Forecasting Food Price Index using SARIMA and Holt-Winters methods

CAS MA 585

Time Series and Forecasting

Joel Choe Yee Hsien U22167011 *(joelchoe@bu.edu)*

1. Introduction

According to the Food and Agriculture Organization (FAO), global food prices reached an all-time high in February 2022 [1]. Food Price Index is a measure which tracks the international prices of goods such as vegetable oils and dairy products [2]. Food Price Index measures the average change over time in the prices that households pay for food [3]. Over the years, forecasting the Food Price Index has become increasingly important due to the changing structure of food and agricultural economies and the important signals the forecasts provide to farmers, processors, wholesalers, consumers, and policymakers.

As such, this paper seeks to leverage on existing time series and forecasting methods to propose various time series models to forecast the Food Price Index for New Zealand into the year 2022. In this paper, we will describe and propose three models - two seasonal ARIMA models and a seasonal Holt-Winters model - that are able to achieve a high level of prediction reliability. From this research, we found that the seasonal Additive Holt-Winters model yields the best result outperforming the other two seasonal ARIMA models.

2. Data

In this project, we will focus on analyzing and forecasting the Food Price Index for New Zealand. The data set used was obtained from Kaggle [3]. The data set captures the change in prices that households in New Zealand pay for different categories of food. For this project, we have chosen to work with the 'General' category across the various categories of food rather than focus on a single category. While the data set captures information from January 1960 to August 2021, we have chosen to focus on the data from January 2000 to August 2021.

Upon a closer inspection of the data, there were no missing values or significant outliers in the data set. This can be seen from the boxplot in Figure 1.

In this project, we chose to split the data into a training and test set with a ratio of 4:1. There are 208 and 52 data points in the training and test set respectively.

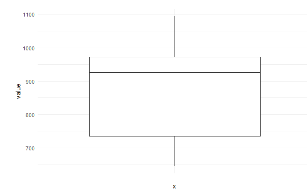

Figure 1: Boxplot of data

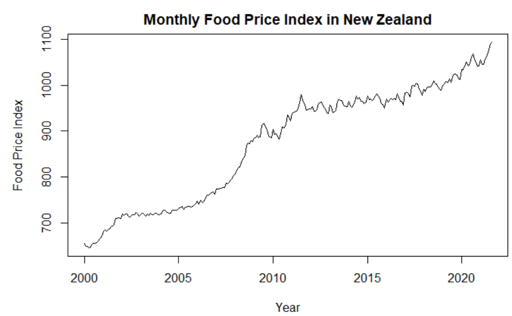

Figure 2: Monthly Food Price Index in New Zealand

3. Data Analysis

Prior to building any models, we examined the data set and observed certain attributes about the time series data. The maximum and minimum of the data are and 645.9184 and 1094 respectively, with a mean value of 874.7066. The plot of the data can be seen in Figure 2.

Observing the data plot (Fig 2), the time series data does not appear stationary. To ascertain the stationarity of the data, the Dicky-Fuller test and KPSS Unit Root test were conducted. The Dicky-Fuller test produced a p-value of $0.6633 > 0.05$, while the KPSS test produced a value of 0.01 < 0.05. Both tests suggest that the data is non-stationary. In the case of the Dicky-Fuller test, we do not have sufficient reason to reject the null hypothesis that the time series is non-stationary. In the case of the KPSS test, we reject the null hypothesis that the time series is stationary.

