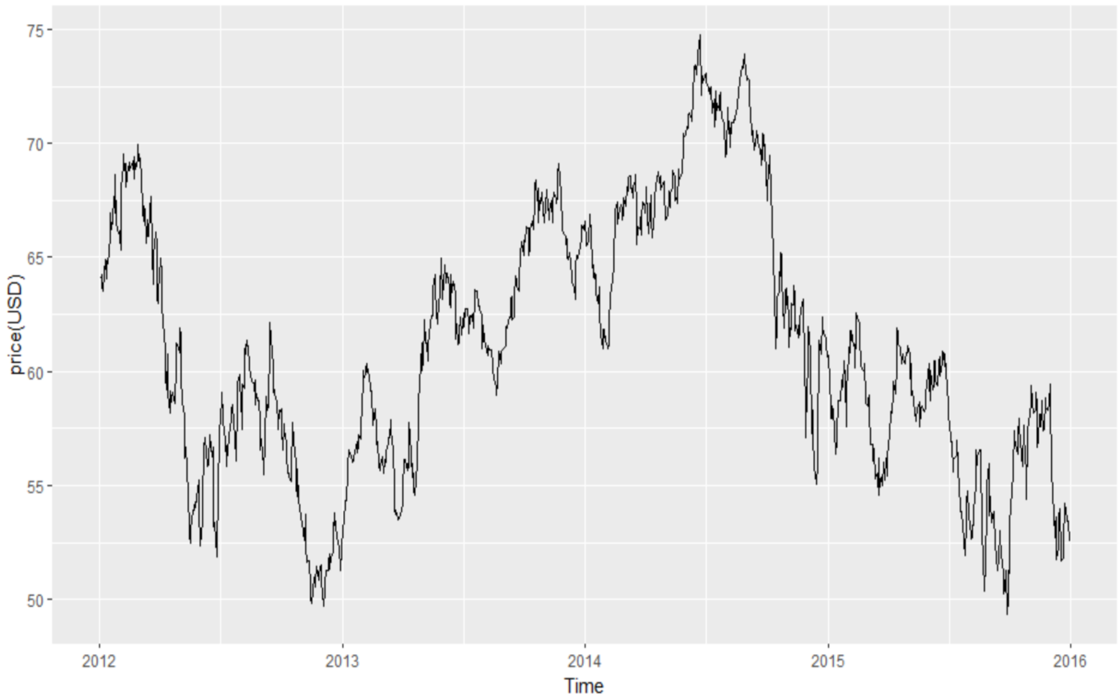


Figure 5. Occidental Petroleum Corporation (OXY) stock price (USD) 2012 – 2015

(Source: Yahoo finance)

Figure 5 is characterized by seasonality due to the different use of energy throughout a year. From the graph, we can see that OXY stock seasonality frequency including monthly, quarterly and yearly seasonality. The choice of frequency is dependent on the need of data usage and analysis.

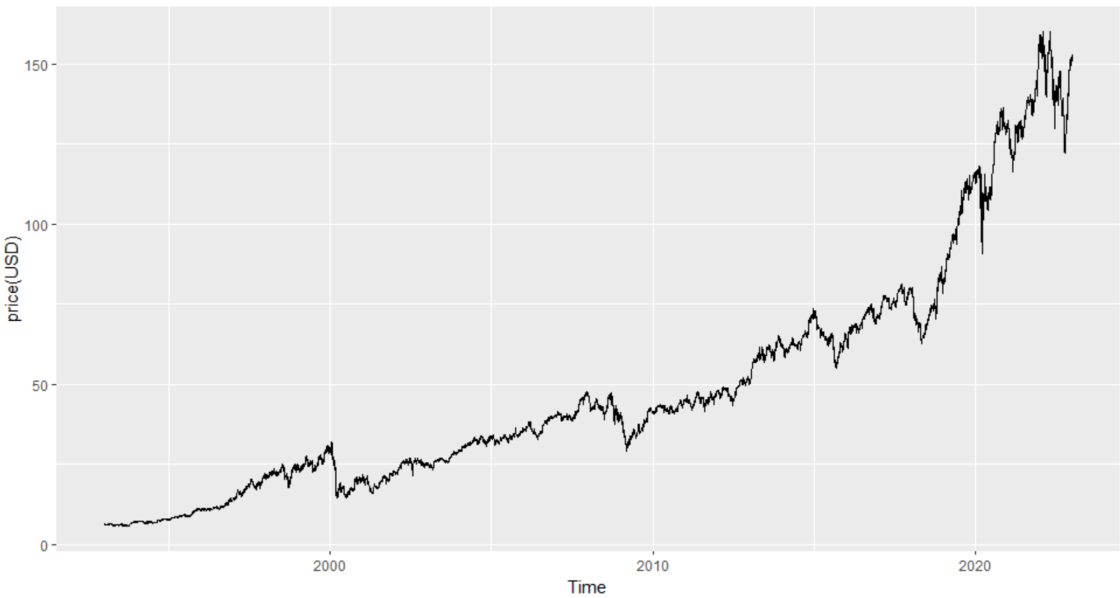


Figure 6. The Procter & Gamble Company (P&G) stock price 1993 – 2023

(Source: Yahoo finance)

The stock price of P&G is characterized by multiple cycles. The first one is from 1993 to 2000, following by a Dot – com recession; the second one is the period 2000 – 2009, following by the 2008 – 2009 world financial crisis (Huddleson Jr., 2020); the third one is continuing to 2016 and stock price is decreased due to the industrial recession of 2016 (Irwin, 2018), and so on.

Stationary tests

The stationary tests go hand – in – hand with the unit root tests. In detail, unit root tests are statistical tests used to determine whether an autoregressive (AR) model has a unit root with an absolute value of one. One commonly used unit root test is the augmented Dickey – Fuller test (ADF test). The null hypothesis of the ADF test is that the process under investigation has a unit root, indicating non – stationarity. The alternative hypothesis suggests that the process is stationary. Another unit root test that is commonly used is the KPSS (Kwiatkowski – Phillips – Schmidt – Shin), with the null hypothesis stating stationarity and the alternative hypothesis indicating the presence of a unit root, which is the opposite of the null hypotheses in the Dickey – Fuller. The third and final test is the Phillips – Perron test, which is similar to the Dickey – Fuller test but has some notable differences. The null hypothesis for the Phillips – Perron test is the presence of a unit root, while the alternative hypothesis indicates stationarity and Phillips – Perron tests. These unit root tests help to assess the stationarity properties of time series data and provide insights into the underlying dynamics of the data – generating process.

3.1.2. Transforming data

Statistical methods work best when the data is normally distributed, or symmetrically distributed with a constant variance. Transformed data shows less skewness and a more constant variance when compared to the original variable will make it easier to see any particular trend or pattern, which helps to choose a suitable model. That is the main reason why data analysts mostly work with transformations of variables, such as log, square root, or other power transformation rather than the original variables. If a transformation removes a dependency between the conditional variance and the conditional mean of a variable, it is called variance stabilising.

Seasonality difference

According to Hyndman (2021), a seasonal difference refers to the computation of the difference between an observation and the corresponding observation from the previous occurrence of the same season. The difference is written as

$$y'_t = y_t - y_{t-m}$$

with m is the number of seasons.

When the seasonally differenced data exhibit characteristics of white noise, then the model becomes

$$y_t = y_{t-m} + \varepsilon_t$$

3.1.3 The Backwards Operator

The backwards operator, denoted by B, is a straightforward. It is commonly employed to represent ARMA (Autoregressive Moving Average) and ARIMA (Autoregressive Integrated Moving Average) models.

$$B_{Y_t} = Y_{t-1} \tag{1}$$

The notation B_{Y_t} signifies that the backward operator B applied on Y_t makes data shift back one time period, which means Y_{t-1}. In a more general sense, B^h indicates that the backwards operator B is applied repeatedly h times, resulting in a shift of h time units.

$$B^h Y_t = Y_{t-h}$$

It is worth noting that, constant value remains unchanged over time, so B_c = c holds true for any constant c. Additionally, B is also referred to as the lag operator, emphasizing its role in shifting the time index of a time series.

3.1.4. Lags detection

Firstly, let Y_t denote the original series and y_t denote the differenced (stationaries) series. When there is no different (d = 0), then

$$y_t = Y_t$$

As stated by Wei (2005), auto-correlation function (ACF) is a complete function that returns auto-correlation values for any series with lagged values. For a stationary process y_t, with mean E $(y_t)= \mu$ and variance Var (y_t) = Var $(y_{t+k}) = \gamma_0$, the ACF at time lag k is denoted by

$$\rho_k = \frac{Cov(y_t,\ y_{t+k})}{\sqrt{Var(y_t)}\sqrt{Var(\ y_{t+k})}} = \frac{\gamma_k}{\gamma_0} \tag{2}$$

with k = 0, 1, 2, …, γ_k is the covariance between Y_t and Y_{t+k} and is defined as

$$\gamma_k = Cov(y_t, y_{t+k}) = E(y_t - \mu)(y_{t+k} - \mu) \tag{3}$$

Partial auto-correlation function (PACF) is an addition for ACF study. Instead of finding all (direct and indirect) correlations between y_t and y_{t+k}, the PACF considers only the direct correlation between y_t and y_{t+k}. In detail, it is partial rather than 'complete' because previously found variations are removed before the next correlation shows up. Thus, it would be useful if any hidden information in the residual becomes obvious to be included in the time series model. PACF at time lag k is defined by

$$\phi_{kk} = corr(Y_{k+1} - P_{\overline{sp}\{Y2,\ldots Yk\}}Y_{k+1}, Y_1 - P_{\overline{sp}\{Y2,\ldots Yk\}}Y_1)\text{ , k>2} \tag{4}$$

3.1.5. Data modelling

Non – seasonal Auto – Regressive Integrated Moving Average (ARIMA) is the most general class of forecasting models for time series that can be stationaries by transformations such as differencing, logging, and or power transform.

For p is order of the autoregressive part, d is the degree of differencing, and q is the order of the moving average part, the ARIMA (p, d, q) model for Y_t can be written as

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_{p+d} Y_{t-p-d} + \theta_0 - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q} + \epsilon_t \qquad (5)$$

where $\varphi(B)$ is the non – stationary autoregressive operator that is assumed to have unit root, $\theta(B)$ is the moving average operator that is assumed to be invertible and

ϵ_t is the error of the series that is $\{\epsilon_t\}$ are i.i.d. N $(0,\sigma_\epsilon^2)$ Gaussian white noise process, θ_0 is the constant of the model and

$$\theta_0 = \mu(\phi_1 - \phi_2 - \cdots - \phi_p) \qquad (6)$$

with μ is the mean of the process y_t.

Let $\phi(B)$ be the stationary autoregressive operator that is assumed to be stationary, and $\nabla = 1 - B$ be the differencing operator, the equation for ARIMA model can also be written as

$$\varphi(B) Y_t = \phi(B) \nabla^d y_t = \theta_0 + \theta(B) \epsilon_t \qquad (7)$$

with

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \qquad (8)$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \qquad (9)$$

Seasonal ARIMA (SARIMA) model

"If d and D are nonnegative integers, then $\{Y_t\}$ is a seasonal ARIMA (p, d, q) × (P, D, Q) $_s$ process with period s if the differenced series

$y_t = (1 - B)^d (1 - B^s)^D Y_t$ is a causal ARMA process defined by

$$\phi(B)\Phi(B^s) Y_t = \theta(B)\Theta(B^s)\,\epsilon_t\,, \{\varepsilon_t\} \sim WN(0, \sigma^2) \qquad (10)$$

where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, $\Phi(B) = 1 - \Phi_1 B - \cdots - \Phi_{\mathrm{P}} B^{\mathrm{P}}$, (11)

$\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B_q$, and $\Theta(B) = 1 + \Theta_1 B + \cdots + \Theta_{\mathrm{Q}} B^{\mathrm{Q}}$." (12)

(Brockwell et al. 2016, p. 203)

3.1.6. Maximum Likelihood Estimation

There are several methods to estimate parameters in the ARIMA model, such as the Method of Moments, Maximum Likelihood, Nonlinear Estimation, Ordinary Least Squares Estimation in Time Series Analysis (Wei, 2006). However, Maximum Likelihood Estimation is one that is widely used and also the one being used as default in R.

"Generally, the conditional and unconditional least-squares estimators serve as satisfactory approximations to the maximum likelihood estimator for large – sample sizes. However, the simulation evidence suggests a preference for the maximum likelihood estimator for small or moderate – sample sizes, especially if the moving average operator has a root close to the boundary of the invertibility region." (Box, et al., 2006, p. 217)

Thus, for a ARIMA (p, d, q) model with $\{\epsilon_t\}$ are i.i.d. N $(0,\sigma_\epsilon^2)$ Gaussian white noise process, the joint probability density of $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$ is given by

$$P(\phi, \theta, \mu, \sigma_\epsilon^2|\epsilon) = (2\pi\sigma_\epsilon^2)^{-\frac{1}{2}n}\, exp\left[-\frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^{n}\epsilon_t^2\right] \quad (13)$$

Let Y = $(Y_1, Y_2, \ldots Y_t)'$. The conditional log – likelihood function is

$$S_*(\phi, \mu, \theta) = \Sigma_{t=p+1}^{n}\varepsilon_t^2(\phi, \mu, \theta|Y) \quad (14)$$

the μ and ϕ being found that minimize the above function will be chosen, thus, is called conditional least – square estimator and will be the value to start with maximum likelihood of observing the given data.

3.1.7. Forecast function

Brockwell & Davis (2016) have created a general solution for the forecast function of ARIMA model.

First, let Y_0 be the random variable that define

$$Y_t = Y_0 + \sum_{j=1}^{t} y_j \quad (15)$$

and $\{Y_t\}$ be the observed process that satisfied the difference equations

$$(1 - B)^d Y_t = y_t \quad (16)$$

where y_t is a causal ARMA (p, q) process, t = 1, 2, … and random vector $(Y_{1-d}, \ldots, Y_0)$ is uncorrelated with y_t (t > 0)

Call P_n the best linear predictor of observations up to time n, and our goal is to optimize the value $P_n Y_{n+h}$, with h being the step forecast ahead.

Finally, we have the forecast function as

$$g(h) := P_n Y_{n+h} \tag{17}$$

satisfies the homogeneous linear difference equations

$$g(h) - \phi_1^* g(h-1) - \cdots - \phi_{p+d}^* g(h-p-d) = 0, h > q \tag{18}$$

where $\phi_1^*, \ldots, \phi_{p+d}^*$ are the coefficients of $y, \ldots, y^{p+d}$ in $\phi^*(y) = (1-y)^d \phi(y)$

The solution to g(h) is

$$g(h) = a_0 + a_1 h + \cdots + a_d h^{d-1} + b_1 \xi_1^{-h} + \cdots + b_p \xi_p^{-h}, h > q - p - d \tag{19}$$

with $a_1, \ldots, a_d$ and $b_1, \ldots, b_p$ are the coefficients calculated from $(p + d)$ taken from $q - p - d \leq h \leq q$

Forecast for SARIMA

Developing from the above function, the SARIMA forecast function is

$$g(h) := P_n Y_{n+h} + \sum_{j=1}^{d+D_s} a_i P_n Y_{n+h-j} \tag{20}$$

3.1.8. Residuals

According to Hyndman (2021), model's residuals are the parts lying around after the model being fitted. Hence, it is defined as

$$e_t = Y_t - \hat{Y}_t \tag{21}$$

When comparing the observed values with the fitted model's predicted values and the fitted model shows suitability, the residuals would then behave in a consistence manner to the model (Brockwell et al., 2002).

