FORECASTING SYSTEM MARGINAL PRICE (SMP) IN THE MALAYSIAN ELECTRICITY MARKET USING TIME SERIES TECHNIQUE

SITI NOOR HIDAYAH BINTI ABDUL RAHMAN

BACHELOR OF ELECTRICAL ENGINEERING WITH HONOURS
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2024

FORECASTING SYSTEM MARGINAL PRICE(SMP) IN THE MALAYSIAN ELECTRICITY MARKET USING TIME SERIES TECHNIQUE

SITI NOOR HIDAYAH BINTI ABDUL RAHMAN

A report submitted
in partial fulfillment of the requirements for the degree of
Bachelor of Electrical Engineering with Honours

Faculty of Electrical Technology And Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2024

DECLARATION

I declare that this thesis entitled "FORECASTING SYSTEM MARGINAL PRICE (SMP) IN THE MALAYSIAN ELECTRICITY MARKET USING TIME SERIES TECHNIQUE " is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in the candidature of any other degree.

Signature : Dayah

Name : SITI NOOR HIDAYAH BINTI ABDUL RAHMAN

Date : 24 JUNE 2024

APPROVAL

I hereby declare that I have checked this report entitled " FORECASTING SYSTEM MARGINAL PRICE (SMP) IN THE MALAYSIAN ELECTRICITY MARKET USING TIME SERIES TECHNIQUE", and in my opinion, this thesis fulfills the partial requirement to be awarded the degree of Bachelor of Electrical Engineering with Honours.

Signature : ______________________

Supervisor Name : DR. INTAN AZMIRA BINTI WAN ABDUL RAZAK

Date : 24 JUNE 2024

DEDICATIONS

To my beloved mother and father who sacrificed everything to ensure that I would have the opportunity of an education, for their unending love, support and encouragement.Thank you for encouraging me to strive for the stars.

ACKNOWLEDGEMENTS

First and foremost, praises and thanks to Allah S.W.T, the Almighty, for His showers of blessings throughout my final year project for completed the research successfully.

I have been in touch with a number of individuals, educators, and practitioners in planning this project. They added to my comprehension and to my learning. In specific, I want to convey my heartfelt gratitude for support, advice and warmth to my project supervisor, Dr Intan Azmira Binti Wan Abdul Razak . I am very eternally thankful for the guidance, advice, and encouragement because without her encouragement and support this project may not have been the same as the one presented here. Next, I thank you for both my panel 1, Dr Mohd Shahril Bin Ahmad Khiar and panel 2, Dr Elia Erwani Binti Hassan for the helpful and encouraging comments during Seminar 1 and Seminar 2.

My deep gratitude always applies to all my friends and those who have on their help and support to complete this report. Their observations and suggestions are very valuable.

ABSTRACT

Forecasting SMP is critical in power systems, allowing market participants and grid operators to made more informed decision. This project is about to analyse system marginal electricity price in Malaysia data via statistical approach. This project used the real data collected from Single Buyer (SB) websites to forecast marginal price. The system marginal price was observed within fourteen months duration from 1 February 2021 until 24 April 2022. This project has three goals which is identifying key inputs that influence future SMP, developing a regression analysis and ARIMA model using Microsoft Excel and MATLAB software, and validating the results using actual data and comparison to other forecasting models. The accuracy of the model was evaluated through mean absolute percentage error (MAPE) and mean absolute error (MAE) calculation to determine the least error of the model. Through a correlation analysis in Microsoft Excel and the time series model development, the project seeks to improve forecasting accuracy beyond that of existing models, thereby providing valuable insights and enhancing comprehension of SMP forecasting. The developed model then used to forecast system marginal price in the future.

ABSTRAK

Ramalan SMP adalah penting dalam sistem kuasa, membolehkan peserta pasaran dan pengendali grid membuat keputusan yang lebih termaklum. Projek ini akan menganalisis data harga elektrik marginal sistem di Malaysia melalui pendekatan statistik. Projek ini menggunakan data sebenar yang dikumpul daripada tapak web Pembeli Tunggal (SB) untuk meramalkan harga marginal. Harga marginal sistem diperhatikan dalam tempoh empat belas bulan dari 1 Februari 2021 hingga 24 April 2022. Projek ini mempunyai tiga matlamat iaitu mengenal pasti input utama yang mempengaruhi SMP masa depan, membangunkan analisis regresi dan model ARIMA menggunakan perisian Microsoft Excel dan MATLAB, dan mengesahkan keputusan menggunakan data sebenar dan perbandingan dengan model ramalan lain. Ketepatan model dinilai melalui pengiraan purata ralat peratusan mutlak (MAPE) dan purata ralat mutlak (MAE) untuk menentukan ralat terkecil model tersebut. Melalui analisis korelasi dalam Microsoft Excel dan pembangunan model siri masa, projek ini berusaha untuk meningkatkan ketepatan ramalan melebihi model sedia ada, dengan itu memberikan cerapan berharga dan mempertingkatkan pemahaman ramalan SMP. Model yang dibangunkan kemudiannya digunakan untuk meramalkan harga marginal sistem pada masa hadapan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATION

SMP	-	System Marginal Price
TNB	-	Tenaga National Berhad
ESA	-	Electricity Supply Act
IPPs	-	Independent Power Producers
EC	-	Energy Commissions
SB	-	Single Buyer
PPAs	-	Power Purchase Agreements
NEDA	-	New Enhanced Dispatched Agreement
SLA	-	Service Level Agreements
IBR	-	Incentive- Based Regulation
SMP	-	System Marginal Price
AR	-	Auto-Regression
MA	-	Moving Average
ARMA	-	Auto-Regressive Moving Average
ARIMA	-	Auto-Regressive Integrated Moving Average
ANN	-	Artificial Neural Network
MAPE	-	Mean Absolute Percentage Error
MAE	-	Mean Absolute Error
RMS	-	Root Mean Square
LSSVM	-	Least Squares Support Vector Machines
GA	-	Genetic Algorithm
AI	-	Aritficial Intelligence
CPP	-	Capacity Purchase Price
SVR	-	Space Vector Regression
DNN	-	Deep Neural Network
CNN	-	Convolutional Neural Networks
GARCH	-	Generalized Autoregressive Conditional Heteroskedasticity
CV	-	Cross Validation
GRUs	-	Gated Recurrent Units
M-LSTM	-	Multiple Long Short-Term Memory
LMP	-	Locational Marginal Price
S-LSTM	-	Single Long Short-Term Memory
RMSPE	-	Root Mean Square Percentage Error
γ	-	Regularization Parameter
σ^2	-	Kernel Parameter
BPNN	-	Backpropagation Neural Network
IABC	-	Improve Artificial Bee Colony

CHAPTER 1

INTRODUCTION

1.1 Background

Tenaga National Berhad (TNB), the national electric utility company, was established in 1990 and went public in 1992.[1]. Based on the Electricity Supply Act (ESA) of 1990, TNB established the Single Buyer in September 2012 as a ring-fenced regulated department to handle electricity procurement services for Peninsular Malaysia and to carry out electricity planning. In a market framework known as the Single Buyer model, power is purchased from generators by a central government-backed organization, frequently through long-term contracts. On the other hand, fully liberalized electricity markets are characterized by a multitude of rival buyers and sellers.[2].

An additional unit of energy produced at any given time at a short-term cost is represented by the SMP, a fundamental indicator in the electricity markets. Market participants, grid operators, and policymakers can make well-informed choices regarding power generation, trading, and consumption by using SMP forecasting.[3]. The proper operation of modern power markets is critical. The prediction model assists energy-generating firms in maximizing generator output and the system marginal price (SMP) by incorporating production costs into the decision-making process with the goal of optimizing earnings. Concurrently, skilled users use price projections as a strategic tool to proactively manage and regulate their consumption patterns, particularly in anticipation of impending price increases.

Significant progress has been made in the development of complex SMP prediction models in recent years, utilizing advances in machine learning and optimization approaches to increase accuracy and resilience. Forecasting techniques such as time series analysis, artificial neural network (ANN), and hybrid models can be proposed. This statistical procedure forecasts future values based on previous values as well as errors.

The management of the electrical networks and the exchange of system marginal price depend on the forecast data. An accurate marginal price forecasting can be used by procedures and consumers to prepare their corresponding bidding strategies.For example, in previous studies for price forecasting average errors in the Spanish market are around 5% and around 3% in the Californian market.[4]. It proved that this technology may possesses an important role in the electricity market and will be crucial part of the electric energy portfolio.

Due to the inherent uncertainty in electricity prices, forecasting electricity prices is more complex and difficult than forecasting the electricity loads. Numerous variables contributes to the complexities of this issue in the context of SMP forecasting models.Weather and demand trends, the nonlinear and volatile nature of electricity prices, and dynamic interactions in electricity power markets are a few examples of exogenous variables.[4].

The SMP shows how much it costs a power system to produce the next unit of electricity at any given time. It is important because of its ability to provide a real-time balance between electricity supply and demand, acting as a critical signal for regulators, system operators, and market participants.Understanding and effectively implementing SMP is becoming more essential as the energy sector undergoes dynamic transformations prompted by shifting consumer behaviour, environmental concerns, and technological advancements.

The data collected are provided by the Single Buyer. The data used in this study are from 1 February 2021 until 24 April 2022. In this project we used many input for forecasting the marginal price. The best value will be compared with the Single Buyer actual marginal price.

1.2 Motivation

This research has the potential for forecasting the system marginal price that can reached the same value as actual marginal price value. It can give a big impact for electricity market participants, such as power producers, retailers, and consumers, that rely on SMP forecasts to make strategic decisions. Accurate forecasts support market participants in developing effective bidding strategies, managing risks, and maximizing economic benefits. Next, the study on the trend of system marginal price data collected will suggest the best time series analysis model that can be used to forecast system marginal price in the future. In conclusion, this study of time series technique on system marginal price can propose the best forecasting value in Malaysian Electricity Market compared to others. The target of government is to achieve 70% of renewable energy by 2050 so forecasting accurate SMP is essential for the effective integration and optimization of renewable energy sources into the power grid.

1.3 Problem Statement

In the dynamic landscape of contemporary energy markets, the determination and comprehension of System Marginal Price (SMP) emerge as critical challenges. Because SMP reflects the cost for generating the next unit of electricity in a power system at a given moment, it is crucial to market operations and economic dispatch. A significant barrier to achieving high forecasting accuracy is the influence of variables which include mix generation forecast, demand forecast, and SMP, in addition to the inherent complexity of SMP forecasting. The forecasts generated by current forecasting frameworks often do not provide the best results since these external factors are not thoroughly analysed.[5]. The impact of mix generation and demand on SMP must be understood and measured in order to increase forecasting accuracy. This necessitates a careful examination of the relationships between the variables and the accuracy of SMP forecasting.[5].

System Marginal Price (SMP) forecasting is critical to the Malaysian electricity market's ability to allocate resources optimally, inform pricing strategies, and preserve grid stability. The intricate temporal dynamics and patterns found in SMP fluctuations may be too complex for conventional forecasting techniques to fully capture. Because of this, it is essential to apply time series forecasting techniques to develop accurate models that can predict SMP trends, allowing for more informed decision-making in the dynamic energy industry. Time series analysis is a powerful tool for data analysts. Its smoothing and seasonality adjustments help to eliminate noise and outliers, which can improve the reliability and interpretability of the data.[6]. Additionally, when employing time series analysis, less data can occasionally be more effective than more without becoming entangled in overly complex models or datasets.[6].

In the Malaysian electricity market, there is no set procedure for selecting time series models that can accurately capture the temporal patterns and dynamics found in SMP data.[7]. The absence of a systematic framework for assessment makes it difficult to choose the optimal model for forecasting. To solve this issue, it is essential to assess the performance of various time series models, ranging from more sophisticated models like Long Short-Term Memory (LSTM), Articial Neural Network (ANN), and other hybrid models and prophet to more traditional methods like Auto-Regressive Integrated Moving Averaged (ARIMA), and regression.[7]. The outcome of this evaluation will help to improve forecasting accuracy and reliability in the Malaysian electricity market by identifying which models best capture the complex dynamics of SMP.

1.4 Objective

1. To analyze the effect of demand, mix generation, and SMP forecasts on the accuracy of SMP forecasting.
2. To develop SMP forecast for Malaysian electricity market using regression analysis and ARIMA model.
3. To evaluate the performance of various time series models in encapsulating the temporal patterns and dynamics of SMP using mean absolute error (MAE) and mean absolute percentage error (MAPE).

1.5 Scope

The main goal of this project is to forecast prices for Malaysia's Peninsula. Its goal is to identify the factors that will affect the System Marginal Price (SMP) in the future so that people can adapt to changes. To do this, historical data on demand, generation mix, and SMP are collected starting in 1^{st} February 2021 until 24^{th} April 2022. The training datasets is from 1^{st} February 2021 until 31^{st} October 2021 and the testing datasets is from 1^{st} Feb 2022 until 24^{th} April 2022. The data used was collected from the Single Buyer (SB) websites. Correlation analysis are used to find the right values for the variables in the method. The best correlation between variables will be selected as input and the data will be train to build a regression and ARIMA model for forecasting. Then the model will be used on the testing data to see the forecasting result. The regression analysis model is implemented using Excel while ARIMA model using MATLAB. Reducing prediction errors is the project's way of improving performance. The accuracy of the forecasting model will be assessed using the Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) metrics. The result from using regression and ARIMA will be compared. From the best forecasting result, Peninsular Malaysian users will be able to predict prices more accurately and gain a better understanding of the factors influencing SMP by examining these inputs and performance metrics.

1.6 Thesis Organization

This thesis is organized into five chapters, including the introduction. Each chapter is different from the others, and each one is discussed along with the theory needed to understand it. The research background, problem statement, objective, and scope of the research are all defined in the chapter introduction in order to better define specific aspects of active suspension system.

The background and basic literature review will be covered in Chapter 2, which will include previous study or research material on the forecasting system marginal price (SMP) and the method that has typically been used in previous research.

The methodology involved in forecasting system marginal price(SMP) using excel, which is all the identified objective function with the related formulation, will be described in detail in Chapter 3. The overall process were also discussed in this chapter.

Chapter 4 shows the simulation and the resulting case using excel. A solution is found using the process. A reliability study, comparison tables, and convergence characteristics are also studied.

Chapter 5 concludes the work performed with significant results. It also provides an insight into the future studies that can be done on this subject.

All costs related to the completed report such as printing, binding, typing, and photocopying will be fully borne by the student.

CHAPTER 2
LITERATURE REVIEW

In this chapter, the content to be discussed are the types of forecasting method, time series technique theory for system marginal price in the Malaysian Electricity Market and several studies that related to the project.

2.1 Overview of Electricity Market in Malaysia

This study focuses at the structure and organisation of the Malaysian power market, focusing special attention to the market's major players, design, and regulatory framework. The research attempts to provide a thorough comprehension of the functioning of the Malaysian power market and the various roles that different entities undertake within it by examining these aspects.

The Energy Commission of Malaysia (EC) is in charge of managing the regulatory framework that determines the structure of the Malaysian power market. The Energy Commission (EC), which was founded by the Energy Commission Act of 2001 is the government agency in responsible for observing and advancing the nation's efficient electricity supply.[8]. The compilation of laws, regulations, and guidelines that control the electricity industry's pricing, market dynamics, and operations is known as the regulatory framework.

The Malaysian power market is governed by generation companies, transmission and distribution companies, independent power producers (IPPs), and electricity retailers. Generation companies use a wide range of energy sources, including thermal, renewable, and hydro power to generate electricity. Customers are supplied with electricity by companies that manage the transmission and distribution infrastructure. Independent power producers enhance the market by privately owned power plants, thereby improving the generation capacity. Conversely, energy retailers are in charge of selling electricity to consumers directly.[9].

Governmental organisations, associations of companies, advocacy for consumers groups, regulatory agencies, and market participants are some of the stakeholders in Malaysia's electricity market. These stakeholders are working together and communicate in order to maintain regulatory compliance, promote the effective operation of the market and address the interests and concerns of consumers and industry participants[10].

This study presents a thorough overview of the dynamics and players involved in the nation's electricity sector through looking at the structure and organisation of the Malaysian power market, including its regulatory framework, market design, significant players, and stakeholders. This comprehension establishes the foundation for additional investigation and evaluation, empowering scholars, professionals in the industry, and decision-makers to make knowledgeable choices and assist in the development of a sustainable and effective power market in Malaysia.

2.2 System Marginal Price (SMP)

The System Marginal Price (SMP) or the cost of generating the last unit of energy needed to meet the system's demand at any given time, is a fundamental concept in the study of electricity markets. The figure that depicts the marginal cost of producing electricity in the current market is dynamic and constantly fluctuates. Understanding SMP is essential for regulators, system operators, and market participants as it impacts pricing, resource allocation, and market behaviour.

The term "single market price" (SMP) refers to the set price at which electricity is traded as a single commodity in a wholesale electricity market.[11]. The balance between supply and demand for electricity is established by the market-clearing price for a specific time frame, usually the day-ahead market. A number of market variables including transmission constraints, electricity demand, petroleum prices, generation capacity, and market laws interact to determine the SMP.

SMP possesses a direct impact on wholesale electricity prices, thus making it an important player in the electricity market. Traders of electricity based on the SMP include generating companies, retailers, and independent power producers. Price discovery and market clearing can be performed simply and effectively with the help of the SMP, which acts as a reference price for bilateral contracts and market transactions. [12].

SMP is a crucial tool for assisting market participants in making decisions that can be justified. Generators use SMP to optimise their production strategies, while retailers and consumers use it to predict and manage electricity costs. SMP is also monitored by regulators and system operators to ensure smooth operation of the power grid and to address any possible abnormalities in the market.

One noteworthy feature of SMP is its temporal dimension, which is dynamic and always evolving in response to the shifting dynamics of the electricity market. Daily and seasonal patterns, as well as unplanned occurrences like equipment breakdowns or extreme weather, all contribute to SMP's dynamic nature. As a result, accurately forecasting SMP presents a challenging but crucial task for market participants, requiring the application of advanced time series forecasting techniques.

In conclusion, the System Marginal Price is an essential component of electricity markets that forms market dynamics and accurately represents the cost of generating electricity. Since the energy industry is so dynamic, its ongoing volatility serves as a reminder of how intricate the electricity markets are and how advanced models as well as procedures are required to ensure equal and efficient pricing.

2.3 Types of Forecasting Method for System Marginal Price(SMP)

The system marginal price (SMP) data analysis method was covered in the topic that follows.

2.3.1 Machine learning and Artificial Intelligence

System marginal price (SMP) forecasting has benefited greatly from the development of machine learning and artificial intelligence (AI) techniques, which enable energy markets to make precise and effective forecasts. Machine learning models are trained using supervised learning techniques using historical data on market prices, generation capacity, and energy demand.[13].(SMP) in several different markets. Although these techniques are more accurate than traditional forecasting methods, they are also more expensive and require a large amount of data to train the models. Deciphering the output of AI and machine learning models can also be difficult. However, these techniques can be helpful in predicting SMP, particularly in markets where time series analysis and other conventional techniques are less reliable.

2.3.1.1 Artificial Neural Network (ANN)

The method used to developed ANN analysis is non-linear function estimation.[13]. ANN analysis works on the principle of the human brain as in the processing of information. Figure 1 shows that schematic diagram for basic structure of back-propagation NN with one input layer, one hidden layer and one output layer.[14]. An ANN analysis needs more than one parameter for forecasting accurately.

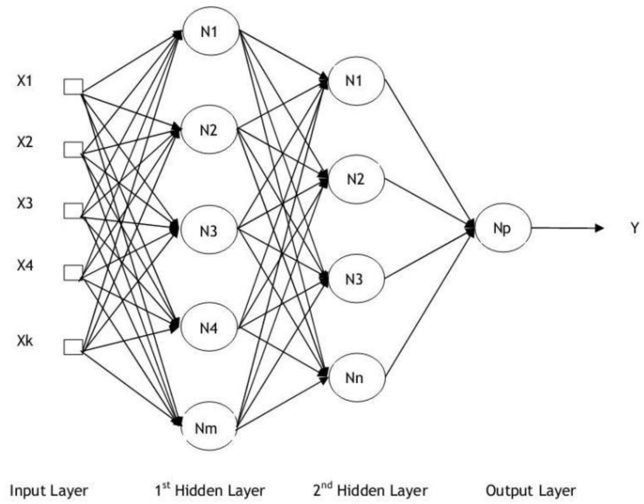

Figure 2.1:Schematic diagram of ANN prediction[15]

Other paper as stated in J. K. Lee, J. B. Park, J. R. Shin, and K. Y. Lee [14] show that ANN model electricity price forecasting proved that this method takes time to constructed the data. The accuracy for the AAN model produces by the evaluation of mean absolute error (MAPE). Besides that, the hidden layer is change one by one and MAPE were evaluated. The author concluded that, 26.052% and 17.015% of MAPE means that the model has high prediction accuracy. If the model has MAPE that is greater than 50%, model forecast was inaccurate. Lastly, the difference ofANN model, needs to be developed using electricity load and hourly cost as input parameter and the accuracy is checked by error measurement analysis.