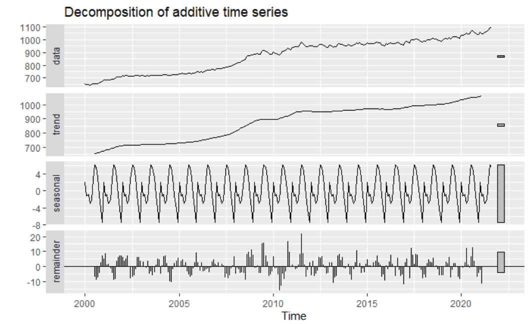

Figure 3: Classical Decomposition of data

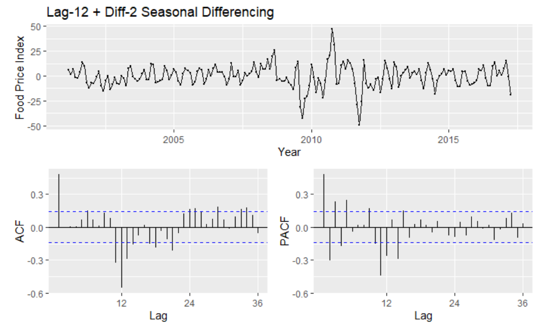

Figure 5: Data after differencing

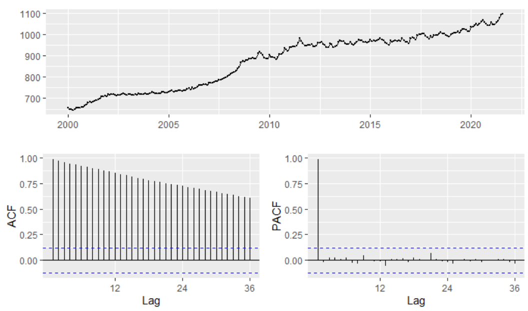

Figure 4: ACF and PACF of original data

From the data plot (Fig 2), there seems to be a quadratic trend which runs into a somewhat linear one around 2010. The trend component is an increasing one, with the possibility of a seasonal pattern. To ascertain on the trend and seasonality of the data, we produced the classical decomposition of the data in Figure 3.

From the classical decomposition of the data (Fig 3), it can be observed that there is a clear increasing trend and seasonality in the data. Concerning the seasonal pattern, we observe that its period is 12 months from the seasonal plot (Fig. 3). This makes sense given that the data is monthly. The seasonal variance seems to be relatively constant, suggesting that a variance stabilizing transformation is likely unnecessary. The ACF and PACF of the original data can be seen in Figure 4.

4. Transformation

To correct for seasonality (deterministic part) and unit root, we applied a difference of order 2 and order 12 to remove the trend and seasonal component respectively. A variance stabilizing transformation was not needed. The result of the above transformation can be seen in Figure 5.

As observed (Fig 5), the trend and seasonality component seems to be eliminated.

To ascertain whether the trend and seasonality component is indeed eliminated, the Dicky-Fuller test and KPSS test is replicated again. The Dicky-Fuller test produced a p-value of $0.01 < 0.05$, while the KPSS test produced a value of 0.1 > 0.05. This suggests that the differenced data is now stationary with its trend and stationary components eliminated.

5. SARIMA Model Fitting

5.1. Initial Models

To forecast future values of the time series, we first propose a seasonal ARIMA model. The model-based approach fits an ARIMA model to the data to then forecast future points.

From the analysis above and the information in the ACF and PACF plot (Fig 5), it seems reasonable to consider a low order ARMA model. The PACF cuts off near lag 3 decaying to 0 and the ACF cuts off after lag 1 decaying to 0. However, from the same plot, although there is an early cutoff, the lag significance returns around 12 months. This periodic pattern is a characteristic of seasonality with period 12 (stochastic part). This confirms what was observed in the seasonal component in Figure 3.

Based on the observation of the ACF and PACF plot in Figure 5, we propose an ARMA (3,2,1) x SARMA $(1,1,1)_{12}$ model. This makes sense because, for the non-seasonal ARMA(p, d, q) model, there is a decaying pattern in the ACF and PACF. Furthermore, for the seasonal ARMA(P, D, Q) model, there is a strong autocorrelation and strong partial autocorrelation around lag 12 and both cut off decaying to 0.

The results relating to the diagnostic checks of the ARMA (3,2,1) x SARMA $(1,1,1)_{12}$ model, containing the Normal Q-Q plot of residuals and various tests, is shown in Figure 6.