The forecast method that is well – produced will set residuals lying within some certain features in order to satisfy the initial assumption for Y_t. The first one is that the residuals are uncorrelated to others if Y_t is white noise WN $(0, \sigma^2)$. There should be information that was not being utilized to build model if there was still residuals correlation. To check if the residuals are correlated or not, ACF or PACF plots can be used to see if there is any significant lag to be included in the ARIMA model. The second property that residuals must follow is normally distributed as Y_t is N $(0, \sigma^2)$. To check for normality, mostly visualization like histogram or QQ – plot is being used.

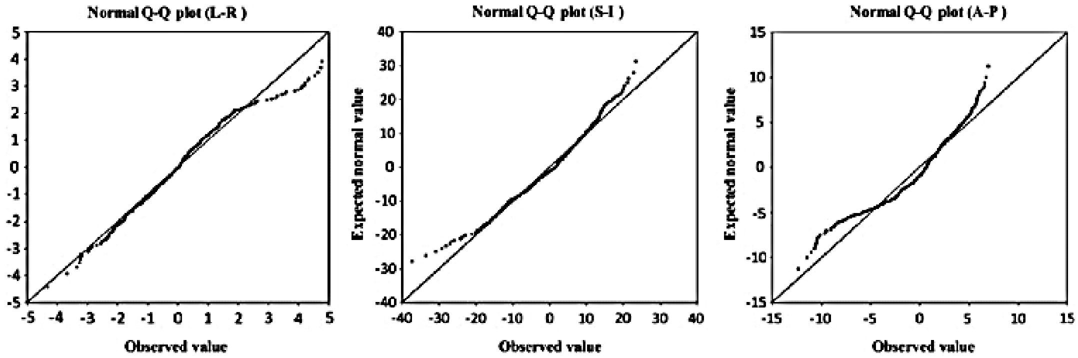


Figure 7. Example of normal Q – Q plot (Source: Researchgate)

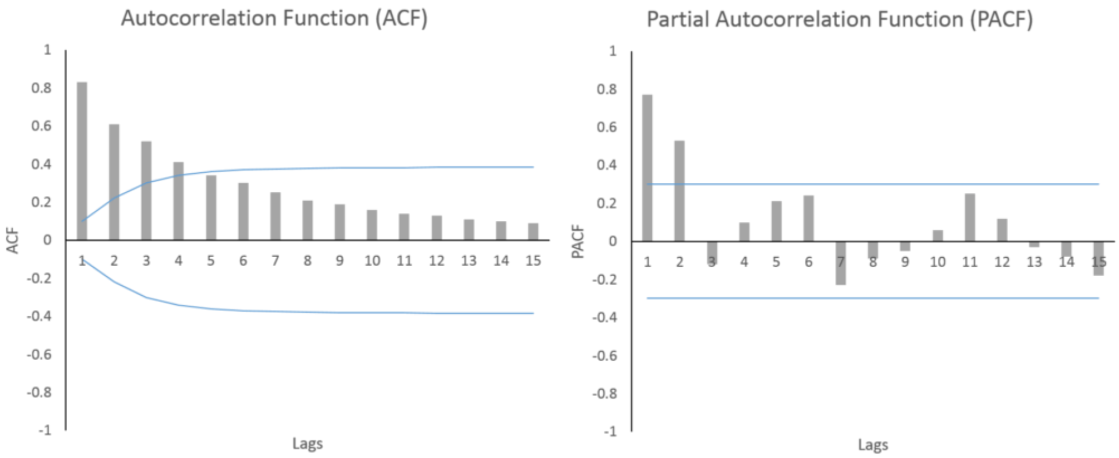


Figure 8. Example of ACF and PACF plots (Source: spureconomics)

To test for autocorrelation, besides ACF/ PACF, a simultaneous test like Ljung – Box test is an alternative approach. It uses a set of null hypotheses to assess if any of them are untrue. In the case of the Ljung – Box test, the null hypothesis (H_0) states that the autocorrelations at various lags, denoted as (1), (2), ..., (K), are all equal to zero. By conducting the Ljung – Box test, if it rejects the null hypothesis, it suggests that at least one of the autocorrelations is non – zero. It is important to note that if all lagged autocorrelations from 1 to K are indeed zero, there is a 1 in 20 chance (given a significance threshold of 0.05) of incorrectly concluding that they are not all zero.

3.1.9. Adjusting outliers

According to Langer et al. (2020), outliers are defined as the points which deviate significantly from the mean of the corresponding random variable. Outliers might or might not have a significant impact on the error assessment, therefore, control the quality of the model's formation.

According to Pankratz (1993), outliers, after testing its impacts, worsen forecast and discover it explains major events that happened to the fluctuation of data. Fox (1972) suggested two types of test models that detect outliers and its level of affection toward variables. Burman (1984) discussed the outlier matter that had not gone under examination from Hillmer, Bell & Tiao (1983), since the researchers concluded that including it resulted in a less stable seasonal pattern of model. However, Ledolter (1988) showed that outliers may worsen forecasts because it may bias parameter estimation.

Some of the methods to detect and solve the problem of outliers accordingly are iterative procedures by Tsay (1986), Chang, Tiao & Chen (1988), Chen & Liu (1993), least squares techniques and residual variance ratios (Tsay, 1988), likelihood ratio and score criteria (Ljung, 1993), a procedure based on a robust estimate of parameters and residuals computed by means of robust filtering (Bianco et al. 2001).

3.1.10. Forecast errors, prediction interval

In accordance with Hyndman and Athanasopoulos (2018), a forecast error calculated by differencing an observed value and its forecast. Let i be the number of lead time. The forecast error at lead time i $e_t(i)$ is written as

$$e_t(i) = Y_{t+i} - \hat{Y}_t(i) \tag{22}$$

where Y_{t+i} is the forecast origin larger than or equal to the length of the series so that the evaluation is based on out sample forecast. T and i are the current time point and numbers of period into the future respectively.

There are differences between forecast residuals and forecast errors. The first is that forecast residuals are calculated on the training set (or in – sample data), while forecast errors are calculated on the test set (or out – of – sample data). The training set is used to estimate the parameters of a forecasting method, while the test set is the part of the data that are being held out to employ accuracy assessment of the model. The test set is typically about 20% of the sample. The second difference is that forecast of residuals are based on one – step, which means that the next one value in the sequence would be predicted on the previous ones and residuals are then obtained by comparing these one – step – ahead predictions with the corresponding value in the training data (actual values). Meanwhile, it involves multi – step to forecast errors. (Hyndman, 2018). In the test phase, multiple forecasting points ahead are being generated and forecast errors are computed by comparing the multi – step – ahead predictions with the actual values in the data group being used as test set. This creates an evaluation of the model's performance in capturing longer-term trends and dynamics.

When generating forecasts, it is important to seek for a level of confidence associated with the predictions. In order to achieve this, one can begin by calculating the variance of the forecast error. By adding or subtracting the standard deviation of the forecast error multiplied by z/2 (the normal upper quantile), a prediction interval with a (1 − α)100% confidence level can be obtained. It is important to note that this approach assumes residuals follow a Gaussian white noise distribution. However, if the residuals exhibit a heavy-tailed distribution, it may raise a concern that needed to be solved.

The most common measurement for errors is based on the absolute errors or squared errors. Namely, some of those measurements are mean absolute error (MAE), root mean squared error (RMSE), percentage errors, mean absolute percentage error (MAPE), scaled errors.

The mentioned statistics to compare models are listed below.

The mean percentage error is $MPE = (\frac{1}{M}\sum_{i=1}^{M}\frac{e_i}{Y_{t+i}})100\%$ (23)

The mean square error is $MSE = \frac{1}{M}\sum_{i=1}^{M} e_i^2$ (24)

Mean absolute error $MAE = mean\ (|e_t|)$ (25)

Root mean squared error $RMSE = \sqrt{mean(e_t^2)}$ (26)

Percentage errors $MAPE = mean\ (|pt|)$ with $p_t = \frac{100e_t}{y_t}$ (27)

3.1.11. Models' comparison

After fitting different statistical models to the data, one has to make decision in choosing the model that is best for the data. Akaike's information criterion (AIC) and Bayesian information criterion (BIC) are the two means that are widely used to achieve a good trade-off between fitness of models and complexity. AIC use maximal log-likelihood value, given here by $\log\{L(\hat{\theta}_{ML})\}$, to assess how well a model fits the data or. The smaller the AIC value, the more maximized $\log\{L(\hat{\theta}_{ML})\}$ is, means the better the model fits the data. However, $\log\{L(\hat{\theta}_{ML})\}$ can be enhanced also when more parameters are added, which may cause problem of overfitting, since increasing model complexity may simply be fitting random noise in the data. As a result, BIC will also be used to compare model complexity as well.

$$AIC = -2log\{L(\hat{\theta}_{ML})\} + 2p \tag{28}$$

$$BIC = -2log\{L(\hat{\theta}_{ML})\} + log(n)\, p \tag{29}$$

where p is the number of parameters in the model selected, and n is the sample size.

4. EMPIRICAL STUDY

4.1. Japan electricity market designed

In the history of the Japanese electricity market, since 1995, the power generation and transmission sector have already been integrated, which allowed entities free entry into the market. After, there was the period of 2000 – 2005 retail sector liberalisation, except low – voltage customers. As a result of the demand, in November 2003, the Japan Electric Power Exchange (JEPX) was founded, and trading began in April 2005. JEPX, during this time, was initially a private entity that worked voluntarily for wholesale exchange. It was not until April 2013 that the electricity system reform policy was adopted. 3 years later in April 2016, the liberalisation of the electric power retailing and generation sectors was finally completed and JEPX was designated as a wholesale electricity market under the provisions of the Electricity Business Act. In April 2020, some legal acts were implemented that changed the market, along with the formation of new transmission and distribution companies from the former general electricity utilities. As of today, the Japanese electrical power industry is divided into three major sectors, including generation, transmission and distribution, and retailing. JEPX operates under the management of the government's Agency for Natural Resources and Energy. Meanwhile, the Organization for Cross – regional Coordination of Transmission Operators (OCCTO) monitor supply/demand and grid operational status on a national scale for the market to run smoothly.

Electricity generation utilities and electricity retailers, which are the primary JEPX participants, accounted for 266 trading members of the marketplace. It is characterised by 4 main types of transaction market. The spot market, which is also the biggest one with the most participants, is the place to trade electricity in 30 – minute increments for next-day delivery. The forward market serves trading for electricity delivery at some point in the future, mostly used by hedgers. The products contained could be monthly 24 – hour products or weekly daytime products. The intra – day market is for correcting unexpected supply and demand misalignments that occur between a spot market transaction and delivery (a minimum of one hour later). Finally, the bulletin board trading market is the place where JEPX facilitates electricity trading for prospective buyers and sellers. A blind single – price auction system is used for trading, in which the system price and trade volume are defined as the intersections of the sell and buy bid curves. High buy bids and low sell bids are typically executed at the system price.

4.2. Application of the forecasting method to JEPX day – ahead price

Electricity price is characterized by seasonality, which could be daily, weekly and monthly. Aspects like weather conditions, energy demand and supply all have an impact on the seasonality of electricity prices. In Japan, summer months (June to August) and winter months (December to February) typically have increased electricity demand and this might result in higher electricity rates.

During summertime, energy demand increases because of hot and humid weather, mostly from use of air conditioning equipment. Because of demand increasing for cooling, the electrical grid may be stressed and potentially leads to higher pricing in electricity.

The need for electricity also rises during wintertime, particularly in colder places in the north, like Hokkaido. The use of electric heaters and heat pumps are strong in everywhere, which also leads to higher electricity rates for the increased demand. Meanwhile, during spring (March to May) and autumn (September to November) months, the weather is milder. It is necessitating less to use heating or cooling systems, which leads to sharp demand in electricity, and eventually, in lower pricing.

It is crucial to also add that while these trends are generally observed, actual electricity prices can vary depending on location, time of day, market conditions, as well as the specific pricing structures of electricity providers. These factors should be considered in choosing data to analyse to have a more accurate forecast of the electricity price.

One critical thing when working with a time series dataset that exhibits excessive or complex seasonality is to adequately detect and comprehend the seasonal patterns, if there is. This entails identifying whether the seasonality is on a daily, weekly, monthly, or other basis by using ACF plots, statistical tests, etc. Then choosing the proper seasonality component to add in the ARIMA model will be much easier. Because when the seasonality in the data is too complex, R will not be able to capture the seasonal parts of the model and it may be necessary to analyse the data at a finer time granularity. One can disaggregate the data into daily, weekly, or monthly components, and just apply the ARIMA model to the relevant disaggregated data subset.

Choosing the appropriate time frame to forecast data with the ARIMA will aid in capturing the correct seasonal patterns, as well as improving forecast accuracy and parameter estimation. Firstly, seasonality can have a substantial impact on time series data behaviour. For example, if the forecasting time frame is not long enough and/or does not include the entire seasonal cycle, the model may fail to accurately capture the seasonal patterns. Therefore, choosing a time span that accurately depicts seasonal variance increases the likelihood of producing more precise and reliable forecasts. Secondly, ARIMA model calculates the necessary parameters and the estimation procedure is as well greatly influenced by the time frame chosen. A longer time frame with numerous seasons allows for more data points for parameter estimation, resulting in more robust and reliable model parameter estimates. Thus, this improves the accuracy of the model's forecasts.

4.2.1. Data description

Understanding the importance of seasonality effects, after careful consideration, this thesis chooses Hokkaido and Kyushu in Japan as the two regions to analyse dates and forecast. Since both places have various meteorological conditions and energy demand patterns, choice of these two to examine electricity might give a possibly contradictory to the forecast result.

See Figure 9, Hokkaido located in northern Japan generally has a colder and longer winters than the rest of the country. This subsequently raises the need for heating systems to counteract the cold weather and drives up demand for electricity in Hokkaido during the winter months. Its location leading to its needs is the main reason for higher electricity price. Else while, summer electricity demand and price in Hokkaido, may be lower because the region's milder environment lessens the need for use of air conditioning.

