

## การพยากรณ์ดัชนีพลังงาน MSCI ด้วยแบบจำลอง SARIMA

วิชญ์เดช นันไชยแก้ว<sup>1</sup>

Received: April 7, 2022

Revised: September 9, 2022

Accepted: September 20, 2022

### บทคัดย่อ

พลังงานถือได้ว่ามีบทบาทอย่างมากในระบบเศรษฐกิจไม่ว่าจะเป็นด้านการบริโภคและการผลิต โดยในปัจจุบันราคาพลังงานมีการเปลี่ยนแปลงสูงมีลักษณะเคลื่อนไหวตามวัฏจักรธุรกิจ เนื่องจากปริมาณความต้องการใช้พลังงานในแต่ละช่วงเวลาไม่เท่ากัน การคาดการณ์ดัชนีพลังงานจึงมีส่วนช่วยทำให้นักลงทุนสามารถวางแผนได้อย่างเหมาะสม โมเดลที่ใช้ในงานวิจัยครั้งนี้ คือ Seasonal Autoregressive Integrated Moving Average model (SARIMA) ซึ่งเป็นโมเดลที่เพิ่มผลกระทบจากฤดูกาล เป็นตัวแบบที่พัฒนามาจาก ARIMA (p, d, q) ของ Box and Jenkins โดยวัตถุประสงค์งานวิจัยในครั้งนี้ คือ 1) เพื่อสร้างแบบจำลองที่เหมาะสมกับข้อมูลดัชนีพลังงานโลก (MSCI World Energy Index) โดยแบบจำลอง SARIMA (Seasonal Autoregressive Integrated Moving Average) 2) เพื่อเปรียบเทียบความถูกต้องของการพยากรณ์ดัชนีพลังงาน MSCI World Energy Index ด้วย SARIMA Model ข้อมูลที่ใช้ในการวิจัยในครั้งนี้ คือ ข้อมูลรายเดือนของดัชนีพลังงานโลก (MSCI World Energy Index) ตั้งแต่ ปี 2548 ถึงปี 2562 (15 ปี) ในขั้นตอนของการวิจัยประกอบไปด้วย การทดสอบความนิ่งข้อมูลด้วยวิธี unit root test การสร้างแบบจำลอง SARIMA การเลือกแบบจำลองที่เหมาะสม และสุดท้ายเป็นการพยากรณ์ข้อมูลตามแบบจำลองที่ดีที่สุด ผลการวิจัย พบว่า ตัวแบบที่ดีที่สุดในการพยากรณ์ คือ SARIMA (2,1,1)<sub>12</sub>×(2,1,3)<sub>12</sub> และเมื่อนำไปผลที่ได้ไปพยากรณ์จะได้ค่าที่ใกล้เคียงกับค่าจริง ซึ่งมีค่าความคลาดเคลื่อน RMSE เท่ากับ 16.68 และ MAE เท่ากับ 12.39 ตามลำดับ

**คำสำคัญ:** การพยากรณ์, ดัชนี, แบบจำลอง, พลังงาน, ฤดูกาล

<sup>1</sup>อาจารย์ประจำโปรแกรมวิชาการเงิน คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏกำแพงเพชร (ผู้รับผิดชอบบทความ, Email: w.nunchaikaew@gmail.com)

## Forecasting MSCI World Energy Sector Index with the SARIMA Model

Wissanudeth Nunchaikaew<sup>1</sup>

Received: April 7, 2022

Revised: September 9, 2022

Accepted: September 20, 2022

### Abstract

Energy plays a crucial role in economic systems in terms of consumption and production. At the present, there is a high fluctuation of energy prices due to the business cycle movement and the differences in business energy consumption demanding in each cycle; therefore, the energy index prediction could help investor plans appropriately. The model used in this study is Seasonal Autoregressive Integrated Moving Average model (SARIMA). This is a model increasing seasonal effects which was developed from ARIMA (p, d, q) of Box and Jenkins. The purposes of this study are 1) to construct a suitable model for MSCI World Energy Index by using SARIMA (Seasonal Autoregressive Integrated Moving Average), and 2) to compare Forecast Accuracy of MSCI World Energy Sector Index via SARIMA Model. The data in this study is a monthly information from the MSCI World Energy Index from 2005 to 2019 (15 years). In the research methodology, there is a data stationary tested by using the unit root test, and simulating SARIMA model. After selecting the most appropriate model, the data prediction test was operated. In conclusion, the result of this study revealed that the most appropriate model for prediction was SARIMA (2,1,1)×(2,1,3)<sub>12</sub>. The prediction model outcome was very close to the real indices, when the deviation of RMSE was 16.68 and MAE was 12.39 respectively.

**Keywords:** Energy, Forecasting, Index, Model, Seasonal

---

<sup>1</sup>Lecturer of the Finance Program, Faculty of Management Science, Kamphaeng Phet Rajabhat University  
(Corresponding author, Email: w.nunchaikaew@gmail.com)

## Introduction

Energy is considered as an important factor in human life and plays a crucial role in the economic system in terms of production sector and consumption sector. It is like a vein in people's lives driving from the grassroots to the national economy. It is more significant when the world has been developed the energy production has been changed as it needs more technology in production process. There are various kinds of energy sources created from human and natural. Energy sources that human currently use are from fossil, biomass, water, wind, geothermal energy, solar cell, electricity, and nuclear power.

