Prediction in Taiwanese Monthly Export on Customs Basis

by

Student-Nr.: S158800

Exam: Predictive Analytics (CDSCO1005E) - Home assignment (UC)

Study: MSc in Business Administration and Data Science
Date: 18. August 2023
Pages:
Characters incl. spaces: 17379
Dataset:
https://nstatdb.dgbas.gov.tw/dgbasall/webMain.aspx?sdmx/a081201010/2.1..M.&startTime=2003-01&endTime=2023-04

Content

The Content of Tables

The Content of Figures

1. Introduction

The export-led growth hypothesis (ELGH) suggests that expanding exports can act as an engine for overall economic growth, poverty reduction, and development in the country (Hallaert & Munro, 2009; Medina-Smith, 2001). Especially in Taiwan, the outward-orientated economy has led to sustained and high economic growth rates since the 1960s. (Sazanami, 1995) Even today, the Taiwanese economy remains highly dependent on exports. According to Lee (2021), the degree of Taiwanese export dependence[1] was 51.6% in 2020. This fact highlights that fluctuations in Taiwanese exports can significantly impact the economy. Accurate export forecasting is essential to ensure proper policy implementation and boost trade surplus and economic growth.

Mahmoud (1984) indicates simple forecasting methods tend to outperform complex techniques. This research aims to compare the forecasting accuracy of simple time series prediction methods and using the best perform model to make the prediction. This research aims to compare the forecasting accuracy of simple time series prediction methods and use the best perform model to make the prediction. The paper will begin with explorative descriptive statistics to understand the data properties and apply possible transformation techniques to meet the model's assumption. Two univariate time series models- the Seasonal ARIMA and ETS models- will be applied, followed by the model identification and evaluation using the Root Mean Square Error (RMSE). The paper will conclude with a summary of the findings.

2. Data Description

2.1 Data Set Introduction

The dataset utilised in this research paper is derived from the Taiwanese monthly exports on customs basis data from 2003 to 2023, published by the Statistics Department, Taiwan (2023). The dataset is presented in CSV format and comprises two columns: the date and the monthly exports on a customs basis, which are represented in hundred million US dollars. This dataset comprises 244 observations, representing the value of

[1] $Degree\ of\ export\ dependence = \frac{Exports}{GDP}$

monthly exports from January 2003 to April 2023.

2.2 General Pattern

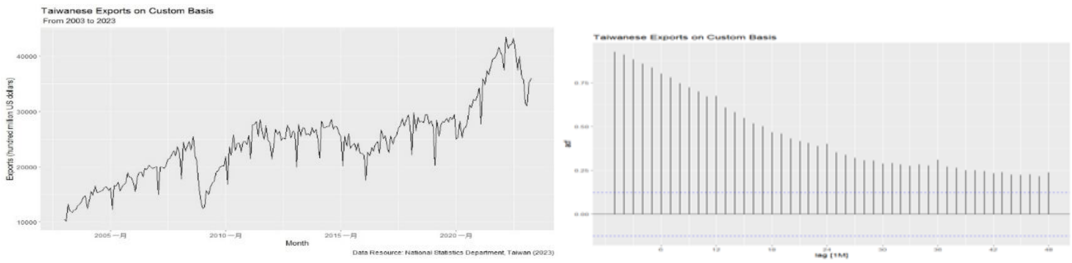

Figure 1 Taiwanese Exports on Custom Basis from Jan. 2003 to Apr. 2023 (left) and its ACF Plot (right).

According to the time series plot of Taiwanese monthly exports (Fig.1, left), the data portrays the overall upward trend with the potential cycle pattern due to the economic conditions. Moreover, the exports also present the potential seasonality and an exponential pattern. Therefore, the box-cox transformation would be applied to transform the data.

The ACF plot (Fig 1. Right) shows that autocorrelation decreases over time, indicating a positive correlation between observation and its lag. However, the autocorrelation peaks at lag 12, 24, 36, and 48 indicate the seasonality in the time series. Although the ACF plot does not show signs of cyclical fluctuations, the cyclical feature might be concealed by the seasonality and trend features.

2.3 Seasonality

Based on the time series plot (Fig 1.), there may be seasonality in the data over a year. The seasonal plot (Fig.2.) shows that, aside from years impacted by global economic events, Taiwanese exports tend to decrease in February, then slightly increase in March, and then fluctuate slightly for the remainder of the year. February in Taiwan has many important holidays, such as Lunar New Year, the 228 ceremony, and the lantern festival, resulting in fewer working days and lower export values. However, in March, the vacation period ends, and companies begin to fulfil orders that were not completed in February, which leads to an increase in exports. The seasonal plot (Fig. 2.) confirms these patterns.

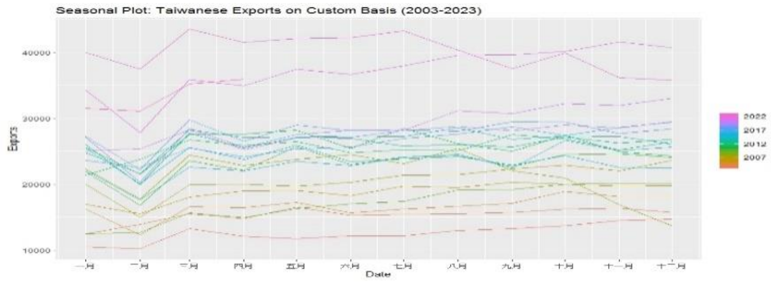

Figure 2 Seasonal Plot of Monthly Taiwanese Exports on Custom Basis from 2003 to 2023

2.4 SEATS Decomposition

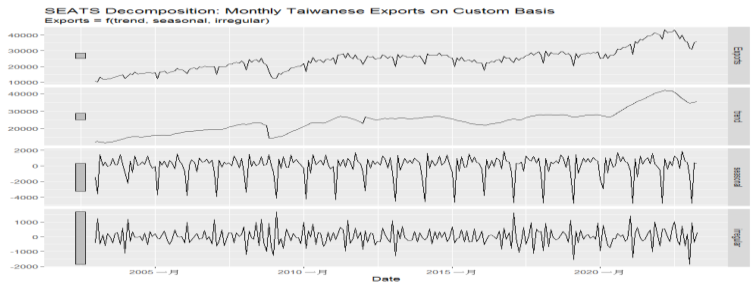

Figure 3 SEATS Decomposition Plot of the Monthly Taiwanese Exports

The monthly export time series were analysed using the SEATS decomposition to adjust for outliers and seasonal fluctuations and to consider specialised economic data features. The trend component was found to be the main contributor to the variation in the time series, as indicated by the scalar bar in the SEATS decomposition plot (Fig. 3). The trend panel, which includes cycle patterns, also suggested the presence of potential cyclical trends in the time series. Although the scalar bar in the seasonal panel was slightly smaller than the trend panel, it still supported the existence of seasonality in Taiwanese exports.

3. Methodology

3.1 Box-Cox transformation

Sometimes the log transformation can be too strong and might distort the data. To address this, the box-cox transformation is employed, which combines the log transformation with the power transformation. The lambda value for the transformation function is automatically selected by the function in R to ensure optimal results. In this case, a lambda value of 0.762 was chosen.

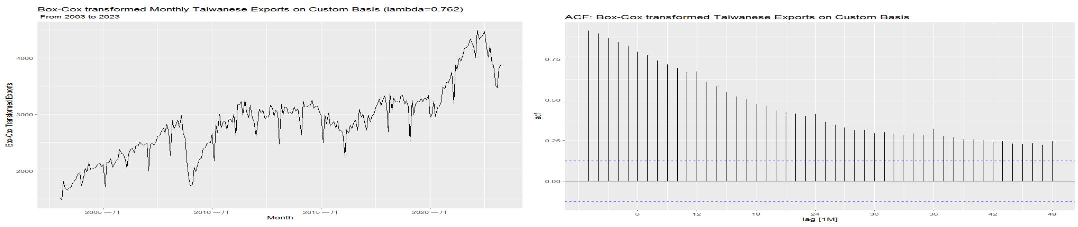

Figure 4 The Time Series (left) and ACF plot (right)of the Transformed Taiwanese Monthly Exports

After the box-cox transformation (Fig.4), the time series is flattened and normalised. The SEATS decomposition plot (Fig.5) also indicates that the variance of the Taiwanese exports has been stabilised. Since the ARIMA model requires stationary input data and a linear relationship between the time series variable, the box-cox transformed data will be utilised for further operations.