However, upon further analysis of the initially hypothesized SARIMA model, we identify that the Ljung-Box Statistic of the model is $0.001478 < 0.05$. Here, we reject the null hypothesis that the residuals are independently distributed, suggesting that the fit of the model might not be ideal. Moving on, we sought to identify another low order SARIMA model with better fit. To do so, we adjusted the parameters to ensure that the Ljung-Box Statistic is satisfied. With some

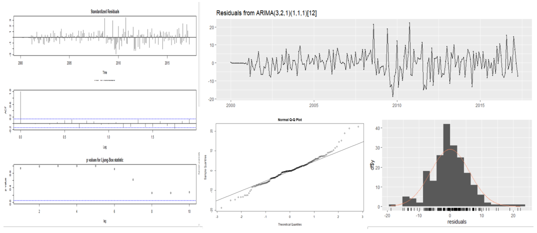

Figure 6: Diagnostic check of ARMA (3,2,1) x SARMA $(1,1,1)_{12}$ model

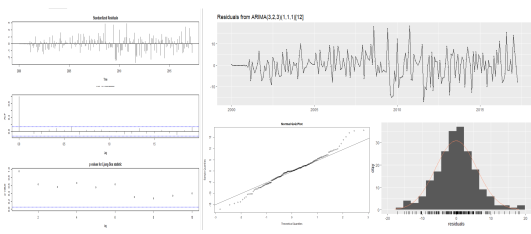

Figure 7: Diagnostic check of ARMA (3,2,3) x SARMA $(1,1,1)_{12}$ model

tuning of the parameters, we eventually chose an ARMA (3,2,3) x SARMA $(1,1,1)_{12}$ model. We will also attempt to fit the AR and MA model here. As such, the three possible model parameters we will consider are:

1. ARMA (3,2,3) x SARMA $(1,1,1)_{12}$
2. ARMA (3,2,0) x SARMA $(1,1,0)_{12}$
3. ARMA (0,2,3) x SARMA $(0,1,1)_{12}$

The results relating to the diagnostic checks of the various models are shown in Figures 7, 8 and 9. The forecasting results of the time series using the different models can be seen in Figure 10.

From these plots, we observe the following:

1. The standardized residuals look sufficiently random
2. The histogram of residuals and Q-Q plot of residuals show that the residuals are enough to be regarded as white noise

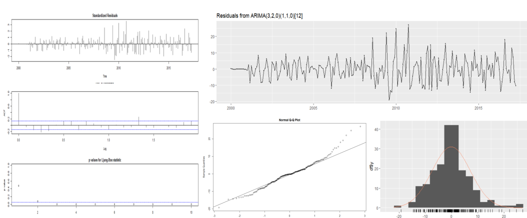

Figure 8: Diagnostic check of ARMA (3,2,0) x SARMA $(1,1,0)_{12}$ model

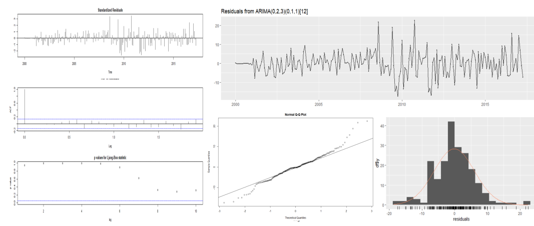

Figure 9: Diagnostic check of ARMA (0,2,3) x SARMA $(0,1,1)_{12}$ model

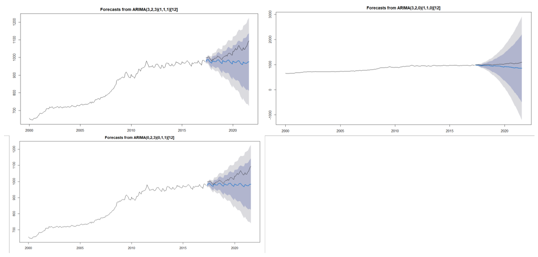

Figure 10: Prediction for the different models

3. The ARMA model's ACF of the residuals cut off after lag 0. However, the AR and MA model's ACF do not cut off after lag 0. Combined with the residuals checking, the AR and MA models may not be satisfactory fits to the data.
4. The p-value for the Ljung-Box Statistic is only satisfied in the case of the ARMA model. The p-value for the Ljung-Box Statistic is $0.08452 > 0.05$, whereas that of the AR and MA models are 8.561e-10 < 0.05 and $0.005267 < 0.05$ respectively. This indicates that the ARMA model will likely be more suited and fitted for forecasting, producing a better performance and result.

