In Partial Fulfillment

of the Requirements for

TIMESER N01

Using ARIMA Modeling to Forecast Peso-Dollar Exchange Rates (Non-Seasonal) and Monthly US Industrial Candy Production Indices (Seasonal)

Submitted by:
Manay, Justin Gabrielle
Marchadesch, Claudynne Kevynne

Submitted to:
Frumencio Co

I. NON-SEASONAL

(Monthly Peso-Dollar Exchange Rates)

1. Introduction

The (peso-dollar) exchange rate is simply the price of one dollar in peso terms. The Bangko Sentral ng Pilipinas (BSP) outlines several reasons as to why exchange rates (and by extension, exchange rate forecasting) are important to take note of.

For one, the peso-dollar exchange rate links domestic prices to the price of goods and services abroad. Knowledge of exchange rates is of great importance in conducting business with international clients. Many such transactions are often settled in the future and as a business owner or employee, one must be able to project exchange rates for the future in order to properly evaluate the benefits and costs of foreign dealings. Moreover, certain firms make a profit from speculating on exchange rate trends, thus necessating the creation of exchange rate forecasting models.

Exchange rate movements have also been linked to inflation, as a weaker peso can increase the domestic value of imported goods and services. Certain commodities such as rice and gasoline are primarily imported from other countries, so weakening pressure on the peso can potentially increase the prices of these goods, making life more difficult for the average Filipino. As such, the BSP also need exchange rate models to avoid the exchange rate from weakening so much as to affect inflation severely.

With regards to our dealings with other governments, exchange rates can also affect the cost of servicing foreign debt and can also impact the competitiveness of our exports. This makes it all the more imperative for the government to have exchange rate forecasting models in place in order to adequately plan for the future.

Thus, this study aims to develop a preliminary forecasting model for peso-dollar exchange rates using time series analysis. Meese and Rogoff (1983) were the first to suggest that a random walk model forecasts exchange rates better than other economic models. This follows from the weak form of the efficient markets hypothesis (EMH), which states that past asset prices will completely reflect all available information about a certain asset in an efficient market.

Empirical evidence, however, has gone on to suggest that financial markets are in fact inefficient, making the random walk model an ineffective forecasting tool. Hsieh (1988) determined that exchange rates tend to be leptokurtic/fat-tailed; that is, large changes in the exchange rates are more probable than they would have been if exchange rates were normally distributed. Moosa and Burns (2014) also suggest that while the random walk outperforms other sophisticated exchange rate models in terms of forecasting accuracy, the other models tended to perform better in terms of direction accuracy and profitability. Ca'Zorzi and Rubaszek (2018) also suggest that while the same models fail to beat the random walk model over longer periods, they are able to outforecast the simple random walk model at medium term horizons.

Despite these claims, the use of ARIMA models (which the random walk model is a subset of) still persists in the estimation of exchange rate models. Hence, in our case, we will employ ARIMA models in order to build a satisfactory exchange rate model.

2. Data Description

The data is from the Bangko Sentral ng Pilipinas and it includes Philippine peso-US dollar exchange rates at the end of every month, from January 2004 to January 2019. While the primary determinants of exchange rates (e.g., money supply, interest rates) tend to exhibit seasonal behavior, empirical studies have largely shown that these seasonal determinants generate exchange rate data that is largely non-stationary (Jiménez-Martín & Flores de Frutos, 2009). Moreover, empirical studies have demonstrated the effectiveness of using random walk models (which are non-stationary time series models) in forecasting exchange rates.

3. Results and Discussion

Table 3.1: Initial Indicators

Name of Variable = exch	
Mean of Working Series	47.7012
Standard Deviation	4.502206
Number of Observations	181

From Table 3.1, the total number of observations of the variable `exch` is 181. It is recommended that the number of lags be at least 25% of the series length. Thus, we choose $\frac{181}{4} = 45.25$ lags. We round this number up to 48 observations in order to observe at least 3 seasonal lags. We also find that the mean is 47.7012 and the standard deviation 4.502206.

Table 3.2: Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	899.39	6	<.0001	0.974	0.944	0.912	0.880	0.848	0.813
12	1456.81	12	<.0001	0.779	0.744	0.710	0.676	0.639	0.602
18	1751.75	18	<.0001	0.568	0.540	0.512	0.479	0.447	0.414
24	1871.38	24	<.0001	0.382	0.350	0.319	0.289	0.261	0.236
30	1902.63	30	<.0001	0.211	0.186	0.162	0.138	0.114	0.090
36	1905.26	36	<.0001	0.072	0.057	0.045	0.032	0.019	0.005
42	1908.35	42	<.0001	-0.011	-0.025	-0.038	-0.051	-0.060	-0.068
48	1919.23	48	<.0001	-0.076	-0.085	-0.089	-0.091	-0.090	-0.084

We check if the defects are white noise by looking at the results of the autocorrelation check for white noise, shown in Table 3.2. All the p-values are less than 0.05. Therefore, we conclude that the defects are not white noise.

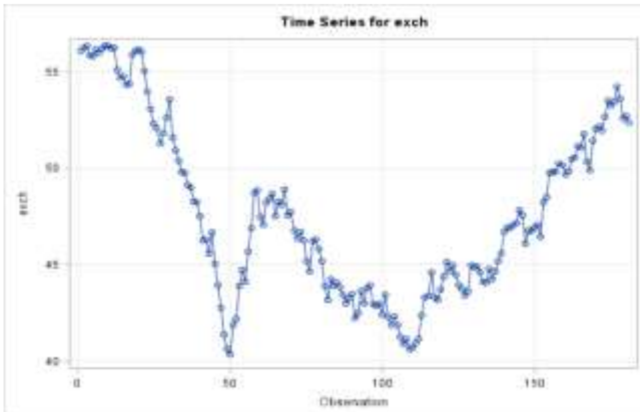

Figure 3.1: Time series plot for exch

Since exch is not a white noise, we can proceed with fitting a model. In order to fit a model, we look at the time plot to check for stationarity. As per the time plot in Figure 3.1, we see that it has a trend and is therefore non-stationary.

Table 3.3: Augmented Dickey-Fuller unit root tests for exch

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-0.1102	0.6569	-0.53	0.4873		
	1	-0.1217	0.6543	-0.52	0.4893		
	2	-0.1266	0.6532	-0.52	0.4904		
Single Mean	0	-3.6852	0.5718	-1.66	0.4505	1.45	0.7024
	1	-4.4750	0.4840	-1.78	0.3874	1.66	0.6489
	2	-4.9243	0.4383	-1.85	0.3568	1.77	0.6205
Trend	0	-2.4449	0.9567	-1.06	0.9315	3.18	0.5423
	1	-3.1371	0.9299	-1.24	0.8987	3.03	0.5724
	2	-3.4567	0.9146	-1.32	0.8806	3.05	0.5679

Therefore, we cross-reference with the Augmented Dickey-Fuller unit root test whose results are shown in Table 3.3 and find that the tau p-values (for trend and single mean, lag = 0) are greater than 0.05, indicating that the series is not stationary.