At present, The International Energy Agency (IEA) revealed the number of world energy situations that in 2020 the need in energy around the world reduced 6%. It was the lowest level in the past 70 years or after World War 2. It was like the loss in energy need all over India, which is the third largest country in energy production. Even if social distancing is relieving in the next few months and the economy is recovering, the overall picture of energy need in this year will reduce 3.8% if businesses recover quickly. When the situation of COVID-19 infection is better in many areas, the energy need will be reduced over 6% around the world.

According to the IEA report, it indicated that in the first trimester the energy needs around the world reduced 3.8% compared to the previous year. The energy source from coals received the high impact from of pandemic

in China, which is the world biggest energy consuming country. At the same time, the oil needs drastically dropped around 60% from the reduction of crude oil needs to the situation of transportation and airlines. People tended to stay home, and most flights were cancelled. This directly affects to the need in oil consumption as well as electricity. IEA expected that the need in electricity would reduce around 5%. Therefore, the need in natural gas in producing electricity will reduce as well after it increased rapidly in the past 10 years (CNBC News, 2020).

Changes in global energy prices have been affected the operation of the energy business. The energy sector index is one of the benchmarks that play a huge role for investors and mutual fund managers, who invest in energy stocks around the world to use the energy sector index to compare investment performance or investment efficiency in the portfolio of mutual funds. MSCI is usually used as a benchmark for investment in the portfolio among global energy stocks.

Investors expecting a high return poses a risk of a different return than expected. To reduce investment risks, investors and fund managers should study the information and consider tools for decision-making. By analyzing the movements of the index, investors and fund managers can reduce their investment risk (Bunnun et al., 2019).

Consequently, MSCI is important and beneficial to investors and mutual fund managers who invest in stocks of the global

energy group. On the other hand, the investors already know the behavior or movement of the energy index, it is beneficial in decision making and investment in energy stocks for both domestically and internationally. From the mentioned situation, the study of Energy Sector Index changes is very necessary in investor and fund manager in order to set the direction and policy on investment. One of the most popular world indexes which reflects the energy price is energy index. MSCI Index is a reference index of Morgan Stanley Capital International (MSCI) which is the world - leader in stock price. This index was created to be the criteria and standard for foreign investors in selecting stocks and returns.

MSCI Index is categorized into different kinds of assets. In terms of stock, it is divided by countries such as the US, China and Thailand under the name of MSCI Thailand Index. It is also divided by regions such as America, Europe, and Asia or by developing markets and emerging markets. The large stocks with liquidity are selected and cost 85% of all stock prices in the country as one of the country index components. Apart from the stock, there are other kinds of assets that can measure such as bonds, money market and real estates. MSCI Index that relates to energy is MSCI World Energy Index. It was designed to gather large and medium groups in 23 developed countries. All properties in the index are considered in the energy group, according to standards in classifying types of industry all over the world (Morgan Stanley Capital International, 2019).

Forecast and prediction are important parts for investors in making any decisions. One of the most famous techniques used at present is time series analysis and prediction. SARIMA Model is added as a seasonal variable. The MSCI World Energy Index was another set of data in the past which has a movement with the seasonal relation. The reason is that the amount of energy use in some periods is different. However, the trends in many years showed that the monthly and seasonal amount of energy use was quite similar.

From the importance of the energy index movement, this research aims to investigate a forecast model of the MSCI World Energy Index which was developed from SARIMA (Seasonal Autoregressive Integrated Moving Average). This is a model that add seasonal impacts to find the equation for further decision-making, planning, and investing.

### **Reserch Objective**

The objectives of this research article were 1) to construct the appropriate model for MSCI World Energy Index by using SARIMA (Seasonal Autoregressive Integrated Moving Average) and 2) to compare of Forecast Accuracy of MSCI World Energy Sector Index with SARIMA Model.

### **Literature Review**

Lee et al. (2007) mentioned to compare the forecasting performance of a neural network (NN) model and a time-series (SARIMA) model in Korean Stock Exchange. In particular,

we investigate whether the back-propagation neural network (BPNN) model outperforms the seasonal autoregressive integrated moving average (SARIMA) model in forecasting the Korea Composite Stock Price Index (KOSPI) and its return. Forecasting performance is evaluated by the forecasting accuracy of each model. KOSPI data and its return data over past 390 weeks period (89 months) extending from January 1999 to May 2006 to be analyzed. We discovered as followed: first, the SARIMA model generally provides more accurate forecasts for the KOSPI than the BPNN model does. This relative superiority of the SARIMA model over the BPNN model is pronounced for the mid-range forecasting horizons. Second, the BPNN model is generally better than the SARIMA model in forecasting the KOSPI returns. However, the difference in forecasting accuracies of the two models is not statistically significant. These results hold for both weekly and monthly data, and are robust across different measures of forecasting accuracy.