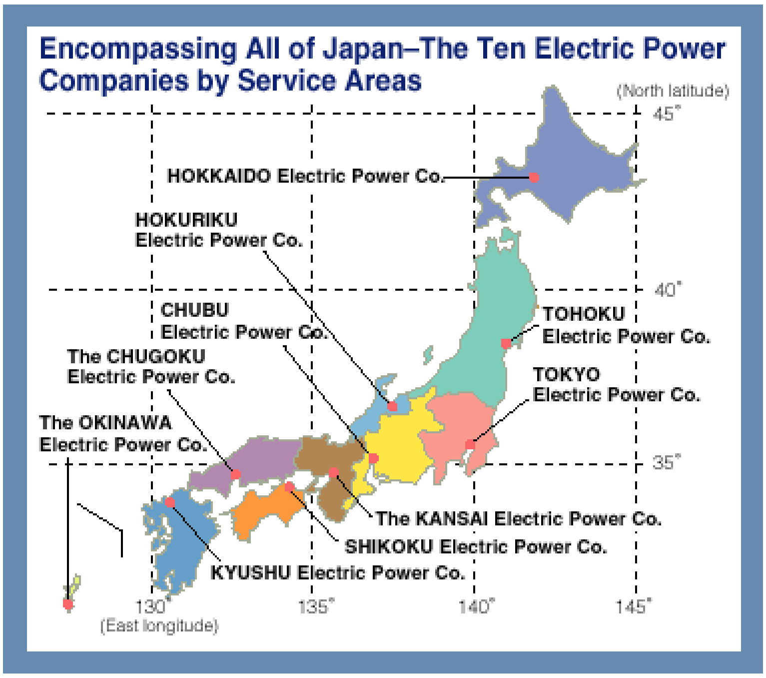


Figure 9. Map of electricity generator over Japan

(Source: Global energy network institute)

Kyushu, on the other hand, has a warmer weather thanks to its location in south. This also means that electricity price during summers in Kyushu is higher because the weather is hot and humid, resulting in a rise in demand on electricity that is necessary to run air conditioning equipment. The winter months of Kyushu can be said to be not as severe as those in Hokkaido, which has meaning in a potential lower electricity demand, that explains for the frequent lower price of the region during this time.

This thesis examines daily electricity prices from the day – ahead market in JPY/kWh observations within 2022 – 2023. In particular, four data sets within this timeframe are selected which shows the obvious distinction of electricity price movement within two regions of the same country in the same timeframe. After inspection from price movement from April 1st 2020 to the current time (see Figure 10), it is shown that electricity price in Japan during springtime is always the lowest.

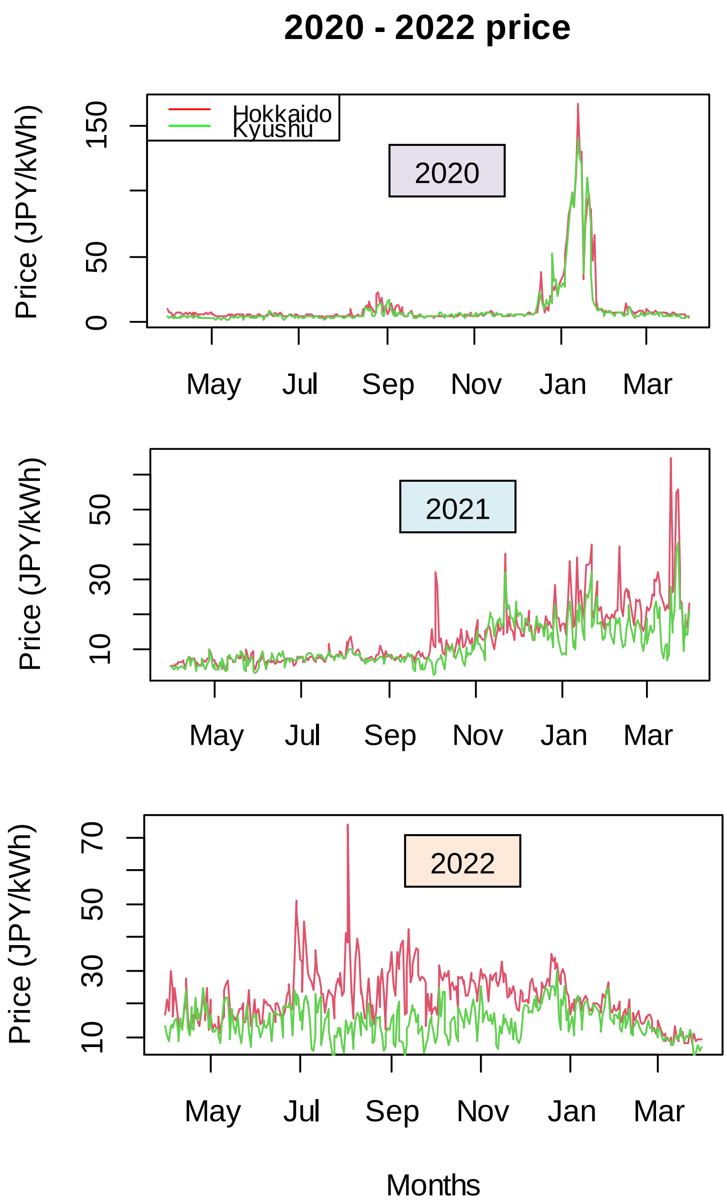


Figure 10. Electricity price from 2020 to 2022 (JPY/ kWh)

Hence, the first data set chosen is electricity price in Hokkaido during spring season from March 1st 2023 to 24th May 2023. The second data set is Kyushu's electricity price from March 1st 2023 to 24th May 2023. Meanwhile, according to Honjo et al. (2018), summertime has always gotten a higher demand for electricity and price in Japan, even though both summer and winter are characterised with high price in historical data. Therefore, the third and fourth data set chosen for the study are a timeframe from 1st June 2022 to 24th August 2022 for both Hokkaido and Kyushu. The data sets were acquired from the Japanese Energy Power Exchange (JEPX).

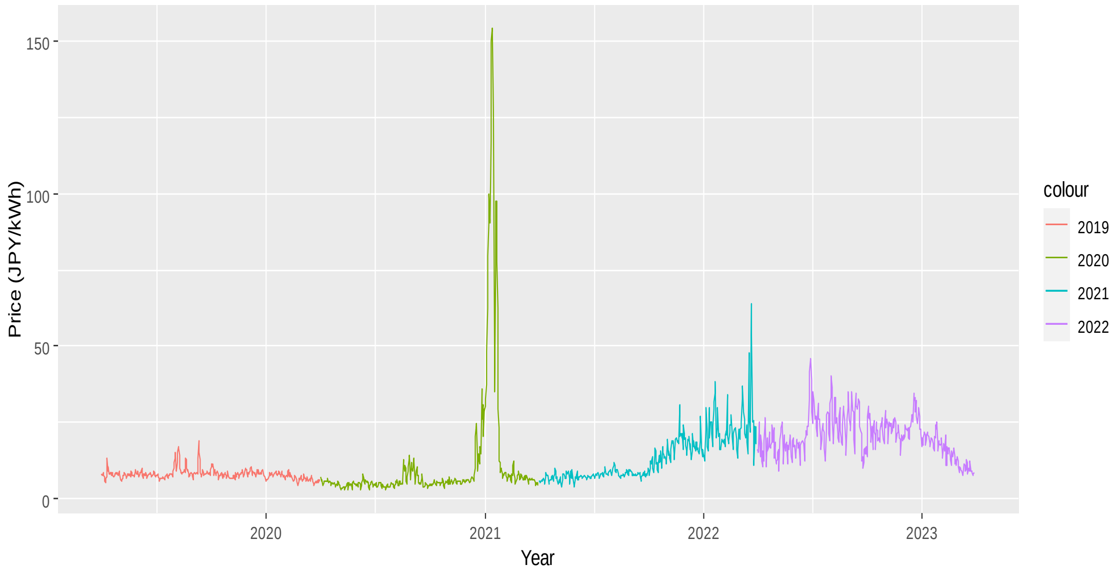


Figure 11. Plot of 2019-2023 system price

The analysis of historical pricing revealed that the electricity price in the Japanese electricity market has not risen considerably (see Figure 11). The only increase was linked to rising fuel prices and rising expectations for higher future CO2 pricing (Schmittmann, 2023). A particular peak period was observed in winter 2020/2021 due to the unusual cold spell, the reactivation of industry following the global pandemic in 2020 and fear of increasing inflation affected worldwide (Kyodo, 2021). The average power price decreased by 3% from 2022 and 2023 following the plan to achieve sustainable energy by 2050 (Blaker, 2023).

Descriptive statistics

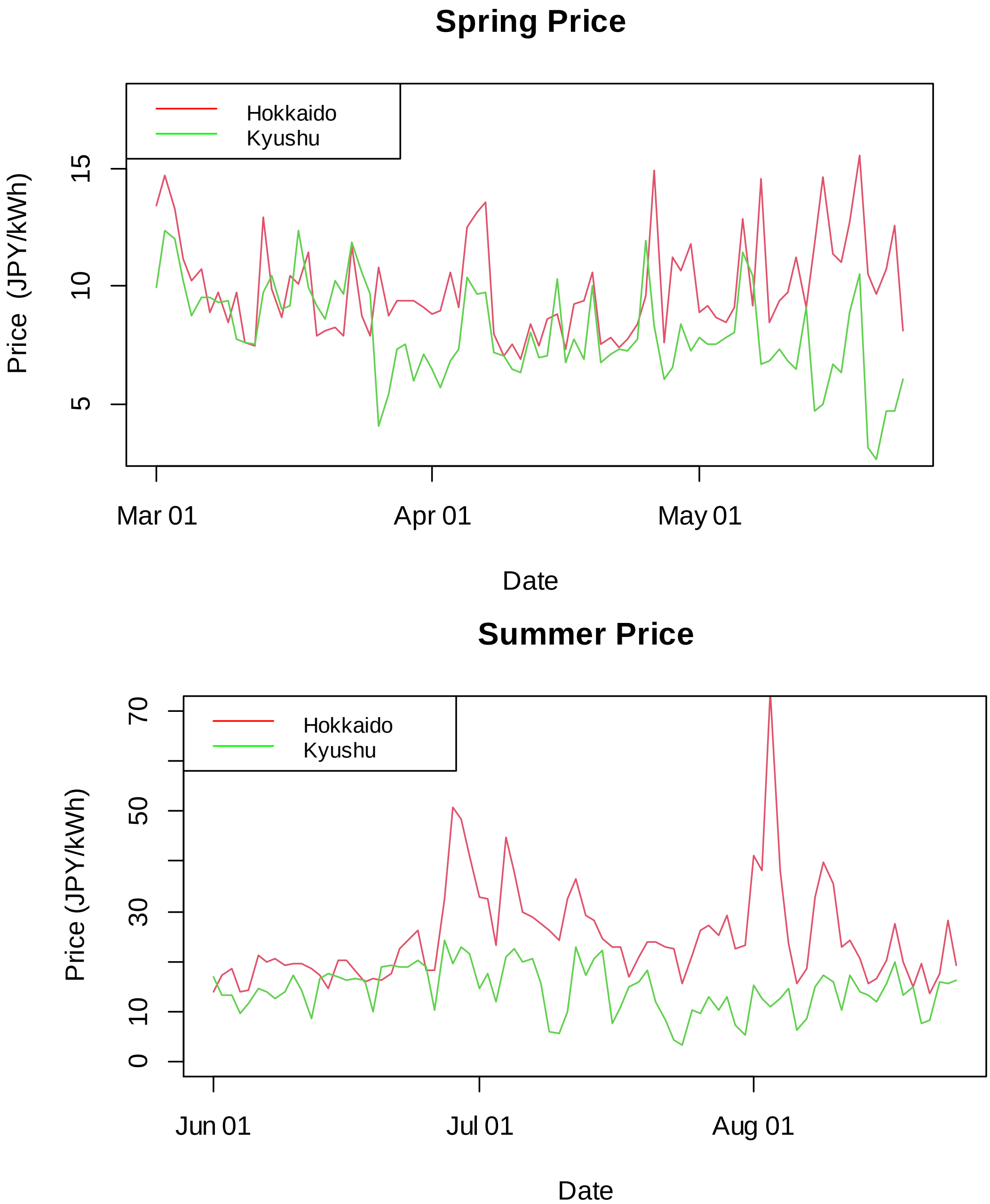


Figure 12. Time series plots of four data sets to be analysed.

For autoregressive integrated moving average (ARIMA) models, the rule of thumb is that you should have at least 50 observations (Box and Tiao, 1975). Therefore, each dataset chosen has 85 observations should be enough for short – term forecast.

4.2.2. Data analysis

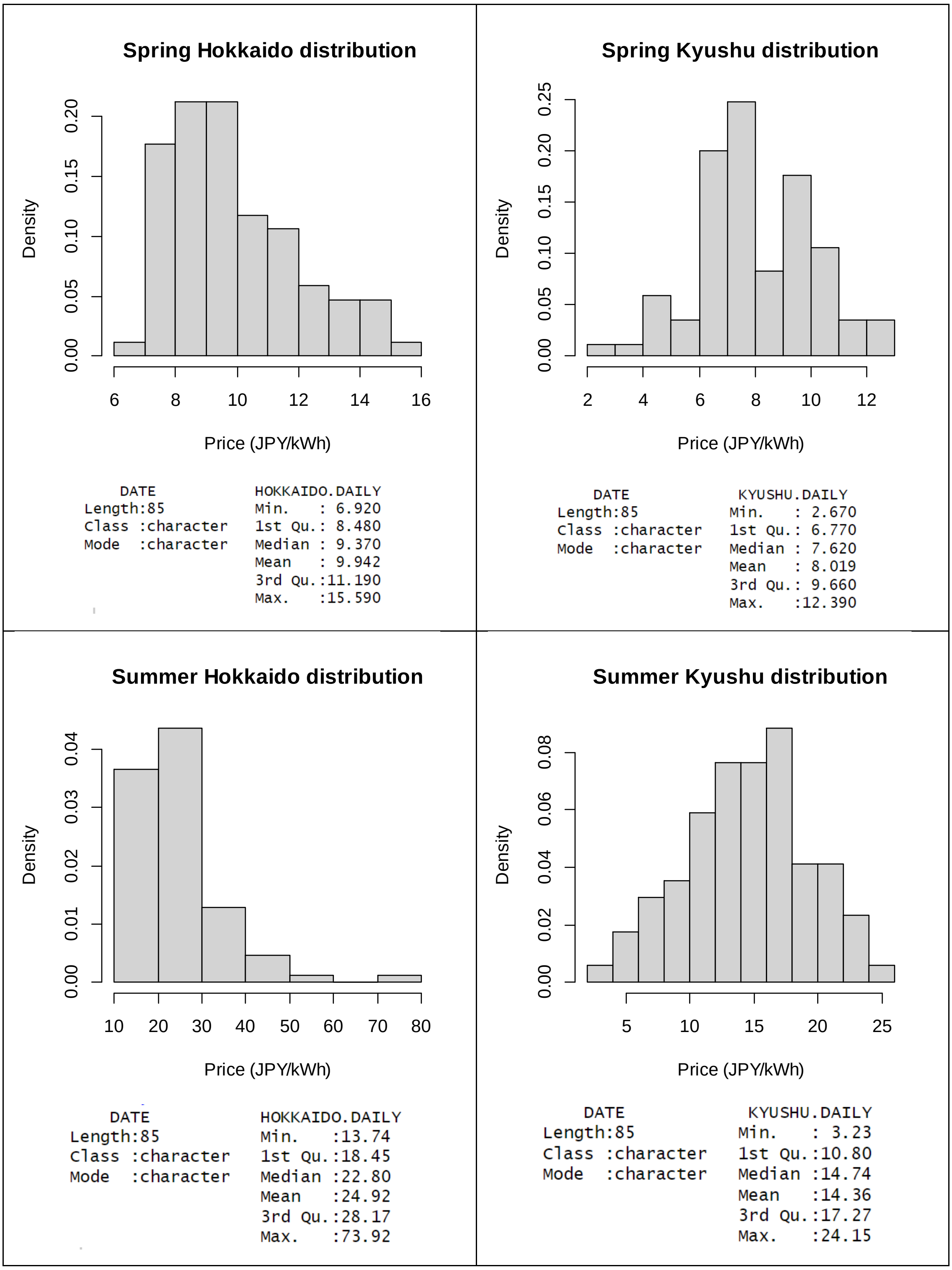


Figure 13. Histogram of four dataset being analysed

The Hokkaido (HKD) time series in the Spring dataset presents a minimum value of 6.92 JPY/kWh. The first quartile of the data is 8.48, implying that 25% of the observations fall below this value. The mean (average) of the price is 9.94. The third quartile (75th percentile) of the data is 13.61, suggesting that 75% of the observations lie below this value. Finally, the maximum value in the time series is 15.59, which is pretty much still not the highest peak that price could reach. The standard deviation (sd) is approximately 5.15, measures the average amount of dispersion or spread of the data points from the mean of 9.94 JPY/kWh, which shows that data is moderately dispersed around the mean. The variance value of 26.49 JPY/kWh suggests that there is some variability within the dataset, although it is not as high as in some other datasets.

