Econometrics: Methods and Applications

Erasmus University Rotterdam

Peer-graded Assignment: Test Exercise 6

Goals and skills being used:

- Experience the process of practical application of time series analysis.
- Get hands-on experience with the analysis of time series.
- Give correct interpretation of outcomes of the analysis.

Questions

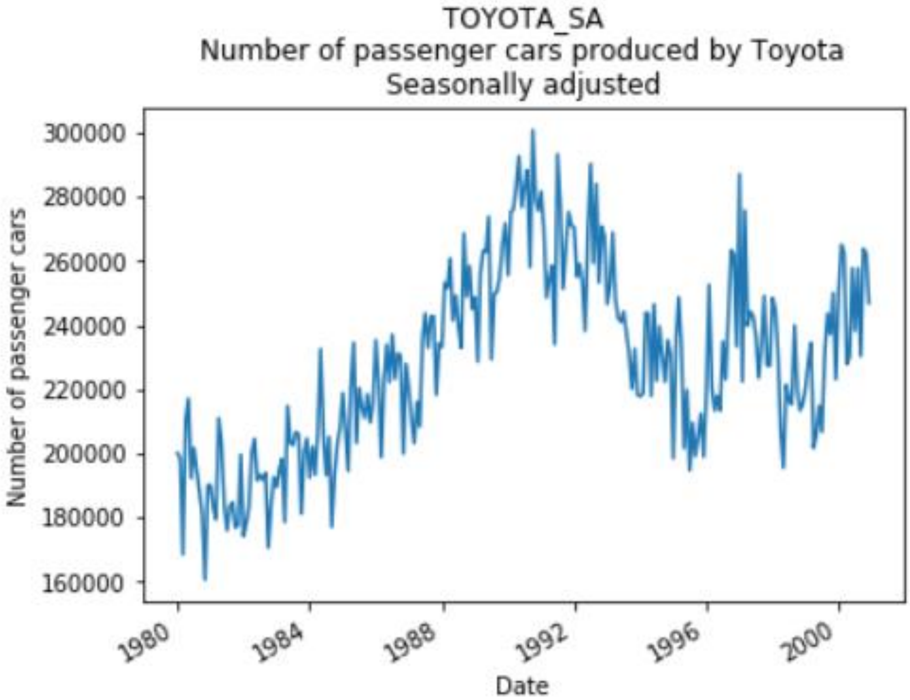
This test exercise uses data that are available in the data file TestExer6. The question of interest is to model monthly production of Toyota passenger cars and to investigate whether the monthly production of all other brands of Japanese passenger cars has predictive power for the production of Toyota. Monthly production data are available from January 1980 until December 2000. The data for January 1980 until December 1999 are used for specification and estimation of models, and the data for 2000 are left out for forecast evaluation purposes.

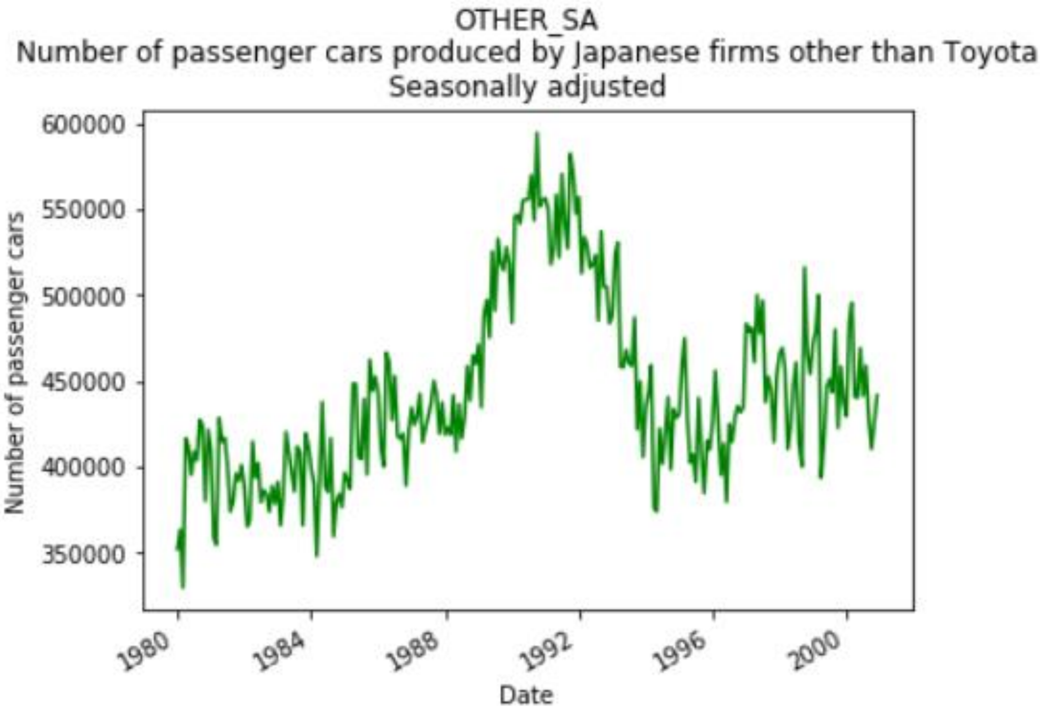
In answering the questions below, you should use the seasonally adjusted production data denoted by '**toyota-sa'** and **'other-sa'**. We will denote these variables by y = toyota-sa and x = other-sa.

