Econometrics Week 6 Test

William Schill

This file was originally created using Jupyter with Python and HTML and has been saved as a pdf rather than exported to a pdf due to errors in exporting and latex, because of this formatting is a little wonky.

Out[1]: Toggle Code.

Were going to import the data and then using *pandas* in python to set the 'YYYY-MM' column to an actual date-time index.

Out[75]:

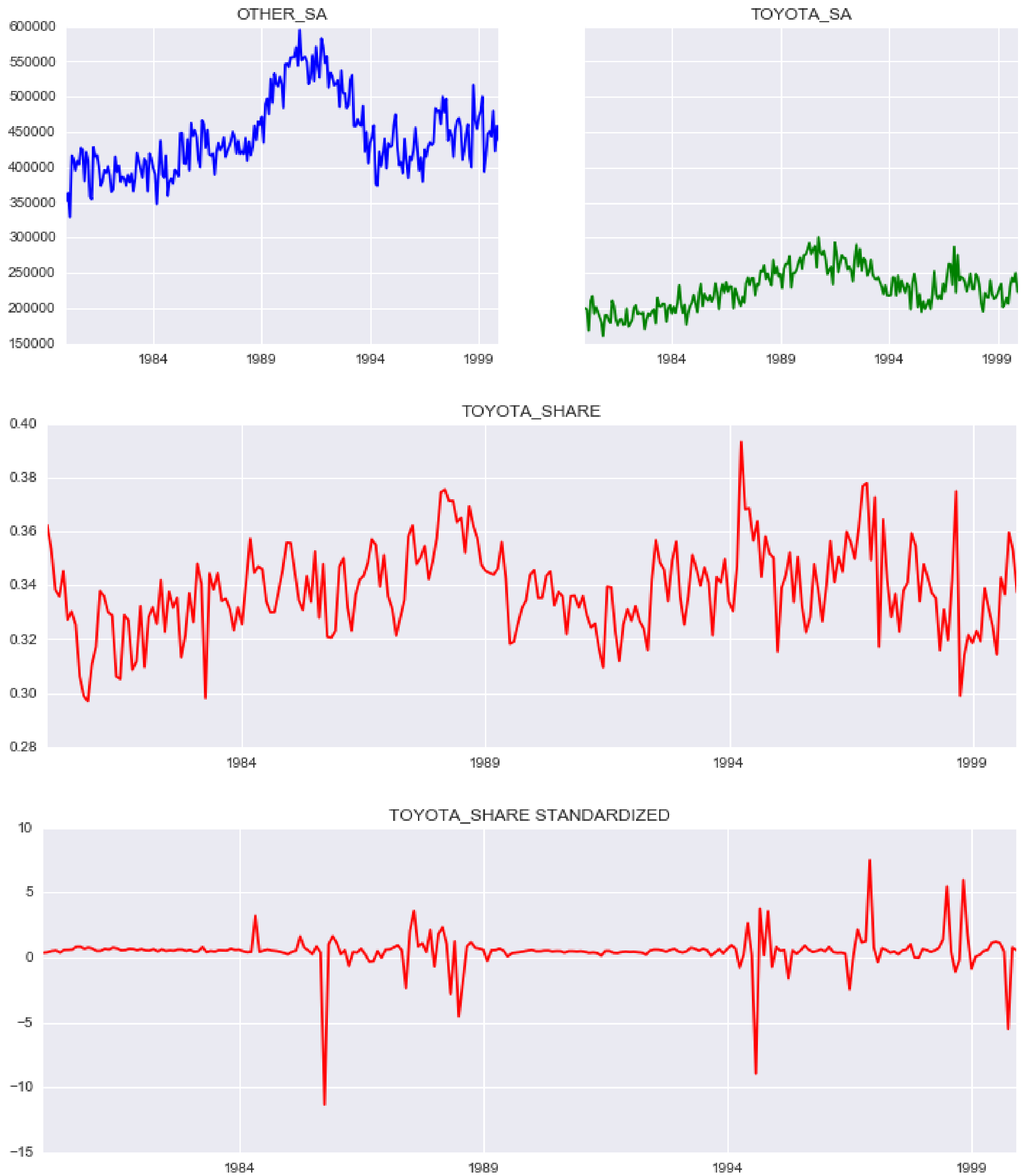
	YYYY-MM	TOYOTA	OTHER	TOYOTA_SA	OTHER_SA
0	1980M01	175734	315111	200015.142857	352054.297619
1	1980M02	200479	377893	198443.190476	363393.202381
2	1980M03	200373	385236	168488.047619	329570.059524

