Time Series Analysis Spring 2019 Midterm

104304033 統計四 劉書宏

Q1.

(a)Describe three different techniques that can be used to determine if a time series is nonstationary.

## 1. Augmented Dickey-Fuller test

## 2. Wavelet Spectrum Test

## 3. The Priestley-Subba Rao (PSR) Test

(b)Describe the techniques that can be used to eliminate the nonstationarity of time series.

Transformations such as logarithms can help to stabilize the variance of a time series. Differencing can help stabilize the mean of a time series by removing changes in the level of a time series, and so eliminating trend and seasonality

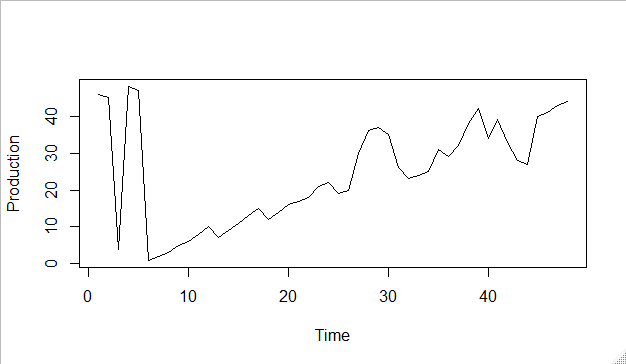
Q2.

Table B.4 contains the US production of blue and gorgonzola cheeses data.

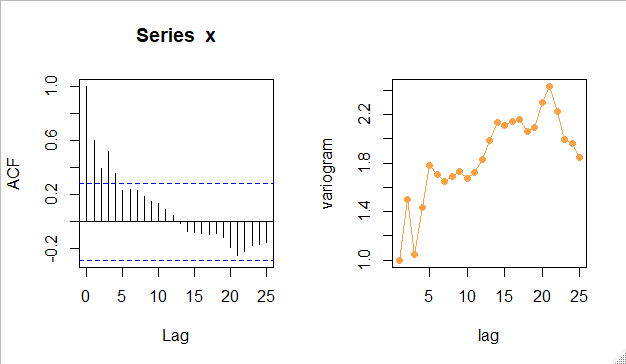
Find the sample autocorrelation function and the variogram for these data.

Is the time series stationary or nonstationary?

Why? If it is nonstationary, what would you do in order to make it stationary?

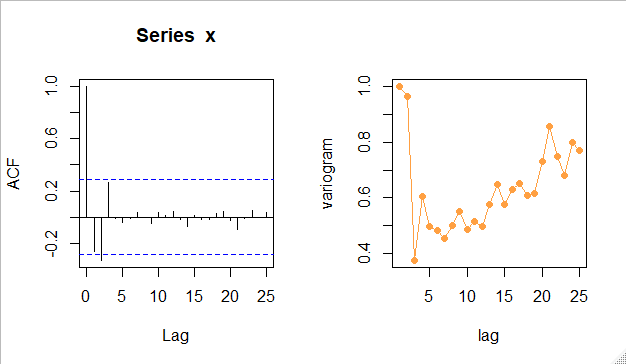


We can that the production has a going-up trend. It’s not stationary



|  |  |  |
| --- | --- | --- |
| lag | ACF | x\_variogram |
| 1.0000 | 0.6029 | 1.0000 |
| 2.0000 | 0.3966 | 1.5019 |
| 3.0000 | 0.5215 | 1.0488 |
| 4.0000 | 0.3594 | 1.4374 |
| 5.0000 | 0.2328 | 1.7818 |
| 6.0000 | 0.2376 | 1.7080 |
| 7.0000 | 0.2340 | 1.6467 |
| 8.0000 | 0.1895 | 1.6889 |
| 9.0000 | 0.1542 | 1.7341 |
| 10.0000 | 0.1329 | 1.6765 |
| 11.0000 | 0.0909 | 1.7252 |
| 12.0000 | 0.0466 | 1.8295 |
| 13.0000 | -0.0146 | 1.9832 |
| 14.0000 | -0.0729 | 2.1342 |
| 15.0000 | -0.0757 | 2.1083 |
| 16.0000 | -0.0892 | 2.1423 |
| 17.0000 | -0.0938 | 2.1571 |
| 18.0000 | -0.0841 | 2.0580 |
| 19.0000 | -0.1188 | 2.0936 |
| 20.0000 | -0.1922 | 2.2982 |
| 21.0000 | -0.2470 | 2.4290 |
| 22.0000 | -0.2226 | 2.2249 |
| 23.0000 | -0.1750 | 1.9917 |
| 24.0000 | -0.1647 | 1.9634 |
| 25.0000 | -0.1510 | 1.8447 |

Because we can see that this data is not stationary, I try to take difference to try to making it stationary.