2.3.1.2 Deep Learning

G. Yang, S. Du, Q. Duan, and J. Su, 2021 [16]. The Multi-Day Load Forecasting model presented in this paper takes into account the effects of Day-Ahead Locational Marginal Price (LMP) and Day-Ahead Cleared and is based on Multiple Long Short-Term Memory (M-LSTM). The difficulty a single deep learning network structure has maintaining the time sequence characteristics between training samples is addressed by the suggested method. The New England Electricity Market (ISO-NE) case studies show that our suggested forecasting technique is superior to S-LSTM. The model's MAPE dropped by 3.33% when compared to Single Variable LSTM (S-LSTM). The RMSE of the model is 457.22 MW, 98.09 MW less than that of the Single Variable LSTM (S-LSTM). Experiments show that the model's prediction accuracy and generalizability are superior.

2.3.2 Hybrid Model and Ensemble Forecasting Technique

In forecasting, a hybrid model is an approach that combines several forecasting techniques or models into a more thorough and efficient method. The strengths of statistical, machine learning, and conventional time series models are frequently combined in this amalgamation.[17]. Hybrid models seek to overcome the shortcomings of single models and offer more dependable predictions in a variety of scenarios by combining the different capabilities of different techniques.[18]. The idea behind hybrid modelling stems from the realization that no single forecasting technique is perfect for every circumstance. Certain models might be better at handling noise, identifying patterns, or adjusting to shifting dynamics. In order to improve overall forecasting performance, hybrid models aim to take advantage of the complementary nature of various forecasting methodologies.[17].

The study introduces a novel hybrid model that improves forecast accuracy for stock price predictions by Q. M. Ilyas, K. Iqbal, S. Ijaz, A. Mehmood, and S. Bhatia.[18]. The proposed method consists of three main components: a machine learning prediction, new features, and a noise-filtering technique. To smooth out the historical stock price data and remove the cyclical portion of the time series, the authors employed a fully modified Hodrick-Prescott filter. A number of novel approaches to stock price prediction are proposed, including the return of the firm, return open price, return close price, change in return open price, change in return close price, and number per total.They look into both traditional and deep learning approaches to using machine learning for prediction. Conventional machine learning employs random forests, auto-regressive integrated moving averages, and support vector regression. Deep learning techniques make use of gated recurrent units and long short-term memory experimented with these machine learning techniques in various ways. The best model they had produced predictions that were accurate 70.88 % of the time, with an error rate of 0.1 and a root-mean-square error of 0.04.

2.3.2.1 CV- LSSVM

The literature review in [19], emphasizes how crucial it is to choose the right hyper-parameter values for the LSSVM (Least Squares Support Vector Machine) algorithm in order to prevent problems with over-fitting and under-fitting. The inherent randomness in hyper-parameter selection makes it a potentially inefficient and unreliable process. Two popular methods for fine-tuning LSSVM hyper-parameters are theoretical methods and the Cross Validation (CV) process in grid search. For short-term meteorological time series data, a study comparing CV-LSSVM with an Artificial Neural Network (ANN) based prediction technique discovered that CV-LSSVM produced lower error rates based on metrics like RMSE, R2, MBE, MAPE, and KS test. Conversely, CV can take a lot of time and produce substandard error rates. Using γ (regularization parameter), σ2 (kernel parameter), MAPE and RMSPE (Root Mean Square Percentage Error) percentages, and other comparisons, IABC-LSSVM, ABC-LSSVM, and BPNN (Back propagation Neural Network) performed better than the other models. The lowest MAPE was 5.6199%, and the RMSPE was 0.0704%.

The literature as a whole emphasises the need for optimisation techniques like CV and the challenges associated with selecting LSSVM hyperparameters.It also highlights the shortcomings and inadequacies of manual selection in addition to the potential unsatisfactory results of CV. The comparison between IABC-LSSVM, ABC-LSSVM, and BPNN shows that IABC-LSSVM performs better in terms of forecasting accuracy, as indicated by lower percentages of MAPE and RMSPE.

2.3.2.2 LSSVM- GA

The benefits of Genetic Algorithms (GA) and Least Squares Support Vector Machines (LSSVM) are combined in a hybrid machine learning model known as Least Squares Support Vector Machines with Genetic Algorithms (LSSVMGA). One supervised learning algorithm that can be used for tasks involving regression and classification is LSSVM. The best solutions to optimisation problems can be found using the GA meta-heuristic algorithm.[3].

Typically, input variables for LSSVM-GA include the historical values of the system marginal price (SMP) as well as other variables that could affect the SMP, such as the supply and demand for electricity. During the training phase of LSSVM-GA, the model must be fitted to historical datasets of SMP values. The validation process involves putting the model to the test on several datasets of SMP values that weren't used for training.

It has been shown that LSSVM-GA outperforms neural networks and time series analysis in terms of SMP forecasting accuracy. A study [3] used daily and weekly forecasts to track LSSVM-GA's performance. The suggested daily forecast model outperformed the SB daily forecast model by 3.54%. Similarly, the suggested weekly forecast model performed 1.19% better than the Single Buyer (SB) forecast. In another study[20], It was demonstrated that LSSVM-GA is superior to neural networks and time series analysis in predicting the SMP in the Spanish market.

2.3.3 Time Series Technique

Another technique for forecasting system marginal price is time series analysis as in [21]– [23].Time series analysis is the process of analyzing and forecasting time series data using statistical and modelling approach. The past values of the data collected, determines the dependent variable for the time series. The historical values used to predict the future values after the model was established.

Auto-regression moving average (ARMA) and Auto-regressive integrated moving average (ARIMA) are time series model that are used to analyse stationary and non-stationary electricity prices data. This model used non-seasonal auto-regression and were applied only to short term data for testing the model [23]. A study in [24] conclude that, time series is the best method for forecasting and become the most common model analysis for modelling and forecasting data. Other than that, ARMA and ARIMA model approached for forecasting electricity prices have proven that this model more effective than other linear and non-linear technique .[22].According to the previous case studies in [22], [24], the most suitable model for forecast the electricity prices is by using time series analysis.Generally speaking, a time series is an accumulation of observations over a long period of time. Generally speaking, observations can be made at set times, randomly sampling during the interval, or for the duration of the interval. Different approaches to data analysis are needed for different types of time sampling. Figure 2 shows the sample for independent uncorrelated variables.

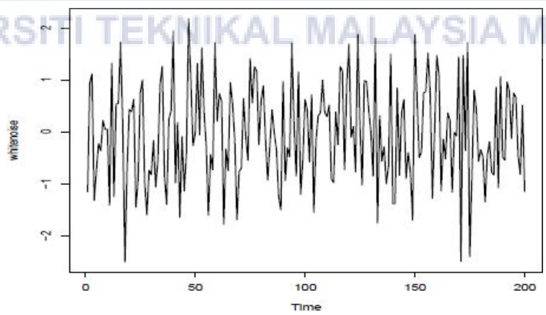

Figure 2.2 : Plot of independent uncorrelated random variables [25]

The observation that can be made is there are no any clear pattern in the data. The best prediction for the following observation is zero, which appears to be the mean. The dependence in the observations that allows for better forecasting of future observations distinguishes time series from classical statistics [25].

2.3.3.1 Box Jenkins Model

The Box-Jenkins method is a technique for time series forecasting. It was developed by George Box and Gwilym Jenkins in the 1970s and is based on the idea of identifying, estimating, and checking a model that describes the underlying structure of a time series [25]. Estimating the parameters for this model analysis can be extremely difficult. Like with other time-series regression models, programmable software is usually used to achieve the best results. Furthermore, the Box-Jenkins Model works well for forecasts that are 18 months or shorter in duration. In many cases, the software will be programmed to automatically use the best fitting forecasting methodology based on the time series data to be forecasted. Due to the stability and minimal volatility, the Box-Jenkins idea is the most used.

The Box-Jenkins Model forecasts data using three principles that are auto-regression, differencing, and moving average. These three principles are known as p, d, and q, respectively. The auto-regression (p) process tests the data for its level of stationarity. If the data used is stationary, it can simplify the forecasting process. If the data being used is non-stationary, it will need to be difference (d).[25]. The data is also tested for its moving average fit which is done in the moving average of the analysis process. In general, preliminary study of the data prepared for forecasting by established the p, d, and q parameters, which are then used to create a forecasting model.

2.3.3.2 ARMA Model

A forecasting model known as an Auto-regressive Moving Average (ARMA) applies the moving average (MA) and auto-regression (AR) analysis techniques to time-series data [25]. The least squares approach can be used to estimate the model parameters if the input and output sequences are measurable.

The model parameters may suffice since this estimate is linear. The output is the only sequence in the model that can be obtained from many spectral estimates. At this stage, it is difficult to get an accurate estimate of the parameters of the ARMA model because the estimates of the parameters are non-linear. Thus, the impact of previous lags along with the residuals is considered for forecasting the future values of the time series [25].

This case studied in [26], performed in Colombian electricity market for the duration monthly frequency from 01/2001 up to 05/2016 historical electricity prices. The author used Cavaliere and Giorgiev testing procedure to conclude that all variables are stationary with level shifts.

The method included the ARMA model equation, stationarity checking, selection of parameters and prediction of the model. Flowchart in Figure 2.6. shown that the flow of the researched paper development throughout analysis.The general equation for ARMA model were define as is (2.1).

$$ARMA\ (p,q) = \emptyset_p(B)y_t = \theta_q(B)z_t + \sum_{i=1}^{k} \beta_i(B)x_{it} + \varepsilon_t \qquad (2.1)$$

The assumption on the level of current observations not only affected by current noise error term, but also by the previous, q noise error known as moving average model of order q. The model is combination between auto-regressive (AR) model of q and moving average (MA) model.[26]. Thus, the observation can be made by the output variable depending linearly on the current and previous values.

Finding from the study stated that all estimated models have 80% and 95% bootstrap confidence-bands. The accuracy is measured by using the error prediction which are the RMSE 0.595 and MAE 0.482. As a conclusion, this paper summarized a simple and briefly step for fitting an ARMA model to historical electricity prices. [26].

2.3.3.3 ARIMA Model

A statistical analysis model called Auto-regressive Integrated Moving Average (ARIMA) makes use of time series data to forecast future trends or to help comprehend the data set. An ARIMA model is a form of regression analysis that evaluates the strength of one dependent variable on other variables that are changing. Every component in ARIMA serves as a parameter and is referred to by a common nomenclature. In ARIMA, every element serves as a parameter with a standard notation. In order to identify the type of ARIMA model being used, a standard notation for ARIMA models would be ARIMA with p, d, and q, where integer values are used in place of the parameters. [25].

To make the data in the ARIMA model stationary, there must be differences. A model that demonstrates stationarity is one that demonstrates the data's consistency over time. Since the majority of energy market data exhibits trends, the goal of differencing is to eliminate any seasonal patterns or trends.[25]Seasonality, or the regular and predictable patterns that appear in the data and repeat over time, may have an adverse effect on the regression model. Many of the calculations made during the process cannot be completed very efficiently if a trend develops and stationarity is not obvious.

This study used ARIMA model to forecast electricity day-ahead prices. The originally hourly day-ahead prices 24 hours data for German electricity market is used in this researched.[27]. The data collected starting from 2000-2011 has been used. Trading interval data is used to create a time series of daily arithmetic means, producing 3836 observations for this reference market.

The ARIMA model indicated that the data must be in stationary. The author mentioned, non-stationary time series needed transformation operations like natural log, by taking a difference, or by taking residuals from a regression for forming the time series stationary. The equation for ARIMA model shown in (2.6). The model depends on the values of autoregressive (p), moving average, (q) and model differential, (d). [27]. The general ARIMA method is formulated as following:

$$Y = \emptyset(B)p_t = \emptyset(B)\varepsilon_t \quad (2.2)$$

Where Y in dependent variable ,p_t is the price at time t, $\emptyset(B)$ and $\emptyset(B)\varepsilon$ are functions of the backshift operator$B: B_p^1 = p_{t-1}$ and ε_t is the error term. The order of auto-regressive lags term is p, differencing and moving average lag terms are represented by d and q respectively. While conditional lease squares and other methods can be used to estimate model parameters, maximum likelihood is the method used by SPSS.[27]. This paper showed ARIMA statistical modelling to forecast the electricity prices. The analysis is to establish the model's parameters and verified using a variety of criteria.

The result of the analysis that performed in this paper indicates that after forecasting the values, the accuracy of the model tested using software. The best fitted ARIMA model is (3,0,3) (1,1,1). It takes three days to forecast the price for the following day. The seasonal component is understood and appropriately accounted for. The model's mean percentage error (MAPE) is 3.55%. The periodicity of the workdays and weekends has been identified. 33.1% is the maximum absolute percentage error (MaxAPE). [27].

2.3.3.4 Regression Analysis Model

A statistical method used in forecasting to model the relationship between a dependent variable and one or more independent variables is regression analysis. In the context of forecasting, regression analysis is used to estimate or forecast future values of the dependent variable based on known values of the independent variables.[28]. In order to develop a predictive model, it is essential to assume that the variables have a linear relationship.

Regression collinearity, according to the study [29], is the existence of a strong correlation between two or more independent variables in a multiple regression model. On the other hand, the relationship between the independent variables is linear. This may make analysing the model and performing regression analysis difficult.Collinearity can be separated into two categories which is perfect multicollinearity and high multicollinearity. When all of the independent variables in a regression model can be perfectly predicted from one another, this can be known as perfect multicollinearity. This occurs often when the independent variables have a mathematical relationship, that include when one variable is a constant multiple of another. High multicollinearity is more prevalent when there is an important but imperfect linear relationship between independent variables. Even though the variables are correlated in this case, the coefficients can still be estimated.

This case studied in [30], The Maluku-Papua system is being used as a test system for demand forecasting until 2050. Electricity demand forecasts serve as the basis for the electrical system planning process and are also used to create financial projections, distribution plans, personnel plans, and other related plans. Regression analysis is a popular tool for estimating energy consumption. An example of a basic linear regression model is shown in (2.3)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon \tag{2.3}$$

Where Y is the dependent variable , X_1 until X_p is independent variables , β_0 is constant (intercept), β_1 until β_p is regression coefficients and ε is error. The dependent variable, such as energy sales or electrical energy consumption, is the one that needs to be searched. Conversely, the factors influencing the dependent variable are known as the independent variables. Among them are the GDP, population, the cost of electricity, and other variables.

In this study, a demand forecast model was created to estimate Maluku-Papua's electricity consumption. While the Maluku-Papua system is the main topic of this paper, other cases can also use the developed method. Between 2020 and 2050, Maluku-Papua's electrical energy consumption is projected to increase by 5.7% in the BaU scenario and 6.7% in the High scenario. In the BaU scenario, peak load rose by 5.6%, while in the High scenario, it increased by 6.5%. Because there are larger customer potentials and a higher GRDP growth assumption than in the BaU scenario, the High scenario grows at a faster rate.

Case studied in [31] used Multiple linear regression. In this study in order to lower the mean absolute percentage error, many predictors are examined. The dates from September 2018 to September 2019 from Turkey's day-ahead electricity market are included in the training data. Mean Absolute Percentage Error (MAPE) utilized to measure error rates for evaluating price estimation , Root Mean Square Error (RMSE) used error rates for evaluating price estimation and Mean Signed Deviation (MSD) used in this paper.The result of the analysis that performed in this paper has been proved that historical electricity which are previous one-day prices, previous one-week prices, and previous moving average one-day prices allow the forecasting model's error rate to decrease because on September 24, 2019, the model hits its lowest MAPE of 1.97%.

2.4 Characteristic That Contribute The Time Series Analysis

It is difficult to quantify system marginal price since it depends on so many different characteristic. Time series forecasting can make it possible in short term prediction such as a data pre-processing. By this statement, there are characteristics that contribute to more robust and reliable forecasts of the system marginal price which depend on suitability of the data collected.

In [23], the study investigated that the multivariate forecast model uses a combination of different parameters. The results were compared with previous uni-variable method and provided that multi-variable approach performance was the best than uni-variable. The effect of combining the variable can produce effective parameter and even greater increase in accuracy for predicting future system marginal price.[23]. The outcomes are a useful prerequisite for applying ensemble multivariate models to practical time series forecasting problems.

In order to forecast the values of the time series, it is intended to apply an automatic selection of models and their parameters in subsequent work. It is practically important to increase the accuracy of electricity price forecasting in the "day-ahead market" since doing so will increase economic efficiency.

For the case study in [32], the researcher also examined the relationships between electricity demand and wind power production. The result concluded that a strong correlation occurred between the parameter used. Beyond electricity demand and wind power, other exogenous variables, such as fuel prices, economic indicators, or policy changes, can impact SMP. Including relevant external characteristic in the time series analysis contributes to a more comprehensive forecasting model.

As a result, by carefully considering these characteristics, time series analysis can provide valuable insights into the temporal patterns of SMP, enabling more accurate and reliable forecasting. The interdisciplinary nature of this analysis, incorporating insights from economics, meteorology, and data science, enhances its effectiveness in capturing the complexities of electricity markets.The suitability of the method and the associated forecast must be well proposed. This is because the results and accuracy of each type of analysis can be varies depend on the parameters chosen and method used.

2.5 Summary

This chapter provides a literature review on electricity price forecasting methodologies. To assess the current state of research in this field, four articles are summarized and compared. J.B. Marin (2018) proposed that estimated models have 80% and 95% bootstrap confidence-bands. The accuracy is measured by using the error prediction which are the RMSE 0.595 and MAE 0.482. Tina Jakasa (2015) showed ARIMA statistical modelling to forecast the electricity prices. The result of the analysis that performed in this paper indicates that after forecasting the values, the accuracy of the model tested using software. The best fitted ARIMA model is (3,0,3) (1,1,1). The model's mean percentage error (MAPE) is 3.55%. The periodicity of the workdays and weekends has been identified. 33.1% is the maximum absolute percentage error (MaxAPE). Alexander Peshkov (2021) showed that the error for ARIMA model was evaluated using MAE which is 26.52% ,MPE with 2.51%, MAPE with 2.77%, RMSE with 33.61% and R^2 with 0.797. T.Ulgen (2020) used Multiple linear regression. In this study in order to lower the mean absolute percentage error, many predictors are examined. Mean Absolute Percentage Error (MAPE) utilized to measure error rates for evaluating price estimation , Root Mean Square Error (RMSE) used error rates for evaluating price estimation and Mean Signed Deviation (MSD) used in this paper. The best result is to use a one-year training set in comparison to other training sets for short-term electricity price forecasting.The result of the analysis that performed in this paper has been proved that historical electricity allow the forecasting model's error rate to decrease because on September 24, 2019, the model hits its lowest MAPE of 1.97%.

However, there are some gaps and limitations in the current literature. To begin, the reviewed articles primarily focus on improving forecasting accuracy rather than extensively addressing model interpretability. Furthermore, there has been little discussion about the potential effects of these forecasting models on energy management systems and policy analysis. The majority of the reviewed studies concentrate on particular geographic areas, like the Colombian and German electricity markets, which restricts the applicability of the suggested models.

The current study attempts to make a contribution by addressing these shortcomings and gaps in light of these findings. It will look into how the suggested model affects policy-making and energy management systems across a wider geographic area. To improve the generalizability of the results, the study will also look into how well the model applies to different energy markets and indices. This study deepens our understanding of electricity price forecasting methodologies, contributing to the body of knowledge already in existence. By offering an interpretable and contextually relevant forecasting model that addresses the shortcomings and gaps found in the reviewed articles, the study seeks to advance the field. The research's multidisciplinary methodology will support policy analysis and the real-world application of electricity price forecasting in energy management systems.

Additionally, studies conducted in more expansive geographic settings will shed light on how the suggested model can be applied to different energy markets and indexes, adding to the body of knowledge in the field.

CHAPTER 3
METHODOLOGY

3.1 Introduction

In this section, the flow of the project is discussed step by step. The detail explanation is as explain the subsection below. There are consist of the flowchart, data preparation, time series model and error measurement analysis.