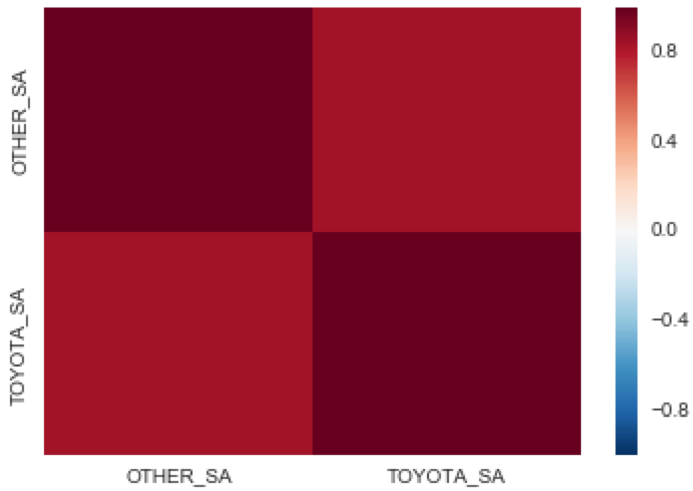
```
DatetimeIndex(['1980-01-01', '1980-02-01', '1980-03-01'], dtype='datetime64[n
s]', freq=None)


            TOYOTA   OTHER      TOYOTA_SA       OTHER_SA
1999-10-01  259871  436813  237267.047619  422691.535714
1999-11-01  263580  470356  249988.809524  458444.535714
1999-12-01  206280  433729  223074.571429  438129.630952
240

            TOYOTA   OTHER      TOYOTA_SA       OTHER_SA
2000-01-01  225841  392589  250122.142857  429532.297619
2000-02-01  267084  499199  265048.190476  484699.202381
2000-03-01  294538  551228  262653.047619  495562.059524
12
```

PART A

```
1980-01-01    352054.297619
1980-02-01    363393.202381
1980-03-01    329570.059524
Name: OTHER_SA, dtype: float64 1980-01-01    200015.142857
1980-02-01    198443.190476
1980-03-01    168488.047619
Name: TOYOTA_SA, dtype: float64
```

Out[130]: <matplotlib.axes._subplots.AxesSubplot at 0x155c5a90>





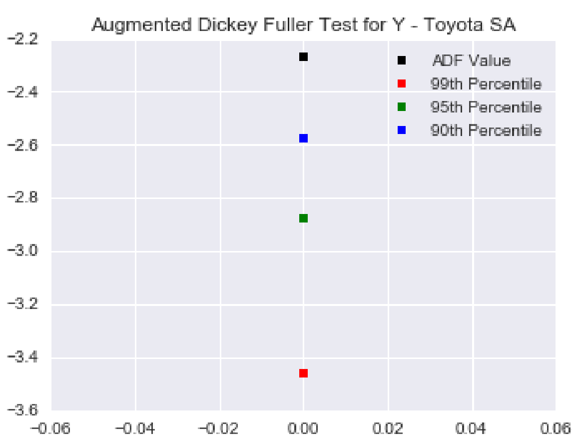
(a):

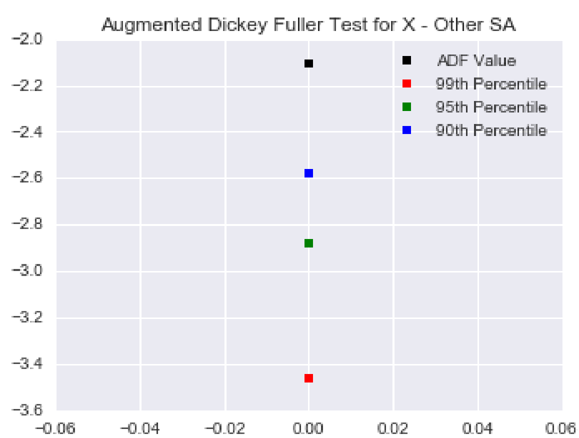
The variables have been set in the python code and plotted above. The toyota share stays with in the 30 to 40 percent range of the other share. However it seems that when comparing the TOYOTA_SA to OTHER_SA, that they did not experience as significant uptake in the early 1990's as some of the companies might have, but never-the-less, both sets of data seem correlated with each other.

The standardized version of the TOYOTA SHARE confirms the suspicion above and even highlights were Toyoata maintaned the same chunk of share for quite some time.