Meanwhile, the time series Hokkaido in the dataset Summer exhibits quite a different statistic. The first quartile (25th percentile) of the data is 13.74 JPY/kWh, confirming the hypothesis that price during summer can be extremely high. The maximum value in the time series, unexpectedly, reaches 73.92 JPY/kWh, representing a huge demand from electricity usage during this time. The standard deviation of the Hokkaido time series is approximately 17.29 JPY/kWh, which technically say that the Hokkaido time series in the Summer dataset exhibits a wide range of values, with the maximum value being considerably higher than the other values. The data points are spread out around the mean value, as indicated by the relatively high variance of 298.80 JPY/kWh.

On the other hand, the sd of the Kyushu (KSH) time series in the Spring dataset is approximately 5.62 JPY/kWh, which is quite near to the value in Hokkaido during Spring. The variance of the Kyushu this time frame is estimated as 31.64 JPY/kWh. With a standard deviation of 5.62 JPY/kWh and a variance of 31.64 JPY/kWh, we can further describe that Kyushu electricity price during Spring has quite less fluctuation. The variance value confirms this, as it is relatively small compared to the range of the data. Overall, the Kyushu time series in the Spring dataset exhibits moderate variability, with most data points concentrated near the mean value.

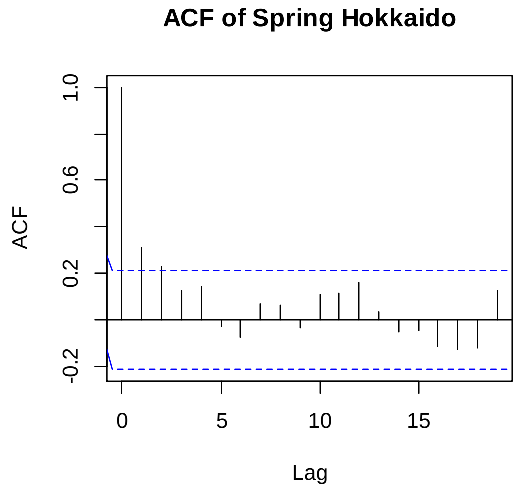
Nevertheless, Kyushu electricity price during summer does show a more volatile characteristic. The first quartile and third quartile of the data are 6.77 JPY/kWh and 9.66 JPY/kWh, respectively. The median of 14.76 JPY/kWh is quite close to the mean of 14.36 JPY/kWh, indicating the central tendency of the distribution as can be seen on the histogram. The maximum value in the time series has a jump of up to 24.15 JPY/kWh. This is a relatively high price for the region during the season, however it is still much lower than the northern Hokkaido of the same season. The sd of Kyushu time series is approximately 11.04 JPY/kWh and the variance (var) is 121.89 JPY/kWh, implying that data points are spread out around the mean value.

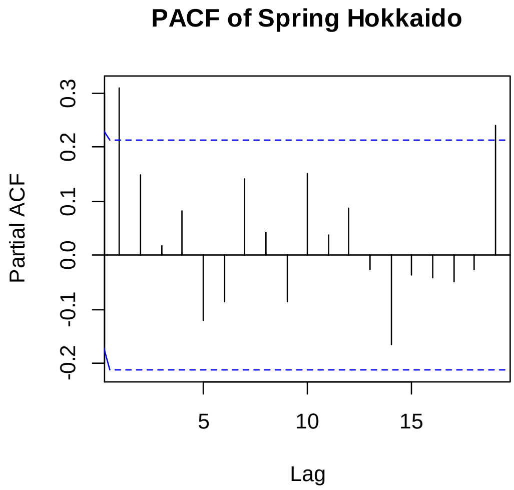
Putting into comparison, the mean price of electricity in Kyushu is considerably higher in the Summer (14.36 JPY/kWh) compared to the Spring season (8.02 JPY/kWh). A greater dispersion of data points around the mean and higher overall variability in the Summer dataset are also observed. Overall, this suggests that electricity price during the summer season in Kyushu shows more fluctuations and potential extreme values than the spring season obviously. If bringing this to compare with the Hokkaido time series in the same dataset, the Hokkaido time series in the Summer dataset exhibits a higher mean and a wider range of values. The presence of a relatively high standard deviation and variance suggests a notable degree of variability of the price for Hokkaido, that is clearly higher than Kyushu.

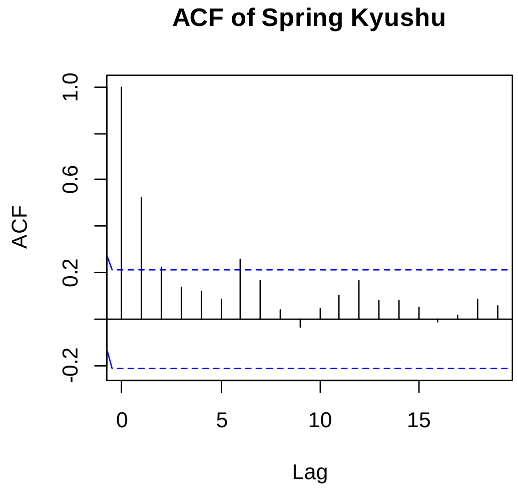
Table 1. Dickey – Fuller test and KPSS test to check stationarity

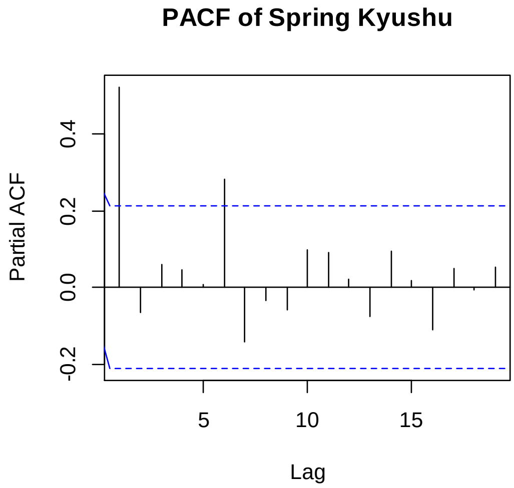
	Spring 01/03/2023 – 24/05/2023		Summer 01/06/2022 – 24/08/2022	
	Hokkaido	Kyushu	Hokkaido	Kyushu
Dickey – Fuller test	-3.996	-3.788	-17.17	-13.04
	P – value = 0.013	P – value = 0.02	P – value = 0.01	P – value = 0.01
KPSS test	0.27	0.947	1.36	0.81
	P – value = 0.01	P – value = 0.01	P – value = 0.01	P – value = 0.01
Conclusion	Non – stationary	Non – stationary	Non – stationary	Non – stationary

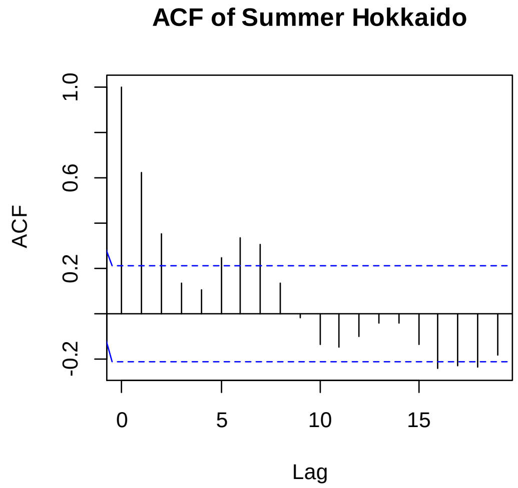
Table 1 presents the results of two test statistics, namely the ADF and KPSS tests. The results indicate that all four datasets exhibit non-stationarity. This outcome can be attributed to the inherent characteristics of the commodity under analysis, electricity. Electricity prices are known to be volatile, often influenced by trends and strong seasonality patterns. These factors contribute to the non – stationary nature of the datasets, suggesting the presence of trends and seasonality that need to be considered in further analysis and modelling.

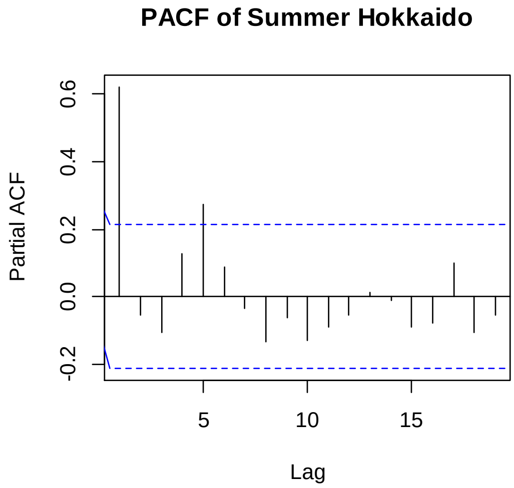












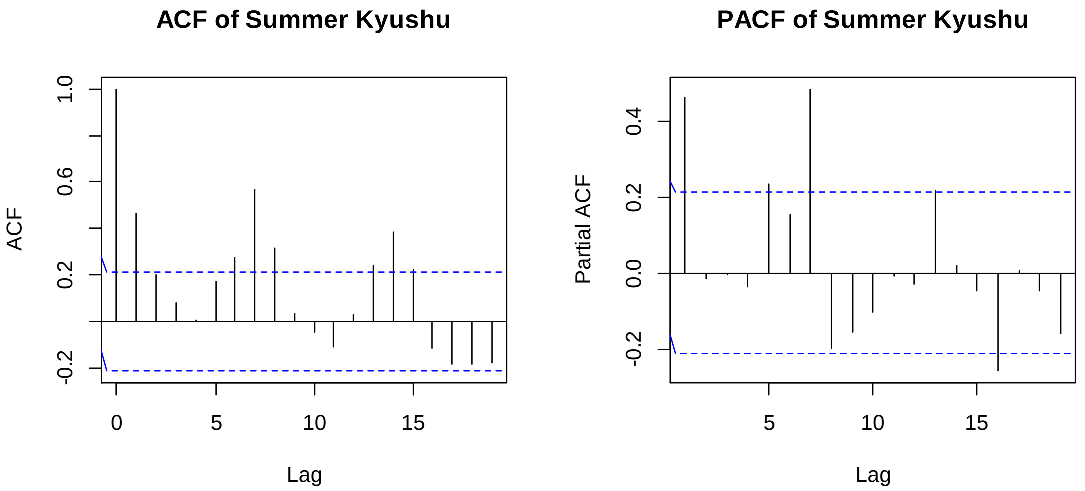
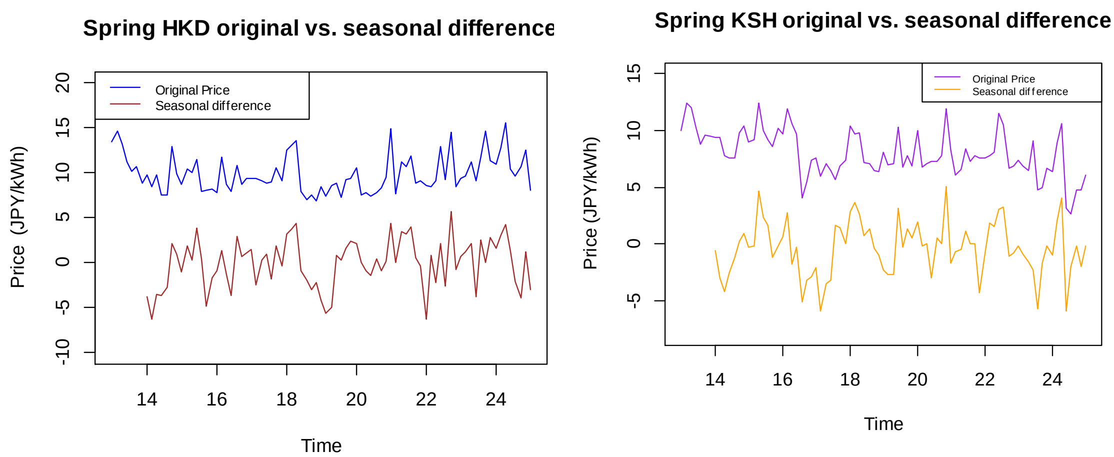


Figure 14. ACF and PACF plots of the four time series being analysed

To determine the appropriate order of an ARIMA model, ACF and PACF plots are commonly used, as shown in Figure 14. In this analysis, the four time series under consideration exhibit an exponential decay in the ACF and show some cutoffs at certain lags in the PACF. This pattern suggests the potential suitability of an AR model. Additionally, the presence of a spike at lag 1 in all models may indicate the need to include a MA component, possibly an MA (1) model.

Furthermore, considering that the daily series exhibit a weekly seasonality, it is necessary to account for this pattern. To address the seasonality, a seasonal difference at lag 7 (representing seven days in a week) is taken to account for the weekly seasonality in the data.