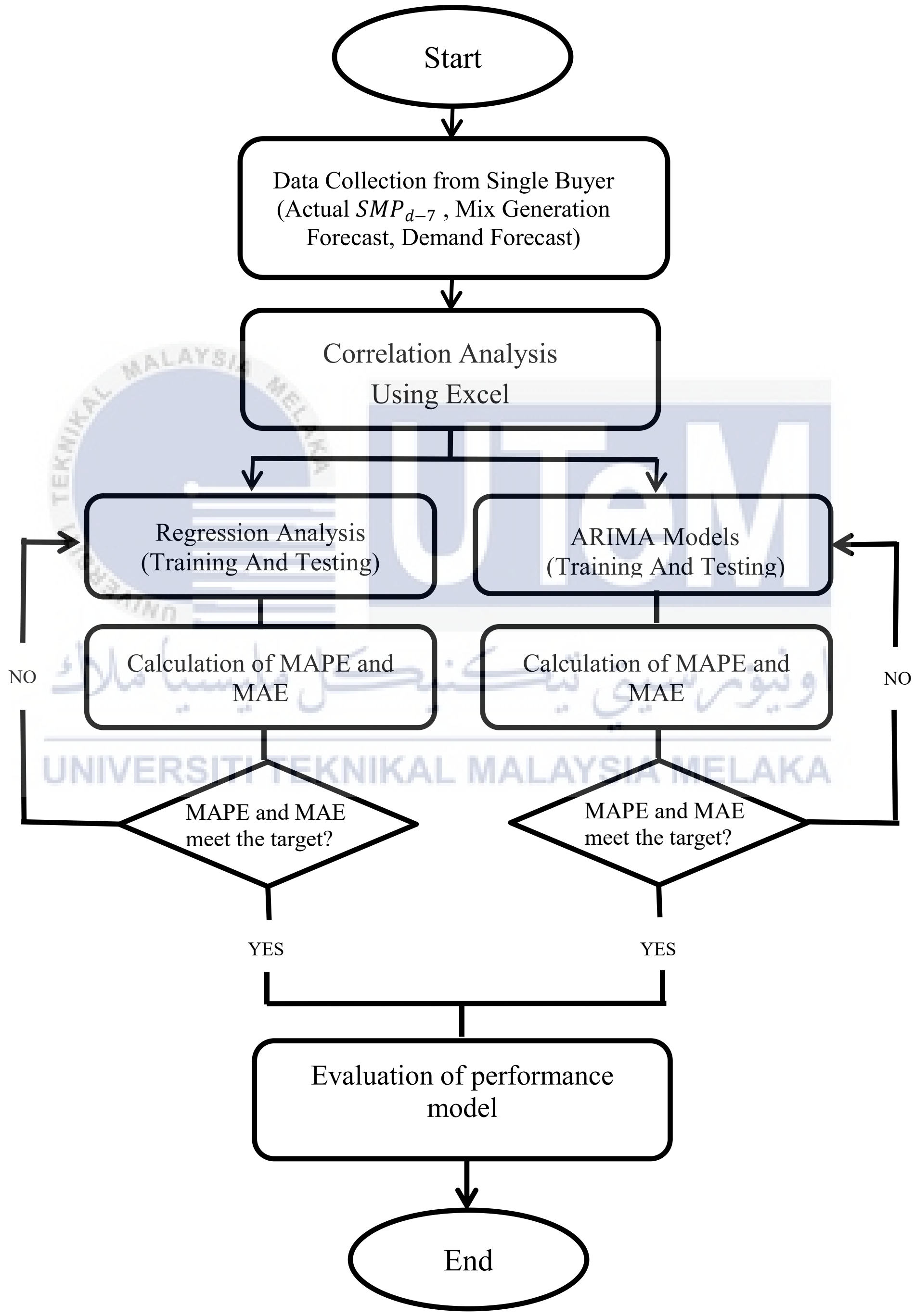

Figure 3.1:Flowchart of process forecasting

3.2 Data Collection

As part of the data collection process for this study, input variables was gathered from reliable sources which is the Single Buyer (SB) website. These organisation is important to the functioning of the Malaysian power market, which makes their data extremely reliable and valuable for analysis. The input variable which is actual System Marginal Price (SMP) represent the hourly electricity price in the market, the Generation Mix data describes the energy sources used to generate power and the demand forecasts offer insight into the expected patterns of electricity consumption among the specific datasets that were obtained. An comprehensive study of price fluctuations and their correlation with patterns in the generation mix and demand was made possible by the hourly data collection, which ensured a thorough understanding of market dynamics. The study ensured the use of accurate and current information by relying on these official sources, resulting in a strong basis for the implementation of the time series method in SMP forecasting for the Malaysian power industry.

3.2.1 SMP Forecasting Data in Malaysia's Power Market

To accurately predict the system marginal price (SMP), one important source of data in the Malaysian power market is the Single Buyer. It is a primary company that purchases energy from power plants and manages the power purchase agreements (PPAs) with various electrical production companies.[33]. Information about the contract's price and power generation is preserved by the Single Buyer. Both the terms and conditions of contracts and the workings of the electricity market can be made clearer with the aid of these records. Researchers and market participants can identify trends, monitor the impact of rule changes on the market, and incorporate contractual information into SMP forecasting models by examining historical data from the Single Buyer.

A possible source of information for SMP forecasting in the Malaysian Power Market is Single Buyer. Not all datasets, nevertheless, will be easily accessible or available. To access and use data from these sources, researchers and analysts must strictly abide by the data security policies, data usage policies, and any agreements or permissions required by the relevant organisations.

3.3 Correlation Analysis Using Excel

System marginal price (SMP) forecasting in electricity markets requires a specific approach to identify the most important factors and features. Although they can still be applied, statistical methods like correlation analysis must be modified to take into consideration the unique characteristics of SMP, such as pricing and market dynamics. Factors influencing the electricity markets include the total amount of energy needed, the amount that can be produced, fuel prices, renewable energy availability, weather, transmission line traffic, and regulatory constraints. The market's dynamics and demands can alter the relative importance of these inputs.

Correlation analysis is one of the most important techniques for figuring out how these sources affect SMP. The impact of variables like mix generation forecast and energy demand forecast on SMP can be observed using correlation coefficients. Positive correlations could mean that higher SMP values were the outcome of adding more of a specific input.[34]. Additionally, it can be very beneficial to look for connections between SMP and other factors, such as the quantity of energy derived from renewable sources or the level of transmission system congestion.Selecting the most significant sources for SMP forecasting is made easier by using ranking techniques for features. However, it's crucial to remember that the parameters used to choose features and determine the relative importance of inputs rely on the particulars of the electricity market under study.

Correlation analysis is a statistical method for determining the pattern and degree of a linear connection between two variables that are quantitative.It decides how closely one variable's change is related to another change. A correlation analysis generated a correlation coefficient, which present by the letter "r." This correlation coefficients was constructed using excel software.The correlation coefficient includes a range between -1 and 1, and it includes all possible linear relationships between two variables. A correlation coefficient of one shows perfect positive correlation, or that modifications in one variable are inversely related to transforms in the other. A correlation coefficient with a value of -1, on the other hand shows a perfect negative correlation, which indicates that when one variable rises, the other decreases proportionally.

When linear correlation between the variables does not have connection as indicated by a correlation coefficient of 0, changes one of them cannot be expected to affect changes in the other. Analysts and researchers may employ the correlation coefficient, that offers a numerical value that reflects the strength and direction of the relationship to determine the degree of association between variables. This range, ranging from -1 to 1, presents a concise and straightforward explanation of the linear relationship's nature and provides illumination on the the relationships and trends observed in datasets from many different areas of study.[35].

$$r = \frac{[\sum (x-\bar{x})(y-\bar{y})]}{\sqrt{\sum (x-\bar{x})^2}\sqrt{\sum (y-\bar{y})^2}} \tag{3.1}$$

Where,

r = correlation coefficient

$\sum$ = sum of

x =value of the x-variable

$\bar{x}$ =mean of the x-variable

y =value of the y-variable

$\bar{y}$ =mean of the y-variable

The result of the correlation can be determined whether variables can be used or not using the correlation coefficient scale.

Table 3.1: Correlation Coefficient Scale

Scale Of Correlation Coefficient	Value
$0<r<0.19$	Very Low
$0.2<r<0.39$	Low Correlation
$0.4<r<0.59$	Moderate Correlation
$0.6<r<0.79$	High Correlation
$0.8<r<1.0$	Very High Correlation

3.4 Development Of Time Series Model

Time series is a sequence of data points collected at regular intervals of time. Time series analysis involves understanding trends and patterns in data over time and using that data processing to make forecasts.

3.4.1 Regression Analysis Model

A key method used in forecasting to model and project future values based on historical data is regression analysis. For forecasting the future value, regression analysis model is employed to determine and measure the relationships among a dependent variable (the variable to be predicted) and one or more independent variables (predictors). Regression analysis develops a statistical model that most accurately fits the historical data, which allow to forecast the best future value.

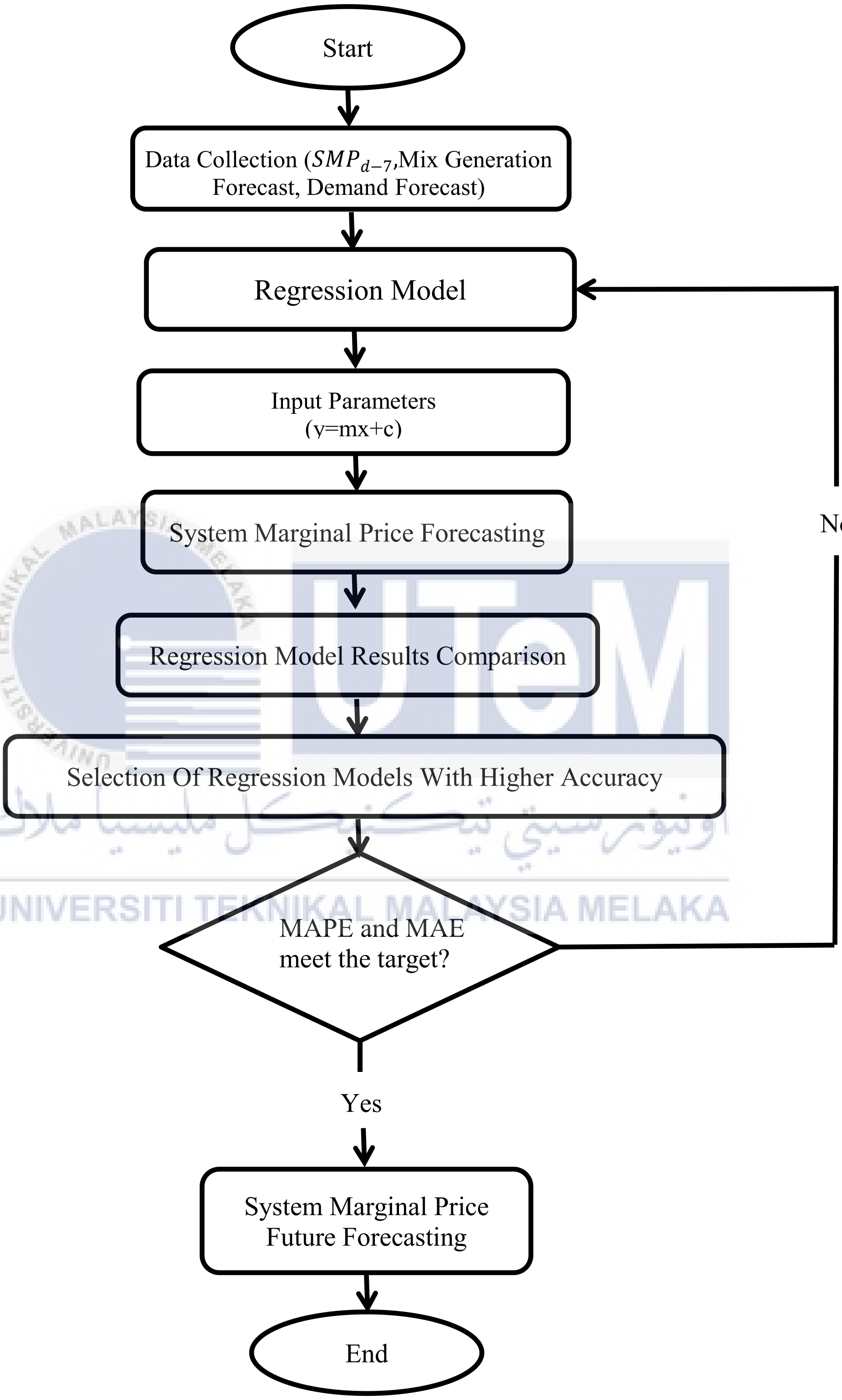

Figure 3.2 : Flowchart of process forecasting using regression model.

One popular forecasting technique used in price forecasting is multiple linear regression. The weighted sum of the input features is how the MLR model always computes. Linearity is among the most notable benefits. It makes process prediction easy. It is also simple to read and comprehend. In this project, the Malaysian Electricity Market prices has been used for estimation. The datasets for this research has been taken from the Single Buyer and all observations are half-hourly data. Some of the features that correlate with system marginal price forecasting to contain like actual SMP seven days prior, mix generation forecast ,demand forecast and actual SMP are added to culminate in forecasting performance. The selection of training and the testing datasets is illustrated in Figure 3.3.

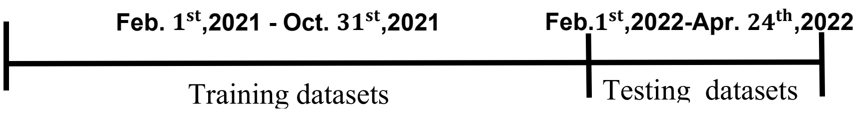

Figure 3.3 : Training and Testing Datasets

The training datasets have been used to find the model and the input actual SMP seven days prior, mix generation forecast and demand forecast were selected in this model while for the output the actual SMP were selected. Equation 3.2 is multiple linear regression for time series analysis.

$$Y = \beta_1 X_1 + \beta_0 \tag{3.2}$$

Where, β_0 is intercept ,β_1 is coefficients SMP seven days prior and X_1 is input data for SMP seven days prior. Then this model equation were used for forecasting system marginal price in testing datasets.

3.4.2 ARIMA Model

Auto-regressive Integrated Moving Average (ARIMA) model is a type of statistical model for analyzing and forecasting time series data. It is widely used model for time series forecasting, and it is particularly useful for analyzing data with trends and seasonality.

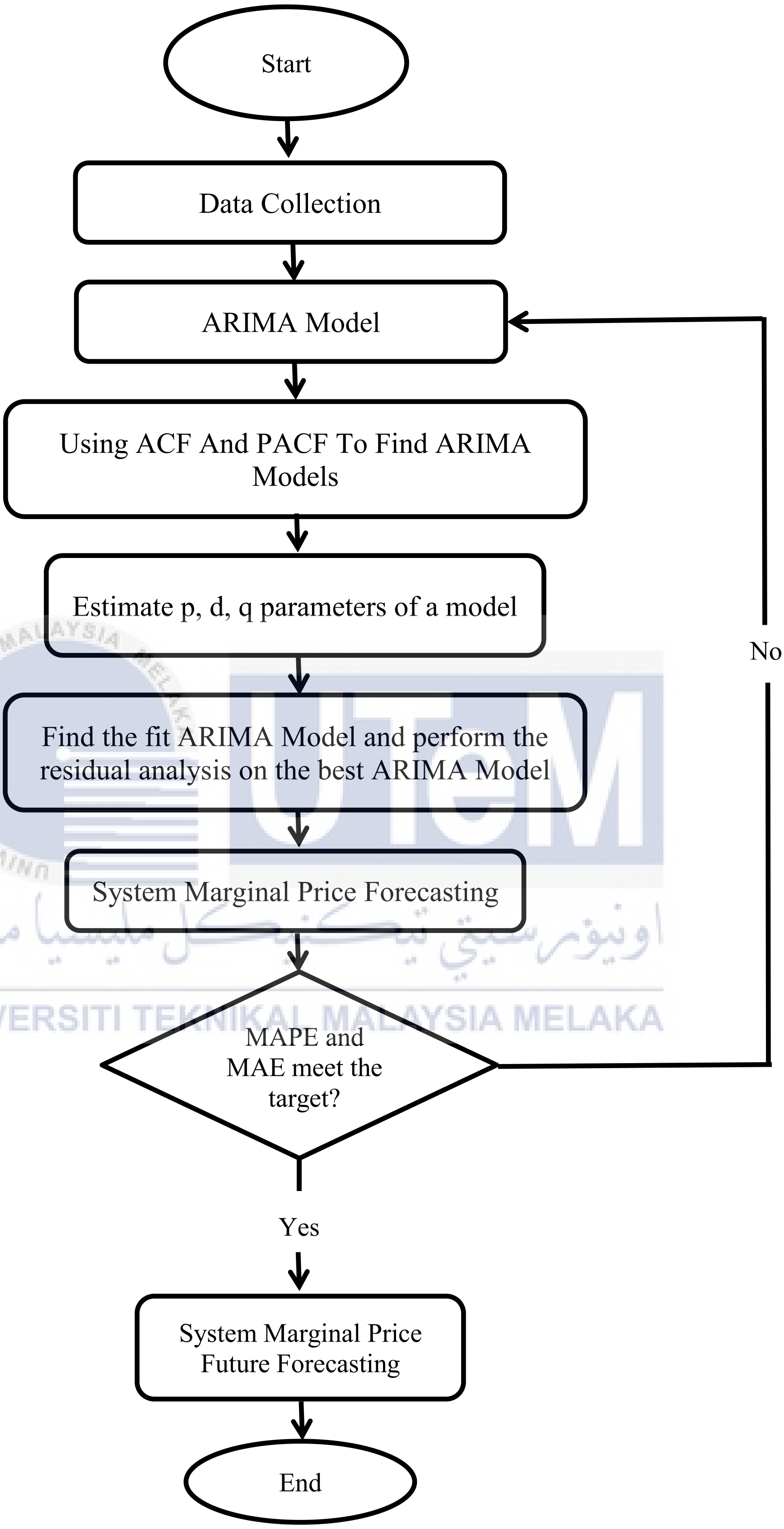

Figure 3.4: Flowchart of process forecasting using ARIMA model

The ARIMA model is a combination of three components Auto-regressive (AR) refers to the dependence of the current value on previous values in the time series. The number of past values used (p) is called the lag. Integrated (I) this term involves differencing the time series to make it stationary. The number of times the time series is differenced (d) is called the degree of differencing. [36].

For moving average (MA) the term involves used the errors or residuals of the previous time steps to predict future values. The number of past errors used (q) is called the order of the moving average. General formula for ARIMA model is given in (3.3)

$$\mathrm{ARIMA(p, d, q)} = \mu + \emptyset_1 y_t + \ldots + \emptyset_p y_{tp} - \theta_1 e_1 - \ldots \theta_q e_{t-q} \quad (3.3)$$

Where $\emptyset$ is auto-regression parameter; $\theta$is moving average parameters; $\mu$is constant; $\emptyset_p y_{tp}$ is the auto-regressive model while $\theta_q e_{t-q}$ is the moving average model; $\theta_1 e_1$ degree of differential; $\emptyset_1 y_t$ is d^{th} a stationary ARMA model. The parameters p, d, and q can be determined using statistical techniques such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC).

3.5 Model Parameter Selection

Model parameter selection determine the performance and accuracy of the models used for forecasting time series analysis.

3.5.1Autocorrelation Function (ACF)

The auto-correlation function (ACF) is a measure of the correlation between a time series at different lag periods. It is used to determine the presence of auto correlation, which is the correlation between observations that are separated by a specific number of time periods [37]. The fundamental of the auto correlation at lag *k* is the correlation between the original data series, $x_t$and the same series moved forward one period represented as x_{t-1}. The auto covariance at lag k is defined in (3.4)

$$\mathrm{cov}(x_t, x_{t-1}) = E(x_t - \mu)(x_{t-1} - \mu) \quad (3.4)$$

where *E* is estimation;μ is the mean of the data observation. Thus, the auto correlation at lag *k* is defined in (3.5)

$$\text{ACF}, \rho_k = \frac{E(x_t - \mu)(x_{t-1} - \mu)}{\sigma_x^2} \tag{3.5}$$

Where σ_x^2 is the variance of stochastic process; E is the estimation; μ is the mean of the data observation; k is specific number of periods. The ACF used in time series analysis to identify the appropriate order moving average (MA) terms in a time series model.

3.5.2 Partial Auto correlation Function (PACF)

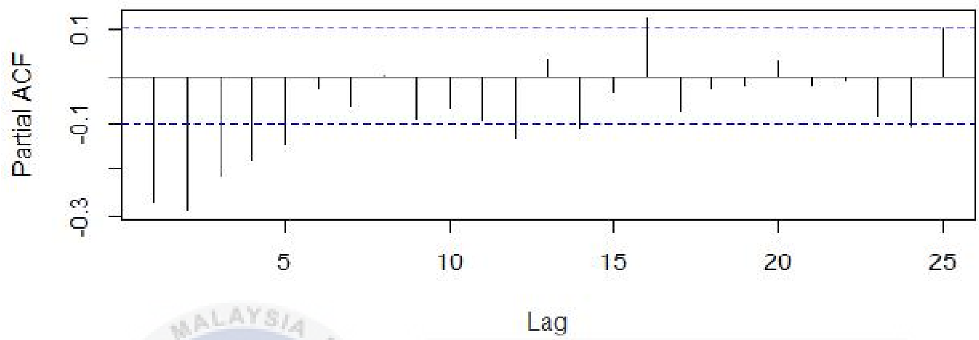

Figure 3.5 : PACF plot of the time series data after differencing [37]

Figure 3.5 shows an example of PACF plot after differencing. PACF is a measure of the correlation between a time series and the lagged values. This function controls the effects of intermediate lags. To compute the PACF, the time series is first transformed into a series of lagged values [27],[37]. The correlation between the original time series and each of these lagged values is then calculated, controlling for the effects of the intermediate lags.