PART B

```
                   Results: Ordinary least squares
===================================================================
Model:              OLS              Adj. R-squared:     0.293
Dependent Variable: d_y              AIC:                5221.9067
Date:               2017-03-01 09:55 BIC:                5239.2258
No. Observations:   236              Log-Likelihood:     -2606.0
Df Model:           4                F-statistic:        25.34
Df Residuals:       231              Prob (F-statistic): 2.05e-17
R-squared:          0.305            Scale:              2.3331e+08
-------------------------------------------------------------------
             Coef.     Std.Err.    t    P>|t|    [0.025    0.975]
-------------------------------------------------------------------
const      19281.8945 8430.4101  2.2872 0.0231 2671.5701 35892.2190
lag1          -0.0832    0.0368 -2.2623 0.0246   -0.1557    -0.0107
d_lag1        -0.5630    0.0699 -8.0565 0.0000   -0.7007    -0.4253
d_lag2        -0.3243    0.0745 -4.3546 0.0000   -0.4710    -0.1776
d_lag3        -0.0639    0.0650 -0.9835 0.3264   -0.1920     0.0641
-------------------------------------------------------------------
Omnibus:              0.924        Durbin-Watson:           2.013
Prob(Omnibus):        0.630        Jarque-Bera (JB):        0.736
Skew:                 0.132        Prob(JB):                0.692
Kurtosis:             3.072        Condition No.:           1948470
===================================================================
* The condition number is large (2e+06). This might indicate
strong multicollinearity or other numerical problems.

Y ADF :

(-2.262284216076055, 0.18442063656746183, {'5%': -2.8738660999177132, '1%': -
3.4583663275730476, '10%': -2.5733390785693766}, <statsmodels.tsa.stattools.R
esultsStore object at 0x0000000012ABF6A0>)


 X ADF:

(-2.1056761420521739, 0.2422383304024896, {'5%': -2.8738660999177132, '1%': -
3.4583663275730476, '10%': -2.5733390785693766}, <statsmodels.tsa.stattools.R
esultsStore object at 0x0000000012ABF668>)
```





```
                      Results: Ordinary least squares
===================================================================
Model:              OLS              Adj. R-squared:     0.293
Dependent Variable: y                AIC:                5221.9067
Date:               2017-03-01 09:55 BIC:                5239.2258
No. Observations:   236              Log-Likelihood:     -2606.0
Df Model:           4                F-statistic:        25.34
Df Residuals:       231              Prob (F-statistic): 2.05e-17
R-squared:          0.305            Scale:              2.3331e+08
-------------------------------------------------------------------
             Coef.     Std.Err.    t    P>|t|    [0.025    0.975]
-------------------------------------------------------------------
ylag1         -0.0832    0.0368 -2.2623 0.0246   -0.1557    -0.0107
ydiff_lag1    -0.5630    0.0699 -8.0565 0.0000   -0.7007    -0.4253
ydiff_lag2    -0.3243    0.0745 -4.3546 0.0000   -0.4710    -0.1776
ydiff_lag3    -0.0639    0.0650 -0.9835 0.3264   -0.1920     0.0641
const      19281.8945 8430.4101  2.2872 0.0231 2671.5701 35892.2190
-------------------------------------------------------------------
Omnibus:              0.924        Durbin-Watson:           2.013
Prob(Omnibus):        0.630        Jarque-Bera (JB):        0.736
Skew:                 0.132        Prob(JB):                0.692
Kurtosis:             3.072        Condition No.:           1948470
===================================================================
* The condition number is large (2e+06). This might indicate
strong multicollinearity or other numerical problems.
```

(b):

The value for y_{t-1} is -0.0832 and the t-value is significant at the 5% level with a value of 0.0246.

