ARIMA Time Series Model for Philippine Exports

VEE Time Series Student Project

Willy M. Tan

INTRODUCTION

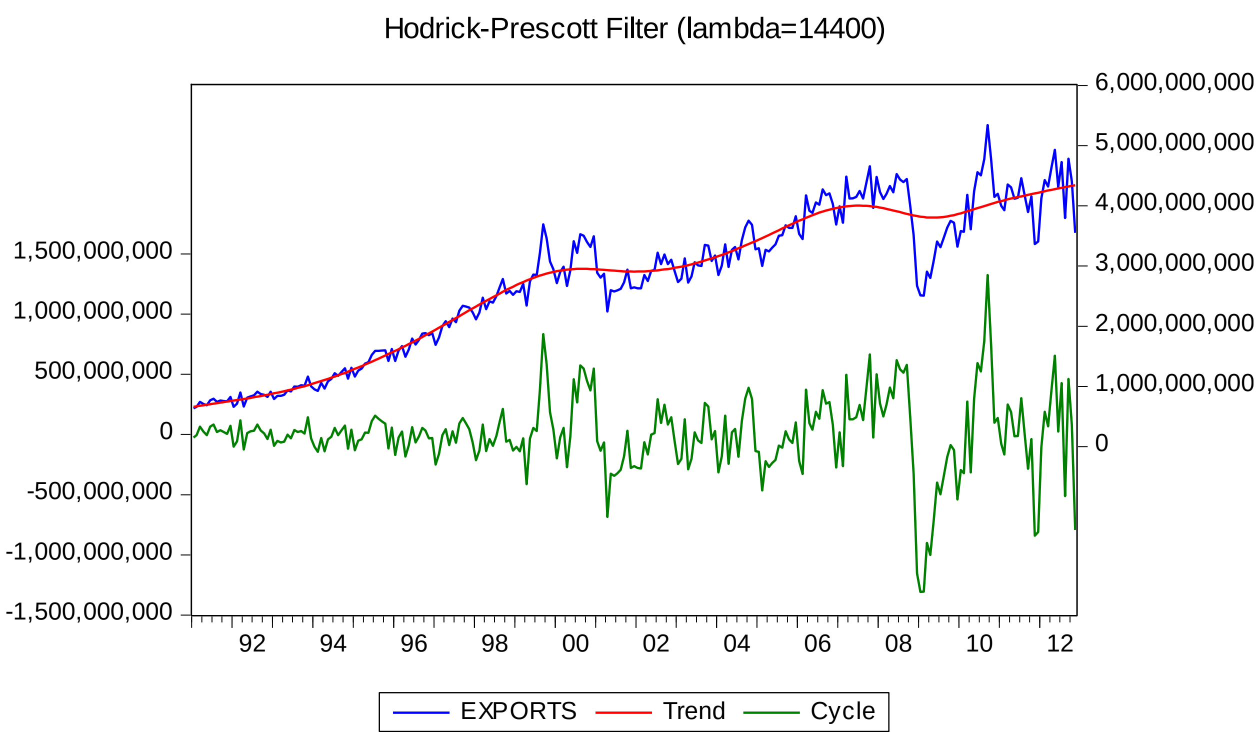
The objective of this project is to use ARIMA time series to model Philippine Exports with the help of EViews software. The process of model specification, fitting, and diagnostics will be demonstrated in this project. The selected model will then be used to generate forecasts with confidence bounds.

DATA SOURCE

The exports dataset is the monthly values of Philippine exports as reported by the National Statistics Office (NSO).

DATA VISUALIZATION

Using the log level of exports, I have extracted the long-term trend of the series and produced a line graph superimposing the actual series and its long-term trend. The Hodrick-Prescott (HP) Filter determines the long-term trend of the series.



The red line in the graph is the HP Filtered long-term trend. As we can see from the graph, the trend is generally upward sloping which signals growth in exports. The green line are outliers in the cycle which show irregular events or shocks.

TEST FOR UNIT ROOT

Since the graph does not tell us exactly if the series has a stochastic trend, a unit root test is necessary to be performed. This is done using the Augmented Dickey-Fuller (ADF) procedure in EViews.