Urrutia et al. (2017) suggested that producing a model that can estimate a three-year forecast of the Philippine Daily Stock Exchange Index. Seasonal Autoregressive Integrated Moving Average (SARIMA) model is applied for a total of 4,927 daily data observations from 1995 to 2014. By then, actual and predicted values was compared through the paired t-test analysis by testing the hypothesis whether the two series of values are having significant difference or not, and other analysis signifying the accuracy of the model to forecast the future values,

that is created through all processes that are included in the evaluation of the model. This study may help the government in generating well-informed decisions regarding the topic discussed and for any improvements they may be able to make in the stock exchanged shares in the global market for the country's sake after learning about possible outcome of the study. The study has shown a favorable outcome by which the future values are expected to rise up higher as time passes. Thus, the investigators recommend further inclusion of the future observations to be compared to the future values of the Opening and Closing Price indices of the Philippine Stock Exchange index. This may be of use in any important discussions of the government for any reforms they can make for the betterment of the Philippine economy.

Ahmed (2018) discovered that most macroeconomic variables such as; inflation, GDP and others have been described by most financial and economics time series analysts to exhibit nonlinear behavior. Therefore, to cater for this behavior, the nonlinear class of models have been largely adopted to model and forecast such time series. In this study, the Keenan and Tsay tests for linearity showed inflation and CIC rates follow threshold nonlinear processes. Hence, the two-regime SETAR model was adopted to accommodate these nonlinearities in the datasets. Using the linear SARIMA model as a benchmark for comparative analysis. Results from both in-sample and out- of- sample forecast

performance using MAE and RMSE measures revealed that, the nonlinear SETAR model outperformed the linear SARIMA model for inflation. This was however different from CIC rates, since the Linear SARIMA model turned to outperform the nonlinear SETAR model. Further analysis of forecast accuracy, using the Diebold-Mariano test showed there was no significant difference between the two models for inflation, but there was a significant difference between both models for CIC rates. Nevertheless, it is recommended that, continuous monitoring of these models, review market conditions and necessary adjustments are vital to make realistic use of these models.

Nwokike et al. (2020) adopted the use of artificial neural network (ANN) which is also a nonlinear model in contributing to the debate by comparing different methods of forecasting frequency of time series data. The choice of ANN is informed by articles in the literature which have shown that Neural Network models outperform some traditional statistical models in modelling time series. The forecasting performance of seasonal autoregressive integrated moving average (SARIMA) model and that of seasonal artificial neural network (SANN) were compared with four forecast performance measures: - Forecast Error (FE), Mean Forecast Error (MFE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Results from the study showed that the SARIMA had lower error indicators in forecast performance and thus was adjudged better than SANN in forecasting frequency

of time series data. A t-test for significant difference showed that there is no statistical significant difference between both forecast values.

The related research by creating the SARIMA model and other models for time series data revealed that the SARIMA model had less error values than the other model. Referring to The Lee et al. (2007), the SARIMA model generally provides more accurate forecasts for the KOSPI than the Back-Propagation Neural Network (BPNN) model. While Ahmed (2018) SARIMA Model is better than SETAR model and Nwokike et al. (2020). SARIMA Model is better than Seasonal Artificial Neural Network (SANN).

All the factors above, this research focused on the SARIMA model in the creation of MSCI World Energy Sector Net USD Index to analyze the most efficient forecasting model which will be beneficial to the investor and fund manager.

### **Reserch Methodology**

This study was quantitative research using secondary data in building SARIMA (Seasonal Autoregressive Integrated Moving Average) and forecasting which consisted of four steps (Wang et al., 2008) as follows;

1. Data collection. Monthly data was collected from secondary data as MSCI World Energy Sector Net USD Index for 15 years from 2005-2020 with 180 data.

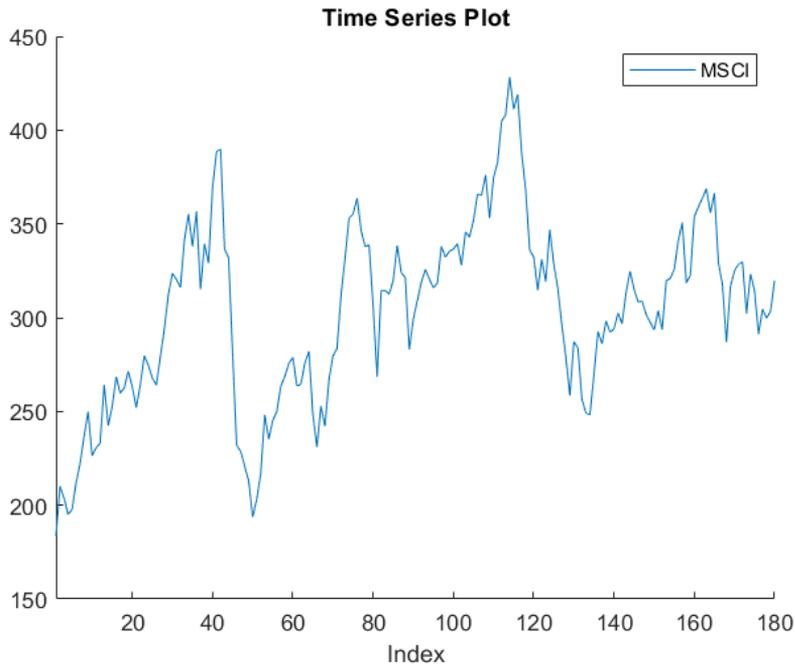


Figure 1 Index data of MSCI World Energy Sector Net USD

From Figure 1, monthly data of MSCI World Energy Sector Net USD was shown during the past 15 years (from 2005-2020) with 180 pieces of data. The index was always fluctuated in the form of business cycle, according to the picture.