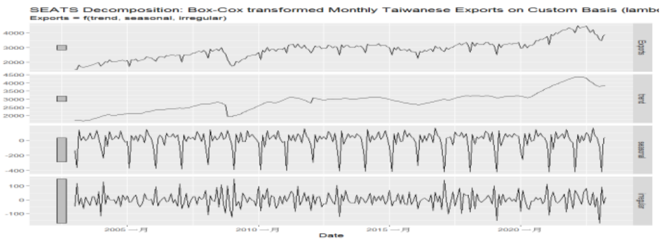

Figure 5 SEATS Plot of Box-Cox Transformed Monthly Taiwanese Exports

3.2 Exponential Smoothing (or Error, Trend, Seasonality) Method (ETS)

Proposed by Hyndman et al. (2002), the ETS method can identify three components in the time series: error, trend and seasonality. The error component can be identified as either ‘A’ (additive) or ‘M’ (multiplicative), seasonality can be chosen as ‘A’, ‘M’ or ‘N’ (None), and the trend can be ‘A’, ‘Ad’ (additive with damped) and ‘N’. To select the optimal ETS model, all three components must be identified correctly, which can be evaluated by minimising the information criteria, such as AIC (Akaike Information Criterion), AICc (The corrected Akaike Information Criterion), and BIC (Bayesian Information Criterion) (Qi et al., 2022).This paper will apply manual and auto selection for model selection, depending on the SEATS decomposition. The model’s accuracy will be evaluated using the information criteria mentioned above and the optimal model will be selected to compare with the seasonal ARIMA model.

3.3 Seasonal Autoregressive Integrated Moving Average (Seasonal ARIMA)

SARIMA, extended by the ARIMA model, is a popular and effective prediction model considering trend and seasonality patterns in the time series. The SARIMA model has three sets of components: nonseasonal components (p,d,q), seasonal components (P, D, Q) and seasonal period. The AR component ('P' and 'p') assesses the impact of the previous value on the current value, capturing the autocorrelation of the time series. The integrated component ('D' and 'd') represents the number of differences necessary to make the time series stationary. Lastly, the MA component ('Q' and 'q') considers the influence of previous forecast errors on the current data. (Shu, 2005) This paper will apply and evaluate both manual model selection based on the ACF and PACF graph and automatic model selection to determine the most accurate prediction model.

4. Model Identification

To train the ETS model and seasonal ARIMA model, the time series has been split around 80% into training set (2003 January to 2018 December) and 20% into the test set (2019 January to 2023 April), following the 80/20 rule of thumb.

4.1 EST Model

	Alpha	Beta	Gamma	Phi	AIC	AICc	BIC	Ljung_Box
ETS(A, A, M)	0.653	0.126	0.00011	NA	**2880.3**	**2883.8**	**2935.6**	0.124
ETS(A, Ad, A)	0.622	0.00014	0.00010	0.979	2880.8	2884.8	2939.5	0.291

Table 1 The parameter and Accuracy of the ETS Guessed Model and Auto Selected Model

Two models were fitted to the time series data, one chosen manually, and one automatically selected by the ETS function in R. The manual model is an ETS(A, A, M) based on the decomposition plot (Fig.5.), which shows a consistent variance of the error component, the additive feature of the trend, and a slightly increasing variance of the seasonal component over time. The automatic model is an ETS(A, Ad, A) as indicated in Table 5. The moderate α in both models shows a balance between sensitivity to recent changes and overall stability while the low values of β (0.126 and 0100014) and γ (0.00011 and 0.00010) shows the forecasted trends and seasonal changes are not influenced significantly by the shorten fluctuation.

4.1.1 Diagnostic Checking

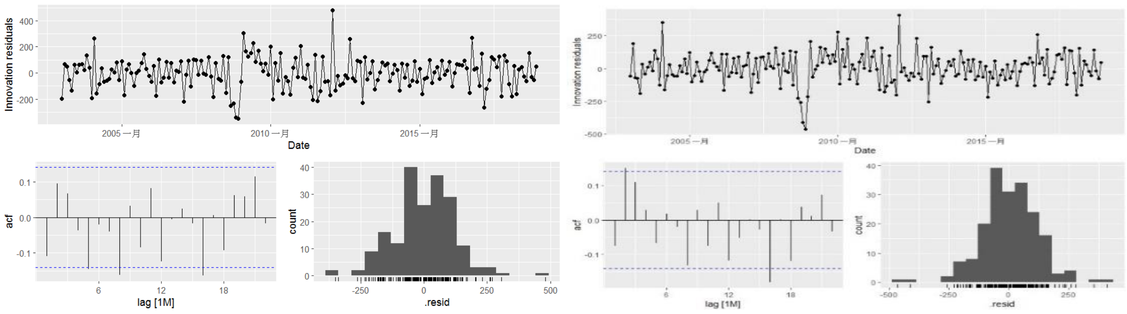

Figure 6 The Residuals of the Guessed model (left) and Auto Selected model (right)

Based on the residual plots (Fig. 6.), the variance in the guessed model is lower than that of the auto-selected model. Both models exhibit autocorrelation spikes at lag 2 and 16, but the guessed model has an additional spike at lag 12. However, the Ljung box test indicates no significant evidence (at a 5% significance level) to reject the hypothesis that residuals in both models are independent distributed. Additionally, the errors present the normal distribution feature in both models. Therefore, since the guessed model has a lower value in the information criterion, it is the preferred choice.

4.2 Seasonal ARIMA Model

4.2.1 Stationary Check

As seasonal ARIMA necessitates stationary input, the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests are employed to detect unit root presence. As per prior visual analysis, the time series exhibits a trend with intercept. The KPSS test employs "tau" type.

	10%	**5%**	**2.5%**	**1%**	**Test statistics value**
KPSS test (tau)	0.119	0.146	0.176	0.216	0.258

Table 2The KPSS Test Result of the Box-Cox Transformed Exports

The test statistics value of the KPSS test is larger than the 1% significant level, indicating the null hypothesis should be rejected. In other words, the result rejected the export time series data as stationary.

ADF test (trend, AIC)	**tau3**	**phi 2**	**phi3**
Critical Statistics Value (significant level = 5%)	-3.43	4.75	6.49

Test Statistics Value	-2.53	3.06	3.26

Table 3 The ADF Test Result of Box-Cox Transformed Exports

Based on ADF, "trend" type with "AIC" lag selection was chosen. The tau3 test statistic (-2.53) exceeds the 5% critical value (-3.43), implying no evidence to reject unit root presence. Phi 2 and phi3 test statistics are below 5% critical values, suggesting non-rejection of null hypothesis. These results indicate the presence of unit root or absence of trend and drift.

4.2.2 Seasonal Differences

According to the unit root testing result, the time series present the random walk feature without the drift. Therefore, the seasonal difference will be applied to detrend the time series.

	10%	**5%**	**1%**	**Test statistics value**
KPSS test (mu)	0.119	0.146	0.739	0.114
ADF test (none)	-1.62	-1.95	-2.58	-2.74

Table 4 The KPSS Test and the ADF Test Results of Seasonal Differenced Box-Cox Transformed Exports

The test statistics value of the KPSS test is smaller than the 5% significant level, indicating the null hypothesis that time series is stationary cannot be rejected. The fact that the test statistics value is smaller than the 1% significant value provides strong evidence to reject the presence of the unit root in the time series, verifying the time series is stationary.

4.2.3 Model Selection

According to the ACF and PACF plot, ACF tails off a feature while PACF presents spikes at lag 1, 2 and 5, suggesting the AR(3) component in the time series. For the seasonality ARIMA components, the PACF plot shows that the time series has spiked in every 12 lags, suggesting the m = 12. The PACF spikes at lags 1, 12 and 24 indicate the presence of seasonal AR(3) in the time series. The initial handpicked model of Seasonal ARIMA is ARIMA[3,0,0][3,1,0]$_{12}$. However, after exploring models[2] with similar components, the ARIMA[2,0,2][2,1,0]$_{12}$ model outperformed the others and was chosen as the handpicked model. The `ARIMA()`

[2] Please see Appendix A For the result of the full version handpicked models' exploration.

function auto selected a seasonal ARIMA[1,1,0][2,0,0]$_{12}$ (with drift) model. This difference might cause by the different differencing methods chosen when making the time series stationary.

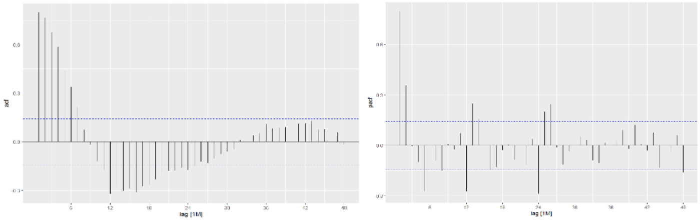

Figure 7 The AFC plot (left) and PACF plot (right) of Seasonal Differenced Transformed Exports

	AIC	AICc	BIC	Ljung Box Test	Normality Test [3]
ARIMA[2,0,2][2,1,0]$_{12}$	**2294.81**	**2295.65**	**232035**	**0.745**	**0.001486**
ARIMA[1,1,0][2,0,0]$_{12}$	2460.03	2460.24	2473.04	0.534	0.0001043

Table 5 The Info Criteria of Guessed Model and Auto Selected Model

4.2.4 Diagnostic Checking

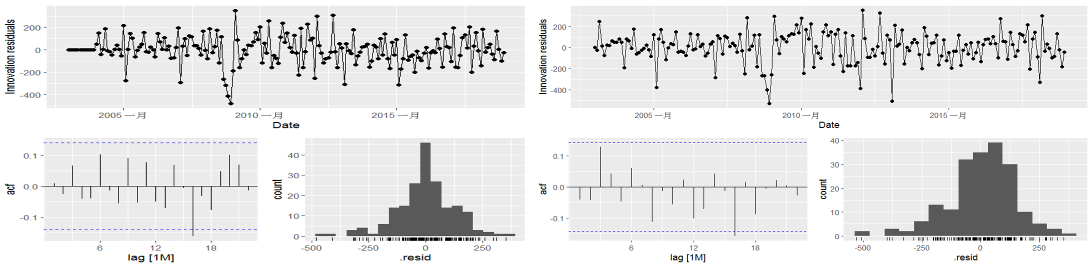

Figure 8 The residual plot of ARIMA(2,0,2)(2,1,0)$_{12}$ (left) and ARIMA(1,1,0)(2,0,0)$_{12}$(right)

Based on the residual plots (Fig. 8.), there is no significant pattern in the residuals of both models. Even though the ACF plot shows a peak at lag 16 in both models, the p-values of the Ljung box test for both models are higher than the 5% significance level, indicating that the autocorrelation in the residuals is not statistically significant. Although the normality test[4]results show strong evidence against the residuals being normally distributed, if the other assumptions, such as the residuals not being autocorrelated, are met, the Seasonal ARIMA model may still provide valid results. Lastly, due to the lower information criterion of the

[3] The values at the Ljung Box test and Normality test column are the p-value of the test.

