Time Series Analysis of Import/Export Price Indexes: Central/South America

Ceyda Çaylak
Middle East Technical University
Ankara, Turkey
ceyda.caylak@metu.edu.tr

Neda Bokaei
Middle East Technical University
Ankara, Turkey
bokaei.neda@metu.edu.tr

***Abstract*— In this paper, we analysed the time series of Annual Import/Export Price Index of Central/South America and applied different forecast models. First, we divided the data into two sets (train and test), stabilized the variance of the system, looked for outliers and cleaned them. Then, we analyzed the unit roots and suggested fits to make our model a stationary process. Lastly, we fitted different forecast models and compared these models based on MAPE values to have the best forecast model.**

Keywords - Forecast, Import/Export price index, time series

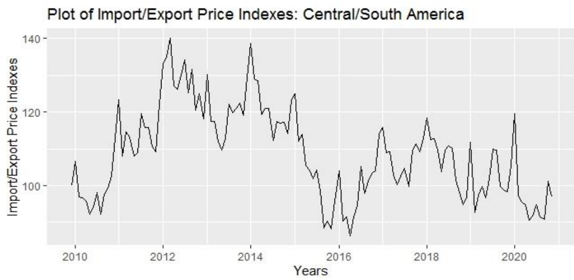
I. INTRODUCTION

BLS International produces Import/Export Price Indexes (MXP) containing data on changes in the prices of nonmilitary goods and services traded between Central and South America. Import price index is also used as an indicator of future domestic inflation. According to the United States Federal Reserve, United States Export Price Index: Import/Export of Central/South America reached a record high of 140.00000 in March of 2012 and a record low of 86.40000 in April of 2016.

The aim of this study is to see how price indexes between Central and South America changed over time and forecast the future values using ARIMA, ETS, TBATS, Neural Network and Prophet models. R Studio Version 1.2.5033 used for this project.

II. DATA DESCRITPION AND PREPROCESSING

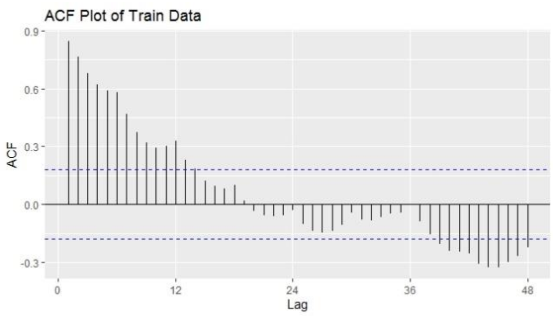
The data is downloaded from https://quandl.com/, which offers different types of data for R. The data starts from December 2009, to December 2020. It's a monthly series, containing 133 observations in total.



Graph 1: *Time Series Plot of Data Set*

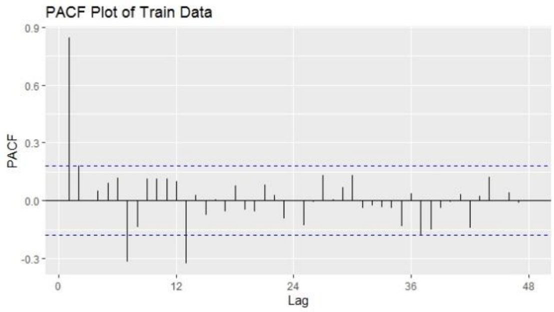
As we can see in the plot of the time series, the data doesn't seem stationary. We see some ups and downs, which might indicates that we have a stochastic trend. Especially during 2015-2016, we see a sharp decresing trend, perhaps it's because of the political reasons. Also we might have a variance problem.

To continue with the process, we will divide the data into two sets: train and test. Train data starts from December 2009 to November 2019, and test data starts from December 2019 to December 2020, keeping the last 12 observations as a test set.



Graph 2: *ACF Plot of Data Set*

As we see in the ACF Plot, the data isn't stationary. The plot follows a sinusoidal pattern.