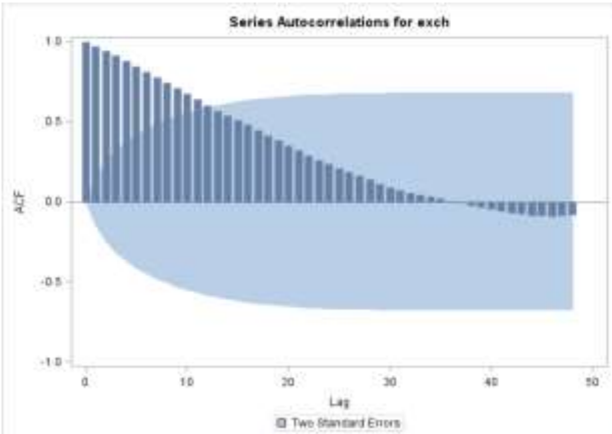

Figure 3.2: ACF plot for exch

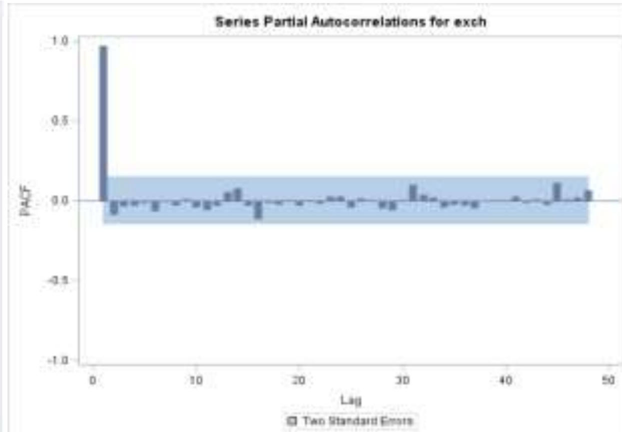

Figure 3.3: PACF plot for exch

In order to determine the ARIMA model to be used, we look at the ACF (Figure 3.2) and PACF (Figure 3.3) plots. The ACF plot exponentially decays slowly with increasing lags. This shows that the time series must be put through ordinary differencing.

Table 3.4: Autocorrelation Check for White Noise for exch(1)

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	4.43	6	0.6182	0.105	0.051	0.027	-0.030	0.093	0.008
12	6.73	12	0.8750	-0.006	-0.022	0.067	0.076	0.030	-0.014
18	16.85	18	0.5334	-0.104	-0.028	0.142	0.033	-0.128	-0.043
24	22.17	24	0.5691	0.013	0.005	0.080	-0.037	-0.057	0.119
30	24.54	30	0.7469	-0.015	0.050	0.021	-0.039	-0.031	-0.073
36	28.48	36	0.8098	-0.054	-0.052	0.062	0.059	0.016	0.066
42	31.02	42	0.8940	0.037	-0.023	0.032	-0.017	-0.035	0.080
48	33.23	48	0.9482	0.009	-0.033	0.028	-0.034	-0.046	0.061

After ordinary differencing, from Table 3.4, we see that the p-values are greater than 0.05. Thus, the residuals are now white noise.

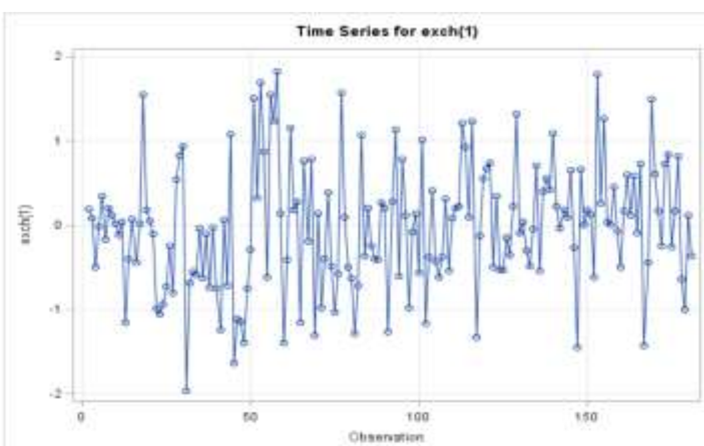

Figure 3.4: Time series plot for exch(1)

Table 3.5: Augmented Dickey-Fuller unit root tests for exch(1)

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-160.122	0.0001	-12.00	<.0001		
	1	-146.542	0.0001	-8.51	<.0001		
	2	-138.082	0.0001	-6.97	<.0001		
Single Mean	0	-160.242	0.0001	-11.97	<.0001	71.65	0.0010
	1	-146.842	0.0001	-8.49	<.0001	36.08	0.0010
	2	-138.461	0.0001	-6.96	<.0001	24.23	0.0010
Trend	0	-165.243	0.0001	-12.27	<.0001	75.33	0.0010
	1	-159.456	0.0001	-8.83	<.0001	38.95	0.0010
	2	-161.773	0.0001	-7.30	<.0001	26.63	0.0010

In addition, looking at the time plot in Figure 3.4, the series now appears to be more stationary. This is confirmed by looking at the Dickey-Fuller unit root test in Table 3.5, wherein all the p-values are less than 0.05, showing that the series is now stationary.

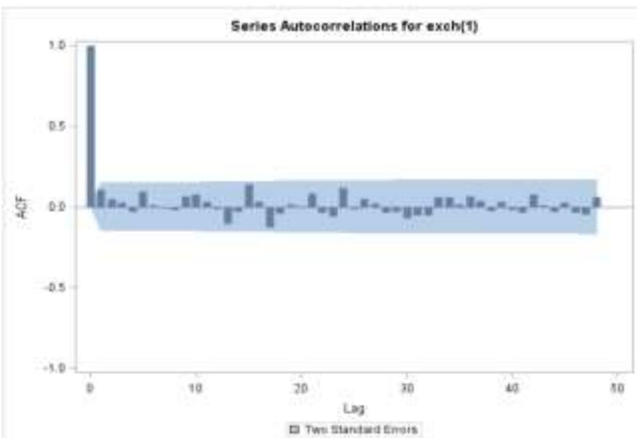

Figure 3.5: ACF plot for exch(1)