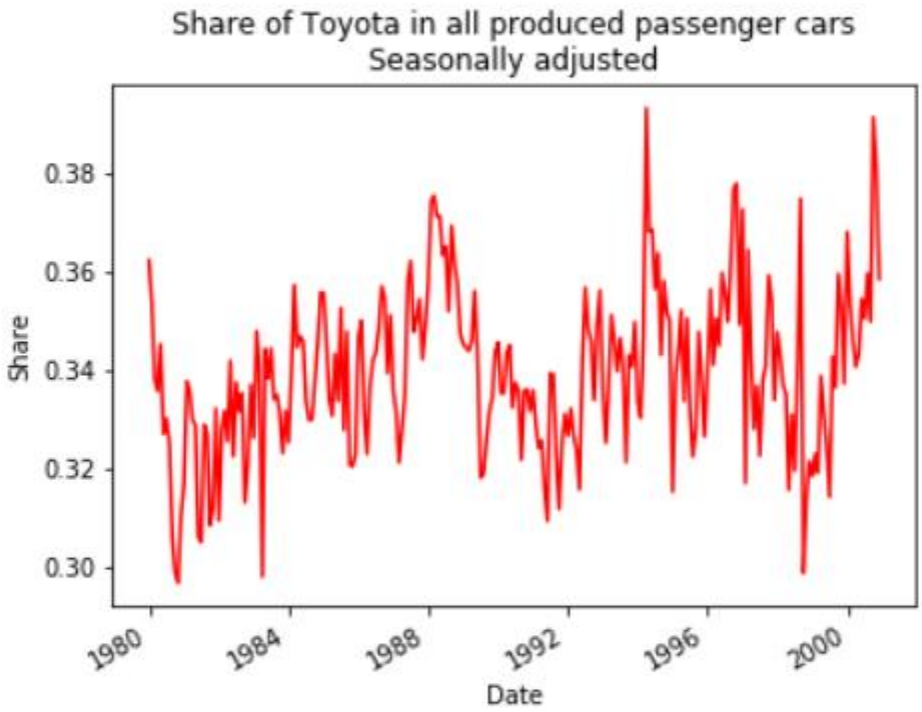
(a) Make time series plots of the variables y_t and x_t , and also of the share of Toyota in all produced passenger cars, that is $y_t /(y_t + x_t)$. What conclusions do you draw from these plots?

series plots of the variables y_t and x_t ,

y = TOYOTA_SA and
x = OTHER_SA







Answer (a)
The share of passage cars produced by Toyota in all produced passenger cars is relatively stable. It fluctuates between 30% and 38%.

(b) (i) Perform the Augmented Dickey-Fuller (ADF) test for y_t . In the ADF test equation, include a constant (α) and three lags of Δy_t , as well as the variable of interest, y_{t-1}. Report the coefficient of y_{t-1} and its standard error and t-value and draw your conclusion.

Calculation (b)i
Augmented Dickey-Fuller (ADF) test for y_t

$$\text{ADF: } \Delta y_t = \alpha + \beta t + \rho y_{t-1} + \sum_{j=1}^{3} \gamma_j \Delta y_{t-j} + \varepsilon_t$$

A time series is said to be “stationary” if it has no trend, exhibits constant variance over time, and has a constant autocorrelation structure over time.