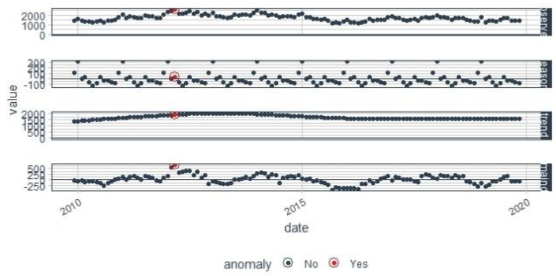


Graph 3: *PACF Plot of Data Set*

We know that the data isn't stationary by just looking at the ACF plot. And when we look at the PACF Plot, we see spikes at lag 1, lag 6 and lag 12. However, interpretting this part isn't necessary.

Then, we check the lambda value to make sure of the stability of the variance of the data. The specific lambda value is 1.698491, not 1, which indicates that we need a Box-Cox transformation.

After conducting the Box-Cox Transformation, we detect if there's any anomalities in the series.

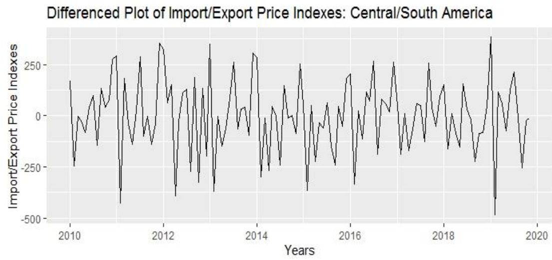


Graph 4: *Anomaly Detection Plot*

As in we see in Graph 4, we can see some outliers in the series. We removed these outliers by using tsclean function from the train set and replaced them with the interpolated values.

Since the data is a monthly data, we need to use Hegy test to detect the existence of the unit root. Hegy test shows that we have one regular unit root (p= 0.1 > 0.05) and no seasonal unit root (p=0.017 < 0.05).

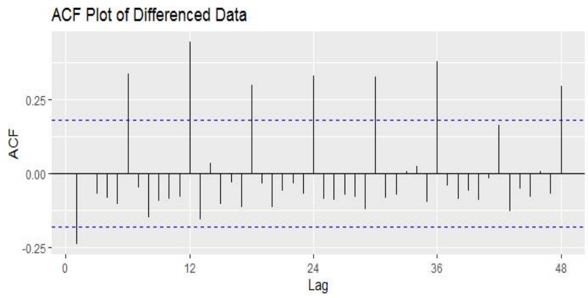
To remove the trend in the series, we took one regular difference to see how it goes. After taking it, we applied one more Hegy test to see if the series is stationary now.



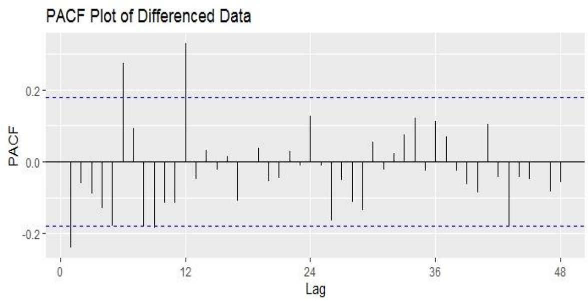
Graph 5: *Time Series Plot of Differenced Data Set*

Hegy test results on the differented series shows that there's no regular unit root (p=0.01< 0.05) and no seasonal trend (p=0.01<0.05) either.

To verify our conclusion, we applied the Canova-Hansen Test on the differenced series, the results show us p value is 0.4869, greater than the alpha. These results made us believe that we have a stationary series.

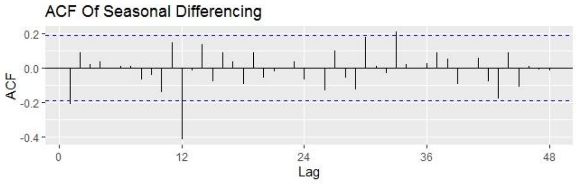


Graph 6: *ACF Plot of Stationary Data Set*

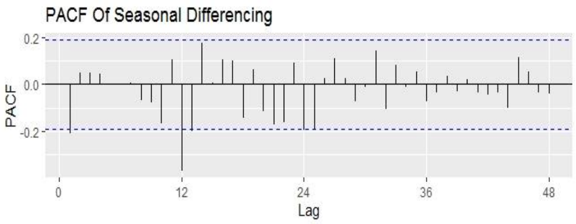


Graph 7: *PACF Plot of Stationary Data Set*

There's something confusing about these plots, the ACF plot of the differenced series gives us a vague answers because it doesn't seem stationary, but we know that the series is supposed to be stationary now by using the Hegy and Conova-Hansen tests. To eliminate this confusion, we took one seasonal difference to see how it goes and fitted the appropriate model ;