3.6 Error Measurement

In time series analysis, error measurement refers to evaluating a forecasting model's accuracy by contrasting its predicted values with the actual observed values. Time series forecasting errors can be evaluated using a variety of metrics and techniques.

3.6.1 Mean Absolute Percent Error (MAPE)

Relative to prediction accuracy, the MAPE measures the average percentage difference between the actual and expected values. It clarifies the extent to which the errors deviate from the actual values. [38]. The formula for calculating MAPE is as follows:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \frac{|\text{Pactual} - \text{Pforecast}|}{\text{Pactual}} \times 100\% \tag{3.6}$$

Where, n= total number of predictions , $\sum =$ the summation symbol and [(Pactual - Pforecast) / Pactual] = absolute percentage difference between the actual value and the predicted value for each observation, multiplied by 100. The MAPE represents the average percentage difference between the actual values and the model's predictions. Because it indicates that the average percentage deviation between the model's predictions and the actual values is smaller, a lower MAPE is indicative of higher accuracy.

3.6.2 Mean Absolute Error (MAE)

Without taking the direction into account, the MAE calculates the average magnitude of the prediction errors in the model. [39]. It calculates the average distance which is the absolute difference between the true and predicted values between the predicted and true values. The MAE formula is given in equation (3.7)

$$\mathrm{MAE} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{Pactual} - \mathrm{Pforecast}) \tag{3.7}$$

Where, n= total number of predictions , $\sum =$ the summation symbol and [Pactual-Pforecast] = absolute difference between the actual value and the predicted value for each observation. The MAE is a representation of the average magnitude of the absolute errors generated by the model. Because it demonstrates that the model's predictions are typically closer to the actual values, a smaller MAE indicates higher accuracy.

3.7 Summary

The current study makes use of a range of data analysis methods to look into the relationship between variables and assess how well the regression analysis and ARIMA model forecast. Data analysis is done to gain a comprehensive understanding of the variables and their interactions. Correlation analysis is used to measure the strength and direction of relationships between variables in order to obtain insight into their dependencies and potential predictive power. This analysis assesses the forecasting model's accuracy in predicting future events and helps select the best input features. Integrating data analysis and correlation analysis yields valuable insights for planning and decision-making. These conclusions are predicated on how well the model forecasts future events.

CHAPTER 4
RESULTS AND DISCUSSIONS

4.1 Introduction

A great deal of knowledge about the electricity markets has been obtained by looking over and choosing important inputs for system marginal price (SMP) forecasting.This study used statistical techniques like regression analysis and ARIMA to increase the accuracy of price forecasts and use correlation analysis to determine the most important factors influencing SMP. Understanding these factors is essential for developing regulations, estimating costs, and strategizing the energy sector. The following section, which focuses on the important inputs and how they affect SMP, presents the analysis's findings. These findings add to the body of knowledge in SMP forecasting, offering the power sector insightful information that helps them make well-informed decisions in the complex environment of the energy market.

4.2 Correlation Analysis

The correlation analysis of the data provided provides useful information about the relationships between the variables and how they affect the system marginal price (SMP) in electricity markets.

Table 4.1: Summarize Correlation Analysis Result

Variable 1 (Target)	Variable 2 (Input)	Correlation Coefficient
SMP_d	SMP_{d-1}	0.223725584
	SMP_{d-2}	0.100172042
	SMP_{d-3}	0.136567541
	SMP_{d-7}	0.136763761
	Mix Generation Forecast	0.419489147
	Demand Forecast	0.388574

The correlation between the SMP_d and the SMP_{d-1} is 0.2237 while the correlation between SMP_d and SMP_{d-7} is 0.1367, which is a very weak positive correlation same as the correlation SMP_{d-2} and SMP_{d-3} . Even though the correlation is close to zero, it shows that there is some link between the real SMP values and the forecasts for them. Even though the link is weak, it still shows that forecasts could be more accurate.

When the correlations between SMP_d and mix generation are examined, it is clear that the good ones are stronger. The correlation between the SMP_d and the mix generation forecast is 0.4194, indicating that the mix generation forecast influences the SMP prediction. Furthermore, the correlation between the SMP_d and demand forecast is 0.3885, indicating that energy demand forecast influences the SMP_d . The fact that these relationships are stronger indicates that the mix generation forecast and demand forecast both play a larger role in SMP forecasting.

4.3 Regression Model Development

From the correlation result, we can conclude that the correlation between input and target are stronger so we can proceed to develop a regression model for future forecasting marginal price.

4.3.1 Multiple Linear Regression

The time series data for forecasting marginal price were divided into training and testing data. The testing data spanned from 1^{st} February 2022 to 24^{th} April 2022 while the training data were from 1^{st} February 2021 until 30^{th} October 2021.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.7410266
R Square	0.5491204
Adjusted R Square	0.5490172
Standard Error	0.0198414
Observations	13104

ANOVA

	df	SS	MS	F	Significance F
Regression	3	6.280908427	2.093636142	5318.107376	0
Residual	13100	5.157216943	0.000393681		
Total	13103	11.43812537			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 90.0%	Upper 90.0%
Intercept	0.0251413	0.001766179	14.23487736	1.21817E-45	0.021679373	0.02860331	0.022236029	0.028046651
SMP d-7	0.7614222	0.006218251	122.4495716	0	0.749233482	0.77361083	0.75119332	0.771650991
Demand Forecast	8.707E-07	1.69033E-07	5.151072349	0.000000263	5.39371E-07	1.202E-06	5.92646E-07	1.14875E-06
Mix Gen Forecast	2.227E-07	1.67877E-07	1.326741822	0.184617227	-1.06334E-07	5.5179E-07	-5.34231E-08	4.98881E-07

Figure 4.1: Summary Output Multi-Regression Analysis

Multi-Regression analysis was carried out using input SMP_{d-7}, demand forecast and mix generation forecast. The figure 4.1 shows the regression analysis output is generated by the training datasets. It shows that adjusted R square is 0.549 which is 55% and suitable for model development. The higher adjusted R square indicates better models. The significance F is 0 which is smalller than 0.05 shows that the value is suitable and can be used for model development. To generate equation the t Stat value chosen must be the value of two or higher and the P-value must smaller than 0.05. From the output, we can conclude that SMP_{d-7} and demand forecast input is the suitable input to generate an equation for a regression model with the t Stat value of SMP_{d-7} is 122.4495 and the P-value is 0 while for demand forecast the t Stat value is 5.1510 and the P-value is 0.000000263 . The equation will be

$$\mathrm{y} = \mathrm{m}_1\mathrm{x}_1 + \mathrm{m}_2\mathrm{x}_2 + \mathrm{c} \tag{4.1}$$

Where m_1 is the coefficients for SMP_{d-7} which is 0.7614222, x_1 is the value of SMP_{d-7} input ,m_2 is the coefficients for demand forecast which is 8.707E-07, $\mathrm{x}_2$is the value of demand forecast input and c is the intercept which is 0.0251413. This equation (4.1) , will be used in testing datasets to calculate error. The model's mean absolute error (MAE) is 0.0249 and the mean absolute percentage error (MAPE) is 8.07%.

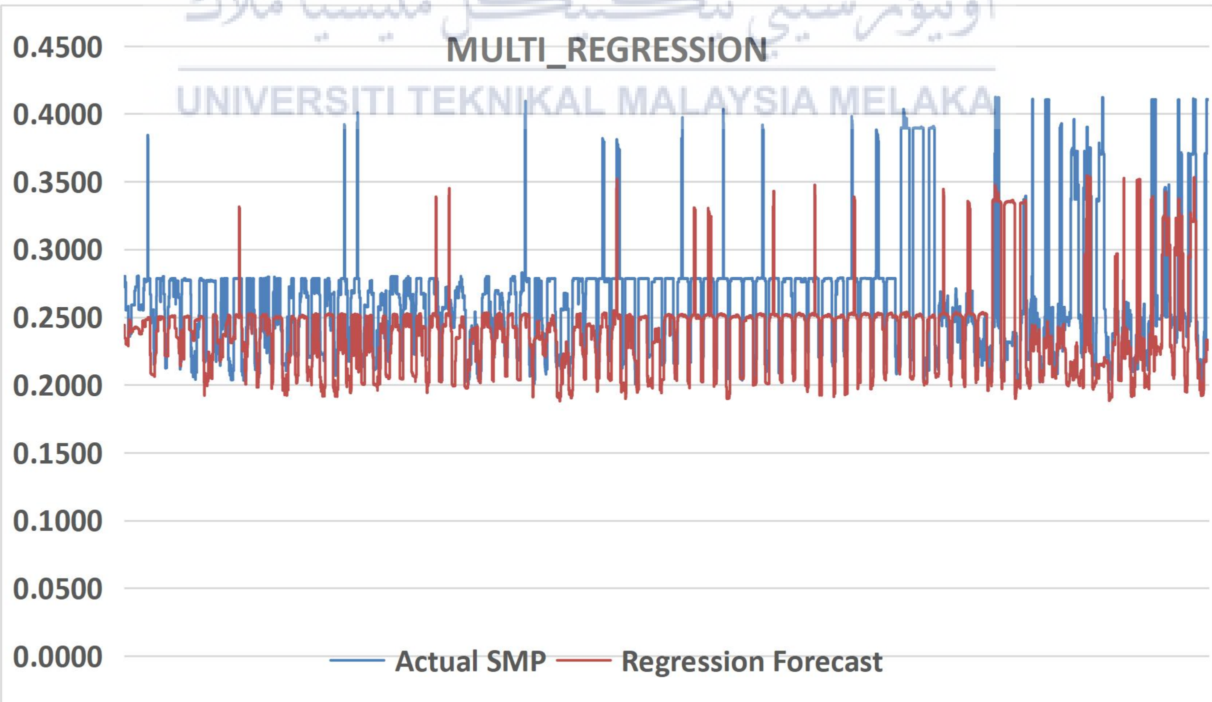

Figure 4.2 : Graph Actual And Forecast Value Using Multi-Regression Model

4.3.2 Simple Linear Regression

A simple linear regression model is a statistical method for modelling the relationship between a dependent variable (also known as the response or outcome variable) and a single independent variable (also known as the predictor or explanatory variable). Simple linear regression seeks the best-fitting linear equation that describes the relationship between these two variables.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.817824381
R Square	0.668836719
Adjusted R Square	0.668811443
Standard Error	0.017003163
Observations	13104

ANOVA

	df	SS	MS	F	ignificance F
Regression	1	7.650238	7.65023824	26461.56	0
Residual	13102	3.787887	0.000289108		
Total	13103	11.43813			

	Coefficients	andard Err	t Stat	P-value	Lower 95%	Upper 95%	ower 90.0%	pper 90.0%
Intercept	0.030363611	0.000839	36.17520417	3.2E-273	0.028718	0.032009	0.028983	0.031744
SMPd-1(31/1)	0.816932618	0.005022	162.6701109	0	0.807089	0.826777	0.808672	0.825194

Figure 4.3 : Summary Output Simple Linear Regression SMP_{d-1}

Figure 4.3 shows the forecasts is generated by the training data from 1^{st} February 2021 until 30^{th} October 2021 for simple linear regression using SMP_{d-1} input. It shows that adjusted R square is 0.668 which is 68% and suitable for model development. The higher adjusted R square indicates better models. The significance F is 0 which is smaller than 0.05 shows that the value is suitable and can be used for model development. To generate equation the t Stat value chosen must be the value of two or higher and the P-value must smaller than 0.05. From the figure 4.3, the t Stat value is 162.6701 and the P-value is 0.The equation for this model is

$$y = mx + c \tag{4.2}$$

Where m is the SMP_{d-1} coefficients which is 0.8169, x is the value of SMP_{d-1} input and c is the intercept which is 0.0303. This equation has been used to test the datasets and the model's mean absolute error (MAE) is 0.0298 and the mean absolute percentage error (MAPE) is 10.70%.

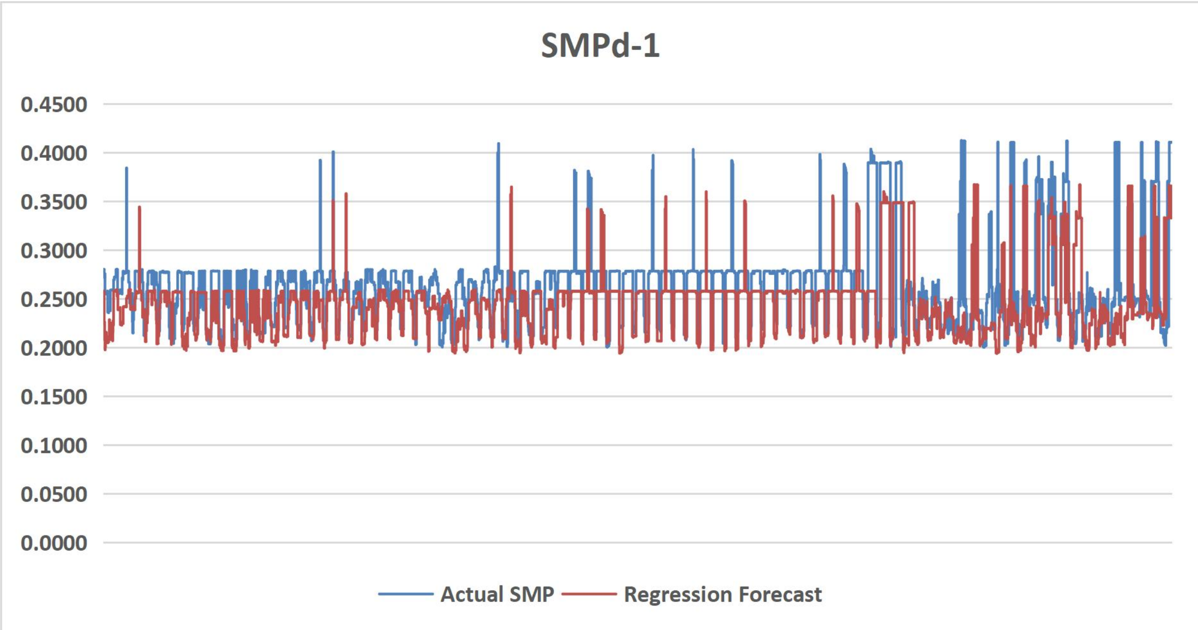

Figure 4.4 : Graph Actual And Forecast Value For SMP_{d-1}

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.738662461							
R Square	0.545622231							
Adjusted R Square	0.545587551							
Standard Error	0.01991669							
Observations	13104							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	6.240895486	6.240895486	15733.0375	0			
Residual	13102	5.197229885	0.000396675					
Total	13103	11.43812537						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 90.0%*	*Upper 90.0%*
Intercept	0.039398594	0.00101436	38.84083423	0	0.0374103	0.041387	0.03773	0.04106719
SMP d-7	0.771118924	0.006147734	125.4314055	0	0.75906847	0.783169	0.76100609	0.78123176

Figure 4.5 : Summary Output Simple Linear Regression SMP_{d-7}

Figure 4.5 shows the forecasts is generated by the training data from 1^{st} February 2021 until 30^{th} October 2021 for simple linear regression using SMP_{d-7} input. It shows that adjusted R square is 0.546 which is 55% and suitable for model development. The higher adjusted R square indicates better models. The significance F is 0 which is smaller than 0.05 shows that the value is suitable and can be used for model development. To generate equation the t Stat value chosen must be the value of two or higher and the P-value must smaller than 0.05. From the figure 4.5, the t Stat value is 125.4314 and the P-value is 0.The equation still using the same equation as (4.2). This value has been used to test the datasets and the model's mean absolute error (MAE) is 0.0333 and the mean absolute percentage error (MAPE) is 11.77%.

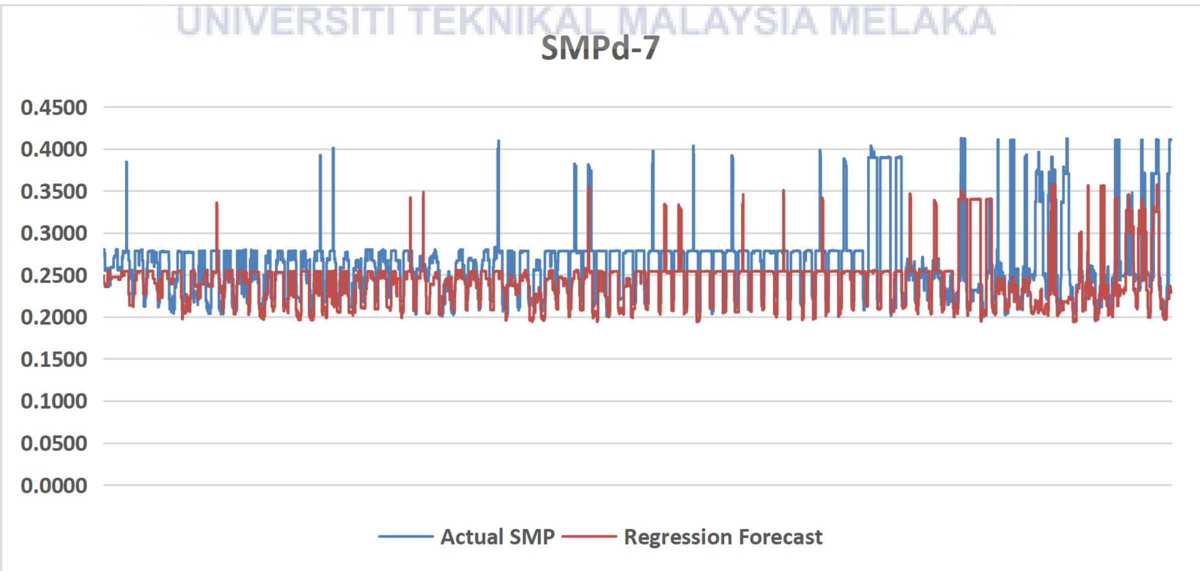

Figure 4.6 : Graph Actual And Forecast Value For SMP_{d-7}

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.164254025
R Square	0.026979385
Adjusted R Square	0.02690512
Standard Error	0.029145375
Observations	13104

ANOVA

	df	SS	MS	F	Significance F
Regression	1	0.308593584	0.308593584	363.2851062	6.51711E-80
Residual	13102	11.12953179	0.000849453		
Total	13103	11.43812537			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	ower 90.0%	Ipper 90.0%
Intercept	0.123038583	0.002202956	55.85159013	0	0.11872047	0.127357	0.119415	0.126662
Demand Forecast	2.90189E-06	1.5225E-07	19.06003951	6.51711E-80	2.60346E-06	3.2E-06	2.65E-06	3.15E-06

Figure 4.7: Summary Output Simple Linear Regression Demand Forecast

Figure 4.7 shows the forecasts is generated by the training data from 1^{st} February 2021 until 30^{th} October 2021 for simple linear regression using demand forecast input. It shows that adjusted R square is 0.026 which is 2.6%. The adjusted R square is too small and the forecasting value will not achieve the suitable value. The significance F is 6.5171E-80 which is smaller than 0.05 shows that the value is suitable . The t Stat value is 19.0600 which is higher than 2 and the P-value is smaller than 0.05 which is 6.5171E-80. The equation still using the same equation as (4.2).This value has been used to test the datasets and model's mean absolute error (MAE) is 0.0966 and the mean absolute percentage error (MAPE) is 35.34%.

