Modeling and Forecasting Brazilian Export Values

Ashwin Bhargava

ST 534 - Applied Time Series

Dr. Donald Martin

Fall 2019

Dataset Introduction

In this analysis, we obtained a data set with measures of Brazilian exports from the years 1995-2018. Export value was measured in terms of Brazil's nominal domestic currency, scaled by millions of units. The data was quarterly in that there were four records per year. This resulted in a total of 96 records overall, with 12 future export values (corresponding to quarters in 2019, 2020, and 2021) forecasted.

For simplicity in re-analyzing the data set, it has been entered using the 'cards' statement in SAS. A link to a txt file containing the data will also be attached.

Prior to analyzing the data, we expected an upward trend in exports. This is due to the growing economy of the country as well as the impact of globalization. We were also expecting a seasonal trend since the data was split into three-month blocks.

Model Fitting

A plot of the raw data and corresponding Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and Inverse Autocorrelation Function (IACF) were displayed to get a better sense of the time series data.

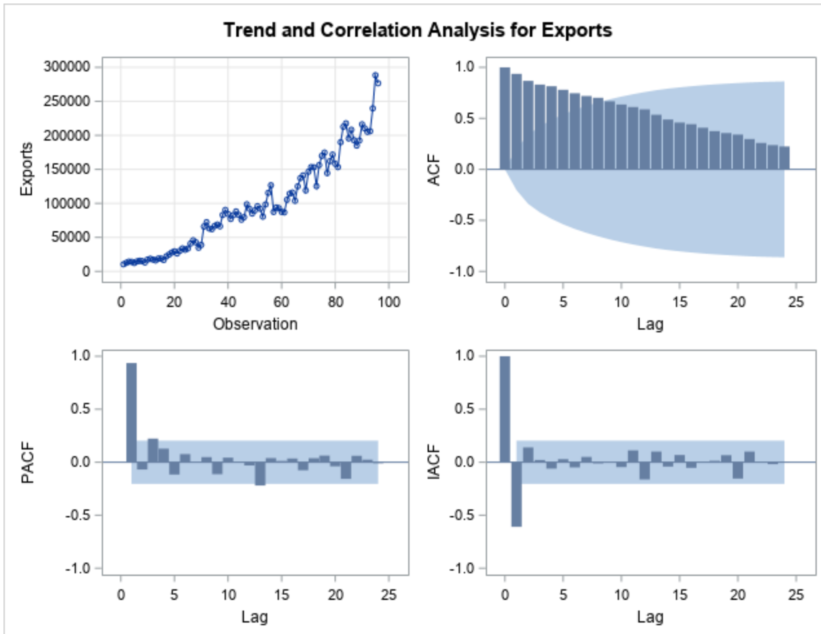

Figure 1: Raw export data and time series diagnostics

The original data has an upward trend, as expected. Notice that as time goes by, the variance in the export time series seems to be increasing, as suggested by the increasing fluctuation in the data. Thus, a log transformation on the exports seems to be appropriate.

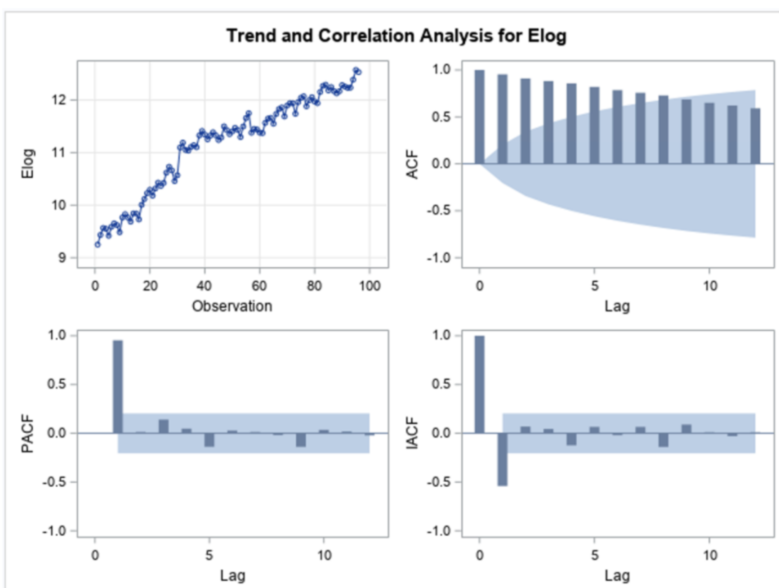

Figure 2: Log transformed data and diagnostics

The logarithmic transformation took care of the problem of unequal variance. However, there is still a slowly decaying pattern in the autocorrelation function, which suggests that we should perform a Dickey-Fuller unit root test to investigate if taking a difference is needed.