Graph 8: *ACF Plot of Seasonal Differenced Data*



Graph 9: *PACF Plot of Seasonal Differenced Data*

SARIMA(1,1,1)(1,1,1)[12], but the sma1 coefficient was -1, which is a undesired situation. So, it's pointless to doubt the Hegy and Canova-Hansen tests. Thus, we decided to use the regular differenced train as our stationary model.

To suggest a model, first, we used auto.arima function, it suggested SARIMA (1,1,0)(2,0,0)[12] and we suggested SARIMA(1,1,0)(1,0,0)[12]. We will examine the accurateness of these models in the next chapter.

III. MODELLING AND DIAGNOSIS CHECKING

In the previous chapter, auto.arima function suggested SARIMA(1,1,0)(2,0,0)[12] and we suggested SARIMA(1,1,0)(1,0,0)[12]. To decide which fit is better than the other one, first we will check the significancy of these models then compare the AICc values.

Table 1: *Summary of Model SARIMA(1,1,0)(2,0,0)[12]*

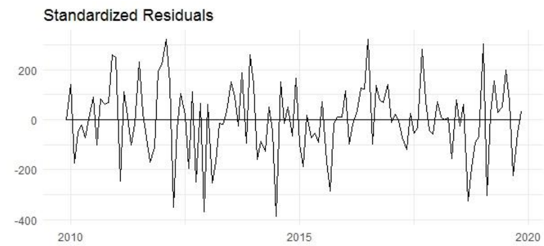
[1] Model 1			
[2] Coefficents	ar1	sar1	sar2
[3]	-0.2264	0.3791	0.2454
[4] s.e	0.0890	0.0873	0.0946
[5] Statistics	-2.54	4.34	2.59
[6] sigma^2 estimated as 21988: log-likelihood=-76475			
[7] AIC= 1537.5 AICc=1537.5 BIC=1548.61			

Table 2: *Summary of Model SARIMA(1,1,0)(1,0,0)[12]*

[1] Model 1		
[2] Coefficents	ar1	sar1
[3]	-0.2187	0.4875
[4] s.e	0.0891	0.0820
[5] Statistics	-2.45	5.94
[6] sigma^2 estimated as 23327: log-likelihood=-767.91		
[7] AIC= 1541.82 AICc=1542.03 BIC=1550.16		

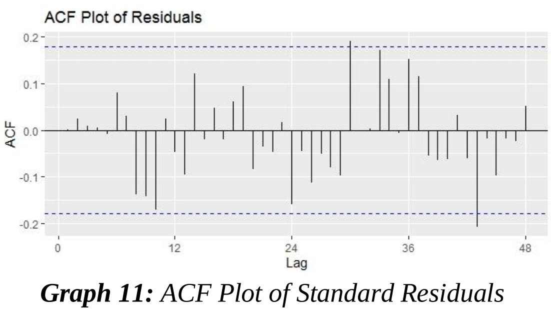
As we can see, by using MLE, both of the fitted models are significant because the statistics values are greater than the absoulute value of 2 in both of the models. Then we choose Akaikie Information Criterion (AIC) to investigate the model with the lowest AIC value. Obviously, SARIMA(1,1,0)(2,0,0)[12] model has lower AIC. Thus, we'll continue with it.

After deciding the best fitted model, we will check the residuals of the model. To have a good model, we need to be sure that the residuals are uncorrelated and have zero mean.