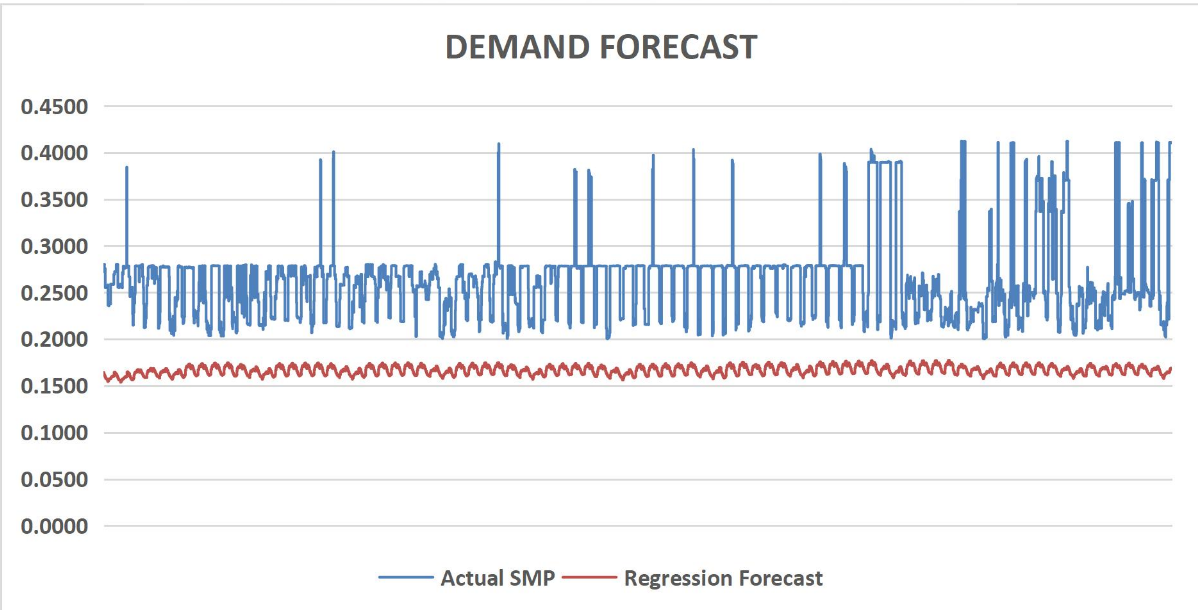

Figure 4.8: Graph Actual And Forecast Value For Demand Forecast

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.177548647
R Square	0.031523522
Adjusted R Square	0.031449604
Standard Error	0.029077239
Observations	13104

ANOVA

	df	SS	MS	F	Significance F
Regression	1	0.360569997	0.36057	426.4649	2.90053E-93
Residual	13102	11.07755537	0.00084549		
Total	13103	11.43812537			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	ower 90.0%	Ipper 90.0%
Intercept	0.118532626	0.002252177	52.6302451	0	0.114118033	0.122947	0.114828	0.122237
Mix Gen Forecast	3.10081E-06	1.50153E-07	20.6510256	2.9E-93	2.80649E-06	3.4E-06	2.85E-06	3.35E-06

Figure 4.9: Summary Output Simple Linear Regression Mix Generation Forecast

Figure 4.9 shows the forecasts is generated by the training data from 1^{st} February 2021 until 30^{th} October 2021 for simple linear regression using mix generation forecast input. It shows that adjusted R square is 0.031 which is 3.1%. The adjusted R square is too small and the forecasting value will not achieve the suitable value. The significance F is 2.9005E-93 which is smaller than 0.05 shows that the value is suitable . The t Stat value is 20.6510 which is higher than 2 and the P-value is smaller than 0.05 which is 2.9E-93. The equation still using the same equation as (3.2). This value has been used to test the datasets and model's mean absolute error (MAE) is 0.0984 and the mean absolute percentage error (MAPE) is 36.08%.

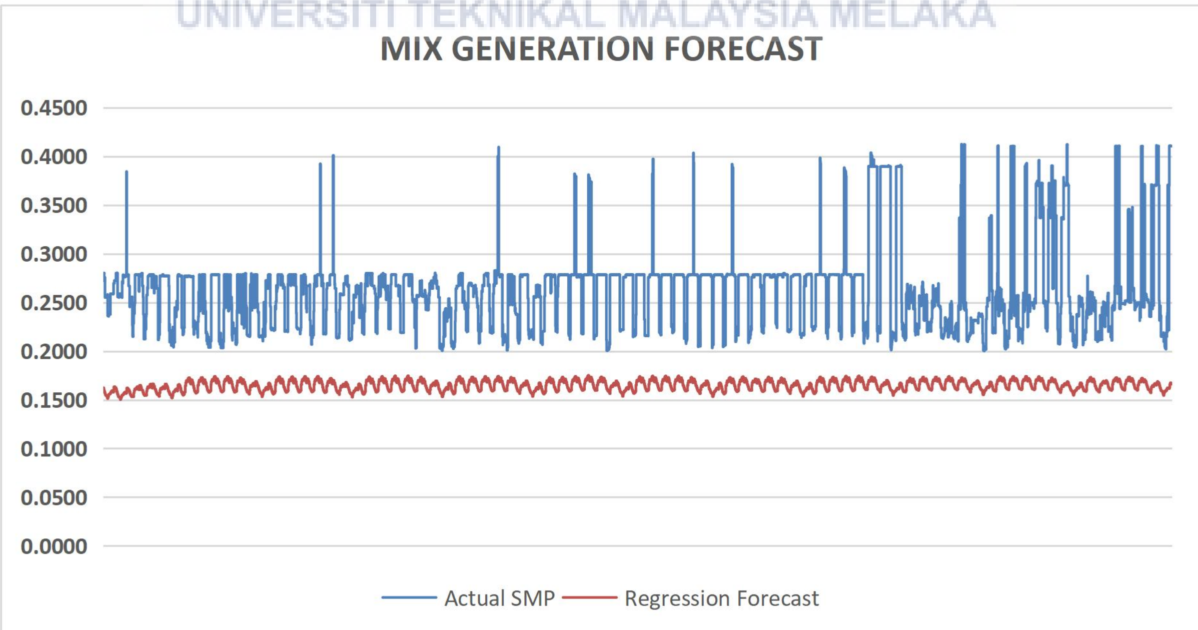

Figure 4.10 : Graph Actual And Forecast Value For Mix Generation Forecast

From the result of regression model showed that even if the correlation between targeted SMP and mix generation forecast and demand forecast is high but their value are not suitable for SMP future forecasting because the model's mean absolute error (MAE) and the mean absolute percentage error (MAPE) is too bigger while the variables with smaller correlation generated the smaller error in regression model and suitable for SMP future forecasting.

4.4 Time Series Plot

Time series plot is a graphical representation of data points collected over a specific period. It displays the values of a variable or multiple variables on the y-axis against corresponding time points on the x-axis. Figure 4.11 until figure 4.14 shows the system marginal price values plotted that allowed the observation of seasonal, pattern and trends in the data.

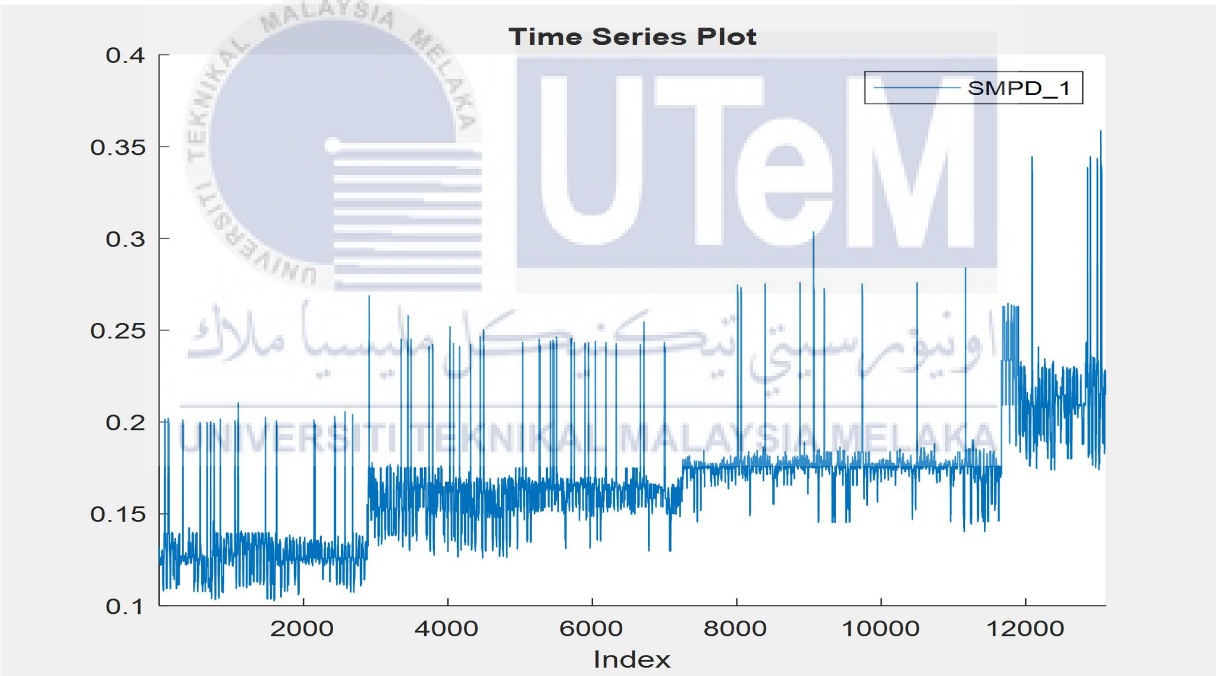

Figure 4.11 : SMP_{d-1} time series plot

Figure 4.11 shows the time series plot that were generated via the Matlab software. The trend helps in understanding various aspects of the data, such as the seasonality. The time series plot illustrates the relationship between the system marginal price one day before which consists of 3984 observations and the time . The x-axis represents the time ,while the y-axis represents the system marginal price. In this case, there is no trend neither a steady rise nor fall over time but it has regular surges that happen at regular intervals recurring patterns imply that there is seasonality in the data for the variable SMP_{d-1} .

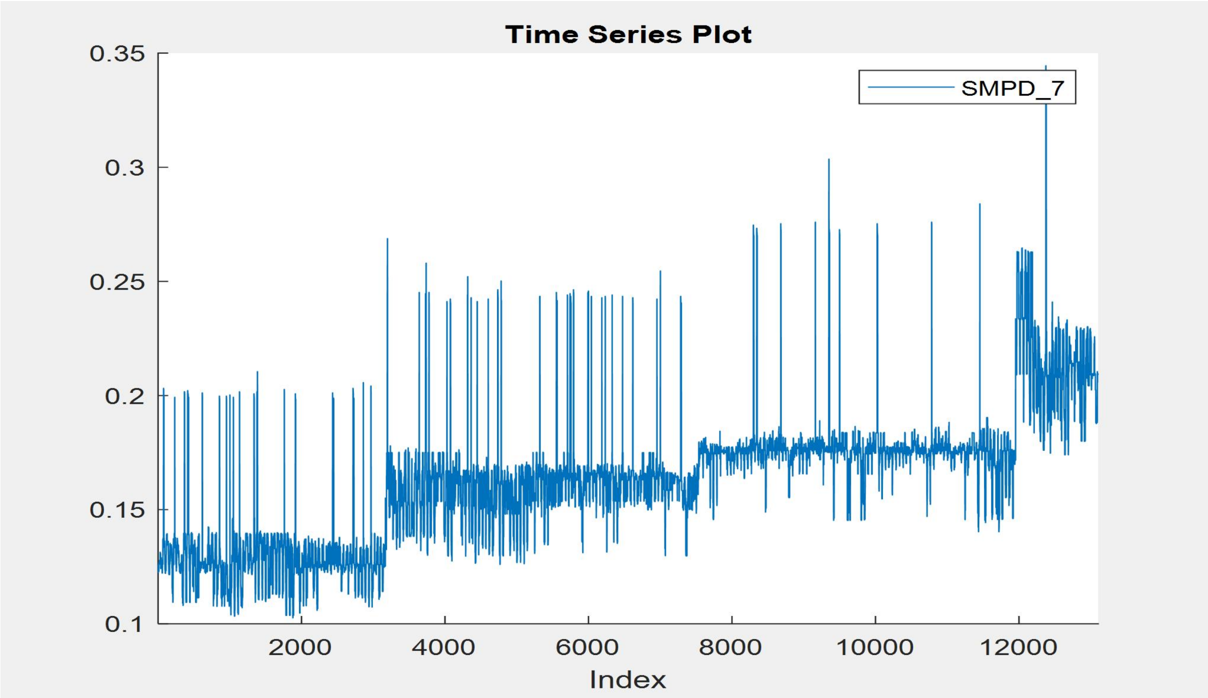

Figure 4.12 : SMP_{d-7} time series plot

Figure 4.12 shows the time series plot that were generated via the Matlab software. The time series plot illustrates the relationship between the system marginal price seven day before which consists of 3984 observations and the time. The x-axis represents the time ,while the y-axis represents the system marginal price. In this case, there is no trend, it can be concluded that there is a seasonality present for the system marginal price seven days before time series plot. It is same as the SMP_{d-1} time series plot because the data used are in the same range.

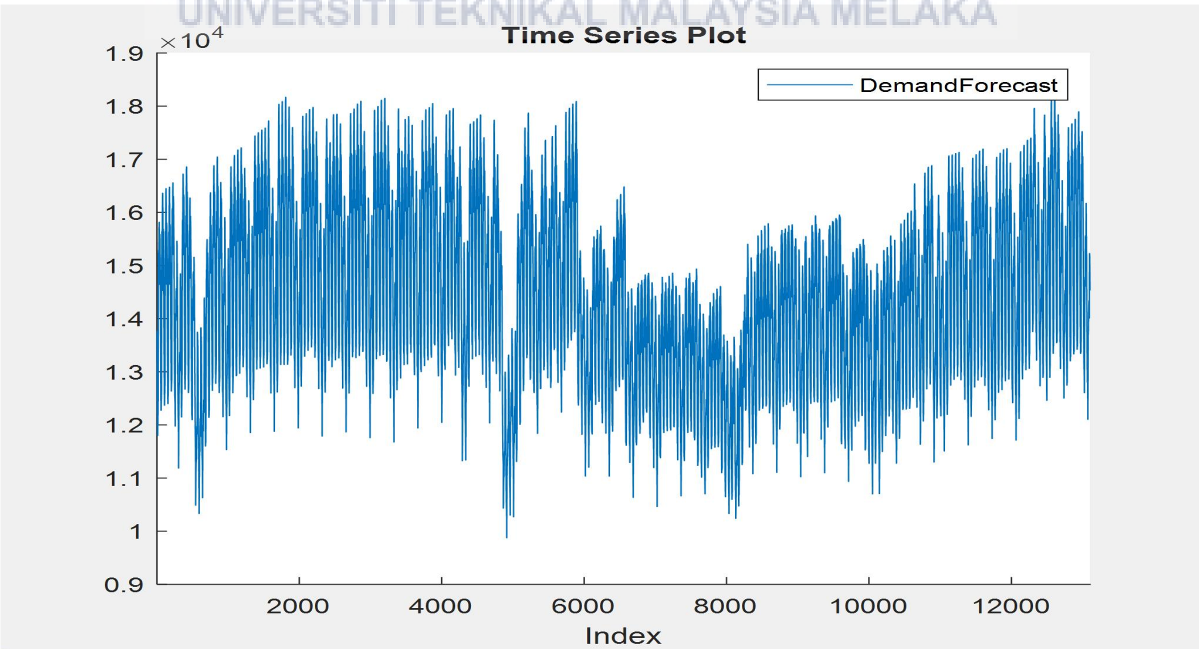

Figure 4.13 : Demand forecast time series plot

Figure 4.13 shows the time series plot that was generated via the Matlab software. The time series plot illustrates the relationship between demand forecast which consists of 3984 observations and the time . The x-axis represents the time ,while the y-axis represents the system marginal price. In this case, there is trend present that demand forecast are trending upward. The values rise steadily as the index rises. This points to a pattern of sustained growth throughout time. Regular spikes that happen at regular intervals demonstrate seasonality. Demand forecast show these recurring trends and this kind of behavior suggests that the dataset is seasonal, most likely due to recurrent occurrences or outside influences.

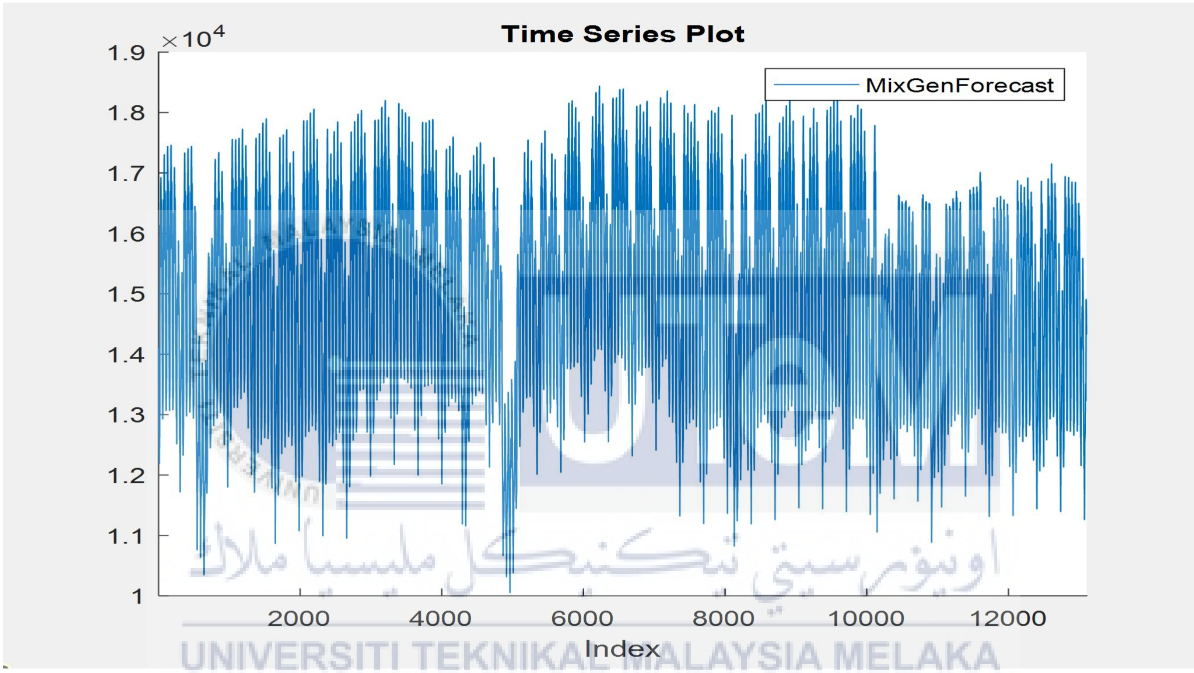

Figure 4.14 : Mix Generation forecast time series plot

Figure 4.14 shows the time series plot that was generated via the Matlab software. The time series plot illustrates the relationship between the mix generation forecast and the time . In this case, the demand forecast graph shows an increasing trend. The values rise steadily as the index rises. This points to a pattern of sustained growth throughout time. Regular spikes that happen at regular intervals demonstrate seasonality. Forecasts of demand show these recurring trends. This kind of behavior suggests that the datasets is seasonal, most likely due to recurrent occurrences or outside influences.

In conclusion, in the time series plot, it can analyze the trend of the data by identifying any consistent direction or movement over time. Additionally, the presence or absence of seasonality by looking for regular patterns or cycles that repeat at fixed intervals. Thus, there is two input with seasonality based on the trend in the time series plot.

4.5 Autocorrelation Function (ACF)

Autocorrelation function graph is used to analyze the correlation between observations in a time series datasets. The ACF graph displays the correlation coefficients between the observations at different lags, indicating the relationship between an observation and its 61 lagged values. The "lag" refers to the number of time steps or observations that separate two data points in a time series. For example, in this analysis a monthly time series was examined. Hence,a lag of 1 corresponds to the time difference between consecutive months. Figure 4.15 until Figure 4.18 show the ACF plot for all the types inputs.

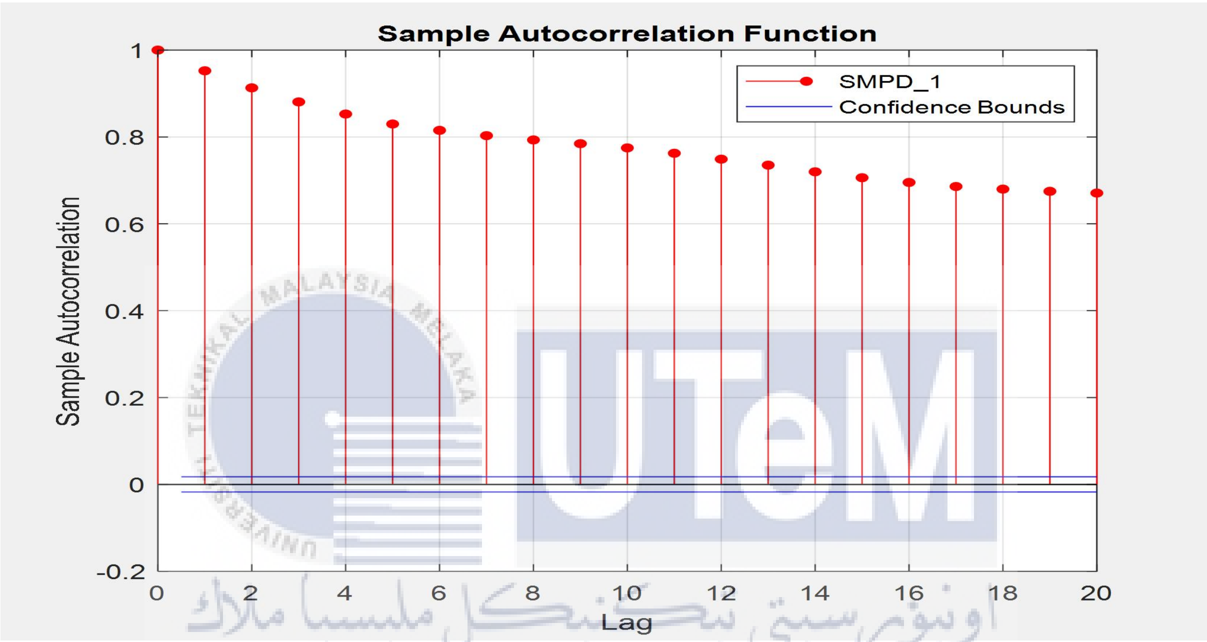

Figure 4.15 :ACF plot for SMP_{d-1}

Figure 4.15 shows ACF graph for SMP_{d-1} , it indicates the correlation between the current observation and its lagged values.The x-axis of the ACF graph represents the lag, which refers to the number of time intervals between the current observation and its lagged values. The y-axis of the ACF graph represents the correlation coefficient, which ranges from 0 to 1. The correlation coefficient measures the strength and direction of the linear relationship between two variables. The ACF graph is typically plotted with the correlation coefficient on the y-axis and the lag on the x-axis.