To ensure that we have indeed chosen the best model of the three proposed models, we calculated the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) metric of the three models. In addition to the AIC and BIC metric, we calculated the the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) of the training data as well. The results are shown in Table 1.

Model	AIC	BIC	RMSE	MAE
(3,2,3) x $(1,1,1)_{12}$	**1296.15**	**1324.58**	**5.928656**	**4.424966**
(3,2,0) x $(1,1,0)_{12}$	1356.13	1372.15	7.287281	5.388077
(0,2,3) x $(0,1,1)_{12}$	1299.48	1315.5	6.152684	4.504208

Table 1: AIC, BIC, RMSE and MAE scores of fitted models

From Table 1, it is clear that the ARMA (3,2,3) x SARMA $(1,1,1)_{12}$ model has the lowest AIC, BIC RMSE and MAE. Adding on to the fact that the Q-Q plot shows that the data comes from a normal population, and diagnostic plots indicates that the ACF of the residuals and p-values are not significant, the model, ARMA (3,2,3) x SARMA $(1,1,1)_{12}$, can be said to be adequate.

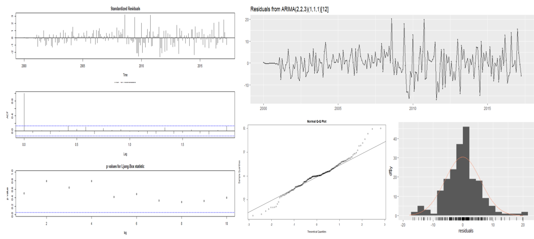

Figure 11: Diagnostic check of ARMA (2,2,3) x SARMA $(1,1,1)_{12}$ model

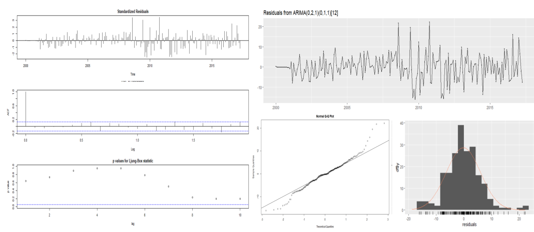

Figure 12: Diagnostic check of ARMA (0,2,1) x SARMA $(0,1,1)_{12}$ model

The estimated parameter values for the ARMA (3,2,3) x SARMA $(1,1,1)_{12}$ model is shown in Table 2.

	AR1	AR2	AR3	MA1	MA2	MA3	SAR1	SMA1
Coef	-0.6248	-0.9531	0.0202	-0.3726	0.4297	-0.9633	-0.1868	-0.6365
SE	0.0837	0.0576	0.0816	0.0477	0.0391	0.0549	0.1209	0.1137

Table 2: Estimated parameter values for the ARMA (3,2,3) x SARMA $(1,1,1)_{12}$ model

5.2. Brute Force Algorithm

Because there are multiple interpretations to take, to find the best model parameters, we used a brute force algorithm to traverse and search the different model parameters and recorded their performance using the AIC metric and BIC metric. The search grid used for the brute force algorithm was $0 \leq p, q \leq 4$ and $0 \leq P, Q \leq 1$.

Using the AIC metric, the best model proposed is the ARMA (2,2,3) x SARMA $(1,1,1)_{12}$ model with an AIC score of 1291.686; while that of the BIC metric is the ARMA (0,2,1) x SARMA $(0,1,1)_{12}$ model with a BIC score of 1305.516. The AIC and BIC scores of the two models beat that of the ARMA (3,2,3) x SARMA $(1,1,1)_{12}$ model proposed in 5.1 (Table 1). The diagnostic plots relating to the two models are shown in Figures 11 and 12. The forecasting results of the time series using the two models can be seen in Figures 13.