2. Data stationary test by using Augmented Dickey - Fuller test (ADF)

In statistics and econometrics, an augmented Dickey-Fuller test (ADF) tests the null hypothesis that a unit root is present in a time series sample. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity. It is an augmented version of the Dickey-Fuller test for a larger and more complicated set of time series models.

In this research, stationary test of the MSCI World Energy Index was used to bring the data to build the forecast model by using data stationary test of the Augmented Dickey - Fuller test (ADF). The equation is as follows (Dickey & Fuller, 1979).

$$\Delta X_t = \theta X_{t-1} + \sum_{i=1}^P \phi_i \Delta X_{t-1} + e_t$$

$$\Delta X_t = \alpha + \theta X_{t-1} + \sum_{i=1}^P \phi_i \Delta X_{t-1} + e_t$$

$$\Delta X_t = \alpha + \beta t + \theta X_{t-1} + \sum_{i=1}^P \phi_i \Delta X_{t-1} + e_t$$

Where;

- $X_t$  = Data information at time t
- $X_{t-1}$  = Data information at time t-1
- $\alpha, \beta, \theta, \phi$  = Parameter,
- $t$  = Trend
- $e_t$  = Random error

To test the hypothesis with Augmented Dickey-Fuller test (ADF), it's the test to see if the variable ( $X_t$ ) has unit root. It can be calculated from  $\theta$ . If  $\theta$  is 0, it means that  $X_t$  has unit root.

The hypotheses testing is

$H_0 : \theta = 0$  The variable information is non-stationary or has unit root.

$H_1 : \theta < 0$  The variable information is stationary or has no unit root.

The Unit root test is a test to prove if time series are integrated of order 0 = I(0) or integrated of order 0 = I(d, d>0). If it rejects  $H_0$ , it means the data is stationary.

3. SARIMA Model and Choose the best model by using the least Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC).

### SARIMA Model

SARIMA (Seasonal Autoregressive Integrated Moving Average) is a model developed from ARIMA (p, d, q) of Box and Jenkins, and added Seasonal component to the model in the form of SARIMA (p, d, q) x(P, D, Q)<sub>s</sub>. p, d, and q are variables not related to seasons. P, D and Q are seasonal variables. s is seasonal factor (Box et al., 2008) The equation is as follows;

$$\Phi(B^S)\varphi(B)(1-B^S)^D(1-B)^d y_t = \Theta(B^S)\theta(B)\varepsilon_t$$

When  $\Phi$  and  $\varphi$  are seasonal and non-seasonal Autoregressive (AR) respectively.  $\Theta$  and  $\theta$  are seasonal and non-seasonal Moving Average (MA) respectively. B is reverse operator in the form of  $B(y_t) = y_{t-1}$  while  $(1-B^S)^D$  is seasonal difference (D) of S season, and  $(1-B)^d$  is a non-seasonal difference (d).  $\varepsilon_t$  is an independently distributed random variable. P and p are orders of Autoregressive process. Q and q are orders of the Moving Average process. Finally, D and d are seasonal and non-seasonal difference terms respectively (Wang et al., 2008).

### Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC)

In finding to the pattern of the appropriate model, various models must be built and tested with the approach in choosing the best model. It can be considered from Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC). The pattern of the model with the least AIC and BIC is the best model. It can be calculated as follows;

Akaike (1973) proposed the criteria in choosing the model which was called Akaike Information Criterion (AIC). It was built from the regression estimate of Kullback Leibler Information from the real model and the appropriate model with accurate quality. If the sample is large, choose the model with the least AIC according to AIC criteria as a model in describing dependent variables. Large samples are well-chosen when using AIC criteria as there is an error in choosing the model with too many independent variables with high possibilities. AIC can be calculated as follows;

$$AIC = n \cdot \ln\left(\frac{SSE}{n}\right) + 2p$$

When n is sample, SSE is a residual squared regression model, p is parameter in a regression model, and ln is the natural logarithm.

Sawa (1978) developed the criteria in choosing the model by adjusting Bayesian Model of AIC criteria which were called Bayesian Information Criteria (BIC). The formula is as follows;

$$BIC = n \cdot \ln\left(\frac{SSE}{n}\right) + \frac{2(p+1)n\sigma^2}{SSE} - \frac{2n^2\sigma^4}{SSE^2}$$

When n is sample, SSE is a residual squared regression model,  $\sigma^2$  is average residual squared regression model, p is parameter in a regression model, and ln is the natural logarithm, the model with the least BIC is chosen as an appropriate model.

4. Forecast accuracy comparison by Root Mean Square Error (RMSE) and Mean Square Error (MAE)

In building the forecast model, the accuracy of the model should be considered based on the residual from the model. At present, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are widely used with the following formulas (Kantananon, 2018).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$

When  $y_i$  and  $\hat{y}_i$  where the actual value and forecast value respectively, the most accurate forecast model was the model with the least RMSE and MAE.