[4] In this paper, the Shapiro-Wilk normality test is used to verify the normal distribution of residuals. The test's null hypothesis is the data follows a normal distribution.

ARIMA[2,0,2][2,1,0]$_{12,}$ the handpicked model has better accuracy than the auto-selected one.

4.3 Structural Break Analysis

In this paper, the Sup-Wald test was employed to verify any potential structural changes in the time series.

	Sup-F	p-value
Sup-Wald Test	9.4011	0.1213

Table 6 The Result of the Sup-Wald Test

The result of the Sup-Wald test shows no significant evidence to justify the structural break in the time series. As such, the time series will remain unchanged, and the selected model will be applied to the test data to make predictions.

5. Result: Forecasting and Model Evaluation

After applying the selected ARIMA and ETS models to the test set, their performance on unseen data was analysed. The ARIMA model performed better than the ETS model based on forecasting error-related criteria such as ME and RMSE. However, both models had a strong autocorrelation in the residual of test set prediction with ACF1 indicator very close to 1. Additionally, both models had MASE, and Theil's U values were greater than 1, indicating worse performance than the Naïve method, such as random walk.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
ARIMA[2,0,2][2,1,0]$_{12}$	**5209.41**	**7608.02**	**5962.01**	**13.18**	**16.18**	**2.17**	**0.94**	**1.95**
ETS(A,A,M)	5864.95	7970.43	6085.23	15.35	16.29	2.21	0.94	2.04

Table 7 The accuracy table of the final selected Seasonal ARIMA Model and ETS Model

According to the comparison of actual data and forecast results (Fig 9. Left), poor model performance can be attributed to the economic shock caused by the Covid pandemic[5]. At the beginning of the COVID pandemic period, the slight decrease in the exports captured.by the model, leading the strong deviation of the forecast from actual data. Despite this, we can still anticipate better performance in the one-year ahead prediction (Fig

[5] Taiwan's exports initially dipped slightly at the pandemic's outset. Yet, due to quarantine measures abroad, Taiwan sustained production and even witnessed a surge in export demand throughout most of the pandemic. However, post-pandemic production recovery and reduced purchasing power led to a subsequent decline in demand.

9., right) trained by the whole time series, as the economic conditions have been relatively stabilise recover after the pandemic.

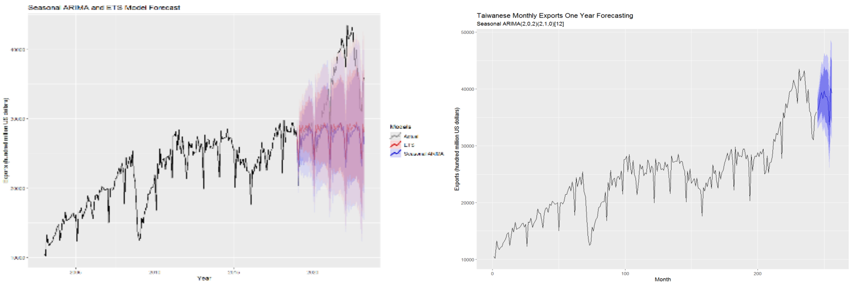

Figure 9 The Seasonal ARIMA and ETS Forecast (Left) The One-Year Ahead Forecasting Based on Entire Data. (Right)

6. Conclusion

Our study revealed the strengths and limitations of seasonal ARIMA and ETS methods. The Seasonal ARIMA excels in capturing intricate seasonal patterns, making it valuable for analysing datasets characterized by complex seasonal variations, such as the exports data used in this paper, while ETS models offer the advantage of interpretability, allowing for a clear understanding of the underlying mechanisms that shape forecasts.

In the context of forecasting Taiwanese monthly exports, the results spotlight a shared challenge. Both models struggle to predict unexpected economic shocks, revealing a gap in their adaptability to real-world dynamics. As such, it's imperative to recognize that economic models, while robust in certain aspects, must be complemented by mechanisms that account for unforeseen disruptions.

To address the limitations posed by unforeseen events, integration of more adaptive methodologies, such as Prophet, with Seasonal ARIMA and ETS models might enhance forecasting precision, enabling the models to effectively capture the nuanced dynamics of economic shocks that unfold.

In conclusion, this paper underlies the need for a holistic forecasting strategy that should consider the models' strengths and their limitations. By bridging traditional methodologies with cutting-edge adaptability, researchers can forge a more robust path towards accurate and dynamic time series forecasting in an ever-evolving economic landscape.

Reference

Bank, W. (1993). *The East Asian Miracle: Economic Growth and Public Policy : Summary*. Washington, D.C. : World Bank.

Ghauri, S. P., Ahmed, R. R., Streimikiene, D., & Streimikis, J. (2020). Forecasting Exports and Imports by using Autoregressive (AR) with Seasonal Dummies and Box-Jenkins Approaches: A Case of Pakistan. *The Engineering Economics, 31*(3), 291–301. https://doi.org/10.5755/j01.ee.31.3.25323

Hallaert, J., & Munro, L. (2009). Binding Constraints to Trade Expansion: Aid for Trade Objectives and Diagnostics Tools. *Social Science Research Network*. https://doi.org/10.2139/ssrn.1521409

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. D. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting, 18*(3), 439–454. https://doi.org/10.1016/s0169-2070(01)00110-8

Lee, Z. Y. (2021). Statistical Release Announce. In *https://service.mof.gov.tw/public/Data/statistic/bulletin/110/%E7%AC%AC19%E8%99%9F_%E5%87%BA%E5%8F%A3%E4%BE%9D%E5%AD%98%E5%BA%A6.pdf* (No. 19). Department of Statistics, Ministry of Finance, ROC (Taiwan).

Mahmoud, E. R. I. (1984). Accuracy in forecasting: A survey. *Journal of Forecasting*, *3*(2), 139–159. https://doi.org/10.1002/for.3980030203

Medina-Smith, E. (2001). Is the export-led growth hypothesis valid for developing countries? a case study of Costa Rica. In *United Nations eBooks*. https://unctad.org/en/docs/itcdtab8_en.pdf

Qi, L., Li, X., Wang, Q., & Jia, S. (2022). fETSmcs: Feature-based ETS model component selection. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2206.12882

Sazanami, Y. (1995). The East Asian Miracle: Economic Growth and Public Policy. A World Bank Policy Research Report. London: Oxford University Press, 1993. xvii, 289 pp. *The Journal of Asian Studies*, *54*(1), 184–185. https://doi.org/10.2307/2058969

Shu, Y. (2005). Wireless traffic modeling and prediction using seasonal ARIMA models. *IEICE Transactions on Communications*, *E88-B*(10), 3992–3999. https://doi.org/10.1093/ietcom/e88-b.10.3992

Appendix

A. Handpicked Model Comparison

	AIC	AICc	BIC	Ljung Box Test	Normality Test
ARIMA[3,0,0][3,1,0]$_{12}$	2302.62	2303.27	2324.97	0.237	0.0003329
ARIMA[2,0,1][3,1,0]$_{12}$	2303.24	2303.9	2325.59	0.173	0.0002583
ARIMA[3,0,0][2,1,0]$_{12}$	2302.22	2302.7	2321.37	0.151	0.0001046
ARIMA[2,0,0][2,1,0]$_{12}$	2301.45	2301.79	2317.41	0.0637	0.000763
ARIMA[2,0,1][2,1,0]$_{12}$	2302.85	2303.33	2322.01	0.103	0.0002873
ARIMA[2,0,2][2,1,0]$_{12}$	**2294.81**	**2295.65**	**232035**	**0.745**	**0.001486**

Table 8 The results of all possible Seasonal ARIMA model

The ACF plot's tail-off might suggest the MA component's (q) presence. However, since the pattern might be covered by the AR pattern, similar models with MA components not equal to 0 have also been considered. The p-value of all candidate models is greater than 0.05, suggesting the lack of evidence to reject residuals are distributed independently, while the low p-value in the normality test of all models shows strong evidence against the residual as normal distribution in the model. However, if the other assumptions of seasonal ARIMA are met, the model might still predict a valid result. Therefore, based on the information criterion, the smaller the value is, the better the model fits the data; ARIMA [2,0,2][2,1,0]12 is selected.