EViews selects the number of lagged values of using automatic selection given an information criterion. The inclusion of the intercept and time trend in the test equation is determined through the significance of the said terms.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOG(EXPORTS))
Method: Least Squares
Date: 03/24/16 Time: 17:33
Sample (adjusted): 1991M03 2012M11
Included observations: 261 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(EXPORTS(-1))	-0.046649	0.021760	-2.143781	0.0330
D(LOG(EXPORTS(-1)))	-0.313979	0.059856	-5.245543	0.0000
C	0.989406	0.451083	2.193400	0.0292
@TREND(1991M01)	0.000209	0.000161	1.298653	0.1952

R-squared	0.130092	Mean dependent var	0.006430
Adjusted R-squared	0.119937	S.D. dependent var	0.087757
S.E. of regression	0.082327	Akaike info criterion	-2.141036
Sum squared resid	1.741865	Schwarz criterion	-2.086407
Log likelihood	283.4051	Hannan-Quinn criter.	-2.119077
F-statistic	12.81114	Durbin-Watson stat	2.025748
Prob(F-statistic)	0.000000		

Since time trend is not a significant variable in the model, we will have to drop it from our model and re-run the test without it.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOG(EXPORTS))
Method: Least Squares
Date: 03/24/16 Time: 17:36
Sample (adjusted): 1991M03 2012M11
Included observations: 261 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(EXPORTS(-1))	-0.021018	0.009175	-2.290729	0.0228
D(LOG(EXPORTS(-1)))	-0.327724	0.058992	-5.555431	0.0000
C	0.463116	0.198358	2.334747	0.0203

R-squared	0.124383	Mean dependent var	0.006430
Adjusted R-squared	0.117595	S.D. dependent var	0.087757
S.E. of regression	0.082436	Akaike info criterion	-2.142158
Sum squared resid	1.753296	Schwarz criterion	-2.101186
Log likelihood	282.5516	Hannan-Quinn criter.	-2.125688
F-statistic	18.32470	Durbin-Watson stat	2.038186
Prob(F-statistic)	0.000000		

All variables are now significant in the test equation. We can now proceed to the interpretation of the result of the ADF test. The null hypothesis is that the series has a unit root.

Null Hypothesis: LOG(EXPORTS) has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=15)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.290729	0.1758
Test critical values:	1% level	-3.455289	
	5% level	-2.872413	
	10% level	-2.572638	

*MacKinnon (1996) one-sided p-values.

Since the p-value of the ADF test is greater than 0.10, we do not reject the null hypothesis that *log(exports)* has a unit root. This means that we need to check for the existence of higher order unit root by applying the unit root test on the differenced series.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOG(EXPORTS),2)
Method: Least Squares
Date: 03/24/16 Time: 17:43
Sample (adjusted): 1991M03 2012M11
Included observations: 261 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LOG(EXPORTS(-1)))	-1.330504	0.059461	-22.37617	0.0000
C	0.008883	0.005163	1.720404	0.0866

R-squared	0.659073	Mean dependent var	-0.000990
Adjusted R-squared	0.657757	S.D. dependent var	0.142063
S.E. of regression	0.083109	Akaike info criterion	-2.129686
Sum squared resid	1.788956	Schwarz criterion	-2.102371
Log likelihood	279.9240	Hannan-Quinn criter.	-2.118706
F-statistic	500.6932	Durbin-Watson stat	2.033226
Prob(F-statistic)	0.000000		

Null Hypothesis: D(LOG(EXPORTS)) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=15)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-22.37617	0.0000
Test critical values:	1% level	-3.455289	
	5% level	-2.872413	
	10% level	-2.572638	

*MacKinnon (1996) one-sided p-values.

After excluding insignificant variables in the test equation, the result of the ADF test is that the differenced series of *log(exports)* does not have unit root. This means that *log(exports)~I(1)*, that is, the *log(exports)* needs a single differencing to make it stationary.

Next, we will conduct a seasonality test by regressing the detrended series with seasonal indicators. The F-statistic of the model captures the information on the amount of variability as explained by the seasonal indicators and its corresponding p-value determines the decision on the hypothesis test. The null hypothesis for this test is that all coefficients of the seasonal indicators are all equal to zero, that is, there is no seasonality in the series.