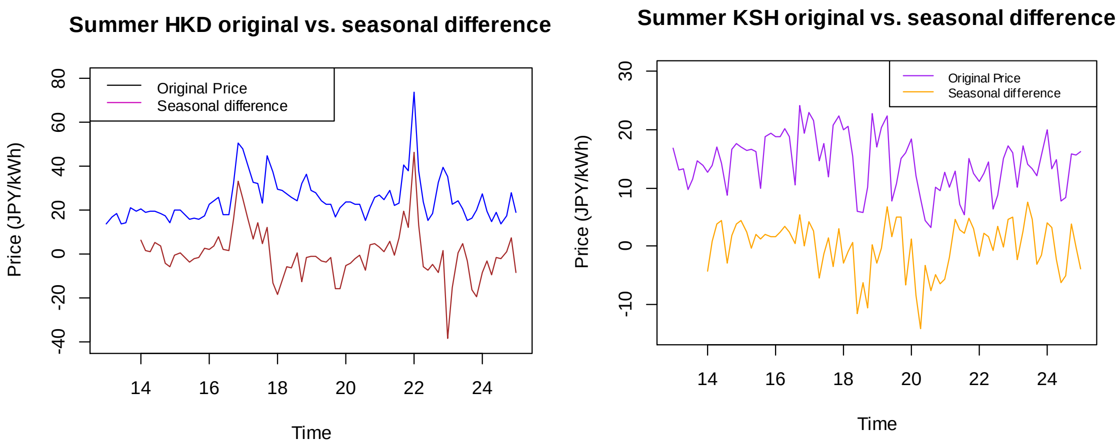
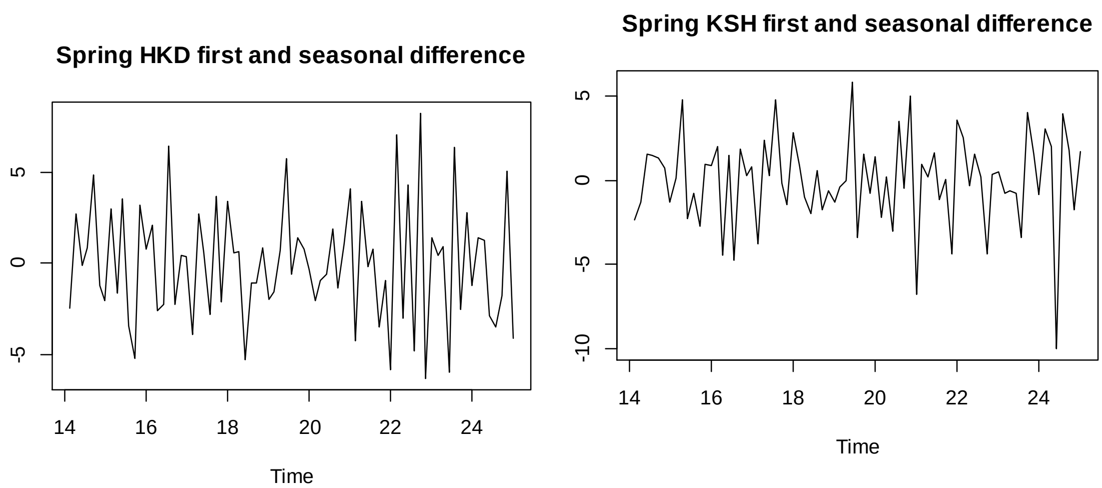


Figure 15. Plot of original time series comparing to the seasonal difference series

After taking the first seasonal difference of the data, see Figure 15, it appears that the transformed series still does not exhibit stationarity. This implies that additional transformations or adjustments may be necessary to achieve stationarity in the time series. Further transformations could include higher order differencing, logarithmic transformations, or other appropriate data transformations based on the specific characteristics and context of the data. For this dataset, based on the observed characteristics of the data and the evaluation of various other transformations, first difference is chosen for initial analysis. Figure 16 displays the plots of all four datasets, and upon visual inspection, they exhibit a satisfactory level of stationarity. The data appears to lack any noticeable trends, seasonality, or other non-stationary patterns. This positive outcome suggests that further analysis can proceed with confidence, as the data meets the key assumption of stationarity required for ARIMA modelling.



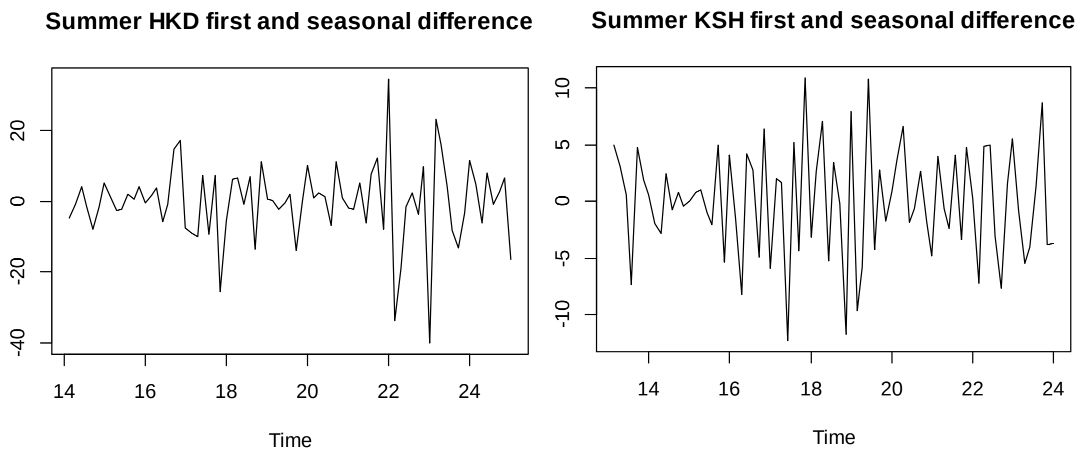
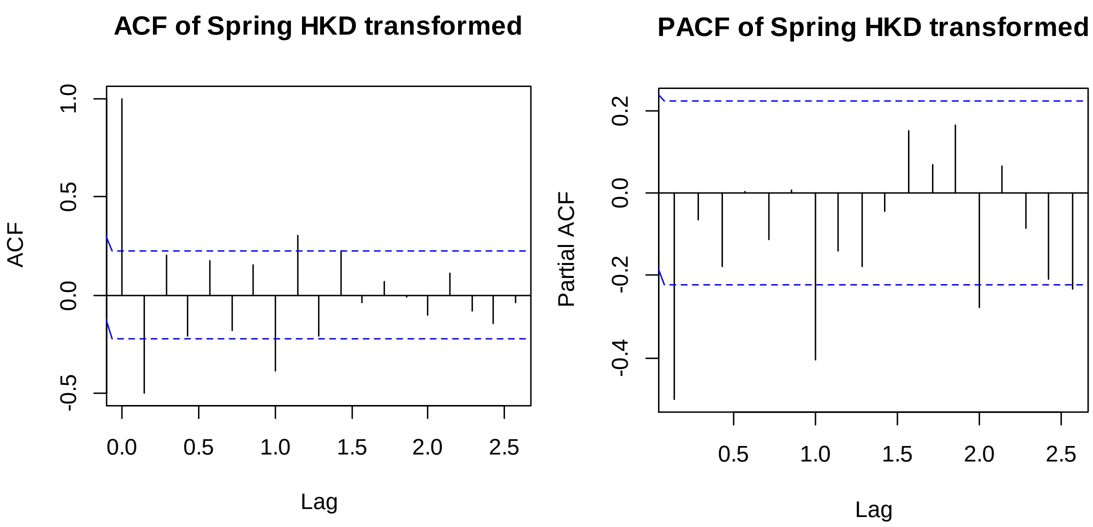


Figure 16 . Plots of the transform data

To select the appropriate order of an ARIMA model, it is common practice to examine the ACF and PACF plots. These plots provide valuable insights into the correlation structure and potential lags that should be included in the ARIMA model. Figure 17 reveals significant spikes at lags 7 and 14 in the ACF plot. These prominent spikes are indicative of seasonality patterns in the data. A spike at lag 7 suggests a weekly seasonality, while a spike at lag 14 corresponds to a bi – weekly seasonality. The presence of such spikes in the ACF plot highlights the need to consider and account for the seasonal component in the ARIMA modelling process.



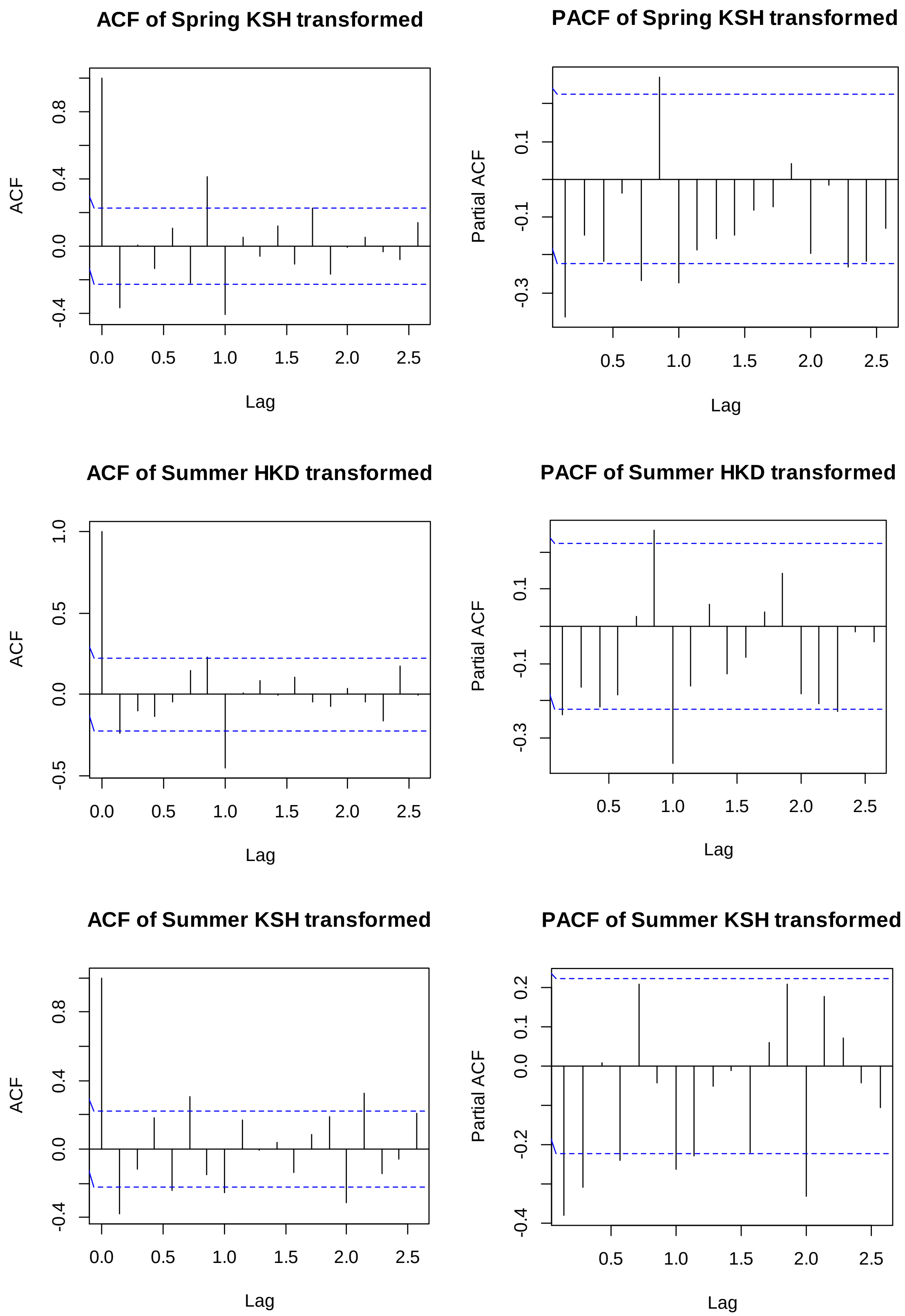


Figure 17 ACF plots of transformed data

In the case of the spring Hokkaido dataset, the presence of a spike at lag 1 in the ACF plot indicates the potential of an MA (1) model. Additionally, the spike at lag 7 in the ACF plot suggests the presence of a seasonal MA (1) component. Consequently, the initial model selection for this dataset would be ARIMA (0,1,1) (0,1,1) $_7$, which take into account also the first and seasonal difference taken previously.

Applying the same logical approach to the other three datasets, one can assess the spikes in the ACF plots and determine the appropriate model orders accordingly. To enhance the model selection process and obtain the best forecast model, the thesis also utilized the "auto.arima" function in R, which automatically identifies and estimates the optimal ARIMA model based on statistical criteria. However, upon closer examination of the model achieved by the "auto.arima" function, the results did not meet the desired level of satisfaction. This conclusion was drawn from an analysis of the data decomposition and AIC tests results, which revealed that there was still valuable information remaining in the data that could be used to improve the forecasting accuracy (see Appendix A).

After analysing the residual information from the model generated by the "auto.arima" function, further considerations were made to improve the forecast accuracy. Leveraging the insights gained from the ACF and PACF analyses of the models' residuals, the final model selections were made. These final models take into account the remaining information in the residuals and aim to capture any remaining patterns or structures that were not adequately captured by the initial model. The specific details and parameters of the chosen final models can be found in Table 2. The formular of each model for Spring Hokkaido dataset, Spring Kyushu dataset, Summer Hokkaido dataset and Summer Kyushu dataset are, consecutively,

ARIMA (1,1,18) (0,1,1) $_7$

ARIMA (1,1,10) (0,1,3) $_7$

ARIMA (1,1,8) (3,1,2) $_7$

ARIMA (2,1,8) (0,1,3) $_7$

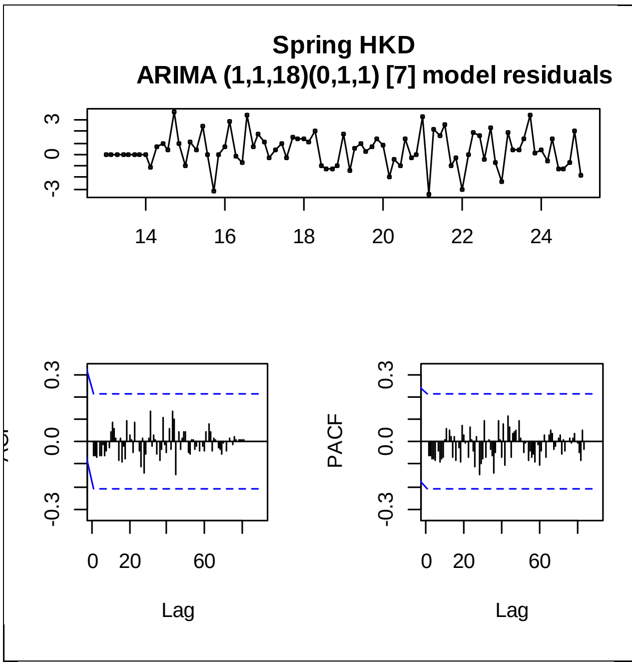
Table 2. Final decision for models fit and their parameters