B. The Collection of Figures in This Paper

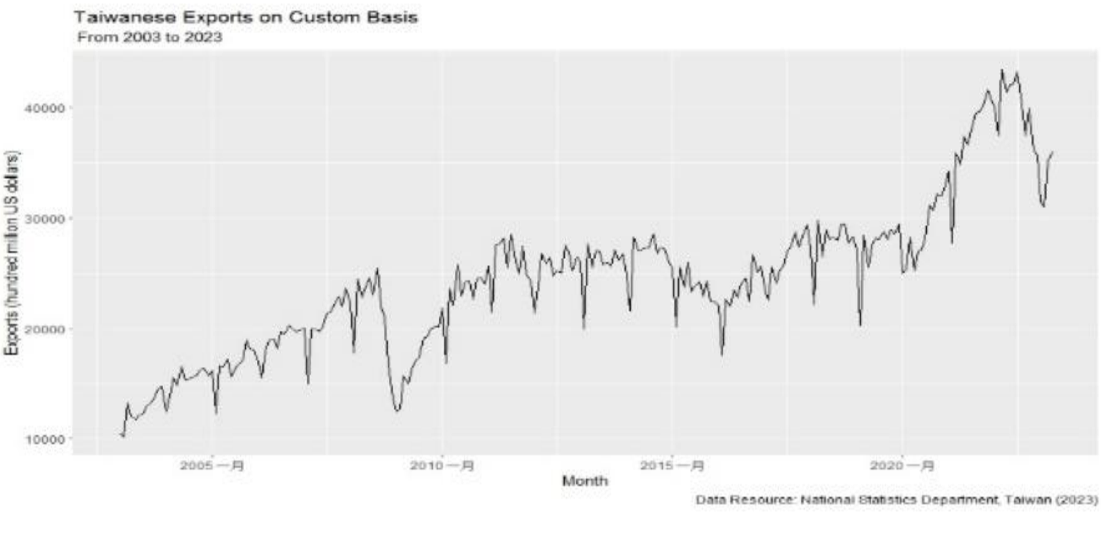

Figure 10 Taiwanese Exports on Custom Basis from Jan. 2003 to Apr. 2023

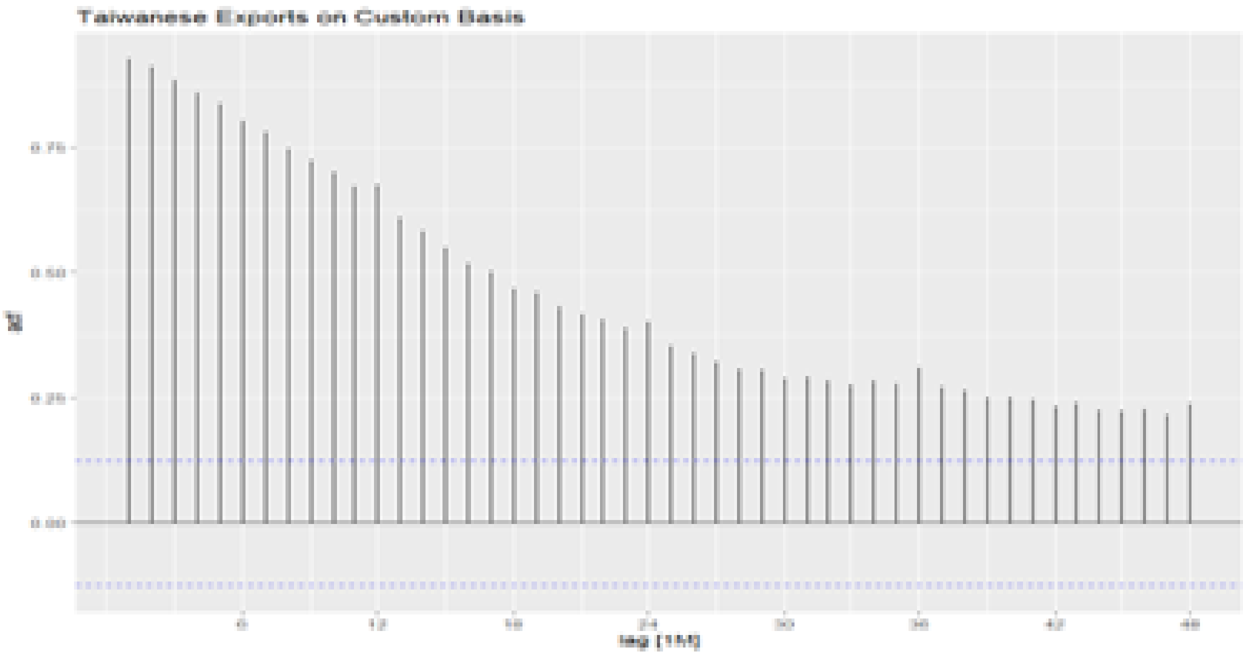

Figure 11 The ACF Plot of Taiwanese Exports on Custom Basis from Jan. 2003 to Apr. 2023

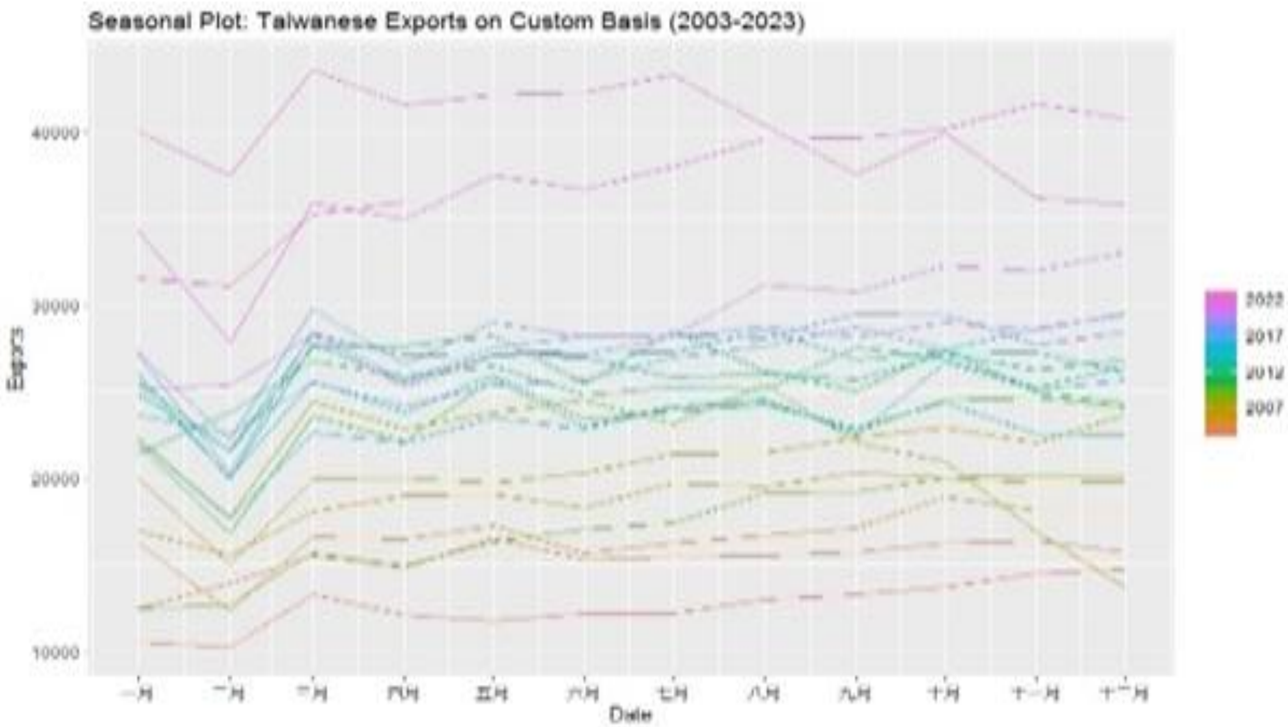

Figure 12 Seasonal Plot of Monthly Taiwanese Exports on Custom Basis from 2003 to 2023