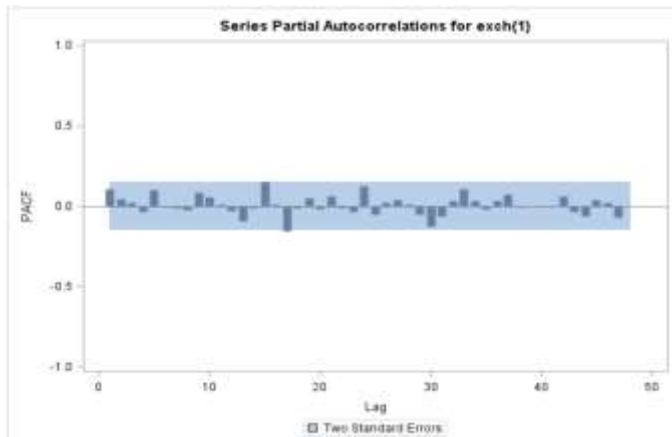

Figure 3.6: PACF plot for exch(1)

Further, we also look at the ACF (Figure 3.5) and PACF (Figure 3.6) plots and find that they are both exhibiting the behavior of white noise, with most lags being within the 95% confidence level. Thus, we now proceed with fitting an ARIMA(0,1,0) model.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.02074	0.05582	-0.37	0.7107	0

Constant Estimate	-0.02074
Variance Estimate	0.560944
Std Error Estimate	0.748962
AIC	407.7508
SBC	410.9437
Number of Residuals	180

Table 3.6: Parameter Estimates for ARIMA(0,1,0) model

Running a plot estimate, we find that the constant term MU (Table 3.6) is insignificant, with the AIC and SBC having the values 407.7508 and 410.9437 respectively. Since the constant term is insignificant, it was taken out of the model.

Table 3.7: Parameter Estimates for ARIMA(0,1,0), θ_0 = 0

Model for variable exch	
Estimated Mean	-0.02074
Period(s) of Differencing	1

Variance Estimate	0.558257
Std Error Estimate	0.747166
AIC	405.8895
SBC	405.8895
Number of Residuals	180

The AIC and SBC (table 3.7) are now less than before and we are left with a stochastic random walk model.

Table 3.8 Forecasts for exch (snippet)

Obs	exch	FORECAST	STD	L95	U95	RESIDUAL
182	.	52.352	0.74717	50.8876	53.8164	.
183	.	52.352	1.05665	50.2810	54.4230	.
184	.	52.352	1.29413	49.8156	54.8884	.
185	.	52.352	1.49433	49.4232	55.2808	.
186	.	52.352	1.67071	49.0775	55.6265	.
187	.	52.352	1.83018	48.7649	55.9391	.
188	.	52.352	1.97682	48.4775	56.2265	.
189	.	52.352	2.11331	48.2100	56.4940	.
190	.	52.352	2.24150	47.9587	56.7453	.
191	.	52.352	2.36275	47.7211	56.9829	.
192	.	52.352	2.47807	47.4951	57.2089	.
193	.	52.352	2.58826	47.2791	57.4249	.

After finalizing the model, we run a simple forecast. Table 3.8 shows the actual and forecasted values from the ARIMA(0,1,0) model. Note that because of the choice of the model, the forecast is equal to the previous actual value (or the previous forecast if no value exists for the actual value). Thus, the multi-step forecast is useful only for February 2019 (observation 182). Based on the model, the forecast for March 2019 should be the actual value of the exchange rate on February 2019, the forecast for April 2019 should be the actual value of the exchange rate on March 2019 and so on.

4. Summary and Conclusions

We find that the final model is an ARIMA(0,1,0) model (i.e., a stochastic random walk) given by

$$\widehat{Z}_t = Z_{t-1}$$

This confirms the findings by Meese and Rogoff (1983) that the random walk model is an accurate representation of exchange rate, given the criteria for ARIMA model selection.

However, this study is limited only to ARIMA models and as such, we cannot evaluate the other models which have been found to beat the random walk model.

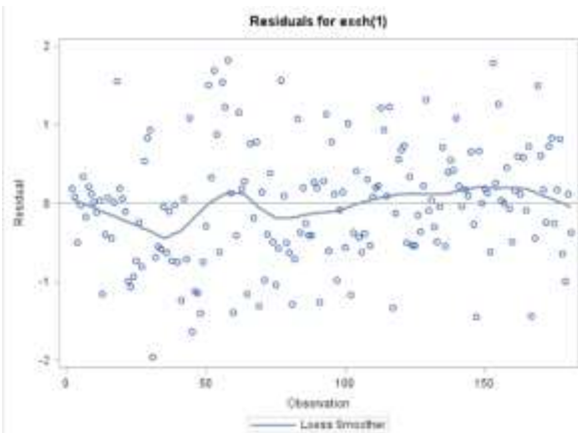

Figure 4.1: Residuals for exch(1) after ARIMA(0,1,0) estimation

As an example, Figure 4.1 suggests that the residuals are not homoscedastic. For this case, we can apply an ARCH model which takes care of the (conditional) heteroscedasticity in the model. However, since we limit the study only to ARIMA models, future research on this area can employ ARCH/GARCH models to take into account the residual heteroscedasticity.

II. SEASONAL

(US Monthly Industrial Candy Production Index)

1. Introduction

The confectionery industry in the United States earns about $36 billion in retail sales and $1.8 billion in exports, generating an economic output of around $44.6 billion in total. Because of its massive size, the confectionery industry needs a large amount of manpower – around 607,660 jobs in total. In fact, due to its sheer size, it is estimated that one US confectionery manufacturing job supports ten other jobs in the US economy, based on a Forbes article by Winstein (2018).

Thus, due to its vital stronghold in the American industry, it is imperative that we track the growth of the candy industry, as a sudden halt can spell disaster for the economy. For one, we can examine trends, and how the candy industry adapts/fails to adapt to them.

According to Grand View Research's January 2018 study, increasing consumer spending and growing urbanization will drive the growth of the candy market in the US, potentially fostering greater industrial candy production. With more and more variants of candy becoming available to the general public, Grand View Research expects the demand for candy to increase particuarly for children and the young population in general, notwithstanding growing health and sustainability concerns among consumers.

However, it is these concerns which Mouhanna (2018) from Market Research primarily warns the candy industry against. Consumers nowadays tend to keep an eye on wellness trends, opting to purchase more nutritious and organic ingredients over the artificial and processed ones that likely go into candy. More eco-friendly consumers have also started holding companies accountable for their sustainability strategies, calling for increased transparency and proactiveness regarding environmentally-friendly measures even in the confectionery industry. As such, these can be causes for concern for the candy industry as these can decrease the overall demand for their products, thereby decreasing industrial production.

The problem with trends, however, is that they tend to be fickle and subjective for the most part. Trends may come and go, but the confectionery industry never stops rolling out candy production. Thus, as a proxy for growth, it may be better to forecast industrial candy production.

Forecasting is an undeniably useful tool in many business operations as it allows business owners to make decisions ahead of time. However, in this case, forecasting is important as it allows us to foresee spikes and troughs in candy production, allowing policy makers to adjust and plan a course of action accordingly, given the confectionery industry's place in the economy.

Thus, the following study aims to determine industrial candy production through time series models (particularly ARIMA models).