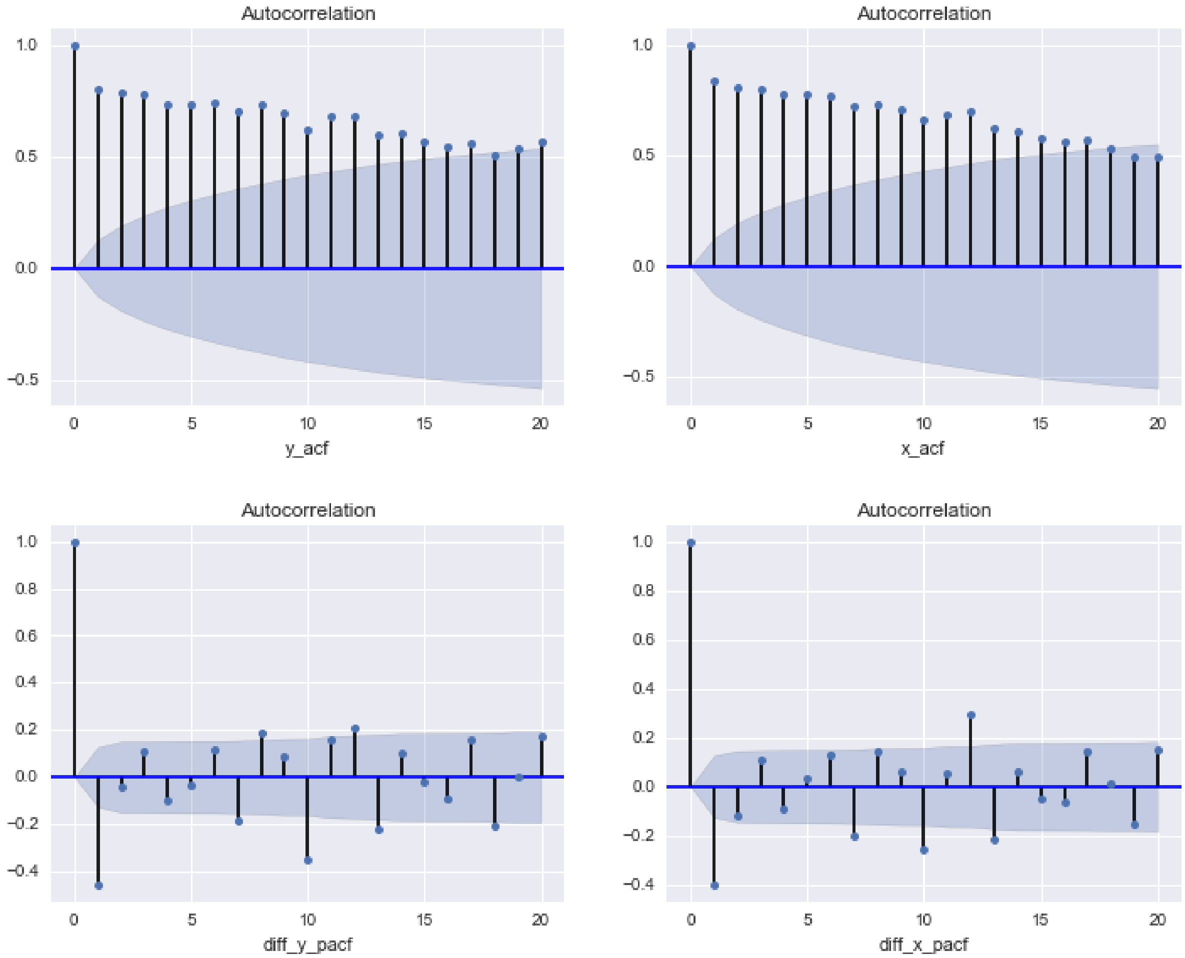
If we manually calculate the OLS regression using the differenced y and its subsequent lags and compare it to the ADF test regression results from statsmodels, we can see that they are nearly identitical. Thus we can use the statsmodels test as set up above and:

(i): **The Augmented Dickey Fuller test for y has a test statistic of -2.2623 and we cannot reject the Null Hypothesis that there is a unit root and the data is not stationary. This can also be confirmed based on the plot above and the p-value of the test statistic.**

(ii): **The ADF test for x has a test statistic of -2.1057 and we cannot reject the Null Hypothesis that there is a unit root and the data is not stationary. This can also be confirmed based on the plot above and the p-value of the test statistic.**

Additionally, based on the Autocorrelation plots below, we have further confirmation.