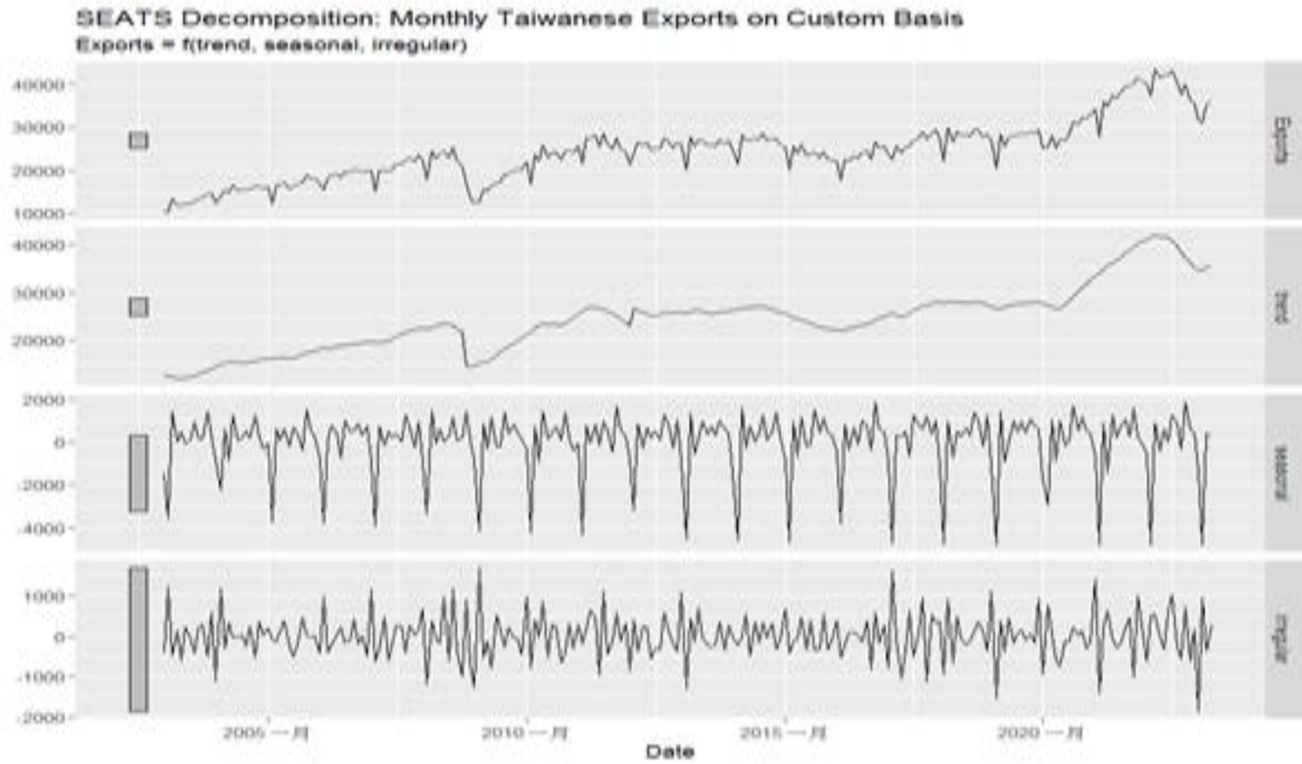

Figure 13 SEATS Decomposition Plot of the Monthly Taiwanese Exports

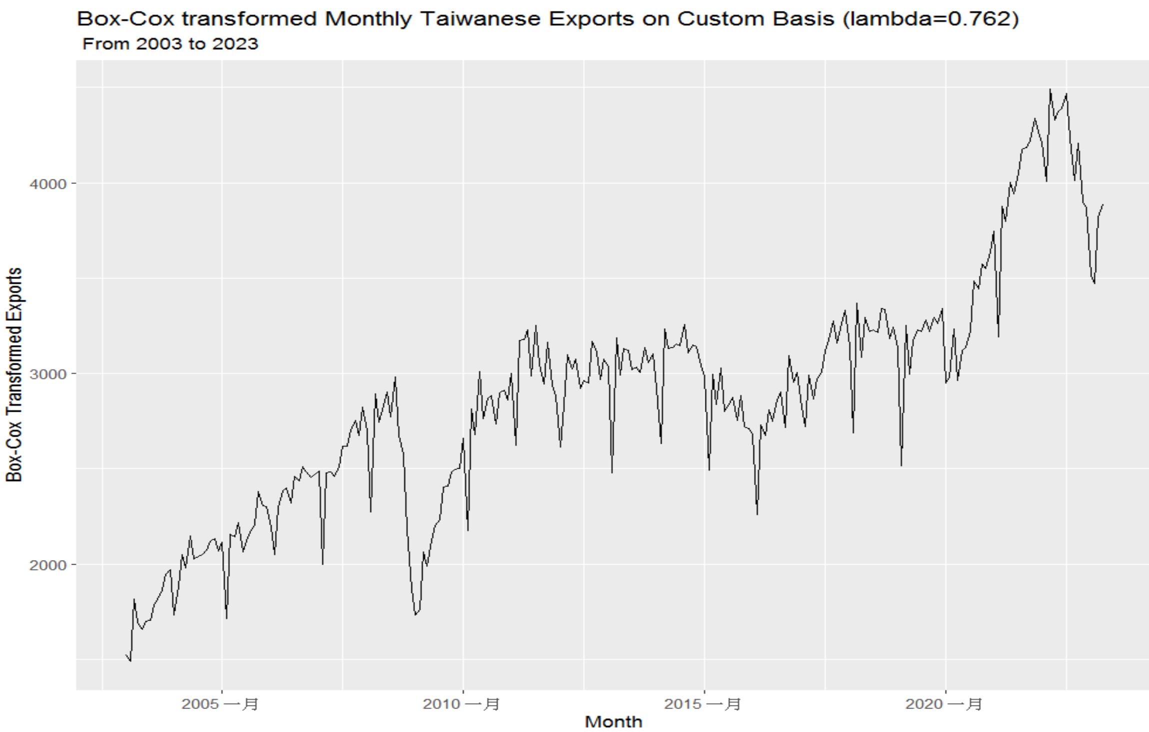

Figure 14 The Time Series of the Transformed Taiwanese Monthly Exports

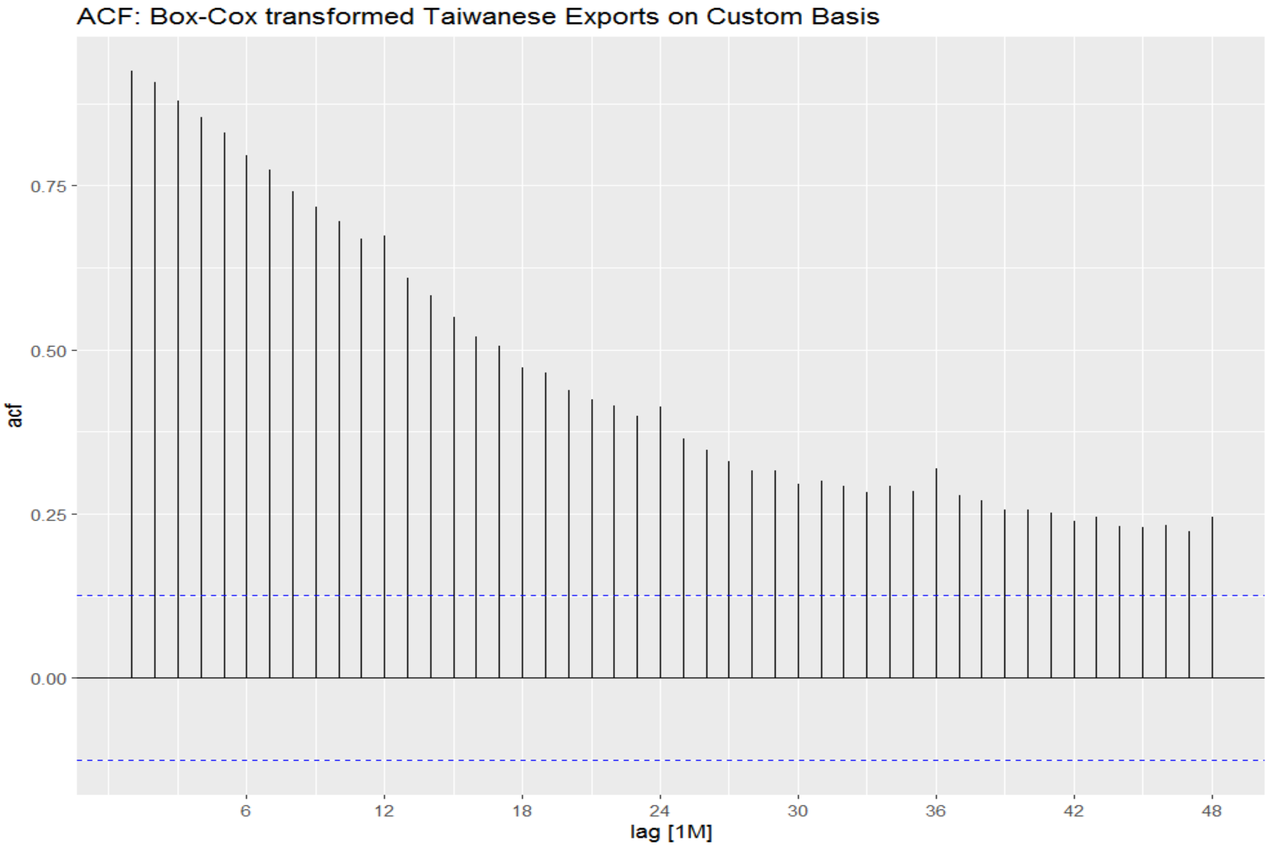

Figure 15 The ACF plot of the Transformed Taiwanese Monthly Exports