One way to test whether a time series is stationary is to perform an augmented Dickey-Fuller test, which uses the following null and alternative hypotheses:

H_0: The time series is non-stationary. In other words, it has some time-dependent structure and does not have constant variance over time.

H_A: The time series is stationary.

If the p-value from the test is less than some significance level (e.g., α = .05), then we can reject the null hypothesis and conclude that the time series is stationary.

```
                     Results: Ordinary least squares
======================================================================
Model:                OLS              Adj. R-squared:     0.293
Dependent Variable:   delta_y          AIC:                5221.9067
Date:                 2023-01-04 15:50 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
y_lag1          -0.0832    0.0368 -2.2623 0.0246   -0.1557    -0.0107
delta_y_lag1    -0.5630    0.0699 -8.0565 0.0000   -0.7007    -0.4253
delta_y-lag2    -0.3243    0.0745 -4.3546 0.0000   -0.4710    -0.1776
delta_y_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.
```

coefficient of y_{t-1} = -0.0832 std error: 0.0368
t-value: -2.2623
p-value: 0.0246

Answer (b)i

We reject H_0 and conclude that the time series y_t is stationary.

(ii) Perform a similar ADF test for x_t .

Calculation (b)ii

```
                      Results: Ordinary least squares
======================================================================
Model:                OLS                Adj. R-squared:       0.262
Dependent Variable:   y                  AIC:                  5469.0478
Date:                 2023-01-04 18:10   BIC:                  5486.3670
No. Observations:     236                Log-Likelihood:       -2729.5
Df Model:             4                  F-statistic:          21.86
Df Residuals:         231                Prob (F-statistic):   2.58e-15
R-squared:            0.275              Scale:                6.6487e+08
----------------------------------------------------------------------
                 Coef.      Std.Err.      t     P>|t|     [0.025    0.975]
----------------------------------------------------------------------
xlag1            -0.0696     0.0331 -2.1057 0.0363    -0.1348    -0.0045
delta_x_lag1     -0.5112     0.0675 -7.5701 0.0000    -0.6443    -0.3781
delta_y_lag2     -0.3614     0.0703 -5.1387 0.0000    -0.5000    -0.2228
delta_y_lag3     -0.1030     0.0645 -1.5961 0.1118    -0.2301     0.0241
const         31540.3617 14808.7693  2.1298 0.0342 2362.8412 60717.8822
----------------------------------------------------------------------
Omnibus:                 2.655          Durbin-Watson:            2.050
Prob(Omnibus):           0.265          Jarque-Bera (JB):         2.495
Skew:                    -0.125         Prob(JB):                 0.287
Kurtosis:                3.438          Condition No.:            3959566
======================================================================
* The condition number is large (4e+06). This might indicate
strong multicollinearity or other numerical problems.
```

coefficient of x_{t-1} = -0.0696

std error: 0.0331
t-value: -2.1057
p-value: 0.0363

Answer (b)ii

The p-value of x_{t-1} is lower than the threshold value of 0.05. We reject H_0 and conclude that the time series x_t is stationary.

(c) Perform the two-step Engle-Granger test for cointegration of the time series y_t and x_t . In step 1, regress y_t on a constant and x_t . In step 2, perform a regression of the residuals e_t in the model $\Delta e_t = \alpha + \rho e_{t-1} + \beta_1 \Delta e_{t-1} + \beta_2 \Delta e_{t-2} + \beta_3 \Delta e_{t-3} + \omega_t$. What is your conclusion?

Calculation (c)
Engle-Granger test for cointegration
Step 1: OLS: $y_t = \alpha + \beta(x_t) + e_t$
Step 2: ADF: $\Delta e_t = \alpha + \rho e_{t-1} + \beta_1 \Delta e_{t-1} + \beta_2 \Delta e_{t-2} + \beta_3 \Delta e_{t-3} + \omega_t$

```
                       Results: Ordinary least squares
===================================================================
Model:              OLS              Adj. R-squared:     0.252
Dependent Variable: y                AIC:                5146.6454
Date:               2023-01-04 19:05 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
delta_e_lag1   -0.2858   0.0785 -3.6396 0.0003    -0.4406   -0.1311
delta_e_lag2   -0.1416   0.0754 -1.8794 0.0614    -0.2901    0.0068
delta_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.
```

coefficient ρ of e_{t-1} = -0.2930 std error: 0.0680
t-value: **-4.3057**
p-value: 0.0000