Dependent Variable: LEXPORTS_D
Method: Least Squares
Date: 03/24/16   Time: 17:55
Sample: 1991M01 2012M11
Included observations: 263

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.021885	0.027387	-0.799106	0.4250
@SEAS(1)	-0.058848	0.038288	-1.536976	0.1256
@SEAS(2)	-0.047138	0.038288	-1.231118	0.2194
@SEAS(3)	0.038298	0.038288	1.000259	0.3181
@SEAS(4)	-0.038939	0.038288	-1.016987	0.3101
@SEAS(5)	0.013765	0.038288	0.359513	0.7195
@SEAS(6)	0.046241	0.038288	1.207702	0.2283
@SEAS(7)	0.055285	0.038288	1.443903	0.1500
@SEAS(8)	0.060479	0.038288	1.579558	0.1155
@SEAS(9)	0.100160	0.038288	2.615925	0.0094
@SEAS(10)	0.081609	0.038288	2.131413	0.0340
@SEAS(11)	0.010715	0.038288	0.279858	0.7798

R-squared	0.138089	Mean dependent var	5.72E-15
Adjusted R-squared	0.100316	S.D. dependent var	0.132315
S.E. of regression	0.125503	Akaike info criterion	-1.268416
Sum squared resid	3.953520	Schwarz criterion	-1.105428
Log likelihood	178.7967	Hannan-Quinn criter.	-1.202915
F-statistic	3.655747	Durbin-Watson stat	0.347980
Prob(F-statistic)	0.000080		

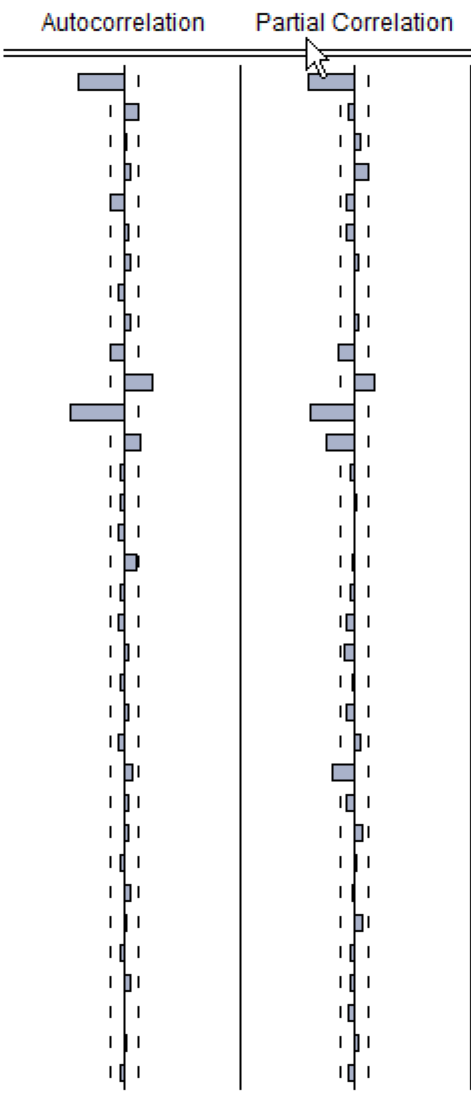
Since the F-test p-value is less than 0.10, we conclude that *log(exports)* has seasonality.

As a result of these tests, the input series will be *dlog(exports,1,12)*.

ARIMA MODELING

The preliminary tests show that *log(exports)* is I(1) and has seasonality. The detrended and deseasonalized series will be *dlog(exports,1,12)*.