Out[131]: <matplotlib.text.Text at 0x15d987f0>



PART C

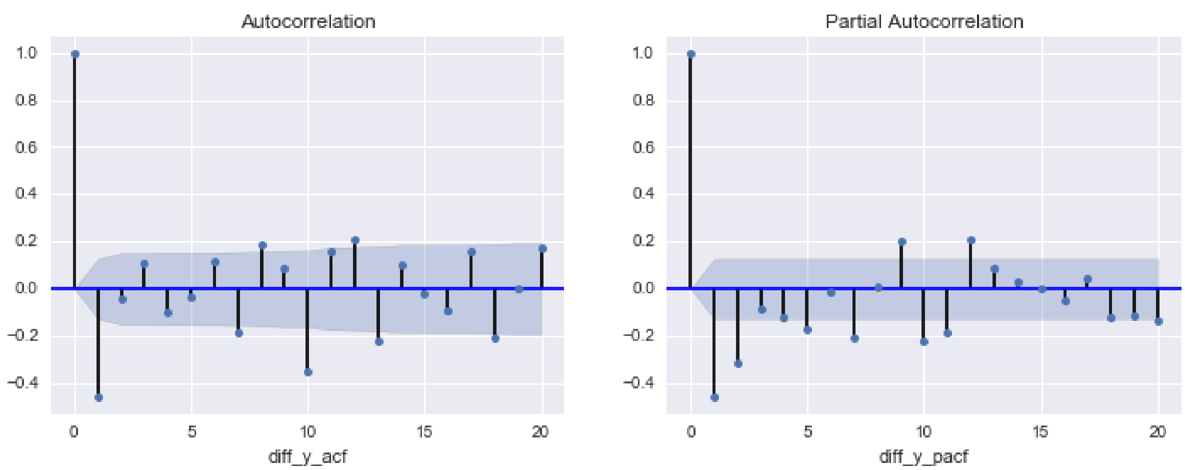
```
                        Results: Ordinary least squares
====================================================================
Model:              OLS              Adj. R-squared:     0.252
Dependent Variable: y                AIC:                5146.6454
Date:               2017-03-01 10:43 BIC:                5163.9646
No. Observations:   236              Log-Likelihood:     -2568.3
Df Model:           4                F-statistic:        20.75
Df Residuals:       231              Prob (F-statistic): 1.26e-14
R-squared:          0.264            Scale:              1.6961e+08
---------------------------------------------------------------------
            Coef.   Std.Err.     t     P>|t|     [0.025      0.975]
---------------------------------------------------------------------
e_lag1     -0.2930    0.0680  -4.3057  0.0000     -0.4270    -0.1589
d_e_lag1   -0.2858    0.0785  -3.6396  0.0003     -0.4406    -0.1311
d_e_lag2   -0.1416    0.0754  -1.8794  0.0614     -0.2901     0.0068
d_e_lag3   -0.0960    0.0657  -1.4607  0.1454     -0.2254     0.0335
const      24.9917  847.8255   0.0295  0.9765  -1645.4676  1695.4510
--------------------------------------------------------------------
Omnibus:              12.771       Durbin-Watson:           2.010
Prob(Omnibus):        0.002        Jarque-Bera (JB):        25.834
Skew:                 -0.214       Prob(JB):                0.000
Kurtosis:             4.563        Condition No.:           19350
====================================================================
* The condition number is large (2e+04). This might indicate
strong multicollinearity or other numerical problems.


ADF RESULTS ON THE RESIDUAL OF X AND Y

(-4.3057053622664112, 0.000434214635598987, {'5%': -2.8738660999177132, '1%':
-3.4583663275730476, '10%': -2.5733390785693766}, <statsmodels.tsa.stattools.
ResultsStore object at 0x000000001492F048>)
```

(c):

Based on the residual in part two, and the results of the ADF test, it appears we can reject the Null Hypothesis that the residuals are not stationary, thus it is not cointegrated between X and y and the linear combination of the two can form a stationary series.

PART D

```
Sigfinicance Rule =  0.129099444874
   Greater than Significance Rule  lags
0                            True     1
1                            True     2
2                           False     3
3                           False     4
4                           False     5
5                           False     6
6                           False     7
7                            True     8
8                            True     9
9                           False    10
10                           True    11
11                           True    12
12                           True    13
13                           True    14
14                          False    15
```

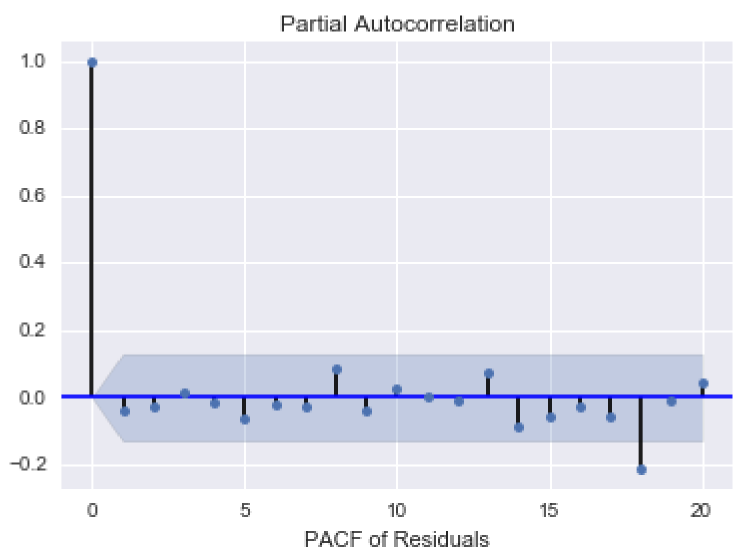


The ACF and PACF plots are created above. However many more lagged terms than "1 to 5, 10, and 12" appear to be significant. We will just go with the ones listed in the instructions of the assignment since handling this selection should essentially take of the rest.