From these plots, we observe the following:

1. Similar to the observations from the initial models, the standardized residuals also look relatively random
2. The histogram of residuals and Q-Q plot of residuals show that the residuals are enough to be regarded as white noise

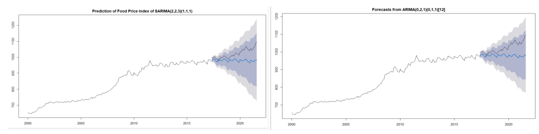

Figure 13: Prediction for the different models (Brute Force Algorithm)

3. Here, the ARMA (2,2,3) x SARMA $(1,1,1)_{12}$ model's ACF of the residuals cut off after lag 0. However, the ARMA (0,2,1) x SARMA $(0,1,1)_{12}$ model's ACF does not cut off after lag 0. Combined with the residuals checking,the ARMA (0,2,1) x SARMA $(0,1,1)_{12}$ model may not be a satisfactory fit to the data.
4. The p-value for the Ljung-Box Statistic is only satisfied in the case of the ARMA (2,2,3) x SARMA $(1,1,1)_{12}$ model. The p-value for the Ljung-Box Statistic is $0.1389 > 0.05$, whereas that of the ARMA (0,2,1) x SARMA $(0,1,1)_{12}$ model is $0.005076 < 0.05$. This indicates that the ARMA (2,2,3) x SARMA $(1,1,1)_{12}$ model will likely be more suited and fitted for forecasting, producing a better performance and result.

Moving forward, we will also consider the ARMA (2,2,3) x SARMA $(1,1,1)_{12}$ model.

The estimated parameter values for the ARMA (2,2,3) x SARMA $(1,1,1)_{12}$ model is shown in Table 3.

	AR1	AR2	MA1	MA2	MA3	SAR1	SMA1
Coef	-0.7915	-0.9988	-0.1563	0.2243	-0.9606	-0.1975	-0.6164
SE	0.0105	0.0032	0.0522	0.0579	0.0650	0.1177	0.1145

Table 3: Estimated parameter values for the ARMA (2,2,3) x SARMA $(1,1,1)_{12}$ model

6. Holt-Winters Model

When the data have both trending and seasonal components, the Holt-Winters model may have adequate performance in fitting and forecasting. Therefore, in this section, we will consider building a Holt-Winters model.

The Holt-Winters model is a smoothing-based method that uses the pattern of the data to extrapolate the forecast using double exponential smoothing. Here, we performed the additive version as the seasonal pattern remained roughly similar across the range of the data. The result of the forecast can be seen in Figure 14.

In Figure 14, the black curve is the original data points for both the training and the test set while the blue curve is the prediction of the Additive Holt-Winters model. Included is the 80% and 95% confidence interval of the forecast. From the figure, the actual values of the test set can also be observed to lie within the 95% confidence interval which seem to suggest that the forecast is reasonable.

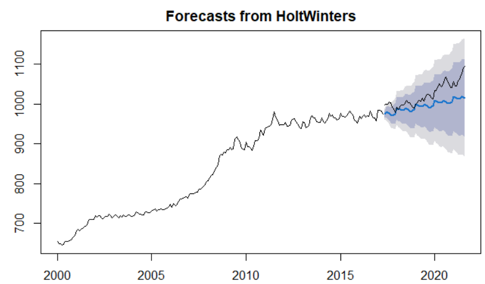

Figure 14: Holt-Winters Forecast

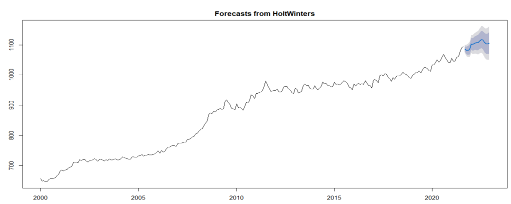

Figure 15: 2022 Forecasts (Holt-Winters Forecasts)