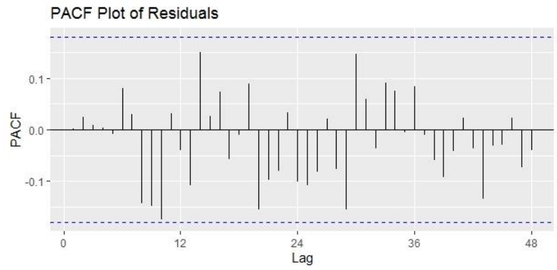


Graph 10: *Plot of Standard Residuals*

The plot seems zero around mean.



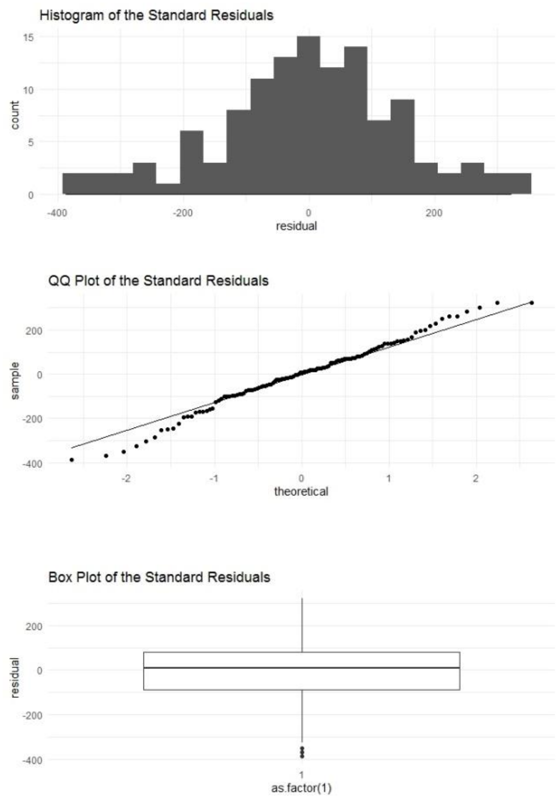
Graph 11: *ACF Plot of Standard Residuals*



Graph 12: *PACF Plot of Standard Residuals*

As we can see in ACF and PACF Plots, all spikes are in the WN, which indicates that the there is no correlation between residuals, but we will conduct the formal test; Ljung-Box test. The test result shows that p value is 0.5395, which is greater than the alpha. So, we don't have a autocorrelation problem.

To check the normality of the residuals, first we will conduct the visual ways, such as Histogram, QQ plot and Box-Plot:



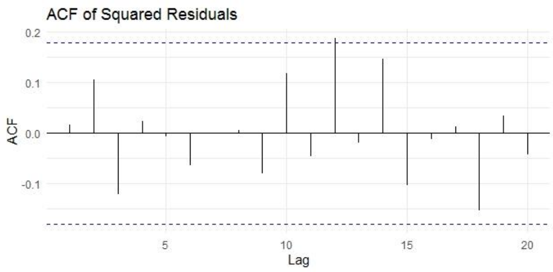
Graph 13: *Histogram, QQ plot, Box-Plot of Standard Residuals, respectively.*

The histogram of the residuals might have a symmetric distribution, the QQ plot suggests heavy-tailed distribution and box-plot shows a symmetric shape but it has some outliers.

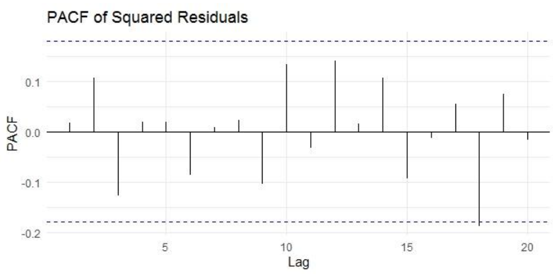
To check the normality of the residuals formally, we will apply Jarque-Bera and Shapiro-Wil tests. The p values for these tests are 0.5032 and 0.3083, respectively. Luckily, we have normal residuals.

And to check if we have a autocorrelation problem in the residuals, we will apply Ljung-Box test, which suggest that p value is 0.5396, greater than the alpha. As we said, in the ACF and PACF plot all the spikes are in the white noise band, so the there's no autocorrelation problem and this test result confirms our opinion.

To check the last assumption, the heteroscedasticity of the residuals, in other words, it's the problem where the variance of the residuals changes over time, we will check the squared residuals.