As an aside, I tried to use the *significance test* from the lecture slides just incase the statsmodels methods were doing something slightly different but this still did not give the expected lag structure. The results can be seen above the ACF and PACF plots.

```
                  Results: Ordinary least squares
===================================================================
Model:              OLS              Adj. R-squared:     0.427
Dependent Variable: d_y              AIC:                4980.5786
Date:               2017-03-01 10:43 BIC:                5007.9782
No. Observations:   227              Log-Likelihood:     -2482.3
Df Model:           7                F-statistic:        25.05
Df Residuals:       219              Prob (F-statistic): 5.67e-25
R-squared:          0.445            Scale:              1.9112e+08
--------------------------------------------------------------------
           Coef.    Std.Err.     t      P>|t|     [0.025      0.975]
--------------------------------------------------------------------
const     561.6126  919.5108   0.6108  0.5420  -1250.6101  2373.8353
dlag1      -0.5983    0.0617  -9.7001  0.0000     -0.7199    -0.4768
dlag2      -0.2634    0.0760  -3.4680  0.0006     -0.4132    -0.1137
dlag3      -0.2273    0.0749  -3.0350  0.0027     -0.3749    -0.0797
dlag4      -0.2297    0.0719  -3.1940  0.0016     -0.3714    -0.0879
dlag5      -0.1520    0.0610  -2.4941  0.0134     -0.2722    -0.0319
dlag10     -0.2683    0.0524  -5.1242  0.0000     -0.3715    -0.1651
dlag12      0.2465    0.0546   4.5121  0.0000      0.1388     0.3542
-------------------------------------------------------------------
Omnibus:                0.156        Durbin-Watson:           2.072
Prob(Omnibus):          0.925        Jarque-Bera (JB):        0.148
Skew:                   -0.060       Prob(JB):                0.929
Kurtosis:               2.964        Condition No.:           25239
===================================================================
* The condition number is large (3e+04). This might indicate
strong multicollinearity or other numerical problems.
```

Out[136]: <matplotlib.text.Text at 0x16932940>



Examining the PACF of the residuals, it can be seen that the model created does cover the majority of the auto-correlation lag structure with the exception of the 18th lag. The model also highlights (based on the p-values) that all of the lag structures added are statsitically significant from a standpoint of inclusion and based on the the calculated t-value.

PART E

```
                  Results: Ordinary least squares
===================================================================
Model:              OLS              Adj. R-squared:     0.436
Dependent Variable: d_y              AIC:                4977.7723
Date:               2017-03-01 10:43 BIC:                5008.5968
No. Observations:   227              Log-Likelihood:     -2479.9
Df Model:           8                F-statistic:        22.87
Df Residuals:       218              Prob (F-statistic): 3.18e-25
R-squared:          0.456            Scale:              1.8797e+08
--------------------------------------------------------------------
            Coef.     Std.Err.     t     P>|t|    [0.025     0.975]
--------------------------------------------------------------------
const     4728.0072  2133.7034   2.2159  0.0277  522.6791  8933.3353
dlag1       -0.5223     0.0706  -7.4009  0.0000   -0.6614    -0.3832
dlag2       -0.1866     0.0833  -2.2403  0.0261   -0.3508    -0.0224
dlag3       -0.1581     0.0809  -1.9552  0.0518   -0.3175     0.0013
dlag4       -0.1847     0.0743  -2.4860  0.0137   -0.3311    -0.0383
dlag5       -0.1331     0.0611  -2.1785  0.0304   -0.2535    -0.0127
dlag10      -0.2737     0.0520  -5.2649  0.0000   -0.3762    -0.1712
dlag12       0.2516     0.0542   4.6402  0.0000    0.1448     0.3585
ECM         -0.1503     0.0696  -2.1599  0.0319   -0.2875    -0.0132
-------------------------------------------------------------------
Omnibus:               0.012        Durbin-Watson:           2.043
Prob(Omnibus):         0.994        Jarque-Bera (JB):        0.095
Skew:                  -0.004       Prob(JB):                0.954
Kurtosis:              2.900        Condition No.:           76171
===================================================================
* The condition number is large (8e+04). This might indicate
strong multicollinearity or other numerical problems.
```