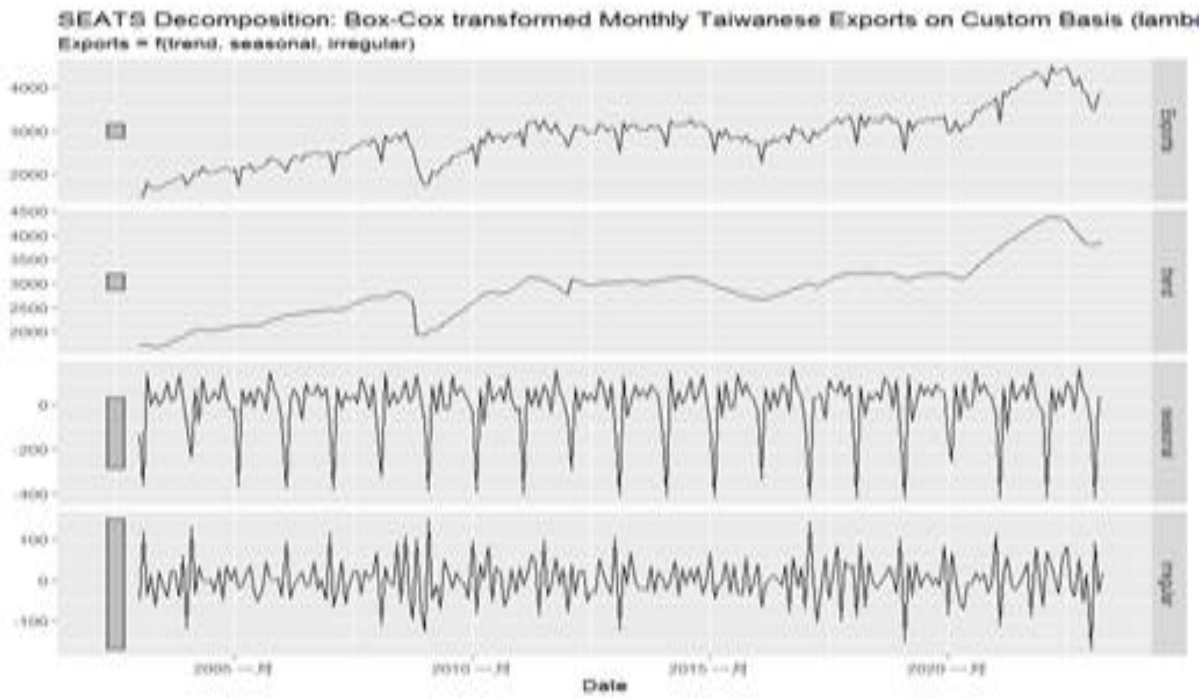

Figure 16 SEATS Plot of Box-Cox Transformed Monthly Taiwanese Exports

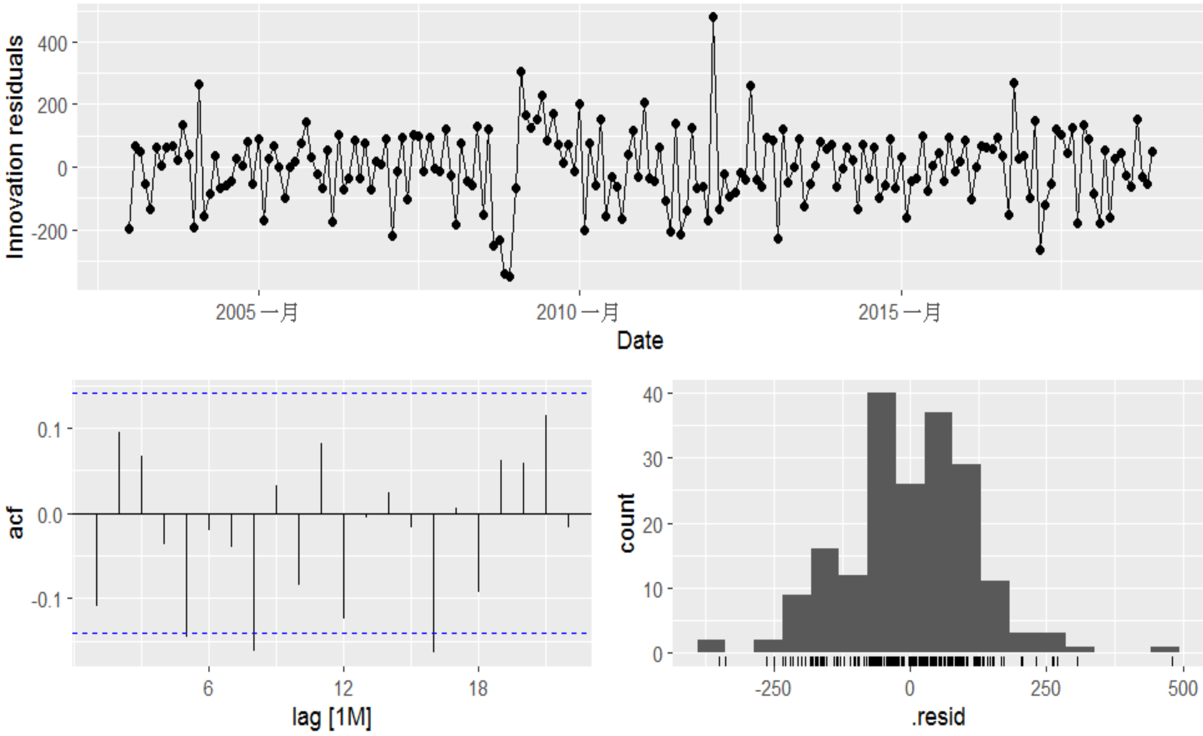

Figure 17 The residual check of the ETS(A, A, M) model

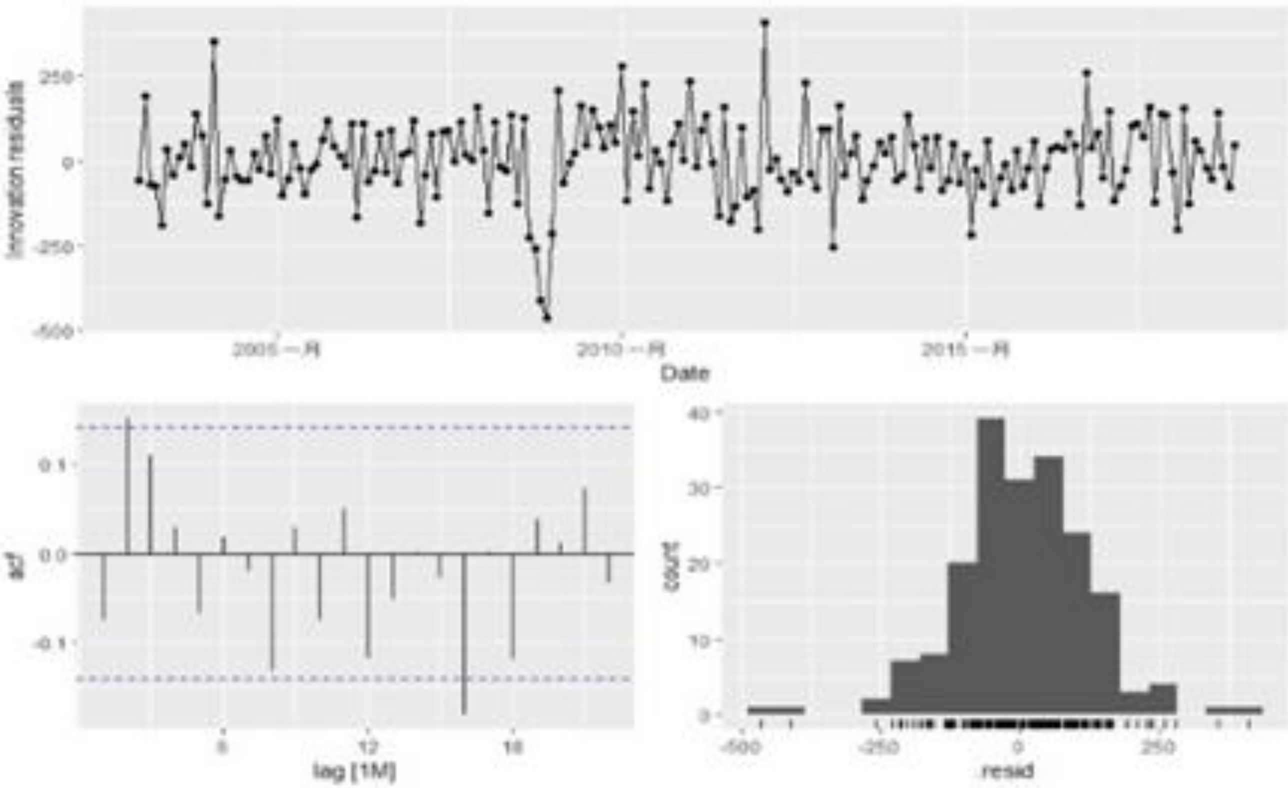

Figure 18 The residual check of the ETS(A, Ad, A) model