Engle-Granger Decision rule (threshold)

- if t-value of $e_{t-1} < -3.8$ → cointegrated
- if t-value of $e_{t-1} > -3.8$ → not cointegrated

Result:
t-value e_{t-1} is **−4.3057** < −3.8 → cointegrated

Answer (c)
The t-value of e_{t-1} is lower than the threshold value of -3.8. Therefore, we conclude that there is cointegration between y_t and x_t and a linear combination of y_t and x_t can form a stationary time series.

(d) Construct the first twelve sample autocorrelations and sample partial autocorrelations of Δy_t and use the outcomes to motivate an AR(12) model for Δy_t . Check that only the lagged terms at lags 1 to 5, 10, and12 are significant, and estimate the following model:
$\Delta y_t = \alpha + \sum_{j=1}^{5} \beta_j \Delta y_{t-j} + \beta_6 \Delta y_{t-10} + \beta_7 \Delta y_{t-12} + \varepsilon_t$
(Recall that the estimation sample is Jan 1980 - Dec 1999).

Calculation (d)

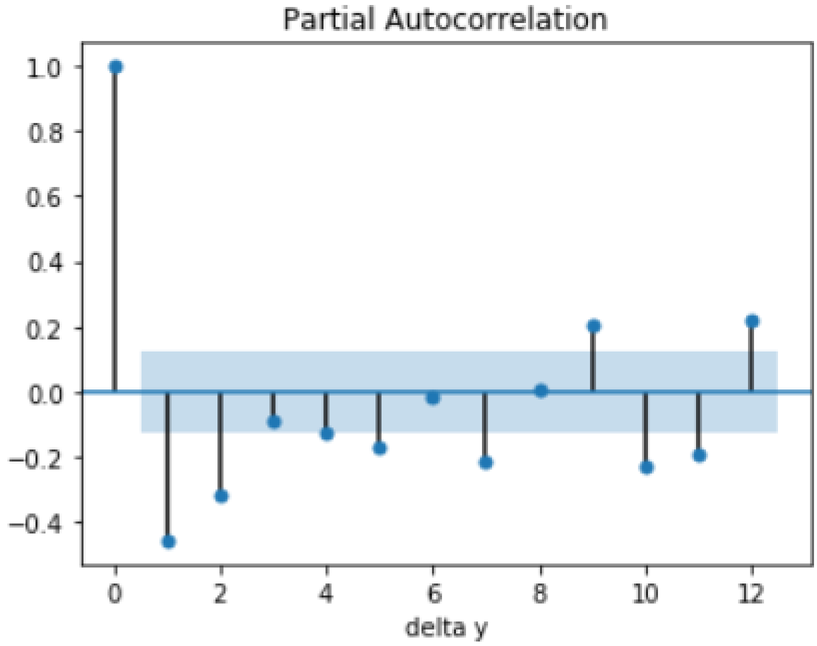
$n = 240$ estimation sample is Jan 1980 - Dec 1999 (monthly data, 20 years)

Significance of the first order sample autocorrelation coefficient. One way to select the orders of an AR model is to test whether the (partial) correlations differ significantly from zero. The null hypothesis of no autocorrelation ($\rho_1 = 0$) can be tested against the alternative $\rho_1 \neq 0$. At (approximate) 5 per cent level, the null hypothesis is rejected if $|r_1| > \frac{2}{\sqrt{n}}$. In this case the first order autocorrelation is significant. (Textbook p.564)

$\frac{2}{\sqrt{n}} = 0.1291$ Significance Threshold estimation sample

These correlations (denoted by r_k) can be estimated from the sample by the sample partial autocorrelation function (SPACF). To select the orders, we can plot the correlations r_k against the time lag k. (Textbook p.548)

The plot of r_k is called the correlogram.



Lag	$\lvert r_k \rvert$	$\lvert r_k \rvert > \frac{2}{\sqrt{n}}$ Significance Threshold
1	0.45808902	> 0.1291
2	0.31796111	> 0.1291
3	0.08675872	
4	0.12326431	
5	0.17378258	> 0.1291
6	0.01803195	
7	0.21442809	> 0.1291
8	0.00786458	
9	0.20881609	> 0.1291
10	0.23168022	> 0.1291
11	0.19678128	> 0.1291
12	0.22153186	> 0.1291

Answer (d)
Only the lagged terms at lags 1, 2, 5, 7 and 9 to 12 are significant.