Table 1: Dickey-Fuller Test Results

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	0.2787	0.7474	2.44	0.9964		
	1	0.2656	0.7441	2.36	0.9955		
	2	0.2618	0.7431	3.78	0.9999		
Single Mean	0	-2.1464	0.7575	-1.52	0.5204	4.53	0.0606
	1	-1.7744	0.8020	-1.29	0.6324	3.93	0.0961
	2	-1.1892	0.8657	-1.43	0.5657	8.82	0.0010
Trend	0	-15.3936	0.1496	-3.02	0.1331	4.86	0.2189
	1	-15.2388	0.1542	-2.80	0.2015	4.11	0.3669
	2	-5.9936	0.7353	-1.77	0.7133	2.08	0.7632

In the trend plot of exports, we can see a clear trend pattern, so we look at the test under "trend" type. The p-value is insignificant at all lags, which means we fail to reject the null hypothesis that there is a unit root in the data. Thus, a difference should be taken.

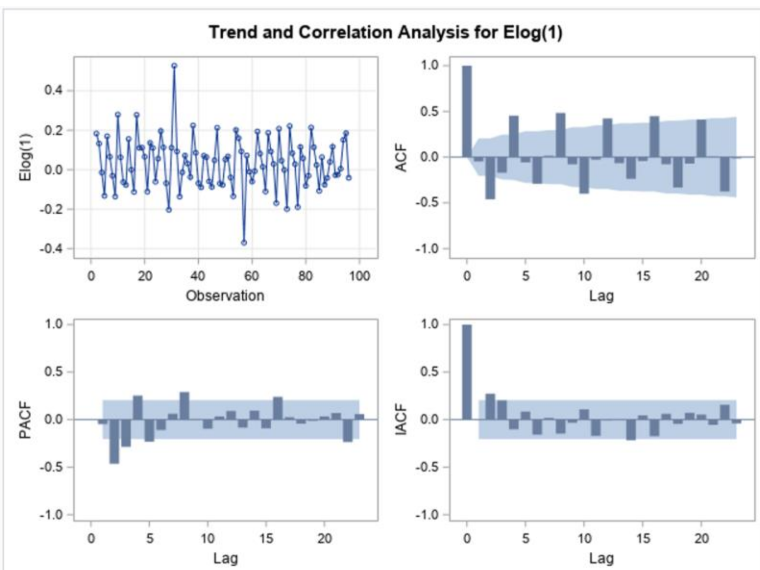

Figure 3: Differenced log-transformed data and diagnostics

The differenced series on logged exports and its corresponding diagnostics are plotted above. The trend plot displays a constant variance pattern, with the mean of the series centered around 0. The autocorrelation function plot suggests a seasonal effect.

Notice that our data is quarterly, so we have 4 observations each year. In the trend plot, it seems like the seasonal length is 4 quarters/1 year since every 4 data points form a whole "cycle." In the ACF, we have significant spikes every other quarter, suggesting that the seasonal length might be 2 quarters. Thus, we will investigate models with these two seasonal lengths and propose the better model using model comparison techniques, such as AIC.