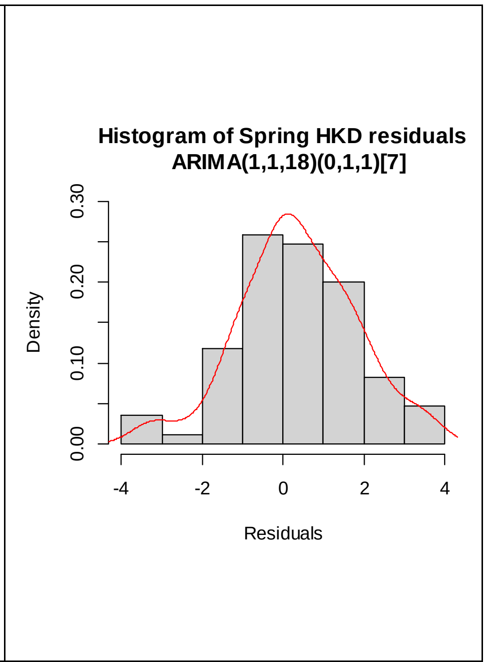
Parameters	Estimation			
	Spring 01/03/2023 – 24/05/2023		Summer 01/06/2022 – 24/08/2022	
	Hokkaido	Kyushu	Hokkaido	Kyushu
ϕ_1	0.3710	-0.4037	-0.4177	-0.9922
ϕ_2	-	-	-	-0.9958
θ_1	-1.1148	-0.0193	-0.0016	0.6469
θ_2	0.3509	-0.5264	-0.2400	0.5049
θ_3	-0.0393	-0.2695	-0.1720	-0.753
θ_4	0.0003	0.1132	-0.1294	-0.6001
θ_5	-0.2675	-0.2024	0.0092	0.1613
θ_6	0.2596	0.1315	0.1523	-0.2651
θ_7	0.0986	0.4084	0.1994	-0.0513
θ_8	-0.0515	-0.3186	-0.2382	-0.6357
θ_9	-0.4445	-0.3495	-	-
θ_{10}	0.6378	0.2976	-	-
θ_{11}	-0.4777	-	-	-
θ_{12}	0.165	-	-	-
θ_{13}	-0.2195	-	-	-
θ_{14}	0.3582	-	-	-
θ_{15}	-0.2482	-	-	-
θ_{16}	-0.1371	-	-	-
θ_{17}	0.0152	-	-	-
θ_{18}	0.3654	-	-	-
Φ_1	-0.9995	-	0.3457	-
Φ_2	-	-	-0.2102	-
Φ_3	-	-	-0.3842	-
Θ_1	-	-1.2568	-	-0.6122
Θ_2	-	0.1191	-1.6191	-0.2778
Θ_3	-	0.1378	0.9881	-0.1100

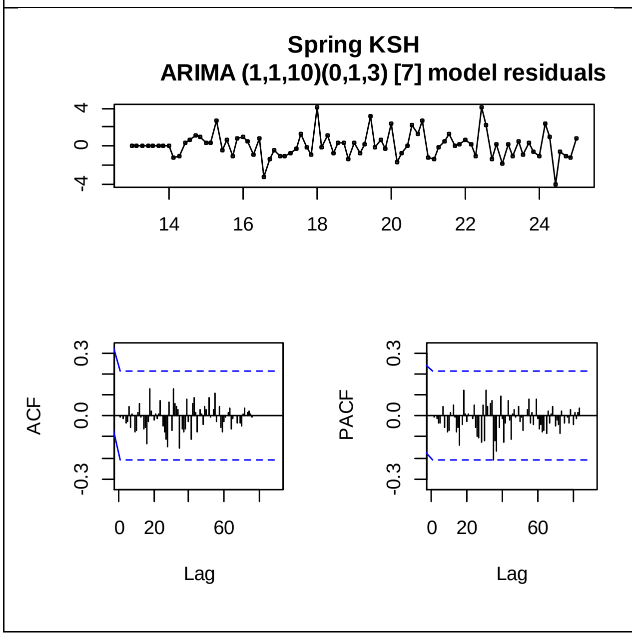
In the given results, the residuals of the models are provided for inspection, allowing for an assessment of their characteristics. Figure 18 shows that the residuals of all the models satisfy the condition of having no significant autocorrelation, which is a positive outcome. This is further supported by the Box – Ljung test results presented in Table 3, providing stronger evidence for the absence of autocorrelation in the residuals.

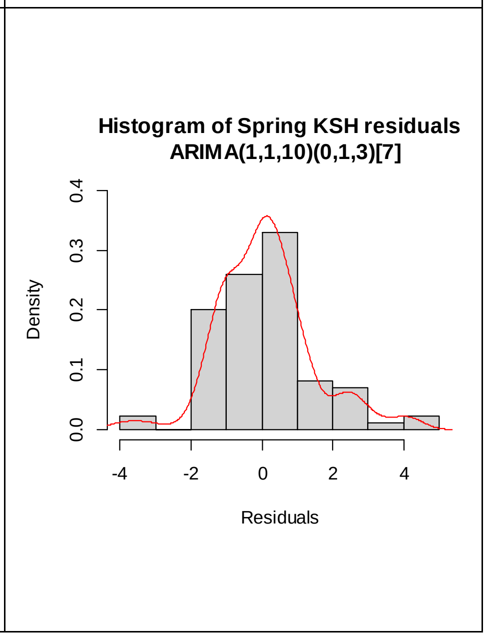
However, it is worth noting that the histogram of the residuals for all four models does not display a normal distribution, indicating potential room for improvement in the modelling process. The departure from normality suggests that the current models may not fully capture all the underlying dynamics of the data.

Nevertheless, it is important to consider the nature of the commodity being analysed, specifically the volatile nature of electricity prices. This volatility introduces challenges in accurately modelling and predicting future prices. While a model may perfectly fit the historical data, it does not guarantee superior predictive performance. Overfitting is a concern, as overly complex models may capture noise and idiosyncrasies in the data, leading to poor generalization to unseen future data. One potential solution to address the non-normality of the residuals could be incorporating an explanatory variable into the model. However, given the limited timeframe of the research, obtaining data for an explanatory variable may present significant challenges. As a result, it may not be feasible to include such a variable in the analysis at this stage. In light of these constraints, it becomes necessary to accept the current result with non – normally distributed residuals.


Spring HKD
ARIMA (1,1,18)(0,1,1) [7] model residuals
-3 0 3
14 16 18 20 22 24
0.3 0.0 -0.3
0 20 60
Lag
PACF
0.3 0.0 -0.3
0 20 60
Lag


Histogram of Spring HKD residuals
ARIMA(1,1,18)(0,1,1)[7]
Density
0.00 0.10 0.20 0.30
-4 -2 0 2 4
Residuals


Spring KSH
ARIMA (1,1,10)(0,1,3) [7] model residuals
-4 0 4
14 16 18 20 22 24
ACF
0.3 0.0 -0.3
0 20 60
Lag
PACF
0.3 0.0 -0.3
0 20 60
Lag


Histogram of Spring KSH residuals
ARIMA(1,1,10)(0,1,3)[7]
Density
0.0 0.1 0.2 0.3 0.4
-4 -2 0 2 4
Residuals

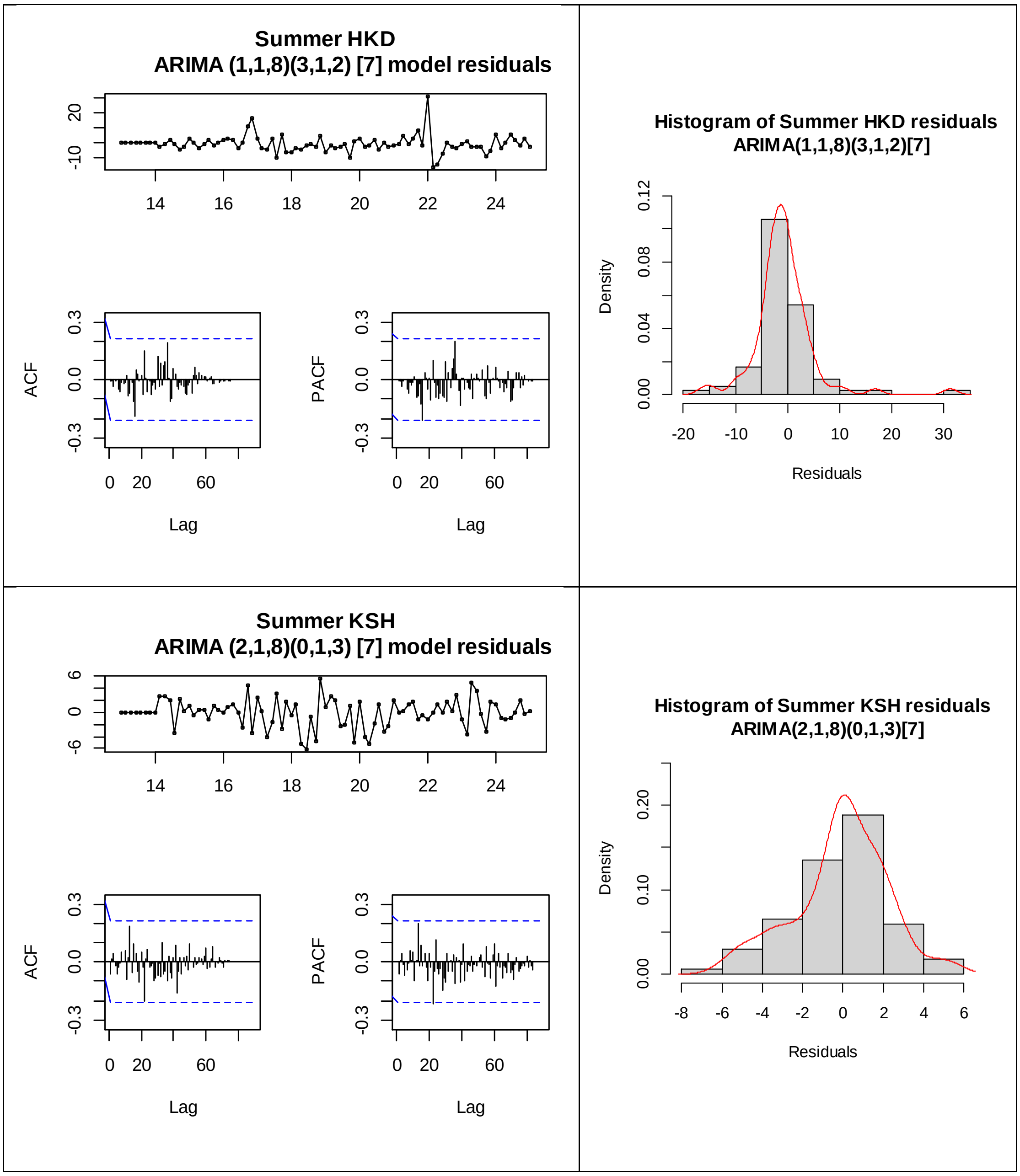


Figure 18. Models residuals analysis

Table 3. Box test for model residuals

	Spring 01/03/2023 – 24/05/2023		Summer 01/06/2022 – 24/08/2022	
	Hokkaido	Kyushu	Hokkaido	Kyushu
Test level	19.92	24.07	27.59	23.88
p – value	0.28	0.40	0.23	0.47

The number of degrees of freedom (fitdf) used are set to the total number of estimated parameters in each SARIMA model. Lags chosen based on the frequency of the seasonality, which is weekly, biweekly and triweekly (Hyndman, 2021).

Figure 19 displays the final forecasts for the four datasets. The exhibited trend and seasonality in the forecasts appear to align well with the observed data, demonstrating a satisfactory level of consistency. Additionally, Table 4 presents the forecasted values for the upcoming seven days. Upon comparing these forecasted values to the real – world data, they demonstrate a favourable level of accuracy. Importantly, all the forecasted values fall within the corresponding confidence intervals, indicating a reliable level of uncertainty estimation. These results suggest that the developed forecasting models are performing well in capturing the underlying patterns and trends in the data, enabling reliable predictions for the near future.

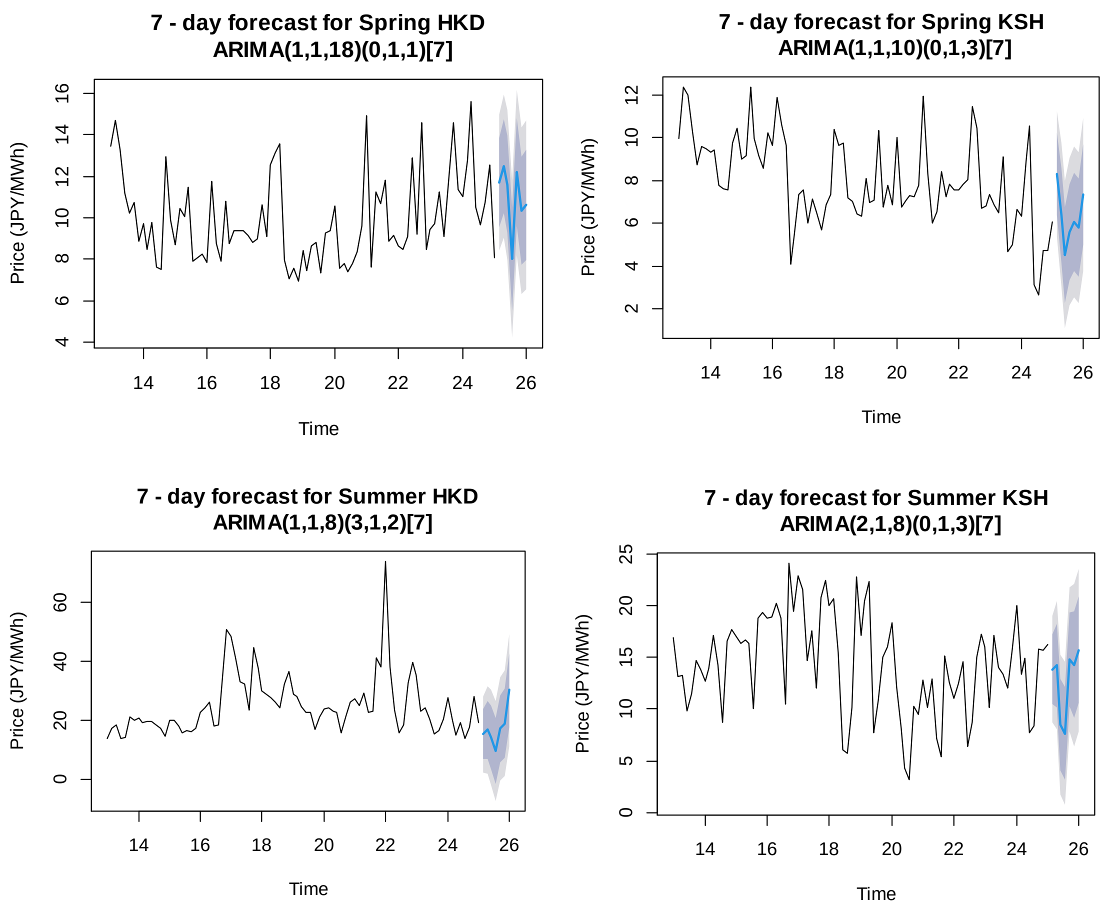


Figure 19. Forecast of the next 7 days of the time series

Table 4. Forecast results and prediction intervals

Date	7 - day forecast of Spring Hokkaido price					7 - day forecast of Spring Kyushu price				
	Point Forecast (JPY/kWh)	Low 80%	High 80%	Low 95%	High 95%	Point Forecast (JPY/kWh)	Low 80%	High 80%	Low 95%	High 95%
25/05/2022	**11.68**	9.53	13.84	8.39	14.98	**8.33**	6.40	10.25	5.39	11.27
26/05/2022	**12.50**	10.25	14.74	9.07	15.93	**6.38**	4.17	8.59	3.00	9.76
27/05/2022	**11.61**	9.24	13.98	7.99	15.23	**4.53**	2.28	6.78	1.09	7.97
28/05/2022	**8.03**	5.54	10.52	4.22	11.84	**5.60**	3.34	7.86	2.15	9.05
29/05/2022	**12.20**	9.58	14.81	8.20	16.19	**6.08**	3.77	8.38	2.55	9.60
30/05/2022	**10.33**	7.71	12.95	6.33	14.34	**5.81**	3.50	8.11	2.28	9.33
31/05/2022	**10.63**	7.97	13.30	6.56	14.71	**7.35**	5.00	9.70	3.75	10.95

Date	7 - day forecast of Summer Hokkaido price					7 - day forecast of Summer Kyushu price				
	Point Forecast (JPY/kWh)	Low 80%	High 80%	Low 95%	High 95%	Point Forecast (JPY/kWh)	Low 80%	High 80%	Low 95%	High 95%
25/08/2022	**15.35**	6.95	23.75	2.51	28.19	**13.87**	10.50	17.24	8.72	19.02
26/08/2022	**16.82**	7.11	26.53	1.97	31.67	**14.26**	10.22	18.30	8.08	20.44
27/08/2022	**14.29**	3.65	24.92	-1.98	30.55	**8.49**	4.08	12.91	1.74	15.25
28/08/2022	**9.60**	-1.48	20.68	-7.34	26.55	**7.66**	3.14	12.19	0.74	14.59
29/08/2022	**17.17**	5.81	28.53	-0.21	34.55	**14.80**	10.26	19.35	7.85	21.76
30/08/2022	**19.00**	7.28	30.72	1.08	36.92	**14.27**	9.13	19.42	6.40	22.14
31/08/2022	**30.18**	17.81	42.55	11.26	49.10	**15.73**	10.57	20.88	7.85	23.61

Models' validation

Table 5. Forecast errors

	Spring 01/03/2023 – 24/05/2023		Summer 01/06/2022 – 24/08/2022	
	Hokkaido	Kyushu	Hokkaido	Kyushu
MAPE	11.61	13.48	11.31	14.21
RMSE	1.49	1.35	1.34	1.32
MSE	0.64	0.68	0.53	0.46

The results obtained for MAPE, RMSE and MSE may not be considered as very good based on traditional benchmarks. However, it is important to note that these measures can vary depending on the specific context and industry (Wheeler, W. 2021).