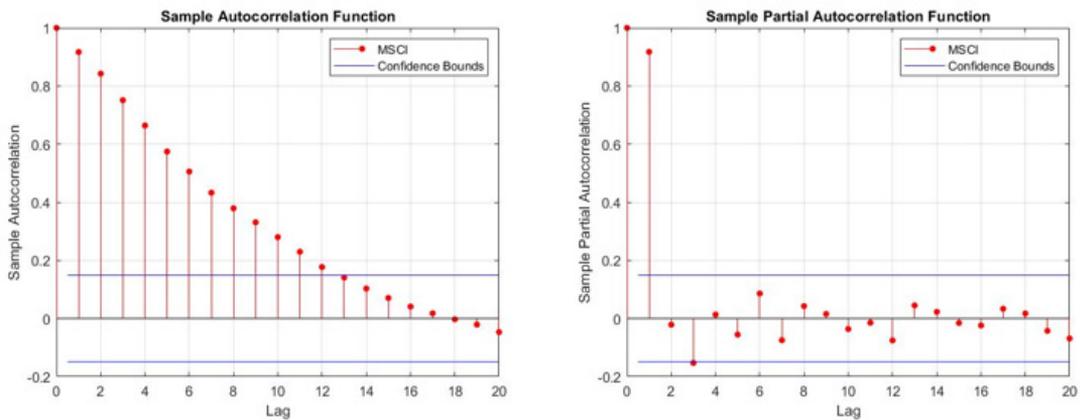
Results

**Table 1** Unit root test by Augmented Dickey-Fuller Test

Test for unit root in	Model	Test Statistic	Critical Value			p-value
			0.10	0.05	0.01	
1 <sup>st</sup> difference	Level	0.0831	-1.6156	-1.9424	-2.5777	0.6811
	Trend Stationary	-3.0945	-3.1423	-3.4369	-4.0128	0.1115
	autoregressive	-13.4456	-1.6156	-1.9424	-2.5778	0.0001
	Trend Stationary	-13.4073	-3.1424	-3.4371	-4.0132	0.0001

From Table 1, it was a stationary test of MSCI World Energy Sector Net USD. It was found that, p-value was higher than significant level (0.05). Therefore, the data was non-stationary at the level. The p-value of the

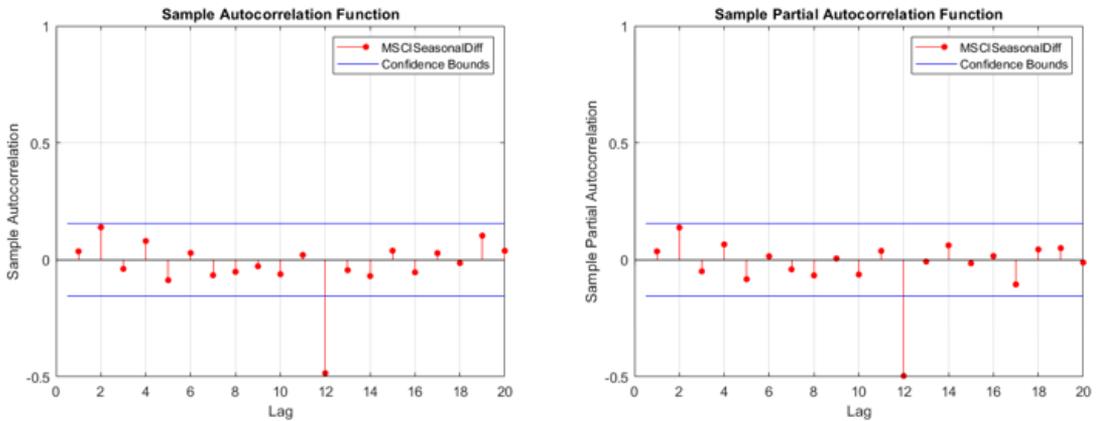
first difference was lower than the Significant level (0.05). It accepted that the data was stationary and had the trend. The data was converted by using the first difference to build the forecast model.



**Figure 2** ACF and PACF of MSCI World Energy Sector Net USD.

From Figure 2, when analyzing Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) with at

level, it was found that the result was over the blue line. It can be concluded that it was not stationary.



**Figure 3** ACF and PACF in case of the data was converted by using the first seasonal differences.

From Figure 3, when analyzing Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) with the first seasonal differences, it was found that the result was in the blue line. Only the data

in the twelfth position was over the limit line. It can be concluded that the data of the past twelve months had an impact on current data or the data had the lag length at 12 months.

**Table 2** Seasonal Unit Root Test

Seasonal Unit Root Test	Test Stat.	Significance Level		
		1%	5%	10%
Frequency 2PI/12 and 22PI/12	11.35290			
Frequency 4PI/12 and 20PI/12	16.31046			
Frequency 6PI/12 and 18PI/12	16.36369			
Frequency 8PI/12 and 16PI/12	12.24799	31.46	8.64	3.87
Frequency 10PI/12 and 14PI/12	17.92481			
All seasonal frequency	15.42933			
All frequency	14.15493			

From Table 2, It's a test for Seasonal Unit Root Test for MSCI World Energy Sector Net USD by Traditional HEGY Method. All

seasonal frequency test statistic is 15.42933. It's larger than 8.64 (5% critical value). So, it's no unit root processes at this data.