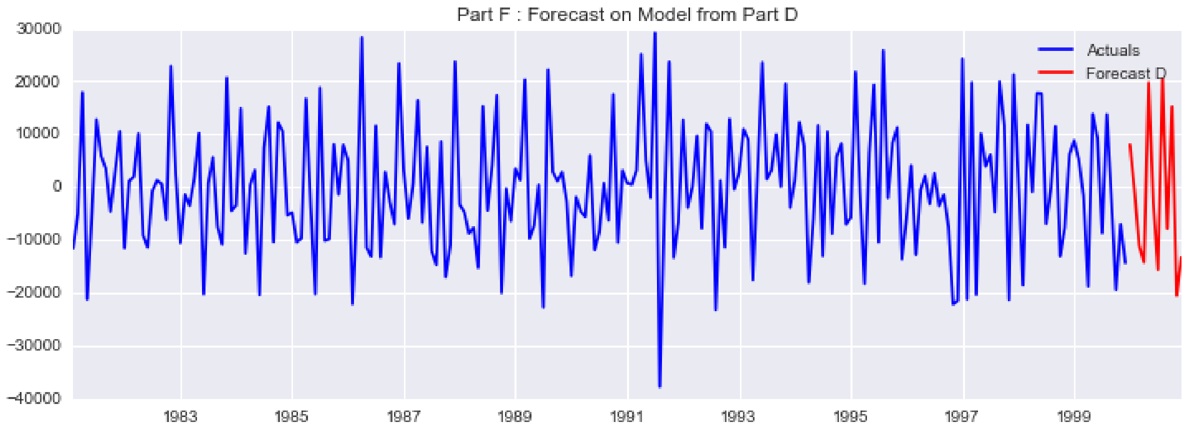
(e):

According to the statsmodels analysis above, at a value of *0.0319*, the Error Correction Term (ECM) is significant at a 5% level but not at a 1% level. For this to have been the case it would have had to have been less than 0.01.

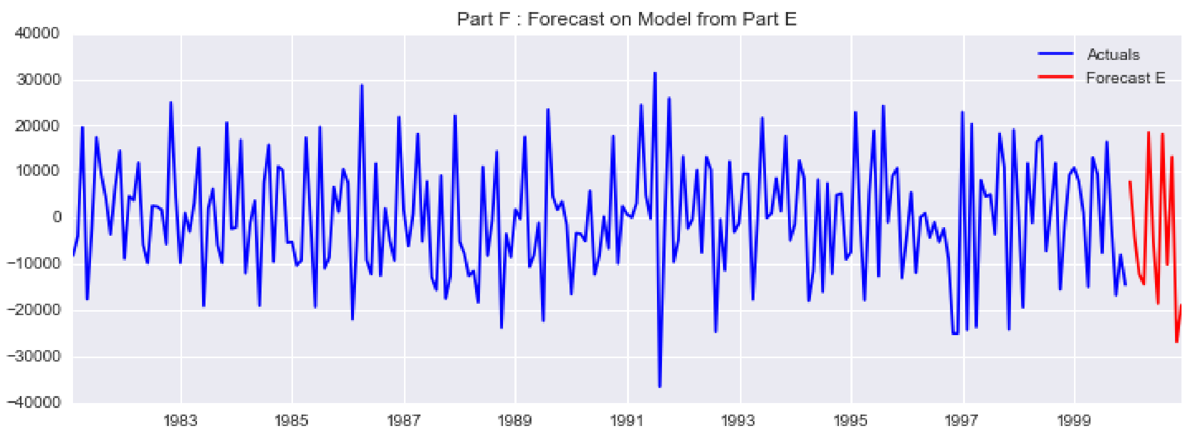
PART F

```
2000-01-01    429532.297619
2000-02-01    484699.202381
2000-03-01    495562.059524
Name: OTHER_SA, dtype: float64
2000-01-01    250122.142857
2000-02-01    265048.190476
2000-03-01    262653.047619
Name: TOYOTA_SA, dtype: float64
```