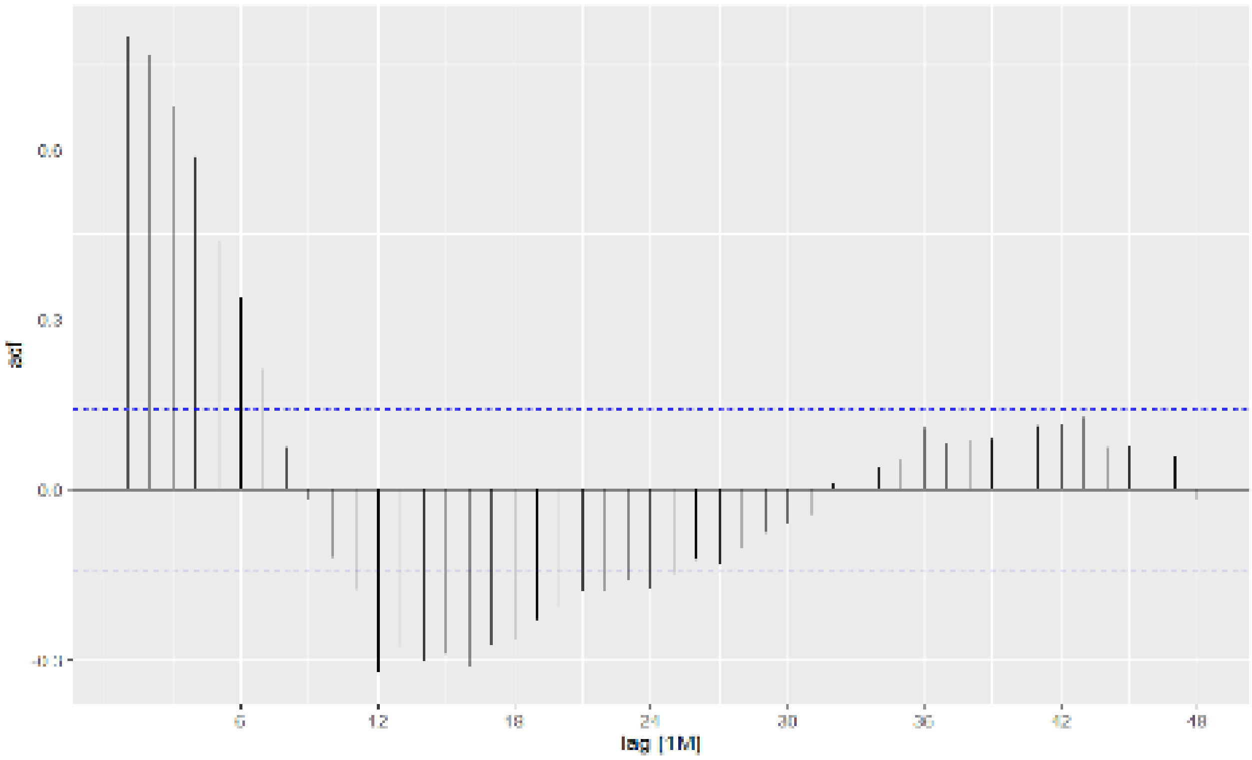

Figure 19 The AFC plot of Seasonal Differenced Transformed Exports

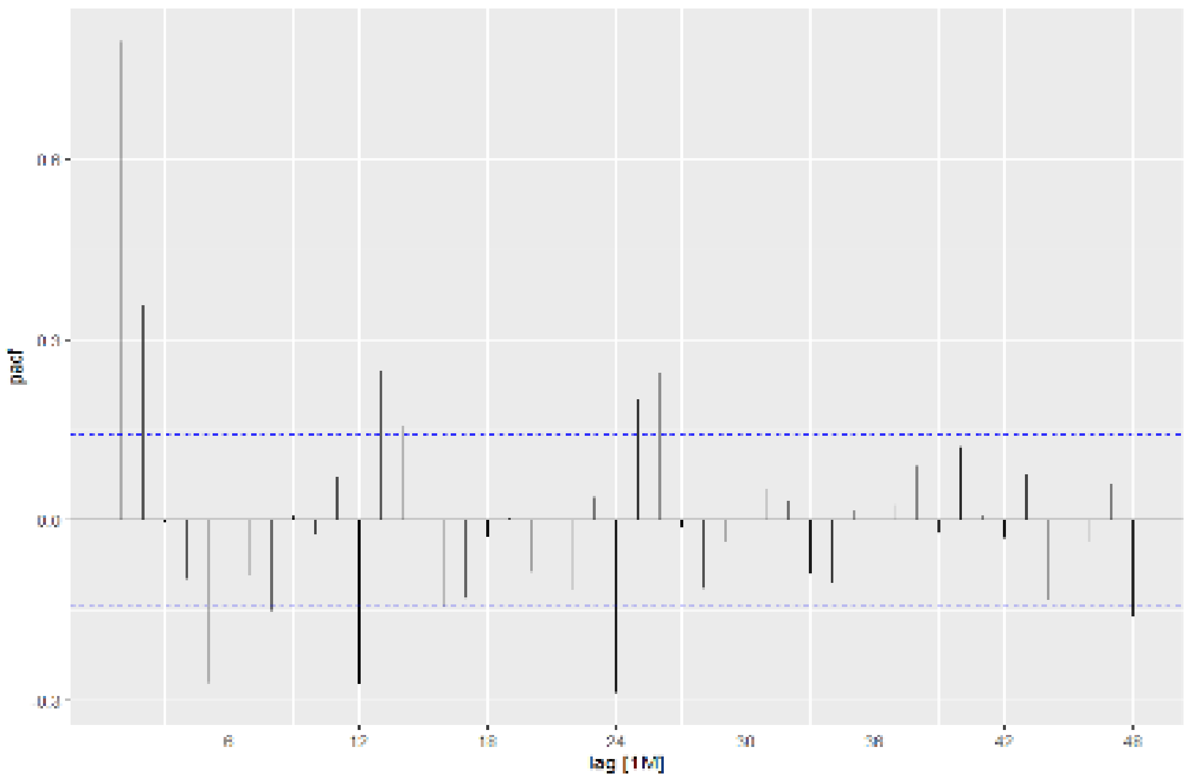

Figure 20 The PAFC plot of Seasonal Differenced Transformed Exports

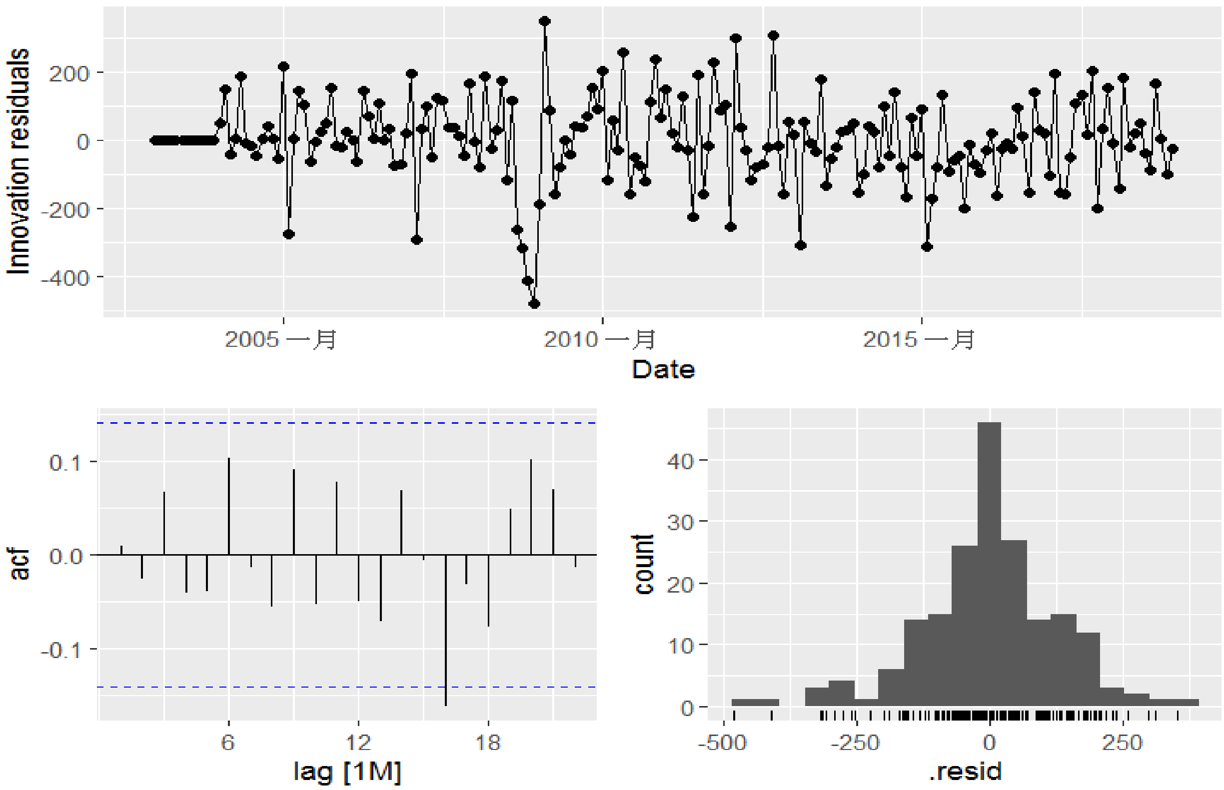

Figure 21 The residual check of Seasonal ARIMA [2,0,2][2,1,0]$_{12}$