In the case of the summer data, the sudden shock of price causing a dramatic increase within a short time can be regarded as outliers that may have influenced the model's performance. These outliers can introduce significant deviations from the expected patterns and contribute to higher errors in the forecasting results. Despite this, given the inherent complexity of the data and the challenges posed by such outliers, the obtained results can still be deemed acceptable within the available limitations and constraints of the research.

One potential solution to address the impact of outliers is to apply more advanced outlier detection methodologies. These techniques can help identify and flag the outliers more effectively, allowing for their consideration and potential inclusion in the modelling process. By incorporating the detected outliers into the model, it may be possible to refine the forecasts and improve the overall performance of the models.

CONCLUSION

In conclusion, the analysis of the heavily fluctuated electricity price with strong seasonality highlights the importance of identifying the seasonality frequency and considering regional variations for accurate analysis, particularly in the context of trading where price differentials between regions are significant. The application of seasonal ARIMA models with seasonal differencing helps capture and model the observed seasonality effectively.

The forecasts obtained from the models show acceptable results, but there are potential paths for improvement. The inclusion of an explainable variable, outlier adjustment techniques, or the incorporation of more advanced seasonal and trend decomposition methods could enhance forecasting accuracy. Additionally, further exploration and refinement of data modelling approaches may lead to better results.

Overall, while the forecasts provide reasonable insights, it is important to continue exploring ways to improve the models and refine the forecasting process. Incorporating additional factors, refining outlier handling techniques, and employing more sophisticated decomposition methods can contribute to more accurate predictions. These enhancements can enable better decision – making and planning, particularly in trading scenarios.

THESIS LIMITATION

Among the mass options of methodology to predict price, the thesis covers only models based on time series analysis, particularly ARIMA model tested with Ljung – Box statistics (Contreras, Espínola, et al., 2003). Certain hypotheses were made and kept throughout the forecasting process. Data's trends and cycles will not be mentioned for the purpose of simplification. The data applied is from Japan, followed by the Japanese energy market type, exchange and regulation (JEPIC, 2022).

REFERENCE

Aggarwal, Sanjeev K., Lalit M. Saini, and Ashwani Kumar. 2009. "Electricity price forecasting in deregulated markets: A review and evaluation." *International Journal of Electrical Power & Energy Systems* 31, no. 1 (Jan): 13-22. https://doi.org/10.1016/j.ijepes.2008.09.003.

Alex. 2020. "Stationarity and Non-stationary Time Series with Applications in R – Boostedml." Boostedml. https://boostedml.com/2020/05/stationarity-and-non-stationary-time-series-with-applications-in-r.html#Non-Stationarity.

A. M. Bianco, M. García Ben, V. J. Yohai, and E. J. Martínez. 2001. "Outlier Detection in Regression Models with ARIMA Errors using Robust Estimates." *Journal of Forecasting* 20 (8): 565-579. https://doi.org/10.1002/for.768.

Athanasopoulos, George, and Rob J. Hyndman. 2021. *Forecasting: Principles and Practice*. N.p.: OTexts.

Badr, EA, Saab S, and Nasr G. 2001. "Univariate modeling and forecasting of energy consumption: The case of electricity in Lebanon." Energy. https://www.researchgate.net/publication/222379773_Univariate_modeling_and_forecasting_of_energy_consumption_The_case_of_electricity_in_Lebanon.

Bell, William R., and Steven C. Hillmer. 1984. "Issues Involved with the Seasonal Adjustment of Economic Time Series." *Journal of Business & Economic Statistics* 2, no. 4 (Oct): 291-320. https://doi.org/10.2307/1391266.

Blakers, A. et al. (2023) Reaching Japan's resourceful solar and Wind Energy Future, East Asia Forum. Available at: https://www.eastasiaforum.org/2023/06/10/reaching-japans-resourceful-solar-and-wind-energy-future/ (Accessed: 07 July 2023).

Box, George E. P., Gwilym M. Jenkins, and Gregory C. Reinsel. 1994. *Time Series Analysis: Forecasting and Control*. N.p.: Prentice Hall.

Box, G. E., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70(349), 70–79. https://doi.org/10.1080/01621459.1975.10480264

Brockwell, Peter J., and Richard A. Davis. 2016. *Introduction to Time Series and Forecasting*. N.p.: Springer International Publishing.

Chinn, M., M. LeBlanc, and O. Coibion. 2001. "The predictive characteristics of energy futures: Recent evidence for crude oil, natural gas, gasoline and heating oil," presented at UCSC Working Paper #409. [Online]. http://people.ucsc.edu/~chinn/energyfutures.pdf.

Chung Chen, and Lon-Mu Liu. 1993. "Forecasting time series with outliers." *Journal of Forecasting* 12 (1): 13-35. https://doi.org/10.1002/for.3980120103.

Chung-Chen, and Lon-Mu Liu. 1993. "Joint Estimation of Model Parameters and Outlier Effects in Time Series." *Journal of the American Statistical Association* 88 (421): 284-297. http://www.jstor.org/stable/2290724.

Contreras, Javier, Francisco J. Nogales, and Antonio J. Conejo. 2003. "ARIMA Models to Predict Next-Day Electricity Prices." *IEEE TRANSACTIONS ON POWER SYSTEMS* 18, no. 3 (Aug): 1014-20. 0885-8950/02.

Deng, S.J., and S. S. Oren. 2006. "Electricity derivatives and risk management." *Energy* 31 (6): 940-953. 10.1016/j.energy.2005.02.015.

DW, Bunn, and Karakatsani N. 2003. "Forecasting electricity prices," Review Paper. London Business School. https://www.semanticscholar.org/paper/Forecasting-Electricity-Prices-Bunn-Karakatsani/7461d5a6792cc365563db917cb43be45fa6cf134.

Eydeland, Alexander, and Krzysztof Wolyniec. 2003. *Energy and Power Risk Management: New Developments in Modeling, Pricing, and Hedging*. N.p.: Wiley.

Falbo, Paolo, Danielle Felletti, and Stefani Silvana. 2014. “Electricity Futures.” In *The World Scientific Handbook of Futures Markets*, 635-60. N.p.: World Scientific. https://doi.org/https://doi.org/10.1142/9789814566926_0018.

“Fig. 3. Normal Q-Q plot of the patient set-up error for each direction. The Q-Q plot compares the observed distribution of set-up errors to a normal distribution.” 2012. Creative Commons Attribution – Non – Commercial 2.5 Generic. https://www.researchgate.net/figure/Normal-Q-Q-plot-of-the-patient-set-up-error-for-each-direction-The-Q-Q-plot-compares-the_fig3_230586019.

Fox, A. J. 1972. “Outliers in Time Series.” *Journal of the Royal Statistical Society. Series B (Methodological)* 34 (3): 350-363. https://www.jstor.org/stable/2985071.

G, Juberias, Yunta R, Gracia J, and Moreno C. 1999. “A New ARIMA Model for Hourly Load Forecasting.” *IEEE Transmission and Distribution Conference* 11-16 (April): 314-19.

G, Li, Chen-Ching Liu, Lawarree J, Gallanti M, and Venturini A. 2005. “State-of-the-art of electricity price forecasting.” In *International symposium of CIGRE/IEEE PES*, 110-9.

Gatzert, Nadine, and Thomas Kosub. 2016. “Risks and risk management of renewable energy projects: The case of onshore and offshore wind parks.” *Renewable and Sustainable Energy Reviews* 60:982-998. https://doi.org/10.1016/j.rser.2016.01.103.

Global Energy Network Institute. n.d. “Map of Japanese Electricity Grid – Japan – National Energy Grids – Library – GENI.” Global Energy Network Institute. Accessed July 7, 2023. http://www.geni.org/globalenergy/library/national_energy_grid/japan/japanesenationalelectricitygrid.shtml.

Global Sustainable Electricity Partnership. 2011. *Strengthening Public-Private Partnerships to Accelerate Global Electricity Technology Deployment*.

GME. n.d. "GME's Markets – Electricity Market – forward electricity market MTE." GME. Accessed July 7, 2023. https://www.mercatoelettrico.org/en/mercati/mercatoelettrico/MTE.aspx.

Greta M. Ljung. 1993. "On Outlier Detection in Time Series." *Journal of the Royal Statistical Society. Series B (Methodological)* 55 (2): 559-567. https://www.jstor.org/stable/2346215.

Gross, G., and F.D. Galiana. 1987. "Short-Term load forecasting." *Proc. IEEE* 75, no. 12 (Dec): 1558-73.

Hagan, M.T., and S.M. Behr. 1987. "Tc." *IEEE Trans. Power System* 2 (Aug): 785-91.

Hammer, Conny, Horst Langer, and Susanna Falsaperla. 2019. *Advantages and Pitfalls of Pattern Recognition: Selected Cases in Geophysics*. N.p.: Elsevier Science.

Hippert, H.S, C.E. Pedreira, and R.C Souza. 2001. "Neural networks for short-term load forecasting: A review and evaluation." *IEEE Trans. Power System* 16 (Feb): 44-55.

Hong, W.C, Y. Dong, C.Y. Lai, S.Y Wei, and L.Y. Chen. 2011. "SVR with Hybrid Chaotic Immune Algorithm for Seasonal Load Demand Forecasting." *Energy* 4:960-77. http://dx.doi.org/10.3390/en4060960.

Honjo K, Shiraki H, and Ashina S. 2018. "Dynamic linear modeling of monthly electricity demand in Japan: Time variation of electricity conservation effect." *PLoS ONE* 13 (4). https://doi.org/10.1371/journal.pone.0196331.

Ih Chang, Chung Chen, and George C. Tiao. 1988. "Estimation of Time Series Parameters in the Presence of Outliers." *Technometrics* 30 (2): 193-204. https://doi.org/10.2307/1270165.

Irwin, Neil. 2018. “The Most Important Least-Noticed Economic Event of the Decade (Published 2018).” The New York Times. https://www.nytimes.com/2018/09/29/upshot/mini-recession-2016-little-known-big-impact.html.

Irwin, Terry. 2018. “The Emerging Transition Design Approach.” 10.21606/dma.2017.210.

J, Contreas, and Santos JR. 2006. “Short-term demand and energy price forecasting.” In *Proceedings of IEEE MELECON*, 924-7. Benalmadena, Malaga: n.p.

JEPX. n.d. “Day Ahead Market | Trading Market Data | Trading Information.” JEPX. Accessed July 7, 2023. https://www.jepx.jp/en/electricpower/market-data/spot/.

Joskow, Paul L. 2001. “California's Electricity Crisis.” *National Bureau of Economic Research Working Paper Series*, (Aug). 10.3386/w8442.

Kaminski V (2013) Energy markets. Risk Books, London

Kern, Timo, Kristin Ganz, and Louisa Wasmeier. 2023. “European day-ahead electricity prices in 2022 - FFE Website.” FfE. https://www.ffe.de/en/publications/european-day-ahead-electricity-prices-in-2022/.

Kim, D.Y., C.J, Lee, Y.W. Jeong, J.B. Park, and J.R. Shin. 2005. “Development of System Marginal Price Forecasting Method Using SARIMA Model.” In *Proceeding of Conference KIEE*, 148-50.

Kyodo (2021) *Japan’s average gasoline price rises to highest level in seven years*, *The Japan Times*. Available at: https://www.japantimes.co.jp/news/2021/10/13/business/japan-gasoline-price/ (Accessed: 07 July 2023).

Leavy, Patricia. 2017. *Research Design: Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches*. N.p.: Guilford Publications.

Maheu, John M., and Thomas H. McCurdy. 2007. "Components of Market Risk and Return." *Journal of Financial Econometrics* 5, no. 4 (Jun): 560-90. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1408923.

Makel, M. C, and J. A. Plucker. 2014. "Facts Are More Important Than Novelty: Replication in the Education Sciences." *Educational Researcher* 43 (6): 304–316. pl.

Matteson, David S., and David Ruppert. 2015. *Statistics and Data Analysis for Financial Engineering: With R Examples*. N.p.: Springer New York.

N, Amjady, and Hemmati M. 2006. "Energy price forecasting – problems and proposals for such predictions." *IEEE Power Energy Mag,* (March-April), 20-9.

Nicolaisen, J.D., C.W. Richter Jr., and G.B. Sheblé. 2000. "Price signal analysis for competitive electric generation companies." *Proc. Conf. Elect. Utility Deregulation and Restructuring and Power Technologies* 4-7:66-71.

Nogales, F.J., J. Contreras, A.J. Conejo, and R. Espínola. 2002. "Forecasting next-day electricity prices by time series models." *IEEE Trans. Power System* 17 (May): 342-8.

Pankratz, Alan. 1993. "Detecting and Treating Outliers in Dynamic Regression Models." *Biometrika* 80, no. 4 (Dec): 847-50. https://doi.org/10.2307/2336876.

R, Weron. 2006. "Modeling and Forecasting Electricity Loads and Prices A Statistical Approach." Scientific Research Publishing. https://www.scirp.org/(S(czeh2tfqyw2orz553k1w0r45))/reference/referencespapers.aspx?referenceid=2492381.

R.E., Abdel-Aal, and Al-Garni A.Z. 1997. “Forecasting Monthly Electric Energy Consumption in Eastern Saudi Arabia Using Univariate Time-Series Analysis.” *Energy* 22 (11): 1059-69. 10.1016/S0360-5442(97)00032-7.