(e) Extend the model of part (d) by adding the Error Correction (EC) term ($y_t - 0.45x_t$), that is, estimate the ECM
$\Delta y_t = \alpha + \gamma(y_{t-1} - 0.45x_{t-1}) + \sum_{j=1}^{5} \beta_j \Delta y_{t-j} + \beta_6 \Delta y_{t-10} + \beta_7 \Delta y_{t-12} + \varepsilon_t$
(estimation sample is Jan 1980- Dec 1999). Check that the EC term is significant at the 5% level, but not at the 1% level.

Calculation (e)

Error correction model

The ADL model can be rewritten in terms of changes of the variables—that is, in terms of the first differences

$\Delta y_t = y_t - y_{t-1}$ and $\Delta x_t = x_t - x_{t-1}$

We consider this reformulation first for the ADL(1,1) model—that is, the model (7.32) with p = 1, r = 1, and q = 0. By subtracting y_{t-1} from both sides of the equation (7.32), we can write this model as

$$\Delta y_t = \beta_0 \Delta x_t - (1 - \phi)(y_{t-1} - \lambda x_{t-1} - \delta) + \varepsilon_t, \qquad (7.33)$$

with $\delta = \alpha / (1 - \varphi)$ and where $\lambda = (\beta_0 + \beta_1) / (1 - \varphi)$ is the long-run multiplier. (Textbook p.639)

Error correction model

```
                    Results: Ordinary least squares
==================================================================
Model:              OLS              Adj. R-squared:     0.436
Dependent Variable: delta_y          AIC:                4977.7723
Date:               2023-01-05 12:04 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
delta_lag1      -0.5223    0.0706 -7.4009 0.0000  -0.6614  -0.3832
delta_lag2      -0.1866    0.0833 -2.2403 0.0261  -0.3508  -0.0224
delta_lag3      -0.1581    0.0809 -1.9552 0.0518  -0.3175   0.0013
delta_lag4      -0.1847    0.0743 -2.4860 0.0137  -0.3311  -0.0383
delta_lag5      -0.1331    0.0611 -2.1785 0.0304  -0.2535  -0.0127
delta_lag10     -0.2737    0.0520 -5.2649 0.0000  -0.3762  -0.1712
delta_lag12      0.2516    0.0542  4.6402 0.0000   0.1448   0.3585
EC_term         -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.
```

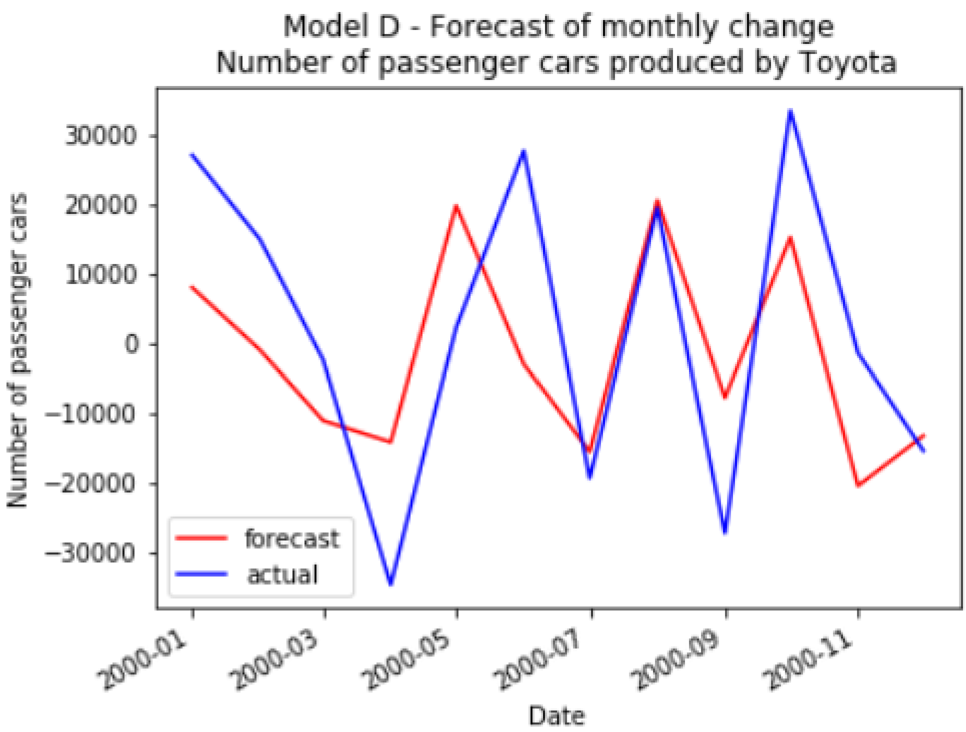
coefficient EC_term = -0.1503 std error: 0.0696
t-value: -2.1599
p-value: **0.0319**