Date: 03/24/16 Time: 16:34
Sample: 1992M02 2012M11
Included observations: 250



	AC	PAC	Q-Stat	Prob
1	-0.415	-0.415	43.492	0.000
2	0.120	-0.063	47.122	0.000
3	0.023	0.060	47.255	0.000
4	0.056	0.115	48.048	0.000
5	-0.116	-0.068	51.479	0.000
6	0.036	-0.067	51.811	0.000
7	0.043	0.040	52.290	0.000
8	-0.059	-0.008	53.191	0.000
9	0.053	0.040	53.928	0.000
10	-0.132	-0.140	58.517	0.000
11	0.252	0.178	75.241	0.000
12	-0.492	-0.396	139.42	0.000
13	0.155	-0.248	145.77	0.000
14	-0.028	-0.043	145.98	0.000
15	-0.027	0.012	146.17	0.000
16	-0.062	0.004	147.19	0.000
17	0.102	-0.009	149.99	0.000
18	-0.047	-0.036	150.59	0.000
19	-0.057	-0.068	151.48	0.000
20	0.030	-0.093	151.73	0.000
21	-0.033	-0.009	152.02	0.000
22	0.030	-0.068	152.27	0.000
23	-0.058	0.062	153.22	0.000
24	0.071	-0.200	154.64	0.000
25	0.030	-0.082	154.88	0.000
26	0.032	0.069	155.16	0.000
27	-0.029	0.009	155.40	0.000
28	0.049	-0.022	156.09	0.000
29	0.022	0.070	156.23	0.000
30	-0.046	-0.033	156.84	0.000
31	0.061	-0.041	157.91	0.000
32	-0.007	-0.065	157.93	0.000
33	0.019	0.032	158.04	0.000
34	-0.036	-0.065	158.42	0.000

At lag 1, the AC and PAC are both found to be significant. We can choose to add either AR or MA term since they are equal at the first lag. In this case, I have added AR(1) in the model to be re-estimated.

Dependent Variable: DLOG(EXPORTS,1,12)
Method: Least Squares
Date: 03/24/16 Time: 16:58
Sample (adjusted): 1992M03 2012M11
Included observations: 249 after adjustments
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.39E-05	0.004153	-0.010570	0.9916
AR(1)	-0.414610	0.057886	-7.162519	0.0000

R-squared	0.171979	Mean dependent var	-9.15E-05
Adjusted R-squared	0.168627	S.D. dependent var	0.101662
S.E. of regression	0.092695	Akaike info criterion	-1.911012
Sum squared resid	2.122297	Schwarz criterion	-1.882760
Log likelihood	239.9210	Hannan-Quinn criter.	-1.899640
F-statistic	51.30168	Durbin-Watson stat	2.046836
Prob(F-statistic)	0.000000		

Inverted AR Roots	-.41

Since the AR(1) term is significant, it will be retained in the model. Next, we need to check the residual correlogram again to see if there are still significant correlations.

Date: 03/24/16 Time: 17:03
Sample: 1992M03 2012M11
Included observations: 249
Q-statistic probabilities adjusted for 1 ARMA term

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.028	-0.028	0.2035	
		2	-0.024	-0.024	0.3452	0.557
		3	0.118	0.117	3.9049	0.142
		4	0.032	0.038	4.1657	0.244
		5	-0.117	-0.112	7.6710	0.104
		6	0.012	-0.007	7.7092	0.173
		7	0.050	0.040	8.3504	0.214
		8	-0.035	-0.007	8.6601	0.278
		9	-0.020	-0.014	8.7647	0.363
		10	-0.034	-0.061	9.0689	0.431
		11	0.043	0.044	9.5476	0.481
		12	-0.494	-0.494	73.888	0.000
		13	-0.038	-0.061	74.265	0.000
		14	0.021	-0.013	74.386	0.000
		15	-0.083	0.018	76.236	0.000
		16	-0.048	-0.008	76.860	0.000
		17	0.089	-0.020	78.976	0.000
		18	-0.044	-0.058	79.504	0.000
		19	-0.090	-0.085	81.725	0.000
		20	-0.003	-0.062	81.728	0.000

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		21	-0.019	-0.016	81.831	0.000
		22	-0.000	-0.033	81.831	0.000
		23	-0.032	-0.006	82.108	0.000
		24	0.088	-0.230	84.240	0.000
		25	0.092	0.040	86.586	0.000
		26	0.046	0.054	87.172	0.000
		27	-0.001	-0.024	87.172	0.000
		28	0.065	0.015	88.369	0.000
		29	0.033	0.050	88.673	0.000
		30	-0.023	-0.071	88.828	0.000
		31	0.059	-0.042	89.836	0.000
		32	0.029	-0.035	90.072	0.000
		33	0.010	0.018	90.099	0.000
		34	-0.015	-0.054	90.167	0.000
		35	0.033	0.009	90.484	0.000
		36	-0.074	-0.183	92.070	0.000