Graph 14: *ACF Plot of Squred Residuals*

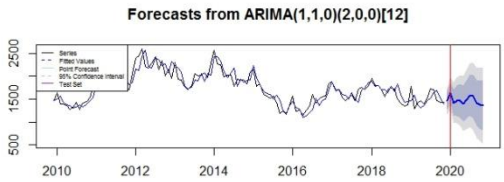


Graph 15: *PACF Plot of Squred Residuals*

As we can see, all the spikes are in the white noise band in both ACF and PACF plots, which suggests us that we don't have a heteroscedasticity problem. But still, we will conduct Engle's ARCH Test. The p value is greater than alpha. Luckily, we don't have heteroscedasticity problem either.

In conclusion, the residuals are zero around mean, normal, have no autocorrelation and heteroscedasticity problem.

After checking the residuals, now we can apply the forecast methods.

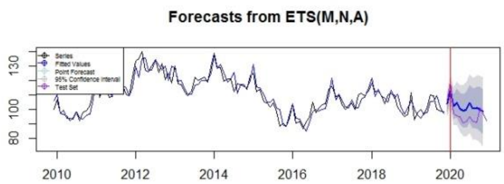


Graph 16: *Forecast Plot of SARIMA*

To have appropriate accuracy results for SARIMA forecasting, we applied back transformation.

Table 3: *Summary of ETS Model*

ETS(M,N,A)
Smoothing parameters:
alpha = 0.807
gamma = 2e-04
Initial states:
l = 104.3616
s = -4.1153 -2.0236 -1.025 -1.8488 2.0289 -2.4917
-3.1629 -1.4935 2.1449 -0.0899 10.315 1.762
sigma: 0.0422
AIC AICc BIC
956.2357 960.8511 998.0481

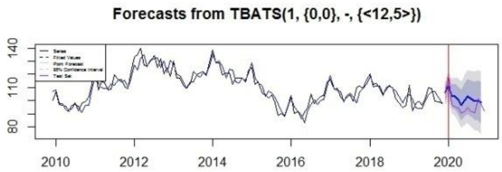


Graph 17: *Forecast Plot of ETS*

It can be seen that th series contain both trend and seasonality, so we applied Holt-Winter's Exponential Method. And our series has multiplicative error and additive seasonality. To check the normality of the residuals, we checked Jarque-Bera and Shapiro-Wilk tests, the p values are 0.4055 and 0.5019, respectively., which indicates that the residuals follow a normal distribution.

Table 4: *Summary of TBATS Model*

TBATS(1, {0,0}, -, {<12,5>})
Call: tbats(y = train)
Parameters
Alpha: 0.7692266
Gamma-1 Values: -0.004659941
Gamma-2 Values: 0.01412067

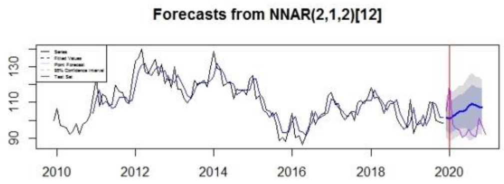


Graph 18: *Forecast Plot of TBATS*

We can see that TBATS(1, {0,0}, -, {<12,5>}) means the series doesn't need any transformation, it's a monthly series, and needs 5 Fourier transformations. Since this is a automatized method, it might fail to capture the seasonality of the series. When we check the normality of the residuals, again, we used Jarque-Bera and Shapiro-Wilk tests, the p values are 0.9127 and 0.9733, respectively. Thus, we have enough evidence to claim that we have residuals with normal distribution.