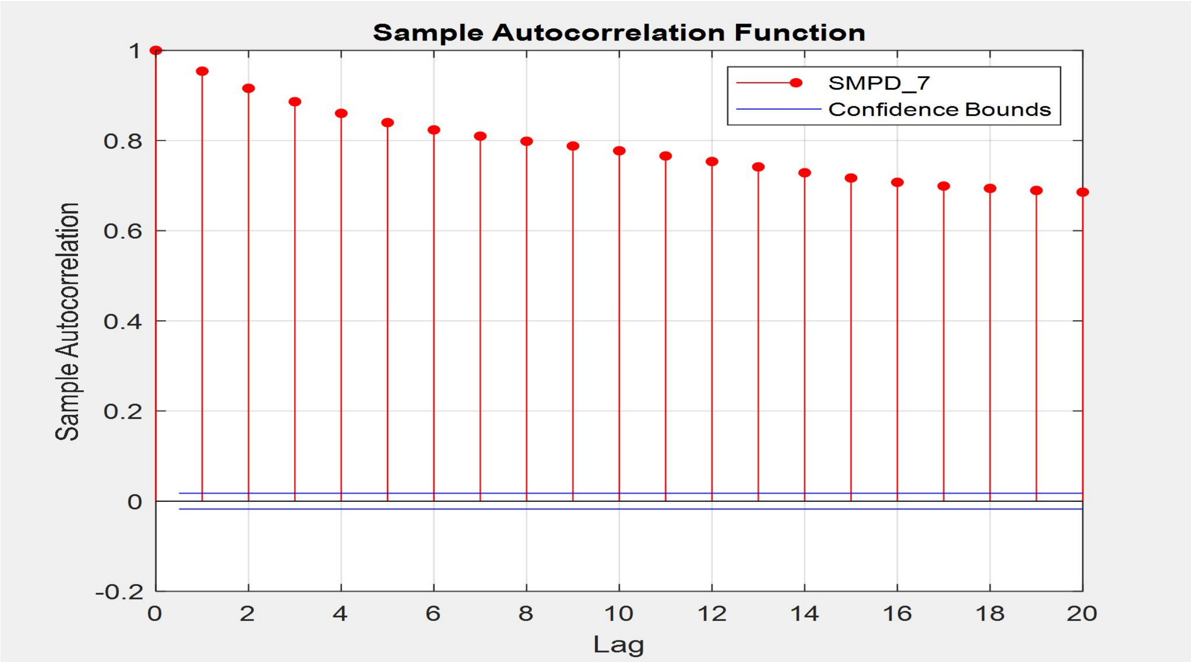

Figure 4.16 : ACF plot for SMP_{d-7}

The ACF graph shows in Figure 4.16 for SMP_{d-7} illustrates the correlation between the current observation and its past values. The x-axis of the ACF graph represents the lag, indicating the time intervals between the current observation and its lagged values. On the y-axis, the ACF graph portrays the correlation coefficient, which varies from 0 to 1 and assesses the strength and direction of the linear relationship between two variables. Typically, the ACF graph is plotted with the correlation coefficient on the y-axis and the lag on the x-axis. The ACF graph is typically plotted with the correlation coefficient on the y-axis and the lag on the x-axis.

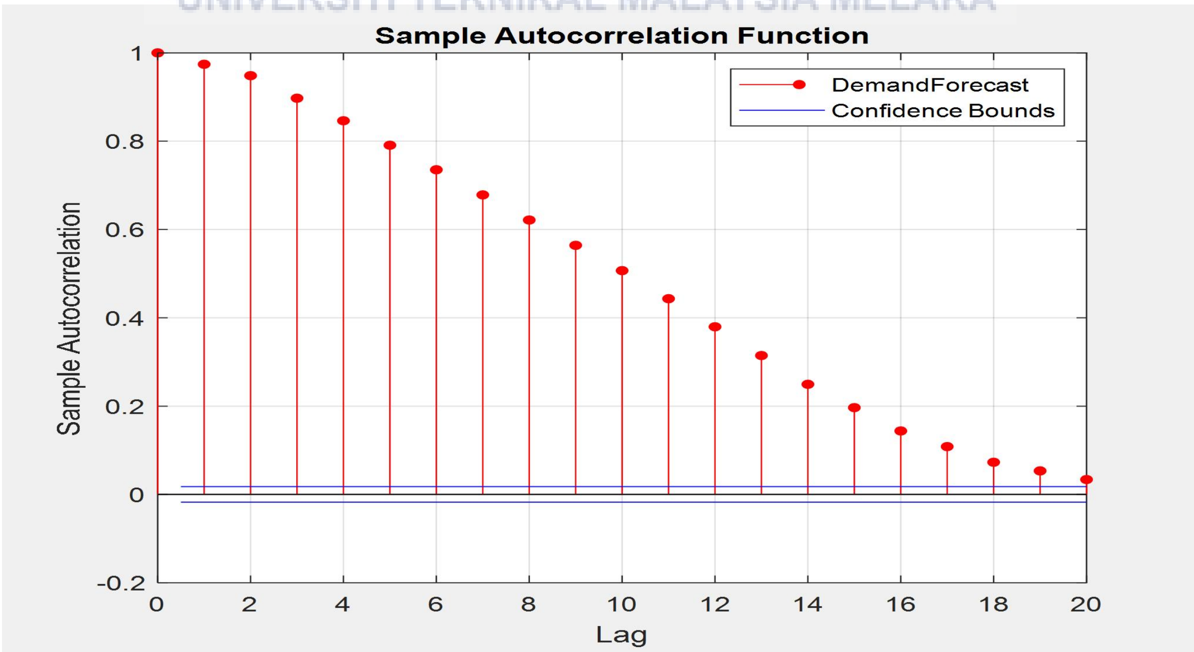

Figure 4.17 : ACF plot for Demand forecast

Figure 4.17 shows an ACF graph for the demand forecast, illustrating the relationship between the current observation and its past values. The horizontal axis of the ACF graph represents the lag, indicating the number of time intervals between the current observation and its lagged values. On the vertical axis, the ACF graph displays the correlation coefficient, which varies between 0 and 1. This coefficient quantifies the strength and direction of the linear association between two variables. Generally, the ACF graph presents the correlation coefficient on the y-axis and the lag on the x-axis.

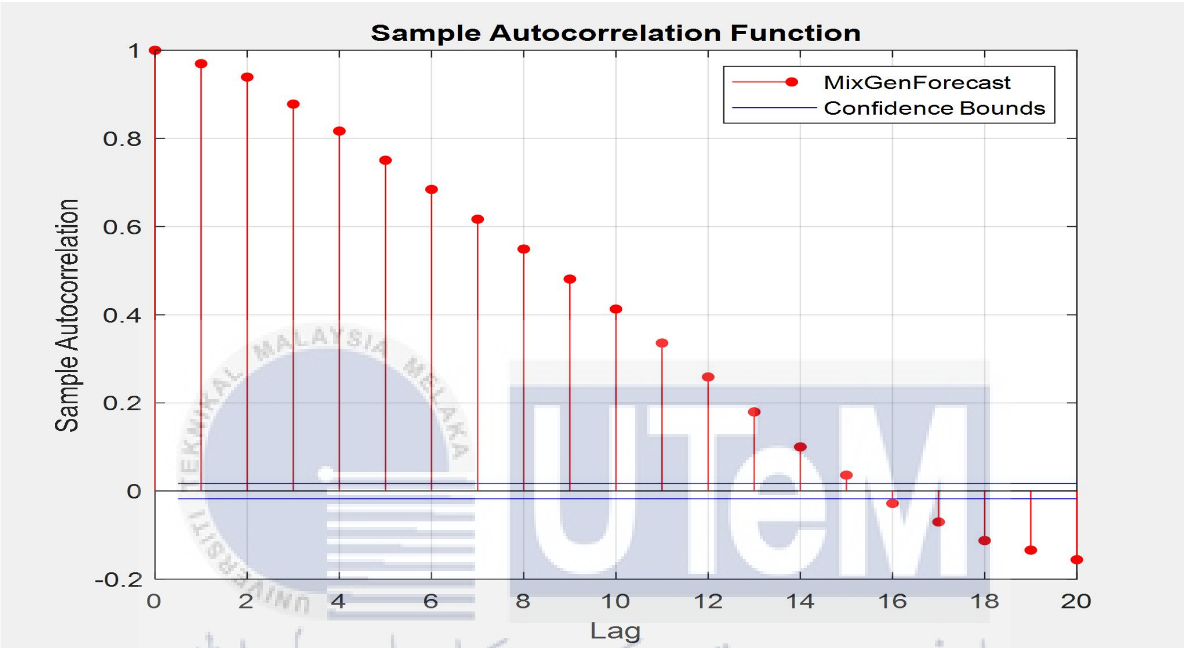

Figure 4.18 : ACF plot for Mix Generation forecast

In Figure 4.18, there is an ACF graph that visualizes the observation values for the mix generation forecast. The ACF graph employs the horizontal axis to represent the lag, denoting the number of time intervals between the current observation and its lagged values. The vertical axis of the ACF graph demonstrates the correlation coefficient, which ranges from -0.2 to 1. This coefficient serves as a measure of the strength and direction of the linear relationship between two variables. Typically, the ACF graph presents the correlation coefficient on the y-axis while plotting the lag on the x-axis.

4.6 Partial Autocorrelation Function (PACF)

Partial Autocorrelation Function, PACF graph is used to analyse the partial correlation between observations in a time series dataset. The PACF graph displays the partial correlation coefficients between the current observation and its lagged values, accounting for the influence of intermediate lags. Figure 4.19 to Figure 4.22 show the ACF plot for all the type inputs.

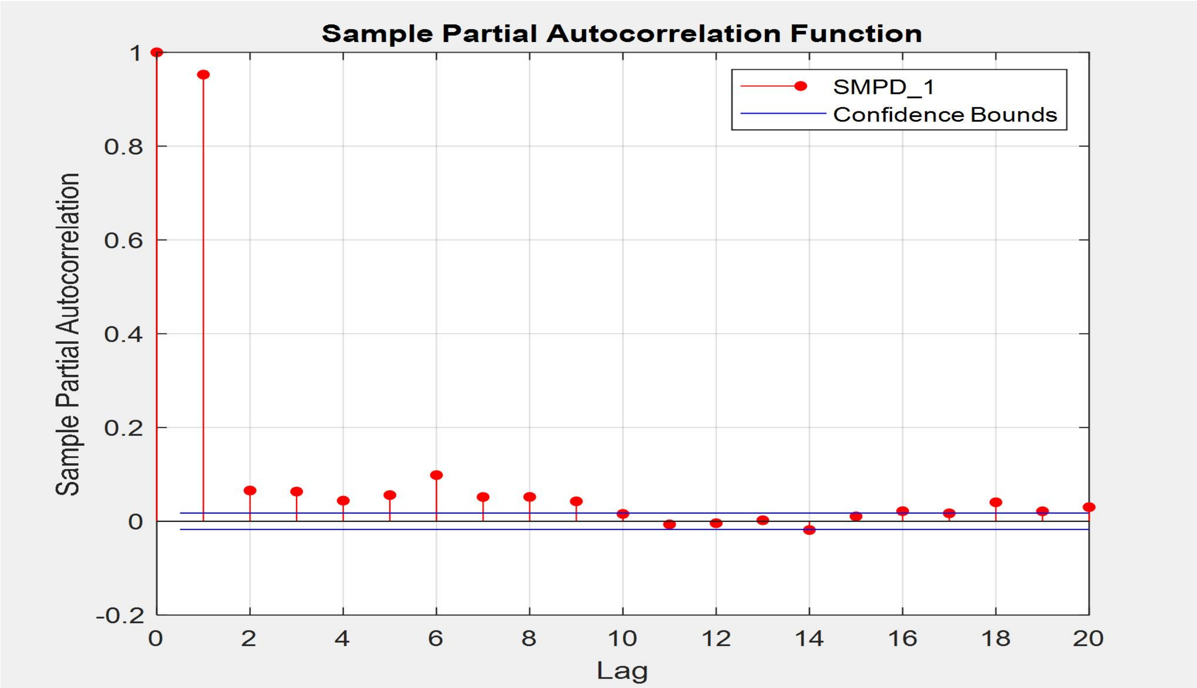

Figure 4.19 : PACF plot for SMP_{d-1}

Figure 4.19 shows PACF graph for SMP_{d-1}, it indicates the partial correlation between the current observation and its lagged values, accounting for the influence of intermediate lags. The x-axis of the PACF graph represents the lag, similar to the ACF graph. It refers to the number of time intervals between the current observation and its lagged values. The y-axis of the PACF graph represents the partial correlation coefficient, which also ranges from -0.2 to 1. The partial correlation coefficient measures the strength and direction of the linear relationship between two variables while controlling for the effects of other variables.

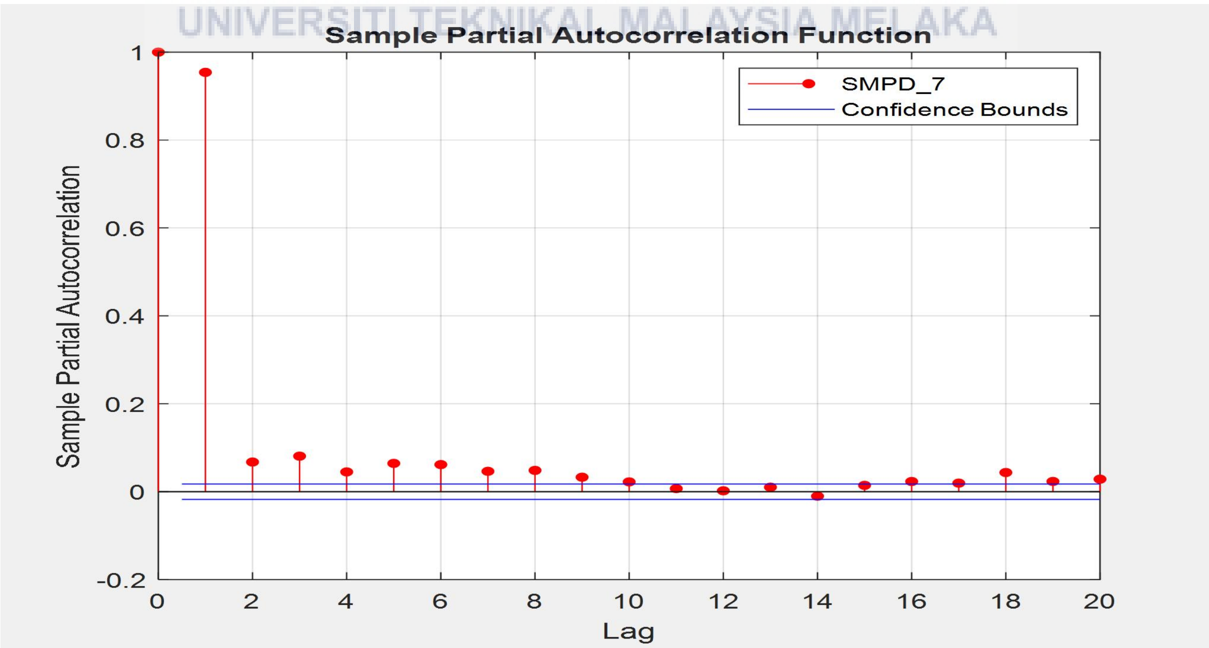

Figure 4.20 : PACF plot for SMP_{d-7}

In Figure 4.20, the PACF graph for SMP_{d-7} illustrates the partial correlation between the current observation and its past values, considering the influence of intermediate time lags. The x-axis represents the lag, indicating the number of time intervals between the present observation and its previous values, similar to the ACF graph. On the other hand, the y-axis displays the partial correlation coefficient, which ranges from -0.2 to 1. This coefficient quantifies the strength and direction of the linear relationship between two variables while taking into account the impact of other variables.

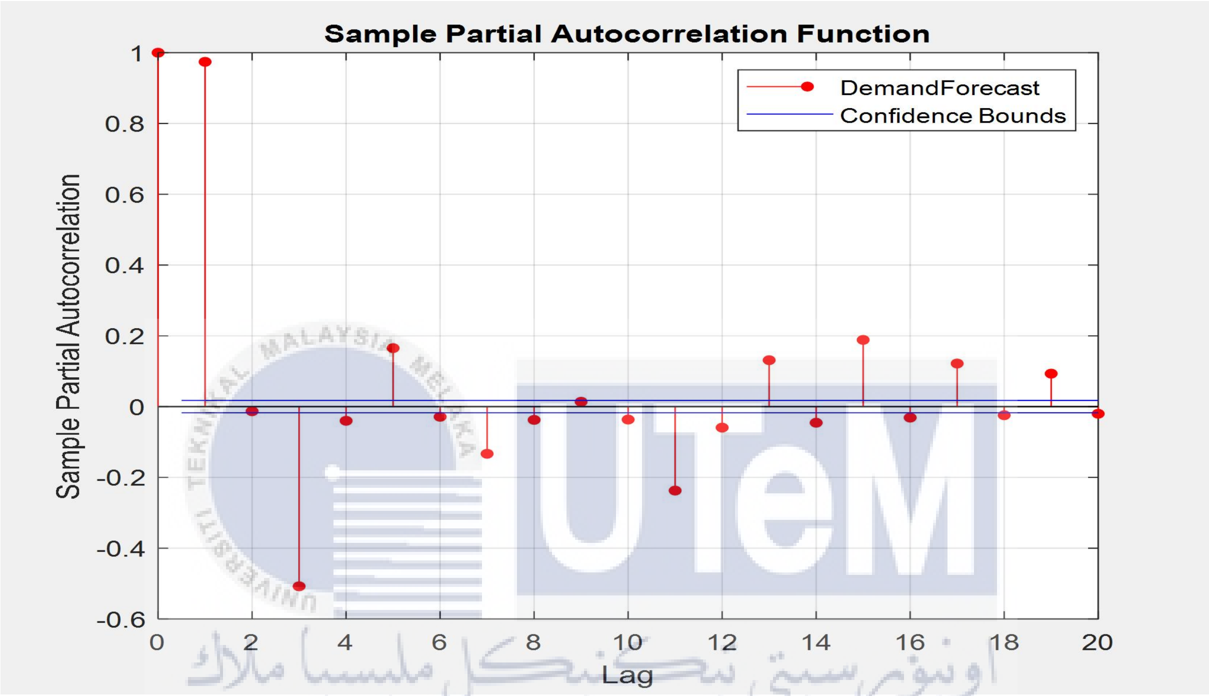

Figure 4.21 : PACF plot for Demand Forecast

The PACF graph for demand forecast in Figure 4.21 shows the partial correlation between the current observation and its previous values, while considering the influence of intermediate time lags. The x-axis represents the lag, indicating the number of time intervals between the current observation and its past values, similar to the ACF graph. Meanwhile, the y-axis illustrates the partial correlation coefficient, which falls within the range of -0.6 to 1. This coefficient measures the strength and direction of the linear relationship between two variables, while also accounting for the influence of other variables.