Answer (e)
The EC term is significant at the 5% level, but not at the 1% level. The p-value of **0.0319** is below 0.05 (95%-Confidence Level), but above 0.01 (99%-Confidence Level).

(f) Use the models of parts (d) and (e) to make two series of 12 forecasts of monthly changes in production of Toyota passenger cars in 2000. At each month, you should use the data that are then available, for example, to forecast production for September 2000 you can use the data up to and including August 2000. However, do not re-estimate the model and use the coefficients as obtained in parts (d) and (e). For each of the two forecast series, compute the values of the root mean squared error (RMSE) and of the mean absolute error (MAE). Finally, give your interpretation of the outcomes.

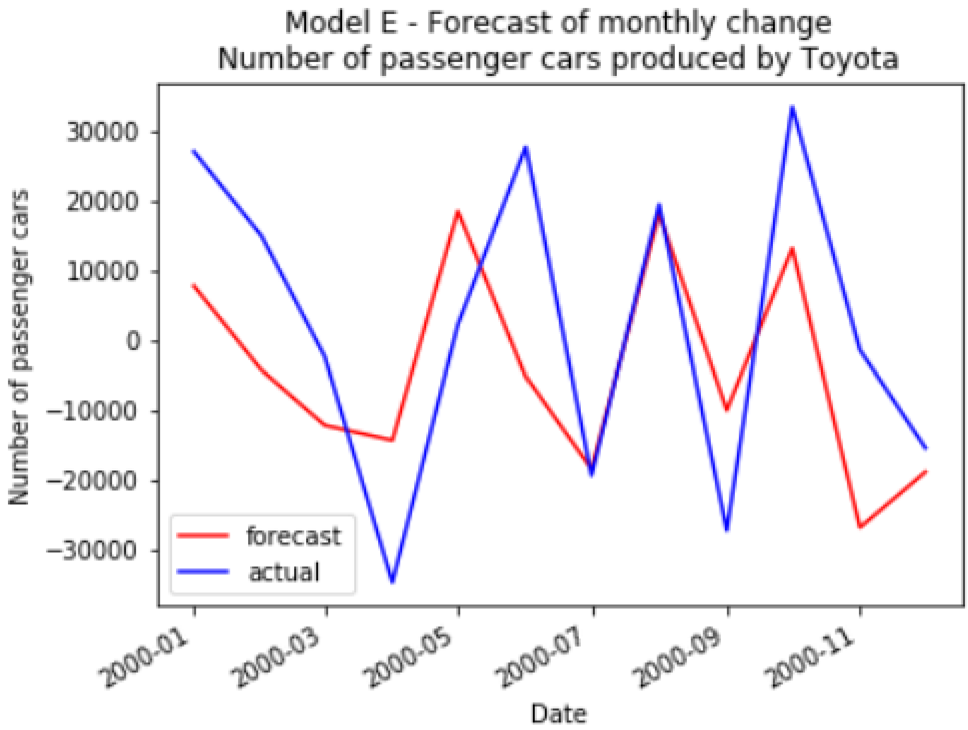
Calculation (f)

Part (d) restricted AR(12) model for Δy_t with lag 1,2,3,4,5,10,12



RMSE = 16991.804 root mean squared error
MAE = 176438.711 mean absolute error

Part (e) Error correction model for Δy_t with lag 1,2,3,4,5,10,12



RMSE =	18204.759	root mean squared error
MAE =	186668.0201	mean absolute error

Answer (f)

The restricted AR(12) model (Model D) produces forecasts with a better fit than the Error correction model (Model E).