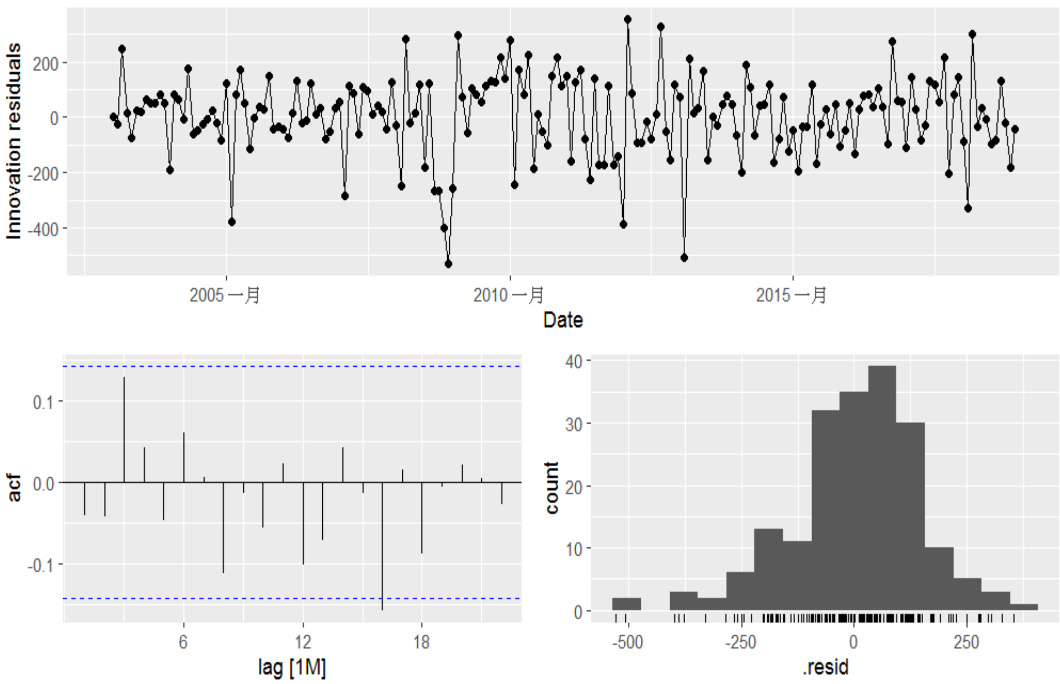

Figure 22 The residual check of seasonal ARIMA [1,1,0][2,0,0]$_{12}$ Model

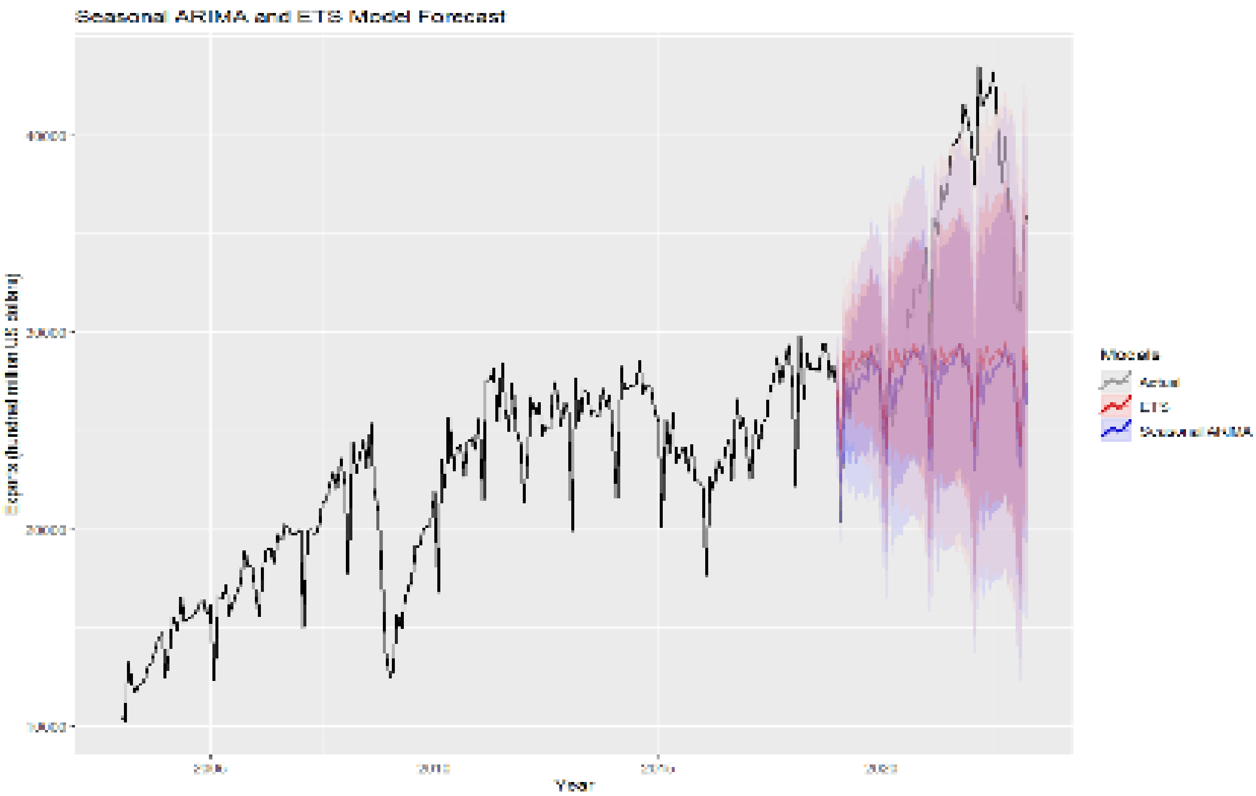

Figure 23 The Seasonal ARIMA and ETS Forecast

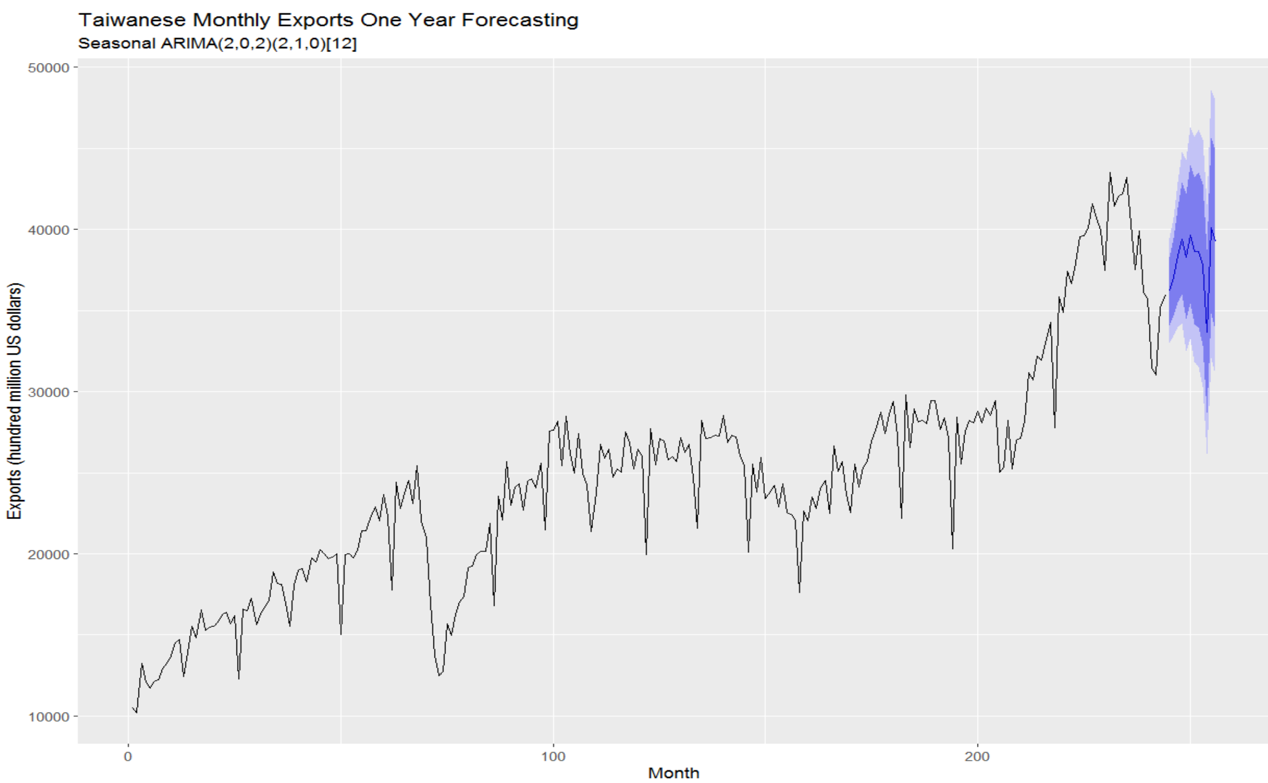

Figure 24 The One-Year Ahead Forecasting Based on Entire Data