Model Using s = 2

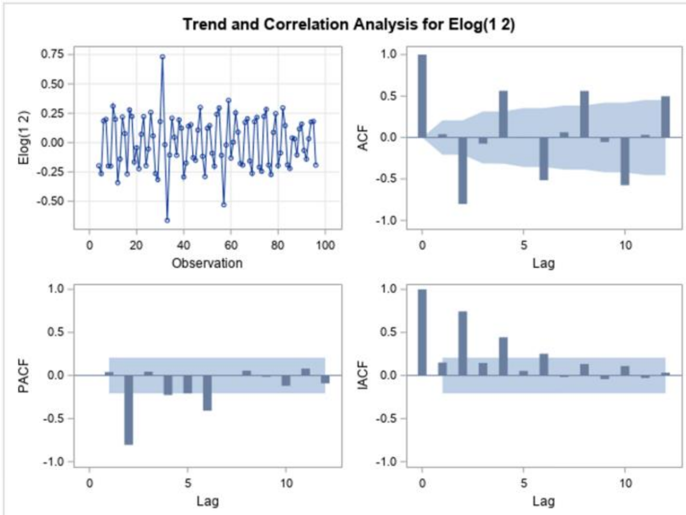

Figure 4: Seasonal length of 2 trend and correlation analysis

After taking a difference of 1 on the original series and taking a difference of 2 to account for the seasonal effect, the series is centered around 0. Notice that within a season (2 quarters), ACF and IACF die out immediately. Across the seasons, ACF and IACF are damping out. Thus, we propose a white noise model at the regular level and an ARMA(1,1) at the seasonal level:

Table 2: Model diagnostics for ARMA(1,1) seasonal (2) model

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.72261	0.08068	8.96	<.0001	2
AR1,1	-0.60254	0.09253	-6.51	<.0001	2

AIC	-131.694
SBC	-126.629
Number of Residuals	93

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.71	4	0.6068	0.080	0.055	-0.113	-0.047	0.044	-0.036
12	15.50	10	0.1147	0.181	0.106	0.070	-0.264	0.008	0.054
18	24.45	16	0.0801	-0.048	0.063	-0.087	0.218	-0.089	-0.093
24	30.48	22	0.1073	-0.041	0.077	0.072	-0.170	0.081	-0.014

The coefficients are significant, and the residuals are white noise, suggesting a good fit. We will continue to try other models to optimize the fit.

Next, we can try a higher-order model at the seasonal level; we propose ARMA(1,2):

Table 3: Diagnostics for ARMA(1,2) seasonal (2) model

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.71295	0.08399	8.49	<.0001	4
AR1,1	-0.99020	0.02207	-44.86	<.0001	2

AIC	-142.724
SBC	-137.659
Number of Residuals	93

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	7.33	4	0.1197	0.116	-0.118	-0.121	-0.037	0.062	0.164
12	11.34	10	0.3317	0.172	0.030	0.009	-0.062	0.040	-0.050
18	16.36	16	0.4284	-0.008	0.202	-0.034	0.049	-0.019	0.002
24	19.58	22	0.6095	-0.042	0.030	0.105	-0.090	0.038	0.053

This time the AIC is -142.724 which is much lower than the previous model, suggesting a better model. Thus, for s = 2, we propose ARIMA (0,1,1) x $(1,1,2)_2$

Model Using s = 4

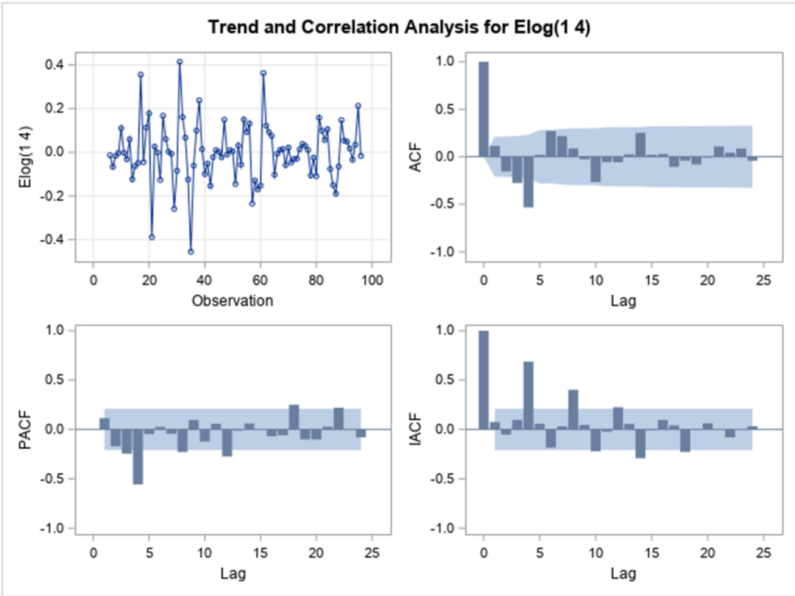

Figure 5: Seasonal length of 4 trend and correlation analysis

After taking a seasonal difference of 4, the trend plot above shows that the constant variance assumption is not ideal, but still appropriate and reasonable; the mean of the series is also centered around 0. Within a season (4 quarters), the ACF and IACF seem to die out immediately. Across the seasons, ACF dies out after season 1, and the IACF is damping out. Thus, we propose a white noise model at the regular level, and an ARMA(0,1):