2. Data Description

The data is from the Federal Reserve Bank of St. Louis and it includes US industrial candy production index at the end of every month, from January 1972 to August 2017. From the website, the industrial production (IP) index "measures the real output of all relevant establishments located in the United States, regardless of ownership." Since candy demand (and possibly candy production) spikes during the months of November (Halloween), December (Christmas) and January (New Year's), we expect the data to be seasonal in nature.

3. Results and Discussion

Table 3.1: Initial Indicators

Name of Variable = prod	
Mean of Working Series	100.6625
Standard Deviation	18.03645
Number of Observations	548

Table 3.1 tells us that the total number of observations of `prod` is 548. It is recommended that the number of lags taken into consideration must be at least 25% of the total observation. Thus, we choose $\frac{548}{4} = 137$ lags. Further, we find that the mean is 100.6625, while the standard deviation is 18.03645.

Table 3.2: Autocorrelation Check for White Noise (snippet)

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	865.22	6	<.0001	0.874	0.683	0.453	0.287	0.173	0.141
12	2050.60	12	<.0001	0.157	0.260	0.412	0.626	0.802	0.903
18	2742.39	18	<.0001	0.792	0.612	0.387	0.229	0.122	0.089
24	3754.57	24	<.0001	0.105	0.208	0.356	0.570	0.743	0.839
30	4342.20	30	<.0001	0.734	0.560	0.341	0.193	0.091	0.061
36	5257.67	36	<.0001	0.082	0.182	0.328	0.536	0.703	0.793
42	5768.17	42	<.0001	0.692	0.518	0.301	0.156	0.047	0.013
48	6511.66	48	<.0001	0.032	0.128	0.289	0.475	0.631	0.723
54	6929.90	54	<.0001	0.629	0.464	0.258	0.118	0.014	-0.018
60	7585.52	60	<.0001	0.000	0.093	0.237	0.436	0.589	0.680

Checking the autocorrelation check for white noise in Table 3.2, we find that all the p-values are less than 0.05. Therefore, the defects are not white noise. Thus, we proceed with the ARIMA procedure.

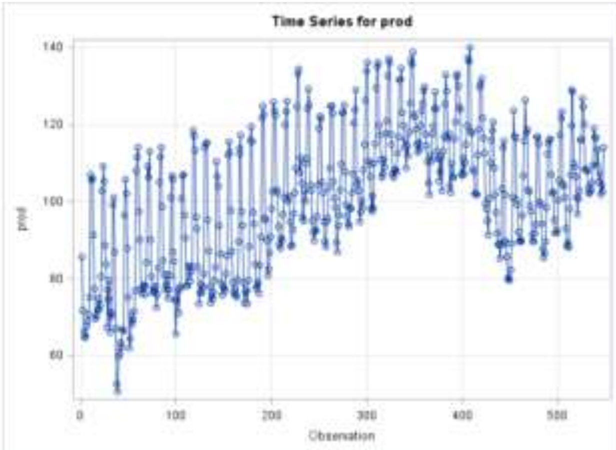

Figure 3.1: Time series plot for prod

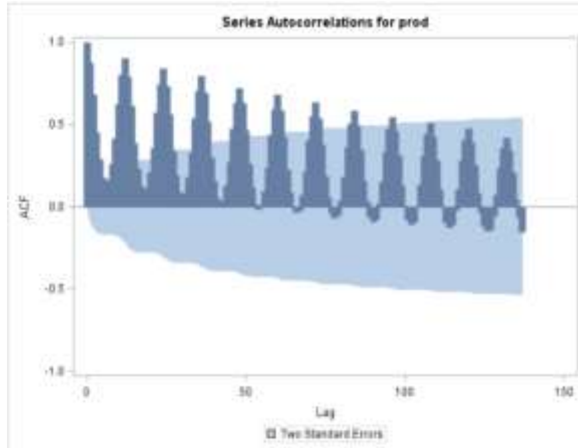

Figure 3.2: ACF plot for prod

We look into the different plots and see that there are too many lags to consider. Thus, it is difficult to make a coherent reading and make a decision on the type of modelling to be used. In the time plot (Figure 3.1), the points are too close together. Meanwhile, the ACF plot (Figure 3.2) shows seasonality but does not clearly show the lags in which it is seasonal. Thus, we decrease the amount of lags to be considered to 36 lags.

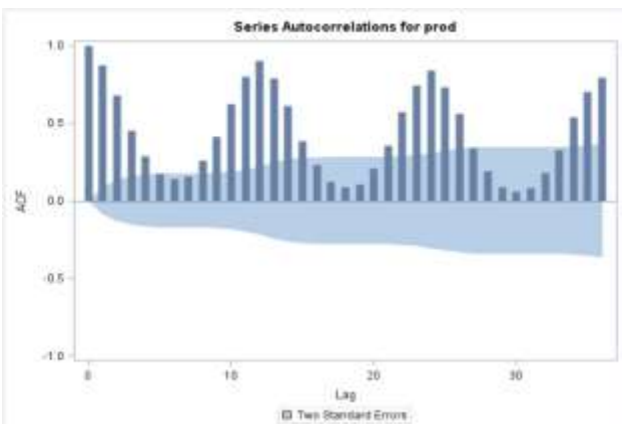

Figure 3.3: ACF plot for prod, lags = 36

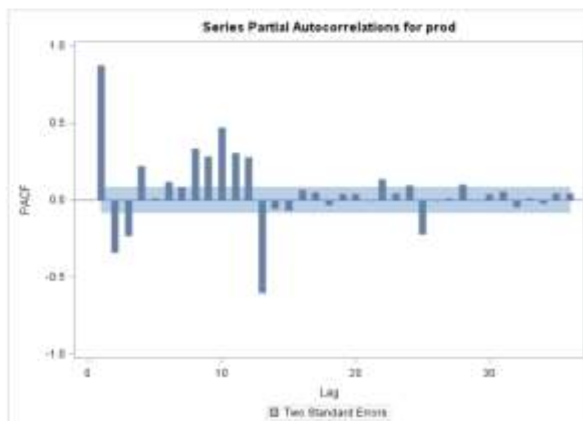

Figure 3.4: PACF plot for prod, lags = 36

After decreasing the number of lags to 36, the ACF and PACF plots become clearer. Looking at the ACF (Figure 3.3), we see the seasonality clearer. The lags are decaying after every 12th lag. Thus, we need to employ seasonal differencing.

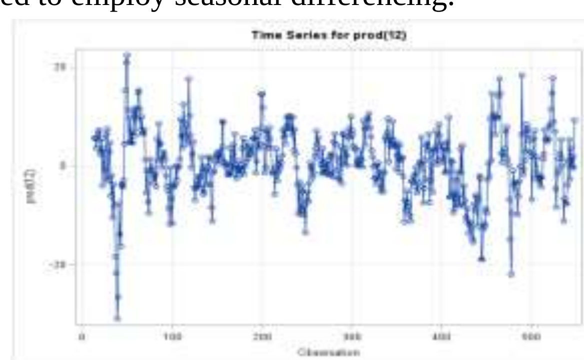

Figure 3.5: Time series plot for prod(12)

After seasonal differencing, we see that the time series plot (Figure 3.5) has become more stationary.