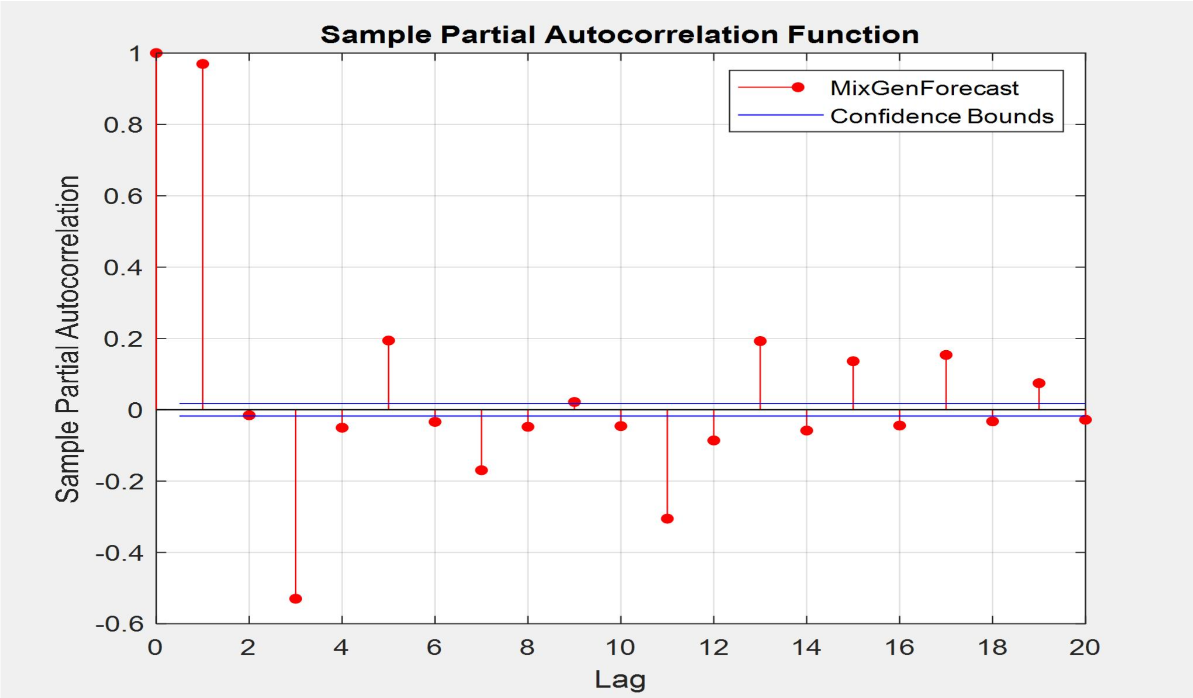

Figure 4.22 : PACF plot for Mix Generation Forecast

Figure 4.22 PACF graph, it indicates the partial correlation between the current observation and its lagged values, accounting for the influence of intermediate lags. The x-axis of the PACF graph represents the lag, similar to the ACF graph. It refers to the number of time intervals between the current observation and its lagged values. The yaxis of the PACF graph represents the partial correlation coefficient, which also ranges from -0.6 to 1. The partial correlation coefficient measures the strength and direction of the linear relationship between two variables while controlling for the effects of other variables.

4.7 Time Series Model Development

The ARIMA model assessed by using p-values to determine the statistical significance of the autoregressive (AR) and moving average (MA) terms. A p-value represents the probability of observing a test statistic as extreme as the one calculated, assuming the null hypothesis is true. In this case, the null hypothesis assumes that the data is suitable for the ARIMA model. Given the p-values for ARIMA model for (1,0,1). The training datasets from 1^{st} February 2021 until 31^{st} October 2021 are used in this model development.

Table 4.2: p-values for ARIMA model

	Autoregression 1 (AR1)	Moving average (MA 1)
SMP_{d-1}	0	1.5484e-50
SMP_{d-7}	0	0
Demand Forecast	0	0.9454
Mix Generation Forecast	0	0.93311

Based on conventional significance levels (such as α = 0.05), these p-values for SMP_{d-1} and SMP_{d-7} are smaller than the significance level. When a p-value is below a significance level (e.g., α = 0.05), it indicates strong evidence against the null hypothesis. Since the p-values for all the AR1 term and the moving average term are 0.00, we reject the null hypothesis. It suggests that there is sufficient statistical evidence to conclude that these terms are significant in the ARIMA model. These terms play an important role in explaining the pattern or relationship in the time series data. In summary, based on the given p-values, we reject the null hypothesis. Therefore, we accept the alternative hypothesis, indicating that the data for SMP_{d-1} and SMP_{d-7} is suitable for the ARIMA model. The significant autoregressive and moving average terms suggest that the ARIMA model can effectively capture the patterns and relationships present in the time series data.

4.8 Forecasting of ARIMA model

The time series data for all solar panel were divided into training and testing data. The training data spanned from 1^{st} February 2021 until 31^{st} October 2021 and the testing datasets is from 1^{st} Feb 2022 until 24^{th} April 2022. An ARIMA(1,0,1) model was selected and fitted using the training set. Figure 4.23, Figure 4.24, Figure 4.25 and Figure 4.26 plotted the forecast ARIMA model.

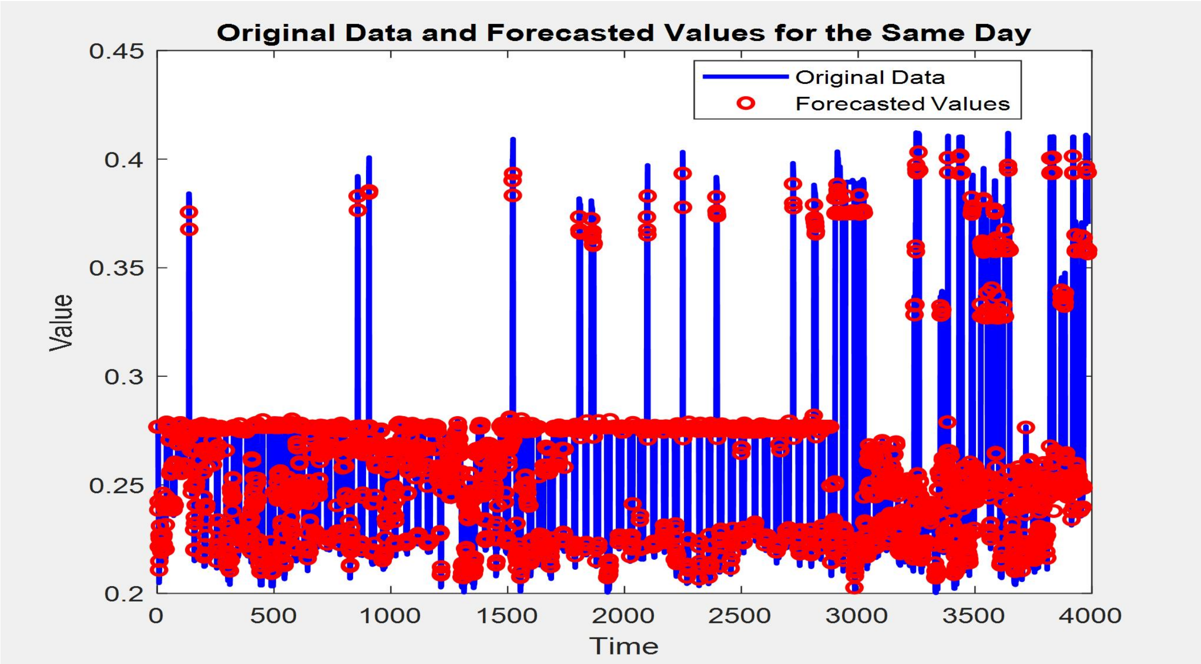

Figure 4.23: Forecast data trend of SMP_{d-1}

Figure 4.23 shows the forecasts is generated by the testing data from 1^{st} Feb 2022 until 24^{th} April 2022. The ARIMA(1,0,1) model produced p-value for AR(1) is 0 and MA(1) is 1.5484e-50 which is smaller than 0.05 tell that the SMP_{d-1} input can be use for forecasting. The accuracy of the ARIMA model were measured by two error metrics. The MAE was determined to be 0.0221, while the MAPE was found to be 8.21% for the SMP_{d-1} data. These error metrics provide insights into the accuracy of the ARIMA(1,0,1) model in predicting the system marginal price data.

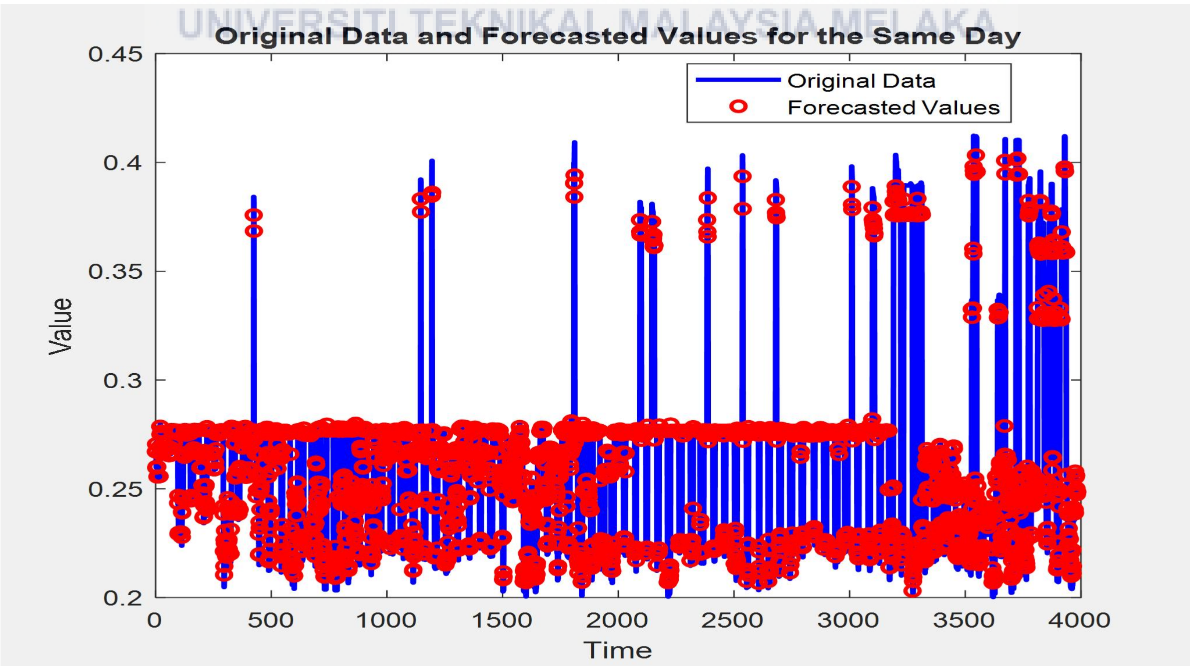

Figure 4.24: Forecast data trend of SMP_{d-7}

Figure 4.24 shows the forecasts is generated by the testing data from 1^{st} Feb 2022 until 24^{th}April 2022. The ARIMA(1,0,1) model produced p-value for AR(1) is 0 and MA(1) is 0 which is smaller than 0.05 tell that the SMP_{d-7} input can be use for forecasting. The accuracy of the ARIMA model were measured by two error metrics. The MAE was determined to be 0.0259, while the MAPE was found to be 9.44% for the SMP_{d-7} data. These error metrics provide insights into the accuracy of the ARIMA(1,0,1) model in predicting the system marginal price data.

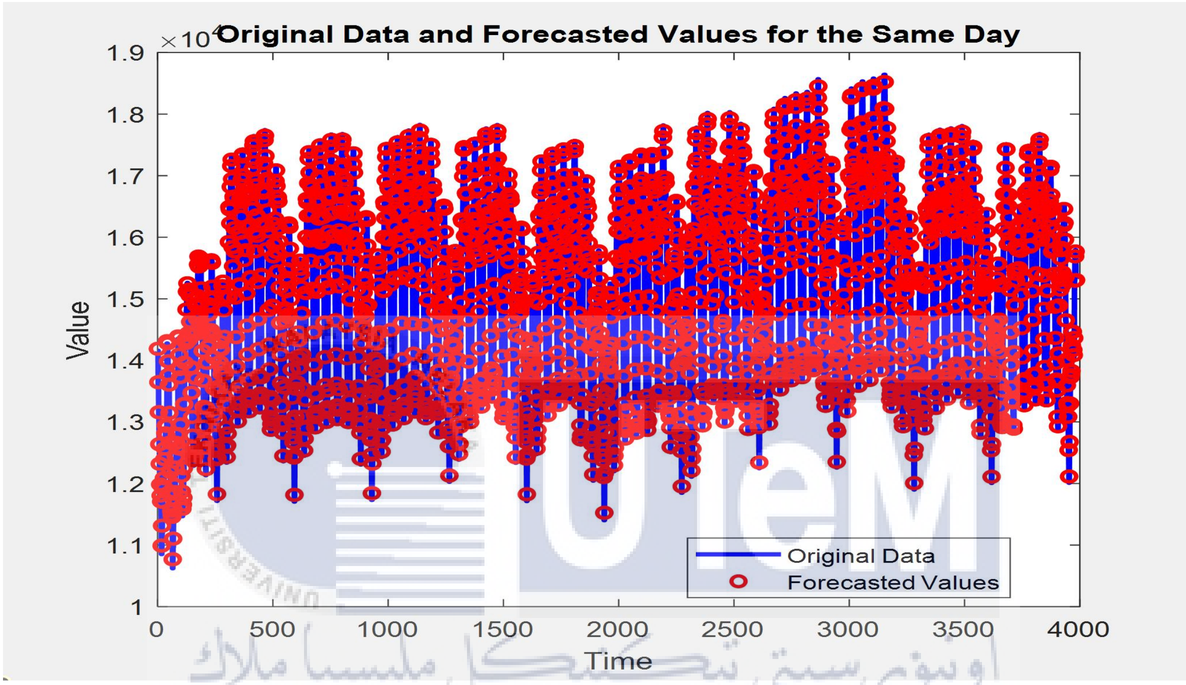

Figure 4.25: Forecast data trend of Demand Forecast

Figure 4.25 shows the forecasts is generated by the testing data from 1^{st} Feb 2022 until 24^{th}April 2022. The ARIMA(1,0,1) model produced p-value for AR(1) is 0 and MA(1) is 0.9454 which is greater than 0.05 tell that the demand forecast input cannot be use for forecasting. The accuracy of the ARIMA model were measured by two error metrics. The MAE was determined to be 1.2443, while the MAPE was found to be 48.07% for the demand forecast data. These error metrics provide insights into the accuracy of the ARIMA(1,0,1) model in predicting the system marginal price data.

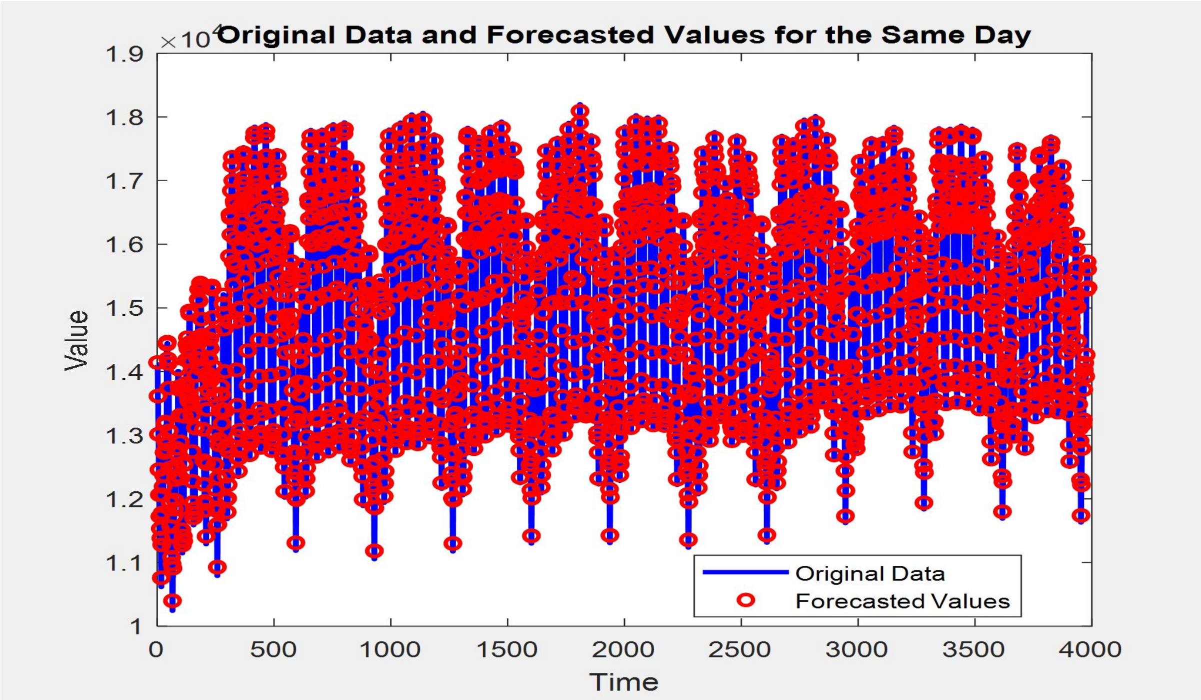

Figure 4.26: Forecast data trend of Mix Generation Forecast

Figure 4.26 shows the forecasts is generated by the testing data from 1^{st} Feb 2022 until 24^{th}April 2022. The ARIMA(1,0,1) model produced p-value for AR(1) is 0 and MA(1) is 0.93311 which is greater than 0.05 tell that the mix generation forecast input cannot be use for forecasting. The accuracy of the ARIMA model were measured by two error metrics. The MAE was determined to be 1.2329, while the MAPE was found to be 47.61% for the mix generation forecast data. These error metrics provide insights into the accuracy of the ARIMA(1,0,1) model in predicting the system marginal price data.

From the result, ARIMA model produces better accuracy for SMP_{d-1} and SMP_{d-7} same as the regression analysis model result. From both result shows that demand forecast and mix generation forecast input cannot be used for forecasting future value as the MAE and MAPE produces was not in the range and not suitable for forecasting SMP in the future.

4.9 The Overall Performance Of the Time Series Technique

Table 4.3 shows the comparison between regression and ARIMA model error measurement with the other method used to forecast the system marginal price in Malaysian electricity market.

Table 4.3 : Model Error Comparison

Testing Datasets	Type Of Model	MAE	MAPE
February-April 2022	ARIMA	0.0221	8.21%
	Regression	0.0298	10.70%
	LSSVM [41]	0.0259	9.28%
	LSSVM-GA [41]	0.0339	12.06%
	Single Buyer [40]	0.0351	13.31%
February 2022 (SMP_{d-1})	ARIMA	0.0156	6.27%
	Regression	0.0165	6.69%
	LSSVM-GA [3]	0.0180	7.09%
	Single Buyer [40]	0.0254	10.63%

From table 4.3, the findings demonstrate that the time series technique produces better accuracy compared to the Single Buyer forecast. From February 2022 until April 2022, the MAE and MAPE for ARIMA model is 0.0221 and 8.21% which is better than the SB forecast where the MAE and MAPE is 0.0351 and 13.31%. Same goes to the regression model where the MAE and MAPE is smaller than the SB forecast which is 0.0298 and 10.70%.

In another comparison, the time series technique was compared with the artificial intelligence technique which is LSSVM and LSSVM-GA. The ARIMA model produces better accuracy in this comparison which is smaller than the LSSVM and LSSVM-GA but for the regression model, the accuracy is bigger than the LSSVM but smaller than the LSSVM-GA. Where the MAE and MAPE for regression is 0.0298 and 10.70% bigger than the LSSVM which is the MAE and MAPE is 0.0259 and 9.28%.

For February 2022 with one input used which is SMP_{d-1}, the time series technique also showed better accuracy compared to the LSSVM-GA method and SB where the MAE and MAPE for ARIMA model is 0.0156 and 6.27% while for regression model is 0.0165 and 6.69%. For the LSSVM-GA model the MAE and MAPE is 0.0180 and 7.09% better than the SB forecast where the MAE and MAPE produces is 0.0254 and 10.63%. It shows that for forecasting using one input, the time series technique is the best model for forecasting.

In conclusion, from the result obtain the time series technique for ARIMA model produces better accuracy compared to other forecasting method such as artificial intelligence. The reason for this improved performance is that ARIMA has a strong framework that helps it identify and identify patterns and trends in the data. ARIMA offers a thorough method of forecasting by taking into account the autoregressive and moving average components as well as differencing to guarantee stationarity. As a result, analysts and researchers across a range of disciplines favour it as a more accurate and dependable tool for forecasting future values in a time series.

CHAPTER 5
CONCLUSION AND RECOMMENDATIONS

5.1 Summary

In summary, the determination and comprehension of System Marginal Price (SMP) emerge as critical challenges. To improve forecasting accuracy, it is imperative to comprehend and measure the impact of mix generation and demand on SMP. This necessitates a careful examination of the relationships between the variables and the accuracy of SMP forecasting. Sophisticated time series forecasting techniques must be applied to develop accurate models capable of predicting SMP trends so that decision-makers in the dynamic energy industry can make more informed choices. Regression analysis and the ARIMA model were utilised in this study to forecast future values, and the model's effectiveness was evaluated using the MAPE and MAE metrics. Due to its ability to recognise underlying patterns and adapt to changing market conditions, it is a helpful tool for participants in the electricity market, empowering them to make well-informed decisions about power generation, trading, and demand response. Furthermore, the model facilitates the integration of renewable energy sources through the power grid system, optimises the supply-demand balance for electricity through accurate SMP forecasts, and facilitates the efficient allocation of resources.