**Table 3** Parameters of five models - SARIMA (p,d,q)×(P,D,Q)<sub>12</sub>

Parameter	SARIMA				
	(0,1,0)×(3,1,3) <sub>12</sub>	(0,1,2)×(3,1,3) <sub>12</sub>	(1,1,2)×(3,1,3) <sub>12</sub>	(2,1,0)×(3,1,3) <sub>12</sub>	(2,1,1)×(2,1,3) <sub>12</sub>
AR {1}	-	-	-0.7963	0.0770	-0.5571
$\phi_1$			(0.0001)	(0.2449)	(0.0004)
AR {2}	-	-	-	0.1347	0.2113
$\phi_2$				(0.0311)	(0.0002)
SAR {12}	-0.9797	-1.0000	-1.0000	-1.0000	-0.9066
$\Phi_{12}$	(0.0001)	(0.00001)	(0.0001)	(0.0001)	(0.0001)
SAR {24}	-0.9562	-0.9244	-0.9132	-0.9180	-0.8005
$\Phi_{24}$	(0.0001)	(0.00001)	(0.0001)	(0.0001)	(0.0002)
SAR {36}	-0.1819	-0.1346	-0.1166	-0.1305	-
$\Phi_{36}$	(0.0076)	(0.0795)	(0.1247)	(0.0901)	
MA {1}	-	0.0827	0.9018	-	0.6754
$\theta_1$		(0.2420)	(0.0001)		(0.0005)
MA {2}	-	0.1276	0.1761	-	-
$\theta_2$		(0.0324)	(0.0097)		
SMA {12}	-0.0879	-0.0346	-0.0187	-0.0333	-0.0271
$\Theta_{12}$	(0.5135)	(0.7874)	(0.8831)	(0.7925)	(0.0281)
SMA {24}	-0.0837	-0.1318	-0.1140	-0.1361	-0.1116
$\Theta_{24}$	(0.4885)	(0.2719)	(0.3292)	(0.2498)	(0.0317)
SMA {36}	-0.7581	-0.7996	-0.7913	-0.8016	-0.8101
$\Theta_{36}$	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0092)
Variance	286.2297	280.3767	276.1246	279.8927	278.3791
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
AIC	1,543.00	1,543.30	1,542.60	1,543.00	1,542.00
BIC	1,563.20	1,569.20	1,571.20	1,568.70	1,562.80
RMSE	16.87	16.74	16.66	16.73	16.68
MAE	12.41	12.39	12.37	12.36	12.39
Ljung-Box Q	14.947	14.953	15.231	14.977	15.547
p-value	0.3749	0.3763	0.3824	0.3771	0.3920

From Table 3, it was an experiment in building SARIMA by choosing five best models from the models with least AIC and BIC and Ljung-Box Q has no significance at 0.05. The number from the table is coefficient

- $\varphi$  was autoregressive parameters of non-seasonal components
- $\theta$  was moving average parameters of non-seasonal components
- $\Phi$  was autoregressive parameters of seasonal components
- $\Theta$  was moving average parameters of seasonal components

The five equations have formed as follows;

SARIMA (0,1,0)×(3,1,3)<sub>12</sub>

$$(1 - \Phi_{12}B^{12} - \Phi_{24}B^{24} - \Phi_{36}B^{36})(1 - B)(1 - B^{12})y_t = (1 + \Theta_{12}B^{12} + \Theta_{24}B^{24} + \Theta_{36}B^{36})\varepsilon_t$$

SARIMA (0,1,2)×(3,1,3)<sub>12</sub>

$$(1 - \Phi_{12}B^{12} - \Phi_{24}B^{24} - \Phi_{36}B^{36})(1 - B)(1 - B^{12})y_t = (1 + \theta_1B + \theta_2B^2)(1 + \Theta_{12}B^{12} + \Theta_{24}B^{24} + \Theta_{36}B^{36})\varepsilon_t$$

SARIMA (1,1,2)×(3,1,3)<sub>12</sub>

$$(1 - \varphi_1B)(1 - \Phi_{12}B^{12} - \Phi_{24}B^{24} - \Phi_{36}B^{36})(1 - B)(1 - B^{12})y_t = (1 + \theta_1B + \theta_2B^2)(1 + \Theta_{12}B^{12} + \Theta_{24}B^{24} + \Theta_{36}B^{36})\varepsilon_t$$

SARIMA (2,1,0)×(3,1,3)<sub>12</sub>

$$(1 - \varphi_1B - \varphi_2B^2)(1 - \Phi_{12}B^{12} - \Phi_{24}B^{24} - \Phi_{36}B^{36})(1 - B)(1 - B^{12})y_t = (1 + \Theta_{12}B^{12} + \Theta_{24}B^{24} + \Theta_{36}B^{36})\varepsilon_t$$

SARIMA (2,1,1)×(2,1,3)<sub>12</sub>

$$(1 - \varphi_1B - \varphi_2B^2)(1 - \Phi_{12}B^{12} - \Phi_{24}B^{24})(1 - B)(1 - B^{12})y_t = (1 + \theta_1B)(1 + \Theta_{12}B^{12} + \Theta_{24}B^{24} + \Theta_{36}B^{36})\varepsilon_t$$