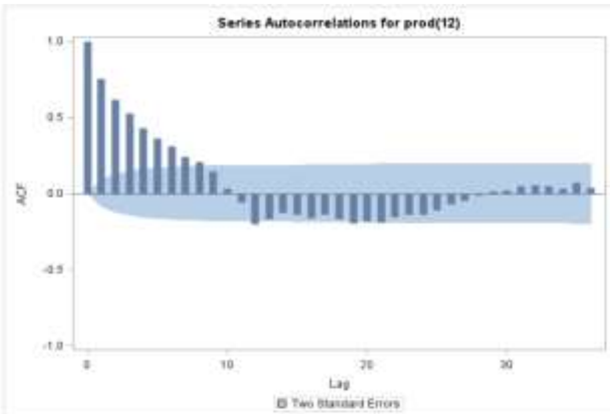

Figure 3.6: ACF plot for prod(12), lags = 36

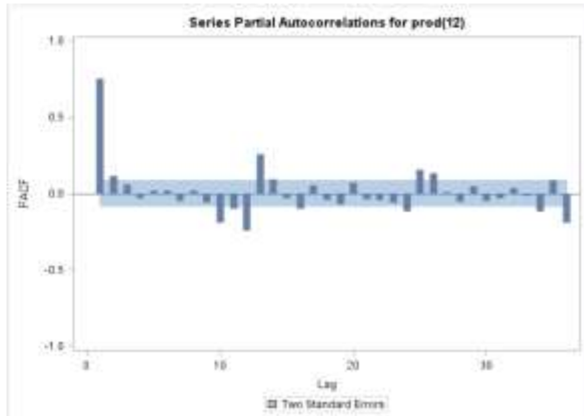

Figure 3.7: PACF plot for prod(12), lags = 36

In the ACF plot (Figure 3.6), the plots are slowly decaying. This suggests that ordinary differencing is also needed.

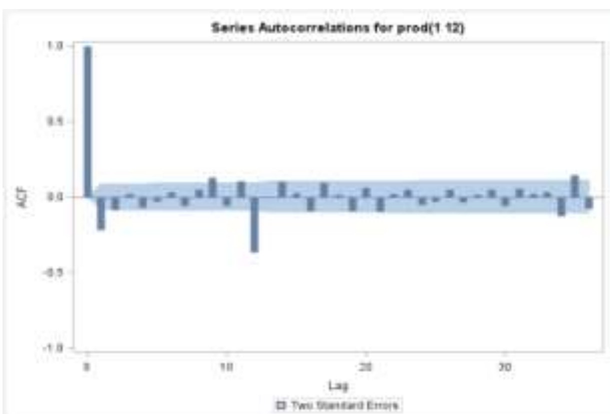

Figure 3.8: ACF plot for prod(1 12), lags = 36

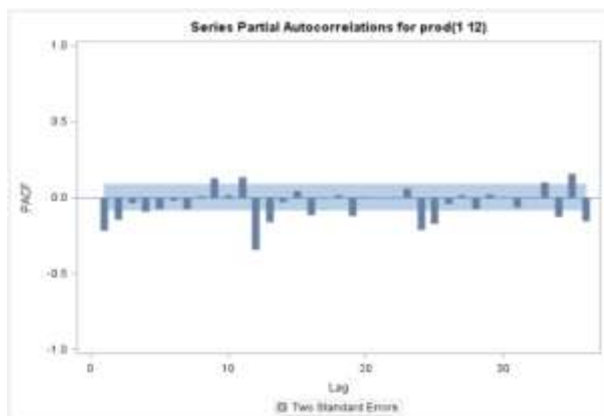

Figure 3.9: PACF plot for prod(1 12), lags = 36

We apply ordinary differencing and check the different plots. Looking at the ACF plot (Figure 3.8), we see that only the first seasonal lag (lag 12) is significant. Meanwhile, the PACF plot (Figure 3.9) shows that the seasonal lags (every 12^{th} lag) exponentially decay.

On the other hand, the ACF plot shows that only the first non-seasonal lag is significant while the non-seasonal lags in the PACF plot are exponentially decreasing. Hence, we fit a seasonal MA(1) and AR(2) model; i.e., an ARIMA$(2,1,0)(0,1,1)_{12}$ model.

Table 3.3: Parameter estimates for ARIMA(2,1,0)(0,1,1)$_{12}$

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.0008600	0.03296	-0.03	0.9792	0
MA1,1	0.72369	0.03082	23.48	<.0001	12
AR1,1	-0.28243	0.04280	-6.60	<.0001	1
AR1,2	-0.18468	0.04285	-4.31	<.0001	2

Constant Estimate	-0.00126
Variance Estimate	14.82855
Std Error Estimate	3.850786
AIC	2964.906
SBC	2982.035
Number of Residuals	535

After running the model, Table 3.3 shows that the constant term is insignificant (since the p-value is greater than 0.05), whereas MA1,1, AR1,1 and AR1,2 are significant. The AIC is 2964.906 and the SBC is 2982.035.

Table 3.4: Correlation of parameter estimates for ARIMA(2,1,0)(0,1,1)$_{12}$

Parameter	MU	MA1,1	AR1,1	AR1,2
MU	1.000	-0.024	0.000	0.003
MA1,1	-0.024	1.000	0.072	-0.083
AR1,1	0.000	0.072	1.000	0.230
AR1,2	0.003	-0.083	0.230	1.000

Looking at the Table 3.4, we see that there is no significant correlation between the different parameters (since none of the values have an absolute value > 0.9).

Table 3.5: Autocorrelation check for residuals for ARIMA(2,1,0)(0,1,1)$_{12}$

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	15.03	3	0.0018	0.006	-0.016	-0.018	-0.147	-0.075	-0.007
12	37.61	9	<.0001	-0.107	0.008	0.123	-0.085	0.047	0.072
18	53.38	15	<.0001	-0.080	0.034	0.004	-0.123	0.076	0.003
24	69.81	21	<.0001	-0.111	0.028	-0.039	-0.065	0.087	-0.054
30	77.48	27	<.0001	-0.019	0.038	-0.065	-0.018	0.078	-0.034
36	93.60	33	<.0001	0.035	0.023	-0.011	-0.065	0.123	-0.083
42	112.21	39	<.0001	0.120	-0.026	-0.126	0.029	-0.014	-0.015
48	132.02	45	<.0001	0.052	-0.011	-0.061	0.125	-0.086	-0.065

However, looking at Table 3.5, we see that the residuals are still not white noise. To determine which model to employ, we cross reference with the previous ACF (Figure 3.10) and PACF plot (Figure 3.11) and see that the first 2 lags of the ACF is significant, while the first 5 lags of the PACF plot are significant (except for the third lag). Thus, we fit a seasonal MA(1), MA(5) and AR(2) model; i.e., an ARIMA(2,1,5)$(0,1,1)_{12}$ model and remove the insignificant parameters through backward selection, if there are any.