7. Results

From our above analysis, we have three candidate models, namely the ARMA (3,2,3) x SARMA $(1,1,1)_{12}$ model, ARMA (2,2,3) x SARMA $(1,1,1)_{12}$ model and Holt-Winters Additive model. While the seasonal ARIMA and seasonal Holt-Winters models look reasonable, forecast accuracy is also an important aspect in determining an appropriate model. To compare the performance of the three models, we ran the models on the test set and compared their RMSE, MAE, and Mean absolute Percentage Error (MAPE) to evaluate the quality of forecast of the different models. The results can be observed in Table 4.

Model	RMSE	MAE	MAPE
$(3,2,3) \times (1,1,1)_{12}$	56.463816	47.006110	4.5170831
$(2,2,3) \times (1,1,1)_{12}$	51.987188	42.906786	4.1211026
Holt-Winters Additive Model	**33.135908**	**27.546730**	**2.6490282**

Table 4: RMSE, MAE and MAPE scores of chosen models

From the results, it can be observed that the seasonal Holt-Winters model performed better than the two seasonal ARIMA models. The best model from our analysis is the Holt-Winters Additive model.

8. 2022 Food Price Index Forecast

To forecast the 2022 Food Price Index, we chose the best model based on the performance of the test set and retrained the model using the full data set. The chosen model was the Holt-Winters Additive model as shown in Section 7. The result of the forecast can be seen in Figure 15 and Table 5.

Period	Point Forecast	95% Low	95% High
Sep 2021	1086.144	1072.299	1099.990
Oct 2021	1081.690	1063.187	1100.192
Nov 2021	1081.446	1059.149	1103.744
Dec 2021	1084.636	1059.017	1110.254
Jan 2022	1102.780	1074.149	1131.411
Feb 2022	1101.285	1069.860	1132.711
Mar 2022	1104.886	1070.830	1138.942
Apr 2022	1106.906	1070.347	1143.466
May 2022	1107.384	1068.423	1146.344
Jun 2022	1112.941	1071.663	1154.218
Jul 2022	1117.949	1074.424	1161.474
Aug 2022	1115.895	1070.181	1161.609
Sep 2022	1108.039	1059.642	1156.436
Oct 2022	1103.584	1053.114	1154.055
Nov 2022	1103.341	1050.833	1155.849
Dec 2022	1106.530	1052.015	1161.046

Table 5: 2022 Forecasted Values and Confidence Intervals

9. Conclusion

The purpose of this time series analysis was to forecast the Food Price Index for New Zealand into the year 2022. In this paper, we have considered both possible seasonal ARIMA models and seasonal Holt-Winters models for forecasting; eventually choosing the best model based on a held out test set. In summary, the chosen seasonal ARIMA models were selected by considering the AIC and BIC values among other ARIMA models and eventually compared against the Holt-Winters method using the RMSE, MAE and MAPE. The results showed that the Holt-Winters Additive model best represented a suitable forecasting method for Food Price Index, serving the purpose of our project.

Moving forward, one can consider other time series models such as higher order seasonal ARIMA models, Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors (SARIMAX) models or Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH) models [4]. This project can also be extended to other countries, food products, and commodities.

References

[1] Emsden, C. (2022, March 4). FAO Food Price Index rises to record high in February. Retrived from https://www.fao.org/newsroom/detail/fao-food-price-index-rises-to-record-high-in-february/en.
[2] -. (-). FAO Food Price Index. Retrived from https://www.fao.org/worldfoodsituation/foodpricesindex/en/.
[3] Muhammad, A. (2021, October 1). Food Price Index: August 2021. Retrieved from https://www.kaggle.com/datasets/intelai/food-price-index-august-2021.
[4] Korstanje, J. (2021, July). Advanced Forecasting with Python. https://link.springer.com/chapter/10.1007/978-1-4842-7150-6$_8$.