Out[140]: <matplotlib.text.Text at 0x161ebef0>

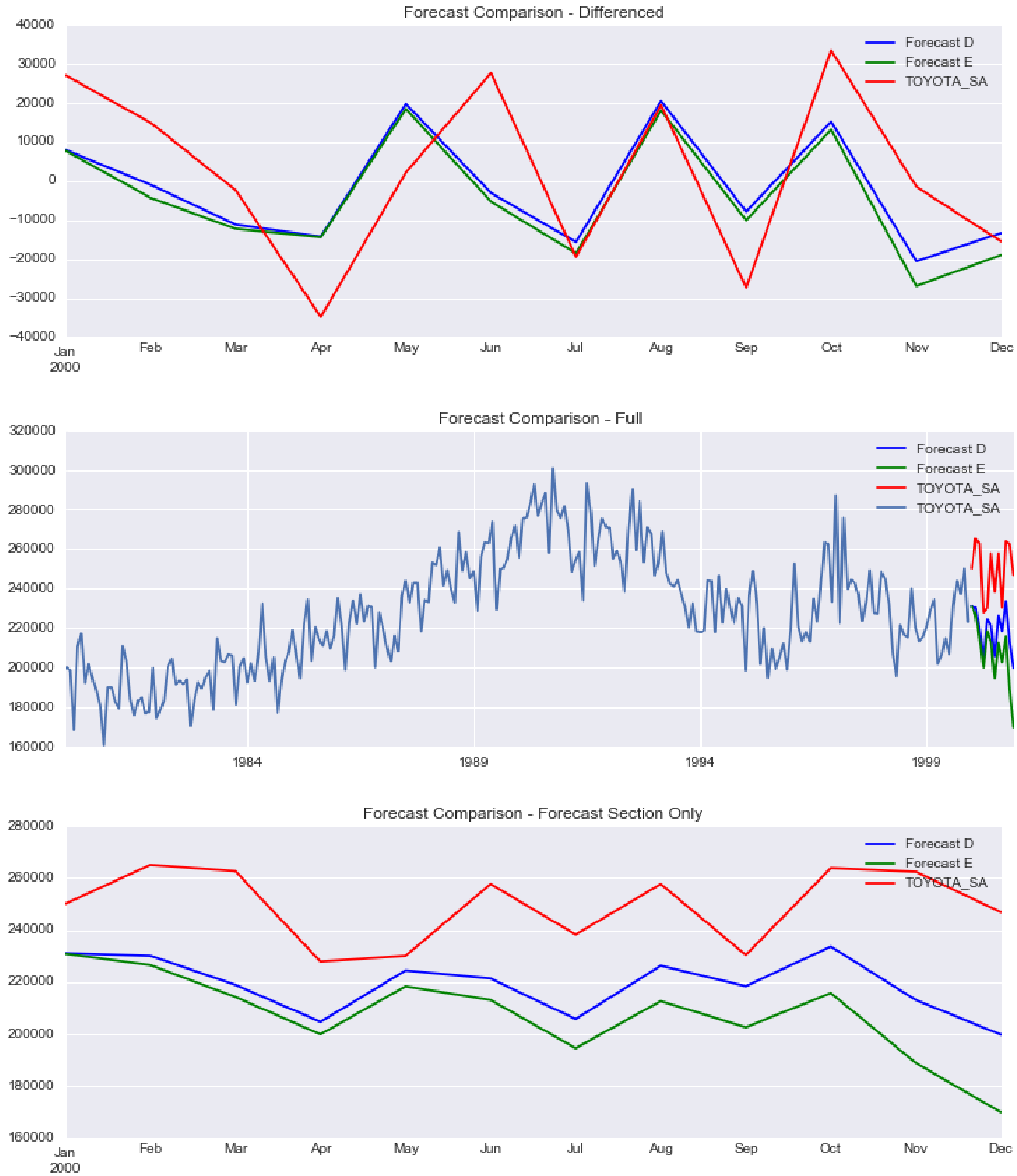


Out[142]: <matplotlib.text.Text at 0x12dd5780>



Out[143]: 223074.57142857101

```
Out[144]: <matplotlib.text.Text at 0x170712e8>
```




```
FIT D  :  RMSE  \  MAE  =  16991.8042351  \  176438.710965

FIT E  :  RMSE  \  MAE  =  18204.7593804  \  186668.020848
```

(f):

The plots above show the differenced forecast, the forecast with the actuals, and finally a zoomed in view of the forecast only reverse differenced. It can be seen that the Error Correction Term actually deviates the forecast farther from the actuals (in red) than the forecast with out the error term. The error ouput shows that the Forecast D is much better at predicting than Forecast E.