Table 3.6: Parameter estimates for ARIMA(2,1,5)$(0,1,1)_{12}$

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.0013509	0.02510	-0.05	0.9571	0
MA1,1	-0.88515	0.04432	-19.97	<.0001	1
MA1,2	-0.56108	0.05775	-9.71	<.0001	2
MA1,3	0.45457	0.05945	7.65	<.0001	3
MA1,4	0.20908	0.05791	3.61	0.0003	4
MA1,5	0.09087	0.04370	2.08	0.0381	5
MA2,1	0.72389	0.03160	22.91	<.0001	12
AR1,1	-1.14765	0.0085244	-134.63	<.0001	1
AR1,2	-0.99197	0.0084350	-117.60	<.0001	2

After running the ARIMA(2,1,5)$(0,1,1)_{12}$ model, we check the parameter estimates (Table 3.6) and see that MA1,1, MA1,2, MA1,3, MA1,4, MA1,5, AR1,1 and AR1,2 are significant as their p-values are less than 0.05. However, the constant estimate is insignificant. Therefore, we remove the constant term.

Table 3.7: Parameter estimates for ARIMA(2,1,5)$(0,1,1)_{12}$, θ_0 = 0

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.88517	0.04428	-19.99	<.0001	1
MA1,2	-0.56113	0.05770	-9.73	<.0001	2
MA1,3	0.45449	0.05939	7.65	<.0001	3
MA1,4	0.20902	0.05786	3.61	0.0003	4
MA1,5	0.09083	0.04366	2.08	0.0379	5
MA2,1	0.72384	0.03155	22.94	<.0001	12
AR1,1	-1.14765	0.0085140	-134.80	<.0001	1
AR1,2	-0.99198	0.0084246	-117.75	<.0001	2

Variance Estimate	13.72698
Std Error Estimate	3.704993
AIC	2927.563
SBC	2961.821
Number of Residuals	535

After removing the constant term, the parameter estimates' p-values are less than 0.05 based on Table 3.7, which shows that all parameters are significant. The AIC is 2927.56 and SBC is 2961.1, both of which are significantly lower than the previous AIC and SBC estimates from the ARIMA(2,1,0)$(0,1,1)_{12}$ model, signifying that this model is better than the previous model.

Table 3.8: Correlations of parameter estimates for ARIMA(2,1,5)(0,1,1)$_{12}$, θ_0 = 0

Parameter	MA1,1	MA1,2	MA1,3	MA1,4	MA1,5	MA2,1	AR1,1	AR1,2
MA1,1	1.000	0.645	0.363	-0.292	-0.103	0.007	0.184	0.098
MA1,2	0.645	1.000	0.699	0.130	-0.284	0.043	0.013	0.097
MA1,3	0.363	0.699	1.000	0.700	0.375	-0.005	-0.079	-0.071
MA1,4	-0.292	0.130	0.700	1.000	0.654	-0.057	0.064	0.008
MA1,5	-0.103	-0.284	0.375	0.654	1.000	-0.015	0.102	0.142
MA2,1	0.007	0.043	-0.005	-0.057	-0.015	1.000	-0.008	0.071
AR1,1	0.184	0.013	-0.079	0.064	0.102	-0.008	1.000	0.585
AR1,2	0.098	0.097	-0.071	0.008	0.142	0.071	0.585	1.000

To continue, Table 3.8 shows that there is no significant correlation between the different parameters (since none of the values have an absolute value > 0.9).

Table 3.9: Autocorrelation check of residuals for ARIMA(2,1,5)(0,1,1)$_{12}$, θ_0 = 0

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	.	0	.	0.007	0.015	0.033	-0.046	-0.038	-0.082
12	8.45	4	0.0764	-0.034	-0.005	0.045	-0.012	0.008	0.023
18	10.53	10	0.3956	0.001	-0.018	-0.025	-0.052	-0.001	-0.012
24	14.93	16	0.5295	-0.051	-0.058	-0.029	-0.015	-0.014	-0.026
30	17.98	22	0.7071	0.021	-0.050	-0.006	0.003	-0.011	0.048
36	32.02	28	0.2737	0.017	-0.035	0.086	-0.097	0.078	-0.011
42	39.37	34	0.2419	0.067	-0.049	-0.049	-0.022	-0.028	0.046
48	44.74	40	0.2795	-0.022	-0.011	-0.015	0.055	-0.061	-0.040

Finally, looking at Table 3.10, we see that the p-values are greater than 0.05 and thus, we can conclude that the residuals are white noise. Thus, our final model for `prod` is the ARIMA(2,1,5)(0,1,1)$_{12}$ model. To check if the model is the best fit for the series, we employ overfitting. We add one non-seasonal AR parameter (Table 3.10), one non-seasonal MA parameter (Table 3.11) and one seasonal MA parameter (Table 3.12) to our final model to see how these would affect the results.

Table 3.10: Parameter estimates for ARIMA(3,1,5)(0,1,1)$_{12}$, θ_0 = 0

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.01075	0.07381	-0.15	0.8843	1
MA1,2	0.20860	0.07021	2.97	0.0031	2
MA1,3	0.91405	0.02883	31.70	<.0001	3
MA1,4	-0.16578	0.06076	-2.73	0.0066	4
MA1,5	-0.04241	0.05178	-0.82	0.4131	5
MA2,1	0.70940	0.03344	21.21	<.0001	12
AR1,1	-0.27611	0.05817	-4.75	<.0001	1
AR1,2	0.01083	0.06648	0.16	0.8706	2
AR1,3	0.87063	0.05814	14.98	<.0001	3

Variance Estimate	13.57781
Std Error Estimate	3.684808
AIC	2922.701
SBC	2961.242
Number of Residuals	535

After adding the AR1,3 term (Table 3.10), we see that MA1,1, MA1,5 and AR1,2 terms became insignificant and that the AIC has reduced in value, albeit the change in value is insignificant (since the change in AIC < 10).

Table 3.11: Parameter estimates for ARIMA(2,1,6)(0,1,1)$_{12}$, θ_0 = 0

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.88956	0.04408	-20.18	<.0001	1
MA1,2	-0.56395	0.05811	-9.70	<.0001	2
MA1,3	0.43256	0.06215	6.96	<.0001	3
MA1,4	0.25339	0.06212	4.08	<.0001	4
MA1,5	0.14947	0.05819	2.57	0.0105	5
MA1,6	0.08098	0.04445	1.82	0.0691	6
MA2,1	0.72999	0.03107	23.49	<.0001	12
AR1,1	-1.15142	0.0065448	-175.93	<.0001	1
AR1,2	-0.99587	0.0063631	-156.51	<.0001	2

Variance Estimate	13.6694
Std Error Estimate	3.697215
AIC	2926.298
SBC	2964.839
Number of Residuals	535

After adding the MA1,6 term (Table 3.11), we find that the parameter MA1,6 is insignificant and the AIC and SBC experienced no significant change.