For the seasonal lags, the AC cuts off rapidly after lag 12, but the PAC decays slowly on the succeeding seasonal periods. This is a special form of the correlogram and can be modeled using MA(12). Instead of adding MA(12), we will add an SMA(12) in the model to indicate that this is a seasonal term. Then we proceed to check the significance of the recently added SMA(12) term and the resulting residual correlogram.

Dependent Variable: DLOG(EXPORTS,1,12)
Method: Least Squares
Date: 03/24/16 Time: 17:18
Sample (adjusted): 1992M03 2012M11
Included observations: 249 after adjustments
Convergence achieved after 3 iterations
MA Backcast: 1991M03 1992M02

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000547	0.000726	-0.752797	0.4523
AR(1)	-0.365336	0.059175	-6.173823	0.0000
MA(12)	-0.848525	0.029429	-28.83304	0.0000

R-squared	0.530274	Mean dependent var	-9.15E-05
Adjusted R-squared	0.526455	S.D. dependent var	0.101662
S.E. of regression	0.069958	Akaike info criterion	-2.469869
Sum squared resid	1.203953	Schwarz criterion	-2.427490
Log likelihood	310.4987	Hannan-Quinn criter.	-2.452811
F-statistic	138.8549	Durbin-Watson stat	2.072736
Prob(F-statistic)	0.000000		

Inverted AR Roots	-.37			
Inverted MA Roots	.99	.85+.49i	.85-.49i	.49-.85i
	.49+.85i	-.00-.99i	-.00+.99i	-.49-.85i
	-.49+.85i	-.85+.49i	-.85-.49i	-.99

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.054	-0.054	0.7317	
		2	-0.028	-0.031	0.9331	
		3	0.097	0.094	3.3410	0.068
		4	-0.023	-0.014	3.4741	0.176
		5	-0.097	-0.095	5.9054	0.116
		6	-0.008	-0.029	5.9237	0.205
		7	-0.047	-0.051	6.4903	0.261
		8	-0.068	-0.058	7.6889	0.262
		9	0.037	0.029	8.0548	0.328
		10	-0.072	-0.074	9.3953	0.310
		11	0.051	0.051	10.066	0.345
		12	-0.086	-0.106	12.009	0.284
		13	-0.013	-0.021	12.055	0.360
		14	0.010	-0.008	12.082	0.439
		15	-0.062	-0.066	13.113	0.439
		16	-0.017	-0.019	13.189	0.512
		17	0.082	0.059	15.007	0.451
		18	-0.057	-0.057	15.894	0.460
		19	-0.068	-0.071	17.147	0.444
		20	0.002	-0.061	17.148	0.513
		21	-0.008	0.002	17.164	0.579
		22	-0.013	-0.015	17.209	0.639
		23	0.003	-0.010	17.212	0.698
		24	0.017	-0.000	17.296	0.747
		25	0.035	0.022	17.639	0.777
		26	0.007	-0.011	17.653	0.819
		27	-0.023	-0.037	17.797	0.851
		28	0.047	0.019	18.418	0.860
		29	0.025	0.039	18.598	0.884
		30	-0.024	-0.022	18.768	0.905
		31	0.020	0.002	18.878	0.924
		32	0.021	0.015	19.001	0.940
		33	0.018	0.029	19.090	0.953
		34	-0.000	-0.014	19.090	0.965
		35	0.041	0.044	19.576	0.969
		36	-0.066	-0.053	20.870	0.962

There are no more bars exceeding the limits. The bar on the third lag is near the Bartlett's Band. We will check the significance of this term in the model. Since AC is larger than the PAC, we will try adding an MA(3) term in the model.