Table 4: Model diagnostics for ARMA(0,1) seasonal (4) model

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.94511	0.04408	21.44	<.0001	4

AIC	-161.962
SBC	-159.451
Number of Residuals	91

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	10.21	5	0.0696	-0.011	-0.183	-0.229	0.001	-0.020	0.139
12	12.90	11	0.2998	0.103	0.075	-0.068	-0.074	0.014	-0.002
18	20.79	17	0.2360	-0.067	0.228	0.004	0.008	-0.114	0.037
24	22.52	23	0.4892	-0.020	-0.016	0.042	-0.063	0.086	-0.019

The coefficient is significant, and residuals are white noise. The AIC is -161.962. Thus, for s = 4, we propose ARIMA (0,1,0) x $(0,1,1)_4$.

Model Selection

The best model in s = 2 has AIC of -142.724; the best model in s = 4 has AIC of -161.962. Since we were able to obtain a lower AIC by modeling the seasonal effect using seasonal length s = 4, that model should be selected. Therefore, our final model is **ARIMA(0,1,0)x$(0,1,1)_4$.**

Forecast

Forecasts for variable Elog				
Obs	Forecast	Std Error	95% Confidence Limits	
97	12.4285	0.0988	12.2348	12.6222
98	12.5818	0.1398	12.3079	12.8558
99	12.6806	0.1712	12.3451	13.0161
100	12.6634	0.1977	12.2760	13.0508
101	12.5614	0.2235	12.1234	12.9994
102	12.7147	0.2466	12.2314	13.1981
103	12.8135	0.2677	12.2887	13.3382
104	12.7963	0.2873	12.2332	13.3595
105	12.6943	0.3075	12.0915	13.2971
106	12.8476	0.3265	12.2076	13.4876
107	12.9464	0.3444	12.2713	13.6215
108	12.9292	0.3615	12.2207	13.6377

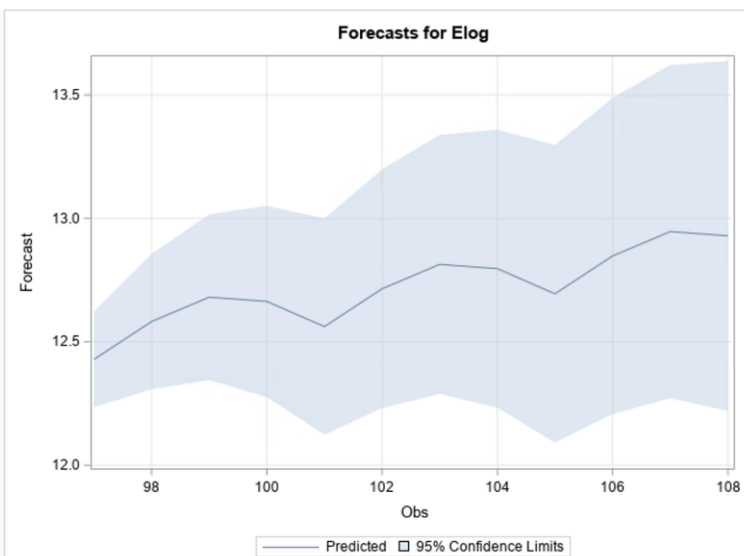

Figure 6: Forecast of log-transformed results using the selected model

Figure 6 above show the forecasted value for the logged exports. To get the original exports forecasts, we need to apply the inverse of the log function:

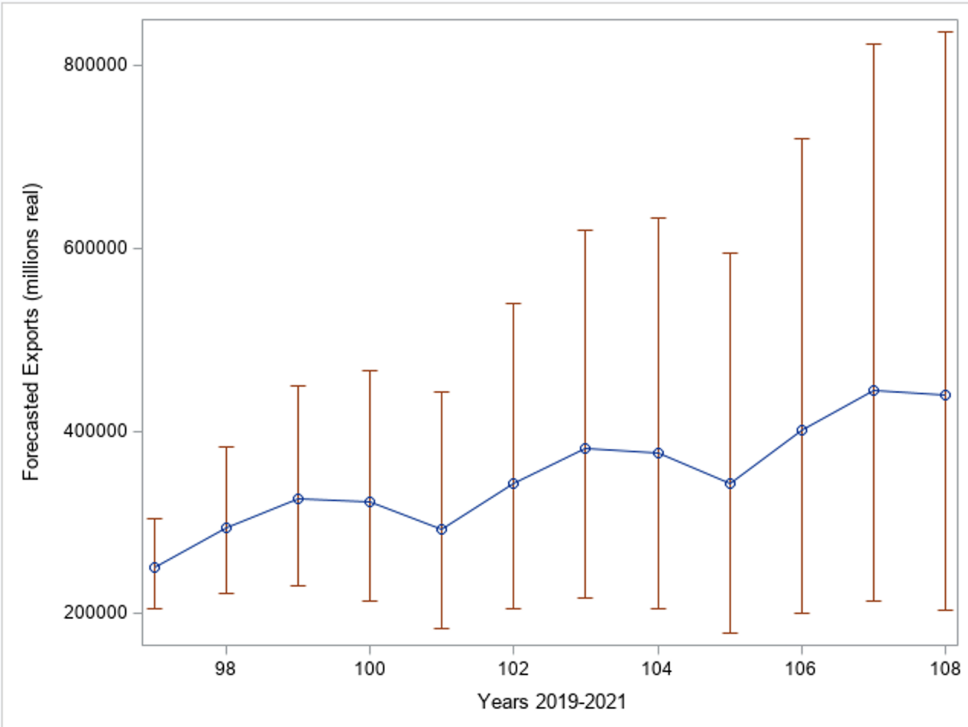

Figure 7: Forecast of Exports between 2019-2021

Table of Forecasted Values in 2019-2021 for Brazilian Exports

Exports (in millions Real)	Lower 95% confidence limit	Upper 95% confidence limit
376094.1	205495.16	633780.41
400681.8	200315.53	720414.32
251045.43	205827.13	303221.28
292545.19	184128.47	442155.24
294074.52	221431.47	382992.93
322209.39	214484.76	465490.21
326188.75	229821.46	449593.84
341667.21	178353.34	595453.59
342876.14	205126.6	539313.82
380529.28	217231.68	620470.58
440010.35	202947.96	837122.75
444941.02	213472.93	823635.34

Conclusion

In this IMF provided dataset of quarterly export data from Brazil, there was a definite, increasing trend in the data. Something of concern we saw in this trend is a ‘megaphone’ shape in its residuals, suggesting that the variance is increasing across the time series. Using a log transformation we brought ‘stability’ to the data. Based on the results of a Dickey-Fuller test with the transformed data, a difference was taken to introduce stationarity.

After comparing the model fit and looking at the trend and diagnostic plots, we concluded that a seasonal effect is present, and the seasonal length is 4 quarters/1 year. The final model has the form of ARIMA(0,1,0)x$(0,1,1)_4$. Forecasted values also show a continued upward trend, with a season length of 4 quarters, which is consistent with our findings.