5.2 Future Work

More advanced modelling techniques, like examining different approaches and including more relevant features like meteorological data, can be applied in future research that uses time series analysis to forecast the marginal price in the Malaysian electricity market. Because the electricity market is dynamic, it is necessary to take into account the quantification of uncertainty, integrate it with external factors, and develop a real-time forecasting system. In order to guarantee the model's relevance and usefulness to decision-makers in the energy sector, user feedback and stakeholder involvement should also direct the model's iterative development.

REFERENCES

[1] Y. Soo Heong, "70 year chequered history in elecricity supply," 2023 3rd International Conference on Computing and Information Technology, ICCIT 2023. [Online]. Available: https://www.bernama.com/en/business/news.php?id=1763105

[2] M. Sibeperegasam, V. K. Ramachandaramurthy, S. Walker, and J. Kanesan,"Malaysia'selectricity market structure in transition," *Util. Policy*, vol. 72, no. September 2020, 2021, doi: 10.1016/j.jup.2021.101266.

[3] I. A. W. A. Razak, W. S. W. Abdullah, and M. F. Sulaima, "Enhanced Short-term System Marginal Price (SMP) Forecast Modelling Using a Hybrid Model Combining Least Squares Support Vector Machines and the Genetic Algorithm in Peninsula Malaysia," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 4, pp. 289–298, 2023.

[4] F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espínola, "Forecasting next-day electricity price by time series models," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 342–348, 2002, doi: 10.1109/TPWRS.2002.1007902.

[5] C. M. Castilho *et al.*, "Time Series Forecasting with exogenous factors: Statistical vs. Machine Learning approaches Dissertation Master in Modeling, Data Analysis and Decision Support System," 2020.

[6] "The Advantages Of Time Series Analysis." Accessed: Nov. 27, 2023. [Online]. Available: https://www.sigmacomputing.com/resources/learn/what-is-time-series-analysis#:~:text=Its strengths include%3A,data more reliable and interpretable.&text=Understanding Data%3A Models like ARIMA,into the data's underlying structure.

[7] R. Weron, *Modeling and forecasting electricity loads and prices: A statistical approach*. 2006. doi: 10.1002/9781118673362.

[8] S. Tenaga, "Overview Energy Comission." Accessed: Nov. 20, 2023. [Online]. Available: https://www.st.gov.my/en/details/aboutus/1

[9] STM, "Malaysia Energy Statistics Handbook," Suruhanjaya Tenaga Malaysia, vol. 1, p. 86, 2021, [Online]. Available: www.st.gov.my

[10] "Analyzing the construct of Stakeholder." Accessed: Nov. 20, 2023. [Online]. Available: https://knepublishing.com/index.php/Kne-Social/article/view/5090

[11] H. Mengi-Dinçer, V. Ediger, and G. Yesevi, “Evaluating the International Renewable Energy Agency through the lens of social constructivism,” *Renew. Sustain. Energy Rev.*, vol. 152, no. May, p. 111705, 2021, doi: 10.1016/j.rser.2021.111705.

[12] “What is System Marginal Price?” Accessed: Nov. 20, 2023. [Online]. Available: https://www.lnrg.technology/2023/03/31/what-is-system-marginal-price-smp/

[13] A. Hernandez-Matheus *et al.*, “A systematic review of machine learning techniques related to local energy communities,” *Renew. Sustain. Energy Rev.*, vol. 170, no. May, p. 112651, 2022, doi: 10.1016/j.rser.2022.112651.

[14] J. K. Lee, J. B. Park, J. R. Shin, and K. Y. Lee, *A system marginal price forecasing method based on an artificial neural network using time and day information*, vol. 16, no. 1. IFAC, 2005. doi: 10.3182/20050703-6-cz-1902.02256.

[15] A. Yadav, A. Sahay, M. R. Yadav, S. Bhandari, A. Yadav, and K. B. Sahay, “One hour Ahead Short-Term Electricity Price Forecasting Using ANN Algorithms,” in *Proceedings of the Conference on the Industrial and Commercial Use of Energy, ICUE*, 2019, pp. 24–27. doi: 10.23919/ICUE-GESD.2018.8635596

[16] G. Yang, S. Du, Q. Duan, and J. Su, “Multi-day Load Forecasting Method in Electricity Spot Markets Based on Multiple LSTMs,” *2021 3rd Asia Energy Electr. Eng. Symp. AEEES 2021*, pp. 708–712, 2021, doi: 10.1109/AEEES51875.2021.9403013.

[17] L. B. Sina, C. A. Secco, M. Blazevic, and K. Nazemi, “Hybrid Forecasting Methods—A Systematic Review,” *Electron.*, vol. 12, no. 9, 2023, doi: 10.3390/electronics12092019.

[18] Q. M. Ilyas, K. Iqbal, S. Ijaz, A. Mehmood, and S. Bhatia, “A Hybrid Model to Predict Stock Closing Price Using Novel Features and a Fully Modified Hodrick–Prescott Filter,” *Electron.*, vol. 11, no. 21, 2022, doi: 10.3390/electronics11213588.

[19] Z. Mustaffa, Y. Yusof, and S. S. Kamaruddin, “Gasoline Price Forecasting: An Application of LSSVM with Improved ABC,” *Procedia - Soc. Behav. Sci.*, vol. 129, pp. 601–609, 2014, doi: 10.1016/j.sbspro.2014.03.718.

[20] A. M. M. AL-Qaysi, A. Bozkurt, and Y. Ates, “Load Forecasting Based on Genetic Algorithm–Artificial Neural Network-Adaptive Neuro-Fuzzy Inference Systems: A Case Study in Iraq,” Energies (Basel), vol. 16, no. 6, Mar. 2023, doi: 10.3390/en16062919.

[21] A. Ciarreta, B. Martinez, and S. Nasirov, "Forecasting electricity prices using bid data," *Int. J. Forecast.*, vol. 39, no. 3, pp. 1253–1271, 2023, doi: 10.1016/j.ijforecast.2022.05.011.

[22] K. Rawal and A. Ahmad, "Day-Ahead Market Electricity Price Prediction using Time Series Forecasting," pp. 1–6, 2023, doi: 10.1109/stpes54845.2022.10006455.

[23] A. Peshkov and O. Alsova, "Development and Research of Multifactor Ensemble Models for Forecasting the Time Series of Electricity Prices," *Int. Conf. Young Spec. Micro/Nanotechnologies Electron Devices, EDM*, vol. 2021-June, pp. 489–492, 2021, doi: 10.1109/EDM52169.2021.9507699.

[24] M. Valipour, M. E. Banihabib, and S. M. R. Behbahani, "Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir," *J. Hydrol.*, vol. 476, pp. 433–441, 2013, doi: 10.1016/j.jhydrol.2012.11.017.

[25] M. Pourahmadi, "A Course in Time Series Analysis," *Am. Stat.*, vol. 56, no. 1, pp. 77–77, 2002, doi: 10.1198/tas.2002.s131

[26] J. B. Marín, E. T. Orozco, and E. Velilla, "Forecasting electricity price in Colombia: A comparison between neural network, ARMA process and hybrid models," *Int. J. Energy Econ. Policy*, vol. 8, no. 3, pp. 97–106, 2018

[27] T. Jakaša, I. Andročec, and P. Sprčić, "Electricity price forecasting ARIMA model approach," *2011 8th Int. Conf. Eur. Energy Mark. EEM 11*, no. July 2015, pp. 222–225, 2011, doi: 10.1109/EEM.2011.5953012.

[28] Y. Huang and W. Liu, "Regression Analysis Model Based on Data Processing and MATLAB Numerical Simulation," *2022 IEEE Asia-Pacific Conf. Image Process. Electron. Comput. IPEC 2022*, pp. 1115–1118, 2022, doi: 10.1109/IPEC54454.2022.9777319.

[29] E. W. Data and O. S. From, "M Ulti - Level R Egression : M Arginal M Odels," pp. 31–55, 2018.

[30] T. Tumiran, S. Sarjiya, L. M. Putranto, E. N. Putra, R. F. S. Budi, and C. F. Nugraha, "Long-Term Electricity Demand Forecast Using Multivariate Regression and End-Use Method: A Study Case of Maluku-Papua Electricity System," *ICT-PEP 2021 - Int. Conf. Technol. Policy Energy Electr. Power Emerg. Energy Sustain. Smart Grid, Microgrid*

Technol. Futur. Power Syst. Proc., pp. 258–263, 2021, doi: 10.1109/ICT-PEP53949.2021.9601144.

[31] T. Ulgen and G. Poyrazoglu, "Predictor analysis for electricity price forecasting by multiple linear regression," *2020 Int. Symp. Power Electron. Electr. Drives, Autom. Motion, SPEEDAM 2020*, pp. 618–622, 2020, doi: 10.1109/SPEEDAM48782.2020.9161866

[32] J. P. González, A. M. S. Roque, and E. A. Pérez, "Forecasting functional time series with a new hilbertian ARMAX model: Application to electricity price forecasting," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 545–556, 2018, doi: 10.1109/TPWRS.2017.2700287.

[33] N. Ziha, Z. Abidin, and J. B. Ibrahim, "Fp_a.6_Roles of Single Buyer in Supporting Renewables- the Malaysia'S Experience"

[34] "Correlation Coefficients." Accessed: Nov. 27, 2023. [Online]. Available: https://latrobe.libguides.com/ibmspss/correlation

[35] M. Flex, "What is Correlation Analysis." Accessed: Nov. 27, 2023. [Online]. Available: https://blog.flexmr.net/correlation-analysis-definition-exploration.

[36] David Abugaber, "ARIMA History." Accessed: Nov. 28, 2023. [Online]. Available: https://ademos.people.uic.edu/Chapter23.html

[37] machine learning Pills, "ARIMA parameters selection." Accessed: Nov. 28, 2023. [Online]. Available: https://mlpills.dev/time-series/parameters-selection-in-arima-models/

[38] Konstantin Rink, "Time Series Error Mesurement." Accessed: Nov. 28, 2023. [Online]. Available: https://towardsdatascience.com/time-series-forecast-error-metrics-you-should-know-cc88b8c67f27

[39] Y. Zakai, "Mean Absolute Error(MAE)." Accessed: Nov. 28, 2023. [Online]. Available: https://deepchecks.com/glossary/mean-absolute-error/

[40] "SINGLE BUYER." Accessed: Nov. 20, 2023. [Online]. Available: https://www.singlebuyer.com.my/

[41] A. Marzuki, "FORECASTING SYSTEM MARGINAL PRICE (SMP) FOR MALAYS," Univ. Tek. Malaysia Melaka, 2023

APPENDIX A

Code For ARIMA Model Forecasting

```
% Load or generate time series data (replace this with your data)
y = xlsread("TestSMPd-7.xlsx");

% Create or estimate ARIMA model
Mdl = arima(1, 0, 1); % ARIMA(1,0,1) model

% Fit ARIMA model
EstMdl = estimate(Mdl, y);

% Initialize array to store forecasted values
yF_same_day = zeros(size(y));

% Forecast for the same day value for each observation
for i = 1:length(y)
   [yF_same_day(i), ~] = forecast(EstMdl, 1, 'Y0', y(1:i));
end

% Display the forecasted values for the same day
fprintf('Forecasted values for the same day:\n');
disp(yF_same_day);

% Plot original data and forecasted values
figure;
plot(1:length(y), y, 'b', 'LineWidth', 2);
hold on;
plot(1:length(y), yF_same_day, 'ro', 'MarkerSize', 5, 'LineWidth', 1.5);
xlabel('Time');
ylabel('Value');
title('Original Data and Forecasted Values for the Same Day');
legend('Original Data', 'Forecasted Values', 'Location', 'Best');
```

Table A1 :Sample Data February SMP_{d-1} For One Day Regression Model

DATE	SMPd-1(31/1)	SMP	REGRESSION FORECAST	MAE	MAPE
2/1/2022 0:00	0.2785	0.2800	0.268706314	0.011293686	0.040334591
2/1/2022 0:30	0.2784	0.2800	0.268647951	0.011352049	0.040543031
2/1/2022 1:00	0.2785	0.2800	0.268706314	0.011293686	0.040334591
2/1/2022 1:30	0.2786	0.2801	0.268764678	0.011335322	0.040468841
2/1/2022 2:00	0.2375	0.2720	0.244777376	0.027222624	0.100083176
2/1/2022 2:30	0.2395	0.2720	0.245944641	0.026055359	0.09579176
2/1/2022 3:00	0.2174	0.2757	0.23304636	0.04265364	0.154710338
2/1/2022 3:30	0.2048	0.2551	0.225692589	0.029407411	0.115277975
2/1/2022 4:00	0.2085	0.2552	0.227852029	0.027347971	0.107162894
2/1/2022 4:30	0.2216	0.2550	0.235497617	0.019502383	0.076479933
2/1/2022 5:00	0.2206	0.2553	0.234913984	0.020386016	0.079851217
2/1/2022 5:30	0.2267	0.2552	0.238474144	0.016725856	0.06554019
2/1/2022 6:00	0.2214	0.2550	0.235380891	0.019619109	0.076937684
2/1/2022 6:30	0.2221	0.2553	0.235789433	0.019510567	0.076422118
2/1/2022 7:00	0.2214	0.2554	0.235380891	0.020019109	0.078383357
2/1/2022 7:30	0.2149	0.2550	0.231587278	0.023412722	0.091814595
2/1/2022 8:00	0.2141	0.2580	0.231120372	0.026879628	0.104184604
2/1/2022 8:30	0.2135	0.2364	0.230770193	0.005629807	0.023814752
2/1/2022 9:00	0.2204	0.2360	0.234797258	0.001202742	0.005096365
2/1/2022 9:30	0.2204	0.2394	0.234797258	0.004602742	0.019226158
2/1/2022 10:00	0.2147	0.2394	0.231470552	0.007929448	0.033122173
2/1/2022 10:30	0.2177	0.2357	0.23322145	0.00247855	0.010515699
2/1/2022 11:00	0.2177	0.2377	0.23322145	0.00447855	0.018841187
2/1/2022 11:30	0.2426	0.2377	0.247753903	0.010053903	0.042296603
2/1/2022 12:00	0.2394	0.2377	0.245886278	0.008186278	0.034439538
2/1/2022 12:30	0.2222	0.2377	0.235847797	0.001852203	0.007792189
2/1/2022 13:00	0.2224	0.2378	0.235964523	0.001835477	0.007718574
2/1/2022 13:30	0.2177	0.2589	0.23322145	0.02567855	0.099183276
2/1/2022 14:00	0.2374	0.2589	0.244719013	0.014180987	0.054773994
2/1/2022 14:30	0.2374	0.2589	0.244719013	0.014180987	0.054773994
2/1/2022 15:00	0.2374	0.2589	0.244719013	0.014180987	0.054773994
2/1/2022 15:30	0.2375	0.2589	0.244777376	0.014122624	0.054548567
2/1/2022 16:00	0.2358	0.2589	0.243785201	0.015114799	0.05838084
2/1/2022 16:30	0.2361	0.2589	0.24396029	0.01493971	0.057704556
2/1/2022 17:00	0.2386	0.2587	0.245419372	0.013280628	0.051336019
2/1/2022 17:30	0.2161	0.2589	0.232287638	0.026612362	0.102790122
2/1/2022 18:00	0.2270	0.2588	0.238649233	0.020150767	0.077862313
2/1/2022 18:30	0.2415	0.2585	0.247111907	0.011388093	0.04405452
2/1/2022 19:00	0.2359	0.2588	0.243843564	0.014956436	0.057791484
2/1/2022 19:30	0.2785	0.2694	0.268706314	0.000693686	0.002574928
2/1/2022 20:00	0.2785	0.2696	0.268706314	0.000893686	0.003314857
2/1/2022 20:30	0.2787	0.2696	0.268823041	0.000776959	0.002881895
2/1/2022 21:00	0.2785	0.2724	0.268706314	0.003693686	0.013559785
2/1/2022 21:30	0.2785	0.2724	0.268706314	0.003693686	0.013559785
2/1/2022 22:00	0.2786	0.2722	0.268764678	0.003435322	0.012620581
2/1/2022 22:30	0.2785	0.2723	0.268706314	0.003593686	0.013197523
2/1/2022 23:00	0.2786	0.2723	0.268764678	0.003535322	0.012983189
2/1/2022 23:30	0.2782	0.2723	0.268531225	0.003768775	0.013840526

Table A2 : Sample Data February SMP_{d-1} For One Day ARIMA Model

DATE	SMPd-1	SMP	ARIMA FORECAST	MAE	MAPE
2/1/2022 0:00	0.2785	0.2800	0.2748	0.0052	0.018571429
2/1/2022 0:30	0.2784	0.2800	0.2751	0.0049	0.0175
2/1/2022 1:00	0.2785	0.2800	0.2753	0.0047	0.016785714
2/1/2022 1:30	0.2786	0.2801	0.2754	0.0047	0.016779722
2/1/2022 2:00	0.2375	0.2720	0.2352	0.0368	0.135294118
2/1/2022 2:30	0.2395	0.2720	0.2423	0.0297	0.109191176
2/1/2022 3:00	0.2174	0.2757	0.2198	0.0559	0.20275662
2/1/2022 3:30	0.2048	0.2551	0.2103	0.0448	0.175617405
2/1/2022 4:00	0.2085	0.2552	0.2152	0.04	0.156739812
2/1/2022 4:30	0.2216	0.2550	0.2273	0.0277	0.108627451
2/1/2022 5:00	0.2206	0.2553	0.2248	0.0305	0.119467293
2/1/2022 5:30	0.2267	0.2552	0.2311	0.0241	0.094435737
2/1/2022 6:00	0.2214	0.2550	0.2251	0.0299	0.117254902
2/1/2022 6:30	0.2221	0.2553	0.2266	0.0287	0.112416765
2/1/2022 7:00	0.2214	0.2554	0.2257	0.0297	0.116288175
2/1/2022 7:30	0.2149	0.2550	0.2194	0.0356	0.139607843
2/1/2022 8:00	0.2141	0.2580	0.2195	0.0385	0.149224806
2/1/2022 8:30	0.2135	0.2364	0.2189	0.0175	0.074027073
2/1/2022 9:00	0.2204	0.2360	0.2257	0.0103	0.043644068
2/1/2022 9:30	0.2204	0.2394	0.2248	0.0146	0.060985798
2/1/2022 10:00	0.2147	0.2394	0.2194	0.02	0.083542189
2/1/2022 10:30	0.2177	0.2357	0.223	0.0127	0.053882053
2/1/2022 11:00	0.2177	0.2377	0.2225	0.0152	0.063946151
2/1/2022 11:30	0.2426	0.2377	0.2469	0.0092	0.038704249
2/1/2022 12:00	0.2394	0.2377	0.2407	0.003	0.012620951
2/1/2022 12:30	0.2222	0.2377	0.2247	0.013	0.054690787
2/1/2022 13:00	0.2224	0.2378	0.2269	0.0109	0.045836838
2/1/2022 13:30	0.2177	0.2589	0.222	0.0369	0.142526072
2/1/2022 14:00	0.2374	0.2589	0.2419	0.017	0.065662418
2/1/2022 14:30	0.2374	0.2589	0.2394	0.0195	0.075318656
2/1/2022 15:00	0.2374	0.2589	0.2397	0.0192	0.074159907
2/1/2022 15:30	0.2375	0.2589	0.2397	0.0192	0.074159907
2/1/2022 16:00	0.2358	0.2589	0.2381	0.0208	0.0803399
2/1/2022 16:30	0.2361	0.2589	0.2386	0.0203	0.078408652
2/1/2022 17:00	0.2386	0.2587	0.241	0.0177	0.068419018
2/1/2022 17:30	0.2161	0.2589	0.2187	0.0402	0.155272306
2/1/2022 18:00	0.2270	0.2588	0.2322	0.0266	0.102782071
2/1/2022 18:30	0.2415	0.2585	0.2446	0.0139	0.05377176
2/1/2022 19:00	0.2359	0.2588	0.2375	0.0213	0.082302937
2/1/2022 19:30	0.2785	0.2694	0.2801	0.0107	0.039717892
2/1/2022 20:00	0.2785	0.2696	0.2746	0.005	0.018545994
2/1/2022 20:30	0.2787	0.2696	0.2755	0.0059	0.021884273
2/1/2022 21:00	0.2785	0.2724	0.2752	0.0028	0.010279001
2/1/2022 21:30	0.2785	0.2724	0.2753	0.0029	0.010646109
2/1/2022 22:00	0.2786	0.2722	0.2754	0.0032	0.011756062
2/1/2022 22:30	0.2785	0.2723	0.2752	0.0029	0.010650018
2/1/2022 23:00	0.2786	0.2723	0.2754	0.0031	0.011384502
2/1/2022 23:30	0.2782	0.2723	0.275	0.0027	0.009915534