Table 3.12: Parameter estimates for ARIMA(2,1,5)(0,1,2)$_{12}$, θ_0 = 0

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.88717	0.04429	-20.03	<.0001	1
MA1,2	-0.56068	0.05777	-9.70	<.0001	2
MA1,3	0.45718	0.05949	7.68	<.0001	3
MA1,4	0.20942	0.05799	3.61	0.0003	4
MA1,5	0.08677	0.04376	1.98	0.0479	5
MA2,1	0.69644	0.04608	15.11	<.0001	12
MA2,2	0.03662	0.04558	0.80	0.4222	24
AR1,1	-1.14860	0.0080557	-142.58	<.0001	1
AR1,2	-0.99307	0.0079605	-124.75	<.0001	2

Variance Estimate	13.73734
Std Error Estimate	3.706391
AIC	2928.95
SBC	2967.491
Number of Residuals	535

After adding the MA2,2 term (Table 3.12), the parameter MA2,2 became insignificant with the AIC and SBC exhibiting insignificant changes.

Thus, after adding the said parameters, the overfitting shows that the final model is the best fit and is parsimonious, as the change in the AIC and SBC after adding these terms were negligible.

We also employ automatic model selection and compare the best model selected through this procedure to our final model.

Table 3.13: Automatic model selection results for prod

Initial Automatic Model Selection						
For Variable prod						
Source of Model	Estimated Model					
Default TRAMO Model	(3,	1,	1)	(0,	0,	1)

Table 3.13 shows that the best model for the data is an ARIMA(3,1,1)(0,0,1)$_{12}$. However, SAS issued a warning that the maximization of the ARIMA model likelihood for `prod` required more than 200 iterations, indicating that the selected model may be inadequate for the data. To check this, we run this ARIMA model and check.

Table 3.14: Parameter estimates for ARIMA(3,1,1)(0,0,1)$_{12}$

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.05409	0.04441	1.22	0.2238	0
MA1,1	0.95163	0.01512	62.95	<.0001	1
MA2,1	-0.57026	0.03679	-15.50	<.0001	12
AR1,1	0.87037	0.04211	20.67	<.0001	1
AR1,2	0.04050	0.05563	0.73	0.4670	2
AR1,3	-0.34409	0.04142	-8.31	<.0001	3

Constant Estimate	0.023433
Variance Estimate	33.36497
Std Error Estimate	5.776242
AIC	3476.892
SBC	3502.718
Number of Residuals	547

From Table 3.14, we find that the constant estimate and the AR1,2 term are insignificant.

Table 3.15: Autocorrelation check of residuals for ARIMA(3,1,1)(0,0,1)$_{12}$

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.45	1	0.0111	-0.007	-0.052	-0.044	0.038	0.010	0.074
12	58.44	7	<.0001	-0.063	-0.033	-0.099	-0.016	0.134	0.244
18	86.93	13	<.0001	0.065	0.043	-0.183	-0.096	-0.039	-0.005
24	285.04	19	<.0001	-0.097	-0.011	-0.206	-0.015	0.027	0.541
30	302.84	25	<.0001	0.012	-0.015	-0.174	-0.007	0.009	0.002
36	388.19	31	<.0001	-0.050	-0.027	-0.086	-0.030	0.125	0.344
42	420.85	37	<.0001	0.106	0.010	-0.196	0.004	-0.074	-0.012
48	547.80	43	<.0001	-0.048	-0.071	-0.207	0.073	-0.020	0.394

Furthermore, table 3.15 shows that the residuals are white noise, signifying that the ARIMA(3,1,1)(0,0,1)$_{12}$ model is inadequate for the data, as the SAS warning suggested.

Table 3.16: Forecasts for prod (snippet)

Obs	prod	FORECAST	STD	L95	U95	RESIDUAL
549	.	118.996	3.70499	111.734	126.258	.
550	.	130.544	4.60364	121.521	139.567	.
551	.	129.263	5.12507	119.218	139.308	.
552	.	128.873	5.53265	118.029	139.717	.
553	.	116.718	5.87595	105.201	128.234	.
554	.	115.679	6.18478	103.558	127.801	.
555	.	112.717	6.52856	99.921	125.513	.
556	.	109.362	6.81497	96.005	122.719	.
557	.	104.777	7.08881	90.883	118.671	.
558	.	107.263	7.39051	92.778	121.748	.
559	.	107.085	7.63947	92.112	122.058	.
560	.	114.789	7.89073	99.324	130.255	.

After confirming that our final model is the best fit model, we then proceed with forecasting. Table 3.16 shows a snippet of the results. Note that we expect output to increase during the months of November, December and January, confirming our earlier hypothesis that candy production peaks in those months due to the increase in demand (from Halloween, Christmas and New Year's). Looking at the other months, candy production remains relatively stable throughout the other months. To observe whether these results are in line with the observations in previous years, we need to look at how the forecasts compare with the rest of the series (shown in the next section).

4. Summary and Conclusions

We find that the final model is ARIMA(2,1,5)(0,1,1)$_{12}$ model, which is represented by:

$$(1-\phi_1 B-\phi_2 B^2)(1-B^{12})(1-B)Z_t$$
$$=(1-\theta_1 B-\theta_2 B^2-\theta_3 B^3-\theta_4 B^4-\theta_5 B^5)(1-\Theta_{12}B^{12})a_t$$

Inputting the parameter estimates from Table 3.6, we get the final model, given by:

$$(1+1.14765B+0.99198B^2)(1-B^{12})(1-B)Z_t$$
$$=(1+0.88517B+0.56113B^2-0.45449B^3-0.20902B^4-0.09083B^5)(1-0.72384B^{12})a_t$$

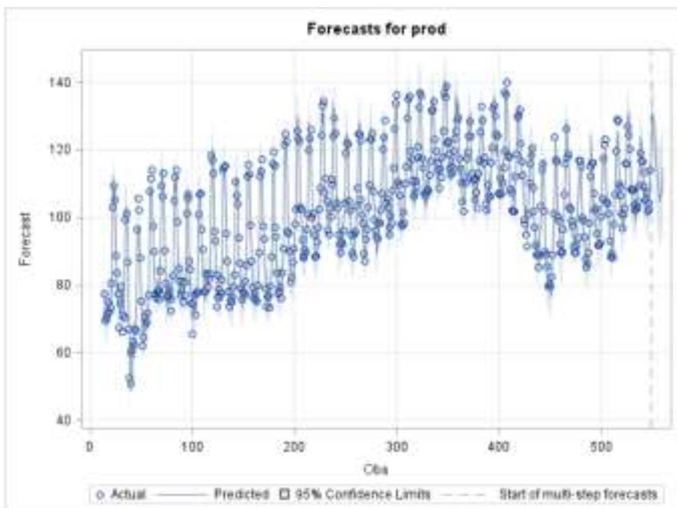

Figure 4.1: Forecasts for prod

From Figure 3.10, we see that the US candy IP index should remain within the usual bounds from the previous years. Thus, we expect the confectionary industry to remain relatively stable for the next year.