The best model was SARIMA 12.39. All variables had Significance level at (2,1,1)×(2,1,3)<sub>12</sub>. AIC was 1,542.00 and BIC was 1,562.80. RMSE was 16.68 and MAE was

and p-value in some model do not contain the parameter like SARIMA (0,1,2)×(3,1,3)<sub>12</sub> do not have AR{1} and AR{2} parameter. Where the parameter has meaning as follows;

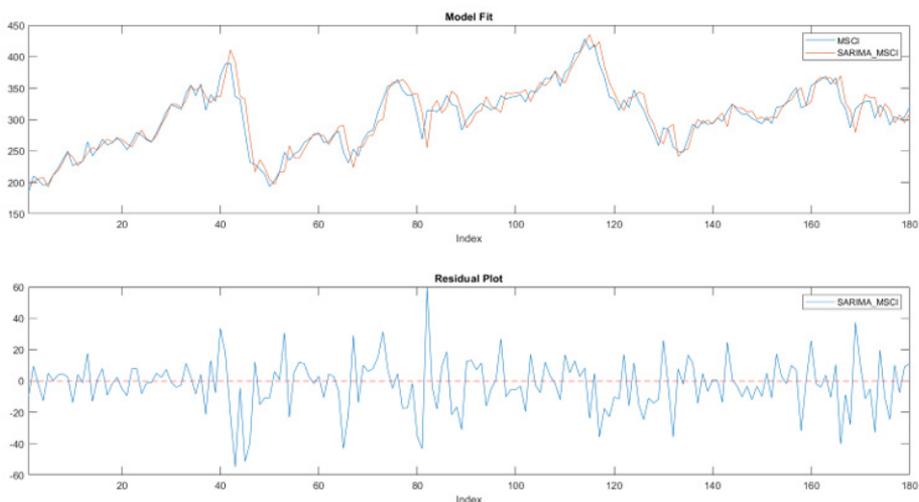


Figure 4 Diagram showing the actual value, forecast value, and residual

In Figure 4, the upper picture was the graph showing the actual value and the forecast value of MSCI World Energy Sector Net USD Index in the past 15 years from 2005 to 2019 with a total number of 180 pieces of

information. The lower picture was residual from forecasting MSCI World Energy Sector Net USD Index in the past 15 years. RMSE was 16.68 and MAE was 12.39 as shown in Table 2.

**Table 4** 24-month forecast from January 2020 to March 2021

No.	Month	Actual	Forecast	No.	Month	Actual	Forecast
1	January 2020	290.73	306.84	13	January 2021	225.54	231.87
2	February 2020	250.05	281.32	14	February 2021	260.97	248.22
3	March 2020	176.58	200.33	15	March 2021	266.94	272.88
4	April 2020	205.19	183.37	16	April 2021	267.95	275.43
5	May 2020	207.99	199.55	17	May 2021	281.68	269.71
6	June 2020	205.81	222.69	18	June 2021	290.14	280.54
7	July 2020	197.99	203.91	19	July 2021	272.07	283.27
8	August 2020	200.87	197.91	20	August 2021	268.73	277.55
9	September 2020	172.96	193.21	21	September 2021	293.94	286.23
10	October 2020	163.75	170.32	22	October 2021	317.59	308.69
11	November 2020	211.36	178.85	23	November 2021	295.64	311.87
12	December 2020	219.20	224.11	24	December 2021	307.08	305.13
						RMSE	14.93
						MAE	12.51

From Table 4, it was in about the 24-month forecast of MSCI World Energy Sector Net USD Index from January 2020 to December 2021. When analyzing the residual from the forecast, RMSE was 14.93 and MAE was 12.51 respectively. The resulting of

comparable between actual value and model forecasting result in time frame 24 months was reflected that RMSE and MAE values from the forecast are approximated. The RMSE and MAE values were obtained by the model.

## Discussion

From the study about Forecasting MSCI World Energy Sector Net USD Index with the SARIMA Model, it revealed the seasonal influences which affected to the changes of the Energy Sector Index. Past seasons had an impact on the indexes in the same period as the need in using energy would be at the same level in each month or season. Therefore, MSCI World Energy Sector Net USD Index forecast with seasonal factors would have less errors.

This study related to the study of Kaewhawong (2015), conducted a study in comparing time series data. The result showed that adding a seasonal factor in the forecast helped reducing the residual. If the past data have a movement with a seasonal relation, the forecast will be more accurate for all kinds of data. Kirdjongrak (2017) used the forecast with a seasonal factor to forecast products, and the residual decreased. According to previous studies, when the data have a movement with a seasonal relation, adding seasonal factors would help the model be more efficient.

The criteria used in comparing errors in forecasting were mean absolute percentage error (MAPE). According to the result of the study, the most appropriate model in forecasting Stock Exchange Index was SARIMA. The reason was that Stock Exchange Index use in each season was related. In addition, the result related to the study of Urrutia et al. (2017). The purpose of the study is to produce a model that can estimate a three-year forecast of the Philippine Daily Stock Exchange Index.

Seasonal Autoregressive Integrated Moving Average (SARIMA) model is applied for a total of 4,927 daily data observations from 1995 to 2014. By then, actual and predicted values were compared through the paired t-test analysis by testing the hypothesis whether the two series of values are having significant difference or not, the results show that forecast by SARIMA Model makes MAPE less.