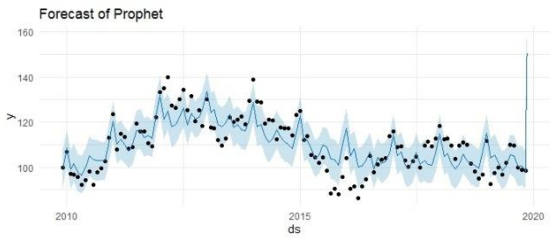
Table 5: *Summary of NN Model*

Series: train
Model: NNAR(2,1,2)[12]
Call: nnetar(y = train)
Average of 20 networks, each of which is
a 3-2-1 network with 11 weights
options were - linear output units
sigma^2 estimated as 34.43

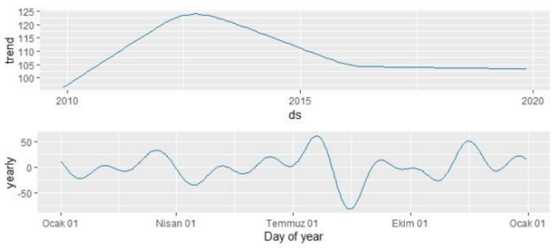


Graph 19: *Forecast Plot of NN*

The summary table of NN tells us that it used 2 non-seasonal lags as input variables, 1 seasonal lag and 2 nodes in the hidden layer. We see that the results aren't very satisfying, the forecasted values aren't even close to the test values. And in the *Table 6* we can see that MAPE value for this method is the highest, which indicates that this isn't the best forecast method to use for our series.



Graph 20: *Forecast Plot of Prophet*



Graph 21: *Components of the Series*

The Prophet forecasting method decomposes the series into three parts: overall growth trend, yearly or weekly seasonality and holiday effects. When we check the residuals of the Prophet model, the residual exhibit a normal distribution. ($p>0.05$)

Table 6: *Summary of Forecasting Models*

	ME	RMSE	MAE	MPE	MAPE	ACF1
ARIMA	-2.589	7.534	6.597	-3.153	6.767	0.574
ETS	-4.651	7.025	6.237	-5.158	6.539	0.418
TBATS	-4.575	6.876	6.085	-5.066	6.355	0.324
NNETAR	-7.783	12.529	11.529	-8.733	11.939	0.521
PROPHET	-5.259	9.190	7.920	-5.921	8.253	0.420

Table 7: *Summary of the Train Accuracy of the Models*

	ME	RMSE	MAE	MPE	MAPE	ACF1
ARIMA	-1.546	145.791	112.408	-0.4285	6.501	0.352
ETS	-0.022	4.414	3.448	-0.110	3.144	-0.008
TBATS	-0.015	4.478	3.510	0.120	3.207	0.295
NNETAR	-0.0001	5.835	4.594	-0.296	4.149	0.386
PROPHET	2.874	6.412	5.459	2.468	5.061	0.291

To decide which forecasting method is the best model for our series, we will consider the MAPE values in *Table 6*. MAPE measures this accuracy as a percentage, can be calculated as the average absolute percent error for each time period minus actual values divided by actual values. Thus, the lower the MAPE value, the better the forecast method. According to Table 6, TBATS has the best and NNETAR has the worst forecasting performance. Also, we can see that ETS follows TBATS as the second best forecasting performance, which also can be seen in the graphs.

IV. CONCLUSION

In this paper, we analyzed the time series of Import/Expory Price Indexes of Central/South America. First, we checked the necessarity of the variance stabilizing transformation, and cleaned the anomalies. To have better forecasting results, it is indeed necessary to make the series statationary. To do this, we used Hegy test, and we saw that taking one regular difference was enough. After taking the difference, we applied Hegy and Canova-Hansen tests to make sure that the series exhibits a stationary process. Then, we checked the ACF and PACF plots of this stationary series, and decided to fit SARIMA(1,1,0)(2,0,0)[12] model. When we checked the residual of this fit, luckily, we realized that the residuals are normal and zero around mean, have no autocorrelation and heteroscedasticity problem. Lastly, we performed different forecasting methods, interpreted them, and decided which one is the best method for our series. Apparently, while TBATS has the best performance, NNETAR has the worst performance.

Overall, we can say that we were lucky because the residuals of the fitted model exhibits a normal distribution, have no autocorrelation and heteroscedasticity problem. And the MAPE values, mean absolute error, for all of the forecasting methods are lower than we expected.

V. REFERENCES

Hyndman, R. J. & Athanasopoulos, G. (2017), Forecasting Principles and Practices, Second Edition.

Quandl, https://www.quandl.com/data/BLSN/EIUIH14243-Import-Export-Price-Indexes-Central-South-America-Dec-2009-100