Dependent Variable: DLOG(EXPORTS,1,12)
Method: Least Squares
Date: 03/24/16 Time: 17:24
Sample (adjusted): 1992M03 2012M11
Included observations: 249 after adjustments
Convergence achieved after 3 iterations
MA Backcast: 1990M12 1992M02

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000557	0.000803	-0.693300	0.4888
AR(1)	-0.385235	0.059211	-6.506100	0.0000
MA(3)	0.134377	0.060755	2.211779	0.0279
SMA(12)	-0.849298	0.029001	-29.28464	0.0000

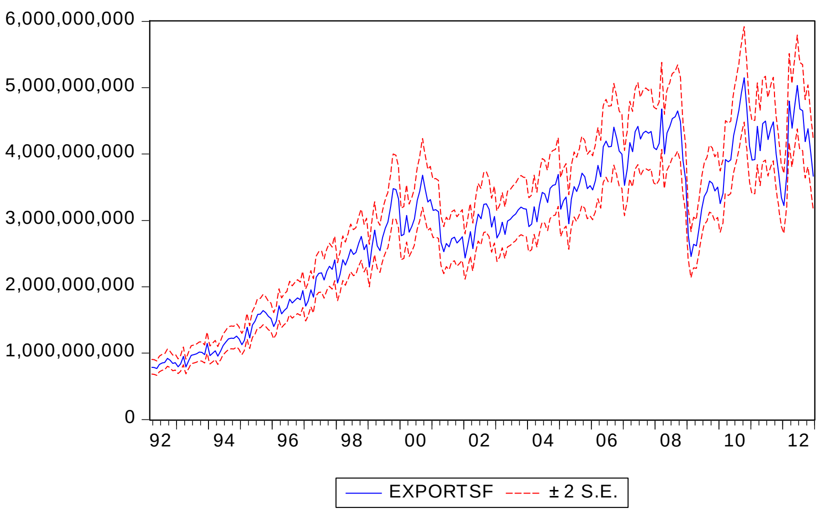
R-squared	0.538379	Mean dependent var	-9.15E-05
Adjusted R-squared	0.532726	S.D. dependent var	0.101662
S.E. of regression	0.069493	Akaike info criterion	-2.479242
Sum squared resid	1.183179	Schwarz criterion	-2.422737
Log likelihood	312.6656	Hannan-Quinn criter.	-2.456498
F-statistic	95.24612	Durbin-Watson stat	2.032039
Prob(F-statistic)	0.000000		

Inverted AR Roots	-.39			
Inverted MA Roots	.99	.85+.49i	.85-.49i	.49+.85i
	.49-.85i	.26+.44i	.26-.44i	-.00-.99i
	-.00+.99i	-.49-.85i	-.49+.85i	-.51
	-.85+.49i	-.85-.49i	-.99	

The MA(3) is significant and will be retained in the model. Our final model is denoted by ARIMA(1,1,3)x(0,1,1).

FORECASTS

We will now forecast the total exports from 2012M12 to 2013M04.



Forecast: EXPORTSF	
Actual: EXPORTS	
Forecast sample: 1991M01 2013M04	
Adjusted sample: 1992M03 2012M12	
Included observations: 249	
Root Mean Squared Error	2.24E+08
Mean Absolute Error	1.57E+08
Mean Abs. Percent Error	5.280547
Theil Inequality Coefficient	0.036157
Bias Proportion	0.000466
Variance Proportion	0.001148
Covariance Proportion	0.998386

QUARTER	POINT FORECAST	90% INTERVAL FORECAST	
		LOWER	UPPER
2012M12	3,616,012,819.271	766,601,182.156	6,465,424,456.386
2013M01	3,198,626,335.670	930,311,479.950	5,466,941,191.390
2013M02	3,315,975,549.009	885,273,246.627	5,746,677,851.391
2013M03	3,616,323,580.809	950,352,395.885	6,282,294,765.733
2013M04	3,319,027,511.416	775,558,444.799	5,862,496,578.033

CONCLUSION

ARIMA modeling was applied to a Philippine exports time series with the help of EViews. The ADF test confirmed that the initial model *log(exports)* has a unit root and needed a single differencing to make it stationary. In addition, the model has been revised to deseasonalize the series. Lastly, the selected model was used to construct a forecast.