Rob J. Hyndman, and Anne B. Koehler. 2006. “Another look at measures of forecast accuracy.” *International Journal of Forecasting* 22 (4): 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001.

Schmittmann, J.M. (2023) ‘The financial impact of Carbon Taxation on Corporates’, Selected Issues Papers, 2023(034), p. 1. doi:10.5089/9798400244223.018.

Szkuta, R., L.A. Sanabria, and T.S. Dillon. 1999. “Electricity price short-term forecasting using artificial neural networks.” *IEEE Trans. Power System* 14 (Aug): 851-7.

T, Niimura. 2006. “Forecasting techniques for deregulated electricity market prices.” *Power Eng Soc General Meet IEEE* 18-22 (June): 1-6.

Thorbecke, Catherine. 2021. “As stocks soar to historical highs, some experts say conditions ripe for correction.” ABC News. https://abcnews.go.com/Business/stocks-soar-historical-highs-experts-conditions-ripe-correction/story?id=80303849.

Tom Huddleston Jr. 2020. “What happened in every U.S. recession since the Great Depression.” CNBC. https://www.cnbc.com/2020/04/09/what-happened-in-every-us-recession-since-the-great-depression.html.

Trevisan, Vinícius. 2022. “Comparing Robustness of MAE, MSE and RMSE | by Vinícius Trevisan.” Towards Data Science. https://towardsdatascience.com/comparing-robustness-of-mae-mse-and-rmse-6d69da870828.

Tsay, Ruey S. 1986. “Time Series Model Specification in the Presence of Outliers.” *Journal of the American Statistical Association* 81, no. 393 (Mar): 132-141. https://doi.org/10.2307/2287980.

Tsay, Ruey S. 1988. “Outliers, level shifts, and variance changes in time series.” *Journal of Forecasting* 7 (1): 1-20. https://doi.org/10.1002/for.3980070102.

United Nations. 2011. “Press Conference on UN-Energy and Global Sustainable Electricity Partnership.” United Nations. https://press.un.org/en/2011/110602_energy.doc.htm.

Varro, Laszlo, Sylvia Beyer, Peter Journeay-Kaler, and Kathleen Gaffney. 2020. “Green stimulus after the 2008 crisis – Analysis.” IEA. https://www.iea.org/articles/green-stimulus-after-the-2008-crisis.

virenrehal. 2022. “Interpreting ACF and PACF plots.” SPUR ECONOMICS. https://spureconomics.com/interpreting-acf-and-pacf-plots/.

Wang, J., and B. Ramsay. 1998. “A neural network-based estimator for electricity spot-pricing with particular reference to weekend and public holidays.” *Neurocomputing* 23:47-57.

Wei, William W. S. 2006. *Time Series Analysis: Univariate and Multivariate Methods*. N.p.: Pearson Addison Wesley.

Weiss, E. 2000. “Forecasting commodity prices using ARIMA.” *Technical Analysis ofStocks & Commodities* 18 (1): 18-19.

Weron, Rafał, and Adam Misiorek. 2005. “FORECASTING SPOT ELECTRICITY PRICES WITH TIME SERIES MODELS.” *Econometrics*, (Apr). https://EconPapers.repec.org/RePEc:wpa:wuwpem:0504001.

Wheeler, W. (2021) Evaluating linear regression models using RMSE and R2, Medium. Available at: https://medium.com/wwblog/evaluating-regression-models-using-rmse-and-r%C2%B2-42f77400efee (Accessed: 07 July 2023).

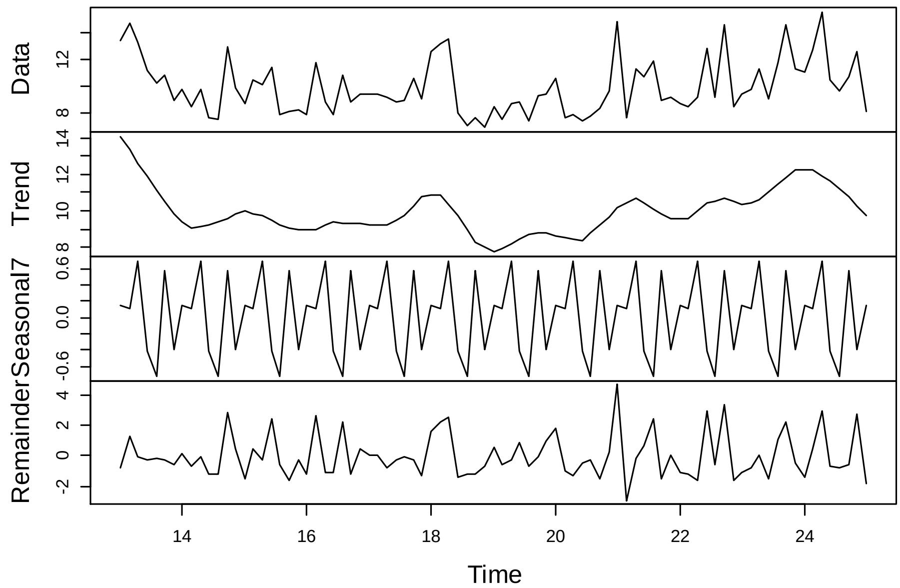
William, Christiaanse R. 1971. “Short-Term Load Forecasting Using General Exponential Smoothing.” *IEEE Transactions on Power Apparatus and Systems* 2:900-11. http://dx.doi.org/10.1109/TPAS.1971.293123.

William R. Bell, and Steven C. Hillmer. 1984. “[Issues Involved with the Seasonal Adjustment of Economic Time Series]: Reply.” *Journal of Business & Economic Statistics* 2 (4): 343-349. https://doi.org/10.2307/1391275.

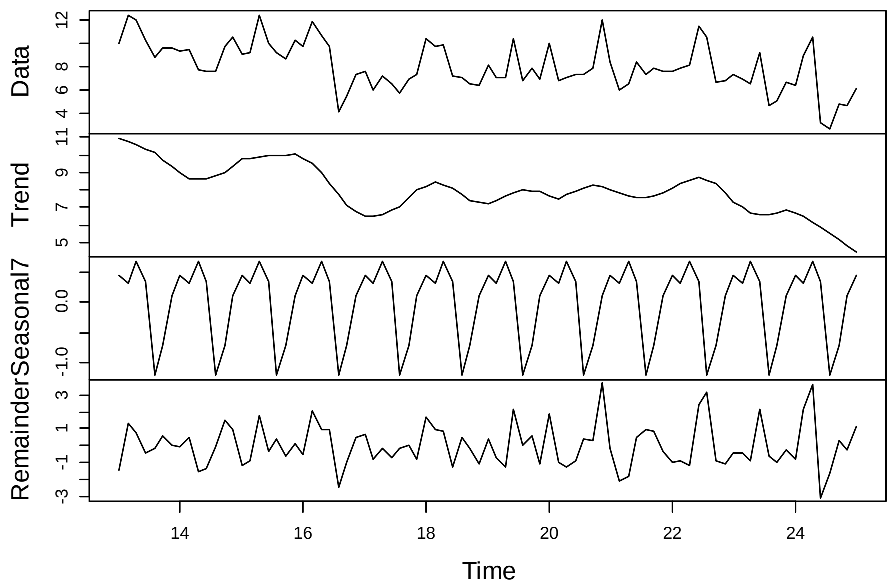
Zweifel, Peter, Georg Erdmann, and Aaron Praktiknjo. 2017. *Energy Economics: Theory and Applications*. N.p.: Springer Berlin Heidelberg.

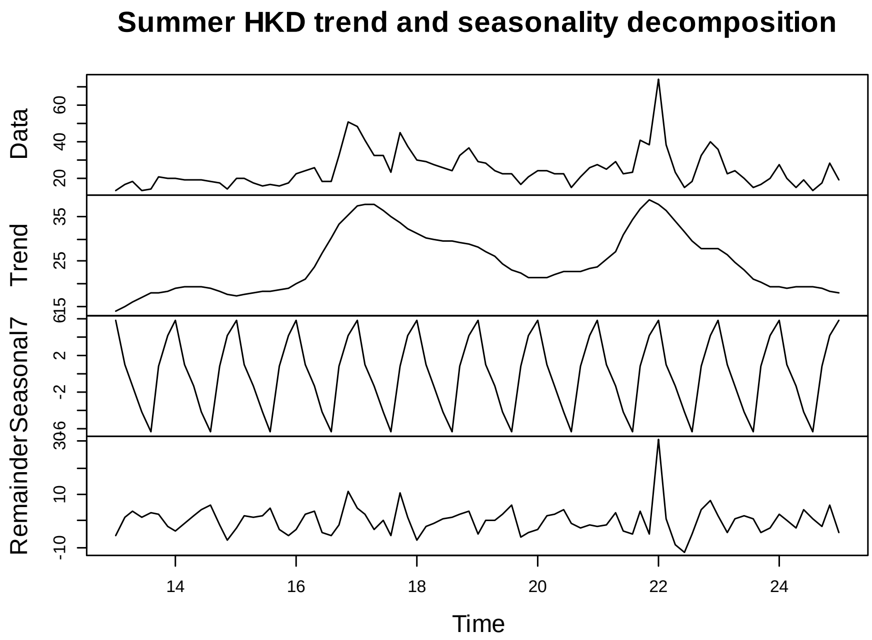
APPENDIX A

Spring HKD trend and seasonality decomposition



Spring KSH trend and seasonality decomposition





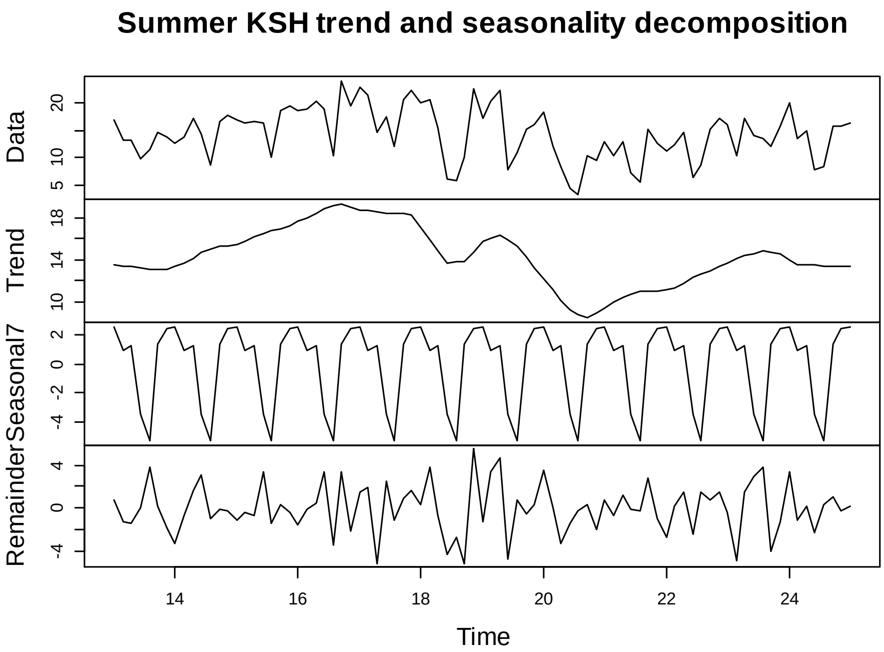


Figure 20. Trend and seasonality decomposition

Table 6. AIC value of difference models

	Spring 01/03/2023 – 24/05/2023		Summer 01/06/2022 – 24/08/2022	
	Hokkaido	Kyushu	Hokkaido	Kyushu
AIC from chosen model	358.63	330.02	564.61	416.03
AIC from (auto.arima)	361.72	335.67	589.86	425.47

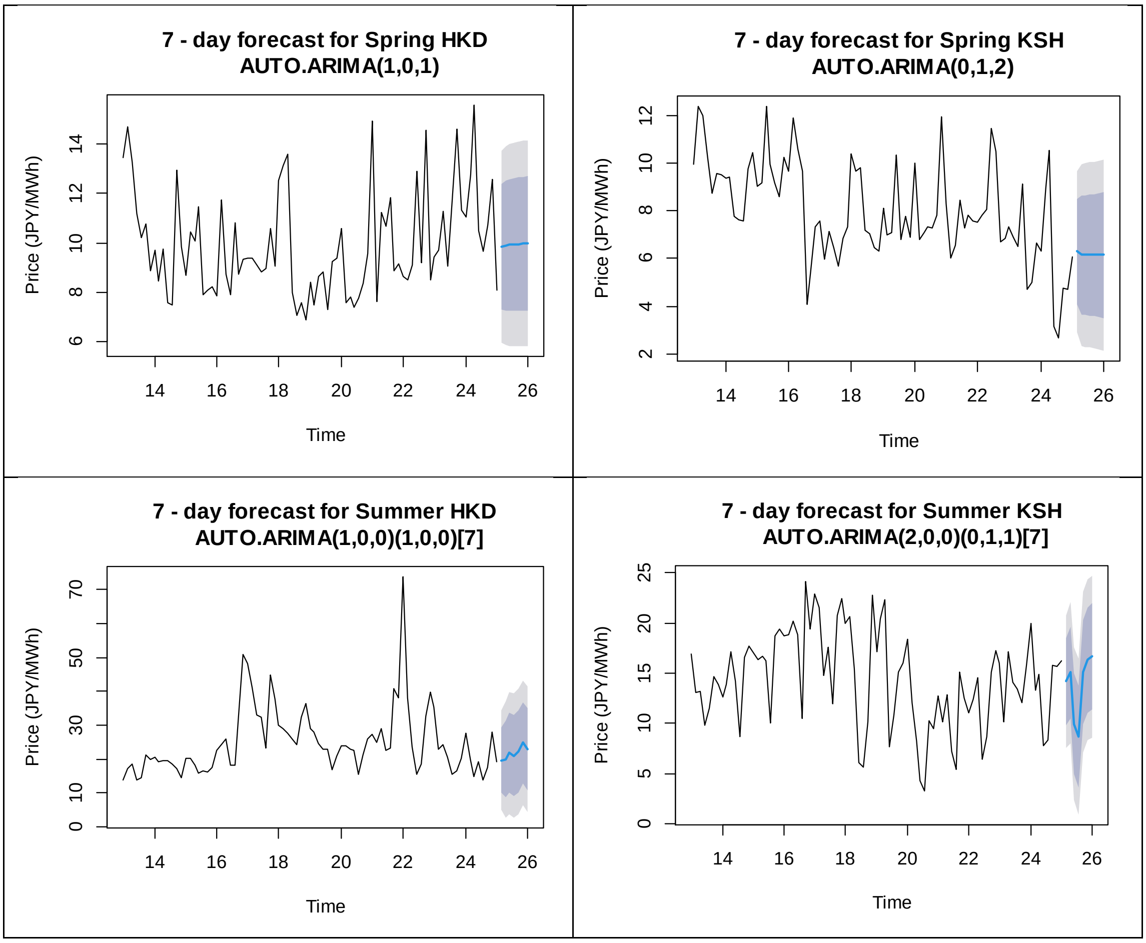


Figure 21. Forecast of (auto.arima) generated models