However, while the final model satisfies many of the selection criteria (significant parameter estimates, low AIC/SBC, low correlation among parameter estimates, white noise residuals), it lacks intuitive appeal and many policy makers who may not be educated on time series models may find the model confusing and opt for simpler models instead.

As such, it may be more useful to fit simple time series models such as seasonal exponential smoothing or seasonal Holt-Winters, as these models are more easily interpretable. Moreover, we can also employ X-12 or X-13 decomposition to deseasonalize the data and make forecasting easier. Again, this study is limited only to ARIMA models, so future research on this area can make use of the methods outlined earlier, as they could result in more intuitive models.

REFERENCES

Bangko Sentral ng Pilipinas. (2010). Exchange rates. Retrieved March 26, 2019 from http://www.bsp.gov.ph/downloads/Publications/FAQs/exchange.pdf.

Ca'Zorzi, M., Muck, J., & Rubaszek, M. (2016). Real exchange rate forecasting and PPP: This time the random walk loses. *Open Economies Review*, *27*(3), 585-609.

Grand View Research. (January 2018). US Candy Market Size Worth $19.6 Billion by 2025. Retrieved March 26, 2019 from https://www.grandviewresearch.com/press-release/us-candy-market-analysis.

Hsieh, D.A. (1988). "The statistical properties of daily foreign exchange rates: 1974-1983." *Journal of International Economics*. 24:129-145.

Jiménez-Martin, J.A., and Flores de Frutos, R. (2009), Seasonal fluctuations and equilibrium models of exchange rate, *Applied Economics*, 41, 2635-52.

Meese, R. and K. Rogoff (1983). "Empirical exchange rate models of the seventies: Do they fit out of sample?" *Journal of International Economics*. 14:3-12.

Moosa, I., & Burns, K. (2014). The unbeatable random walk in exchange rate forecasting: Reality or myth?. *Journal of Macroeconomics*, *40*, 69-81.

Mouhanna, Alecia. (December 19, 2018). 3 Important Candy Trends to Watch in 2019 & Beyond. Retrieved March 26, 2019 from https://blog.marketresearch.com/3-candy-industry-trends-to-watch-in-2019-beyond.

Winstein, Bruce. (Octover 13, 2018). How the Robust Candy Industry is Transforming the Way You Think about Sugar. Retrieved March 26, 2019 from https://www.forbes.com/sites/bruceweinstein/2018/10/13/how-the-robust-candy-industry-is-transforming-the-way-you-think-about-sugar/#4acee8067b6d.

APPENDIX

Note that we imported our datafiles into SAS. In the event that you would like to run our code, replace <DIRECTORY> with the directory of the files "BSP_Exchange.csv" and "candy_production.csv" for the non-seasonal and seasonal models, respectively.

A. Non-Seasonal SAS code

```
proc import datafile = '<DIRECTORY>' out = temp replace;
run;

data exchange_data;
set temp;
date = intnx('month', '1jan2004'd, _n_-1);
format date yymon.;
t = _n_;

proc print data = exchange_data;
id date;
run;

proc arima data = exchange_data plots = all plots(unpack);
identify var = exch nlag = 48 alpha = 0.05 stationarity = (adf);
run;

/*ordinary differencing*/
proc arima data = exchange_data plots = all plots(unpack);
identify var = exch(1) nlag = 48 alpha = 0.05 stationarity = (adf);
run;

/*random walk*/
estimate plot;
run;

/*same as above but no constant term*/
estimate plot noint;
run;

/*forecast*/
forecast lead = 12 alpha = 0.05 out = exch_forecast printall ;
run;

proc print data = exch_forecast;
run;
```

B. Seasonal SAS code:

```
proc import datafile = '<DIRECTORY>' out = temp replace;
run;

data candy_data;
set temp (rename = (observation_date = date IPG3113N = prod));
t = _n_;
run;
```

```
proc print data = candy_data;
id date;
run;

proc arima data = candy_data plots = all plots(unpack);
identify var = prod nlag = 137 alpha = 0.05 stationarity = (adf);
run;

/*decrease lags*/
proc arima data = candy_data plots = all plots(unpack);
identify var = prod nlag = 36 alpha = 0.05 stationarity = (adf);
run;

/*seasonal differencing*/
proc arima data = candy_data plots = all plots(unpack);
identify var = prod(12) nlag = 36 alpha = 0.05 stationarity = (adf);
run;

/*ordinary differencing*/
proc arima data = candy_data plots = all plots(unpack);
identify var = prod(1 12) nlag = 36 alpha = 0.05 stationarity = (adf);
run;

/*seasonal MA(1), AR(2)*/
estimate p = 2 q = (12) plot;
run;

/*seasonal MA(1), MA(5), AR(2)*/
estimate p = 2 q = (1 2 3 4 5)(12) plot;
run;

/*same as above but no constant term*/
estimate p = 2 q = (1 2 3 4 5)(12) plot noint;
run;

/*overfitting*/
estimate p = 3 q = (1 2 3 4 5)(12) plot noint;
run;

estimate p = 2 q = (1 2 3 4 5 6)(12) plot noint;
run;

estimate p = 2 q = (1 2 3 4 5)(12 24) plot noint;
run;

/*forecast*/
estimate p = 2 q = (1 2 3 4 5)(12) plot noint;
run;

forecast lead = 12 alpha = 0.05 out = candy_forecast printall ;
run;

proc print data = candy_forecast;
run;
```

```
/* automatic model selection */
proc import datafile = 'C:/Users/11206438/Desktop/TIMESER
PROJECT/candy_production.csv' out = temp replace;
run;

data candy_data;
set temp (rename = (observation_date = date IPG3113N = prod));
date = intnx( 'qtr', '1jan72'd, _n_-1);
t = _n_;
run;

proc x12 data = candy_data date = date interval = qtr;
var prod;
transform power = 0;
automdl print = unitroottest unitroottestmdl autochoicemdl best5model;
estimate;
x11;
tables;
output out = out a1 d10 d11 d12 d13;
run;

/*(3,1,1),(0,0,1)*/
proc arima data = candy_data plots = all plots(unpack);
identify var = prod(1) nlag = 36 alpha = 0.05 stationarity = (adf);
run;

estimate p = 3 q = (1)(12) plot;
run;
```