If the data have a movement with a seasonal relation, adding seasonal factors helps the model be more efficient. MSCI World Energy Sector Net USD Index was another set of data in the past which has a movement with the seasonal relation. The cause is that the amount of energy used in some periods differently. However, the trends in many years showed that the monthly and seasonal amount of energy used were quite similar.

## Conclusion

In conclusion, this study is a quantitative research focusing on constructing a model used in MSCI World Energy Sector Net USD Index forecast. The result showed that MSCI World Energy Sector Net USD Index information had influences from its own information in the same season from the past. It seems that seasonal information in the past had influences on current indexes. When the seasonal information in the past influences on current indexes, the researcher applied this factor to create SARIMA Model (Seasonal Autoregressive Integrated Moving

Average) and chose the best model. The best model contained the least AIC and BIC and Ljung-Box Q has no significance.

From the result of building the model, the best model is SARIMA (2,1,1) × (2,1,3)<sub>12</sub>. AIC and were 1,542.00 and 1,562.80. RMSE and MAE were 16.68 and 12.39 respectively. All variables passed the Significance level at 0.05. When using the best model to forecast MSCI World Energy Sector Net USD Index 24 months in advance, it was found that the statistic error was low. RMSE and MAE were 14.93 and 12.51 respectively.

### Recommendations

A forecasting becomes one of most useful research at the present. The predicting the changes of interested variables may improve better results in decision making and planning. This research focuses on Forecasting MSCI World Energy Sector Index with the SARIMA Model. It is a beneficial to investors and fund managers who invest stocks in the Energy Fund. Moreover, the results of the forecast can be used as components in decision-making in stock investment in the energy sector. Another beneficial aspect of this study is that the energy index forecast can be used the energy price forecast to estimate business income. For these producing energy companies, forecasting MSCI can be considered as a forecast for income and expenses of the business. The executive can manage policies or plan suitably for their business by using this energy forecast. The

interesting aspect improving the next research is to add other models which are linear and non-linear models. These models should be enhanced the forecast to find the better best model in the index forecast. Other energy price information such as oil and coal, and the relation and energy index should be included to study to provide benefits to investors and business. In addition, adjusting the model, to forecast price stock and energy sector as the changes of energy price, affects the profit and loss in the business. Conclusively, if the socioeconomic context has changed continually, the established model should be revised whether it is still suitable to be used in forecasting to find the best model.

### References

- Ahmed, N. B. (2018). *A Comparative Analysis of Forecast Performance between SARIMA and SETAR Models using Macroeconomic variables in Ghana*. Ghana: University of Ghana.
- Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In B.N. Petrov and F. Csaki (Eds.), *Second International Symposium on Information Theory*, 267-281.
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time series analysis, forecasting and control, 4th ed*. New Jersey: John Wiley and Sons.

- Bunnun, W., Phonkhen, K., Penalawat, K., & Nguatchai, W. (2019). An Analysis of Risk and Return in the Stock Exchange of Thailand: A Case of Banking Sector During Periods Before and After on September 2018. *Kasetsart Applied Business Journal*, 13(19), 19-28.
- CNBC News. (2020). Energy demand, hit by coronavirus crisis, is set to see record drop this year, *IEA says*. <https://www.cnbc.com/2020/04/30/energy-demand-set-to-fall-the-most-on-record-this-year-amid-coronavirus-pandemic-iea-says.html>
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a), 427-431.
- Kaewhawong, N. (2015). Forecasting Electricity Consumption of Thailand by Using SARIMA and Regression Models with ARMA Errors. *Thai Journal of Science and Technology*, 4(1), 24-36.
- Kantananon, K. (2018). *Statistical Forecasting*. Bangkok: SE-EDUCATION Public Company Limited.
- Kirdjongrak, W. (2017). *Forecasting Sea Salt Price*. (Master's Independent Study). Bangkok: Thammasat University.
- Lee, K., Yoo, S., & Jin, J. (2007). Neural Network Model vs. SARIMA Model In Forecasting Korean Stock Price Index (KOSPI). *Issues in Information Systems*, 8(2), 372-378.
- Morgan Stanley Capital International. (2019). *MSCI 2019 Annual Report*. America: Morgan Stanley Capital International.
- Nwokike, C. C., Offorha, B. C., Obubu, M., Ugoala, C. B., & Ukomah, H. I. (2020). Comparing SANN and SARIMA for forecasting frequency of monthly rainfall in Umuahia. *Scientific African*, 3(10), 1-15. <https://doi.org/10.1016/j.sciaf.2020.e00621>
- Sawa, T. (1978). Information Criteria for Discriminating among Alternative Regression Model. *Journal of Econometrica*, 46(6), 1273-1282.
- Tadesse, K. B., & Dinka, M. O. (2017). Application of SARIMA model to forecasting monthly flows in Waterval River, South Africa. *Journal of Water and Land Development*, 12(35), 229-236.
- Urrutia, J. D., Diaz, J. L. B., & Baccay, E. B. (2017). Forecasting Philippine Daily Stock Exchange Index. *Journal of Fundamental and Applied Sciences*, 9(7S), 202-218. <https://dx.doi.org/10.4314/jfas.v9i7s.21>
- Wang, J., Du, Y. H., & Zhang, X. T. (2008). *Theory and application with seasonal time series*, 1st ed. Nankai: Nankai University Press.