SAS Code:

```
data brazil;
input Count $ Quarter $ Exports;
/* Exports are measured in millions of currency units.
   Data is quarterly from 1995-2018
*/
cards;
9  1995Q1   10401.00
10 1995Q2   12493.41
11 1995Q3   14259.50
12 1995Q4   14053.16
13 1996Q1   12305.88
14 1996Q2   14576.31
15 1996Q3   15560.48
16 1996Q4   15084.72
17 1997Q1   13162.00
18 1997Q2   17401.76
19 1997Q3   18528.58
20 1997Q4   17398.23
21 1998Q1   16098.85
22 1998Q2   18804.60
23 1998Q3   18785.81
24 1998Q4   16781.03
25 1999Q1   22153.37
26 1999Q2   24727.75
27 1999Q3   27646.22
28 1999Q4   29511.05
29 2000Q1   26401.85
30 2000Q2   30238.77
31 2000Q3   33767.67
32 2000Q4   31755.78
33 2001Q1   33576.48
34 2001Q2   40837.16
35 2001Q3   45700.99
36 2001Q4   42666.83
37 2002Q1   34811.53
38 2002Q2   38896.19
39 2002Q3   65893.98
40 2002Q4   72261.51
41 2003Q1   63030.91
42 2003Q2   62161.50
43 2003Q3   66753.63
44 2003Q4   68852.29
45 2004Q1   66312.69
46 2004Q2   82949.57
47 2004Q3   90341.26
48 2004Q4   84321.34
49 2005Q1   77093.40
50 2005Q2   82694.06
51 2005Q3   88086.50
52 2005Q4   83006.24
53 2006Q1   76051.28
54 2006Q2   79741.78
```

```
55 2006Q3   98607.88
56 2006Q4   91941.01
57 2007Q1   85073.57
58 2007Q2   89584.57
59 2007Q3   95806.74
60 2007Q4   92082.93
61 2008Q1   80459.15
62 2008Q2   98392.59
63 2008Q3   115435.04
64 2008Q4   126593.99
65 2009Q1   87414.68
66 2009Q2   93899.82
67 2009Q3   92906.41
68 2009Q4   87459.56
69 2010Q1   86801.18
70 2010Q2   105294.33
71 2010Q3   114224.84
72 2010Q4   115899.66
73 2011Q1   103744.07
74 2011Q2   124992.55
75 2011Q3   137086.16
76 2011Q4   141072.22
77 2012Q1   119042.78
78 2012Q2   146496.07
79 2012Q3   153302.60
80 2012Q4   153033.54
81 2013Q1   125319.87
82 2013Q2   156291.55
83 2013Q3   169810.17
84 2013Q4   174629.41
85 2014Q1   144496.77
86 2014Q2   162063.93
87 2014Q3   171654.97
88 2014Q4   158159.34
89 2015Q1   153323.91
90 2015Q2   189821.91
91 2015Q3   212635.31
92 2015Q4   217686.88
93 2016Q1   195483.57
94 2016Q2   208217.49
95 2016Q3   192878.21
96 2016Q4   184997.73
97 2017Q1   192447.13
98 2017Q2   216182.61
99 2017Q3   210281.20
100   2017Q4   205133.56
101   2018Q1   206113.14
102   2018Q2   239688.73
103   2018Q3   288391.86
104   2018Q4   276653.89
;
run;
```

```
data brazil;
  set brazil;
  Elog = log(Exports);
run;

ods listing close;
proc arima data=brazil;
identify var = exports; run;
* unequal variance, take the log;

identify var=elog stationarity =(adf=2) nlag=12; run;
* DF test suggest that difference is needed;

identify var = elog(1); run;
* seems to be seasonal every 2 records / 2 quarters by ACF, or every 4
records by trend plot;

**************** if season = 2 ********************;
identify var=elog(1,2) nlag=12; run;

* notice that (0) is not necessary, just to see the (p,d,q)x(P,D,Q);
estimate p = (0)(2) q = (0)(2) noconstant; run;

estimate p = (0)(2) q= (0)(4)  noconstant; run;
* significant model so far. aic= -142;

**************** if season = 4 ********************;
identify var=elog(1,4) nlag=24; run;
estimate p = (0)(0) q = (0)(4) noconstant; run;
* aic = -161 ==> better model than above;

**************** forcasting ********************;
forecast lead=12 out = out; run;
quit;

* Plot the forecasted value back to original units;
data plot;
  set out;
  y=exp(elog);
  l95=exp( l95 );
  u95=exp( u95 );
  forecast = exp(forecast+ std*std/2 );
  obs = _N_;
  if _N_ > 96;
run;



proc sgplot data=plot noautolegend;
 scatter x = obs y=forecast / yerrorlower=L95 yerrorupper=U95;
 series x = obs y=forecast;
 yaxis label='Forecasted Exports (millions real)';
```

```
 xaxis label='Years 2019-2021';
 Run;

data plot;
  set plot;
  label Forecast ="Exports (in millions Real)"
        L95 = "Lower 95% confidence limit"
            U95 = "Upper 95% confidence limit"
  ;
proc report data=plot;
title1 "Table of Forecasted Values in 2019-2021 for Brazilian Exports";
column Forecast L95 U95;
define forecast / order;
run;
```

Link to the dataset:
http://data.imf.org/regular.aspx?key=61545852

(Change country to Brazil and select "Quarterly" tab - see "Exports of Goods and Services, Nominal, Domestic Currency" row)