|  |  |  |
| --- | --- | --- |
| lag | ACF | x\_variogram |
| 1.0000 | -0.2652 | 1.0000 |
| 2.0000 | -0.3288 | 0.9663 |
| 3.0000 | 0.2652 | 0.3763 |
| 4.0000 | -0.0082 | 0.6052 |
| 5.0000 | -0.0371 | 0.4979 |
| 6.0000 | -0.0063 | 0.4822 |
| 7.0000 | 0.0372 | 0.4535 |
| 8.0000 | -0.0046 | 0.5018 |
| 9.0000 | -0.0470 | 0.5528 |
| 10.0000 | 0.0351 | 0.4850 |
| 11.0000 | 0.0148 | 0.5154 |
| 12.0000 | 0.0431 | 0.4980 |
| 13.0000 | -0.0165 | 0.5775 |
| 14.0000 | -0.0674 | 0.6473 |
| 15.0000 | 0.0115 | 0.5772 |
| 16.0000 | -0.0162 | 0.6288 |
| 17.0000 | -0.0192 | 0.6523 |
| 18.0000 | 0.0295 | 0.6086 |
| 19.0000 | 0.0425 | 0.6142 |
| 20.0000 | -0.0266 | 0.7319 |
| 21.0000 | -0.0979 | 0.8565 |
| 22.0000 | -0.0128 | 0.7503 |
| 23.0000 | 0.0526 | 0.6810 |
| 24.0000 | -0.0038 | 0.8005 |
| 25.0000 | 0.0354 | 0.7713 |

It seems like it become more stationary.

Q3.

(a)There are 6 criteria for choosing a model that provides the best fit to historical data, state and briefly describe these criterian

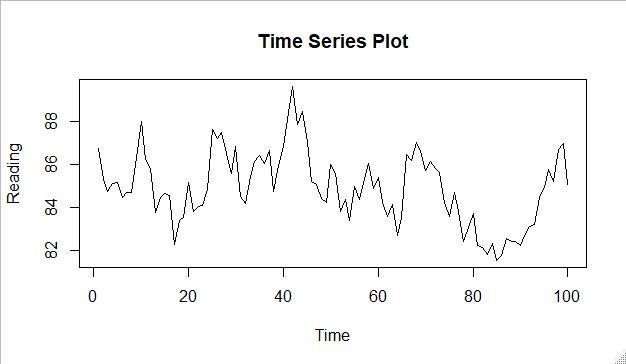
AIC、BUC、Stepwise regression、

(b)Explain the consistency and asymptotically efficiency of model selection criteria. Which criteria are consistent? Which criteria are asymptotically efficient?

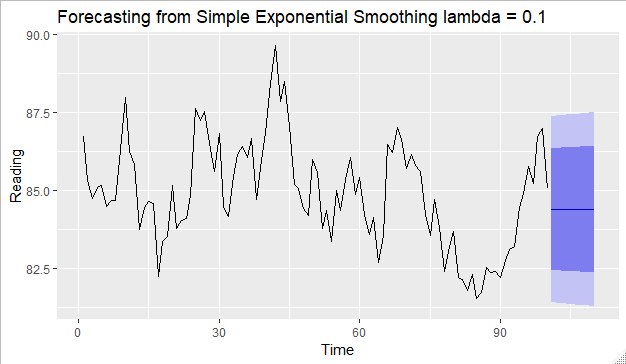
Q4

Time series B.3 contains data on pharmaceutical product sales.

1. Make a time series plot of the data.



1. Use simple exponential smoothing with  to smooth this data . How well does this smoothing procedure work?



1. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors.

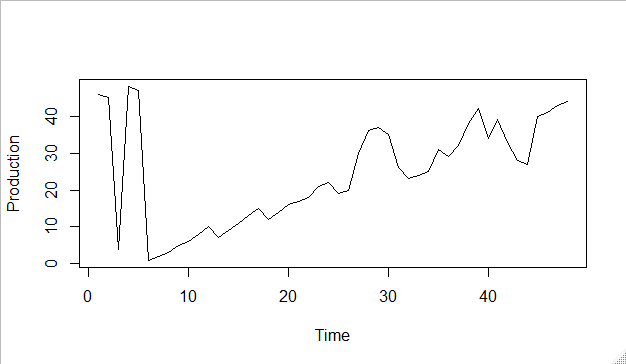
|  |  |  |
| --- | --- | --- |
|  | predict | residuals |
| 1.0000 | 86.7418 | 0.0000 |
| 2.0000 | 86.7418 | -1.4223 |
| 3.0000 | 86.5996 | -1.8641 |
| 4.0000 | 86.4132 | -1.3019 |
| 5.0000 | 86.2830 | -1.1343 |
| 6.0000 | 86.1695 | -1.6920 |
| 7.0000 | 86.0003 | -1.3176 |
| 8.0000 | 85.8686 | -1.1929 |
| 9.0000 | 85.7493 | 0.5676 |
| 10.0000 | 85.8061 | 2.1945 |

1. Find the value of that minimizes the sum of the squared one-step-ahead forecast errors.

Q5

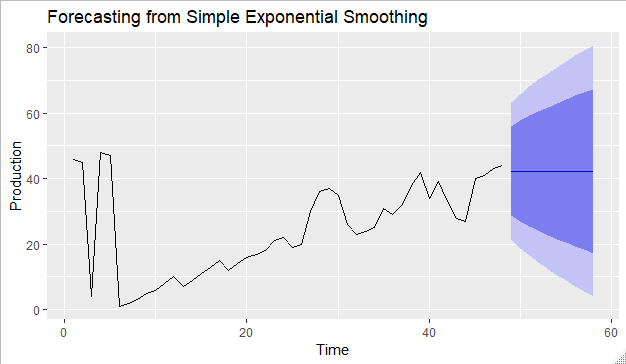
Time series B.4 contains data on the annual US production of blue and gorgonzola cheeses. This data have a strong trend.

1. Verify that there is a trend by plotting the data .



By Q2 , we can see that there is a strong going-up trend .

1. Develop an appropriate exponential smoothing procedure for forecasting.



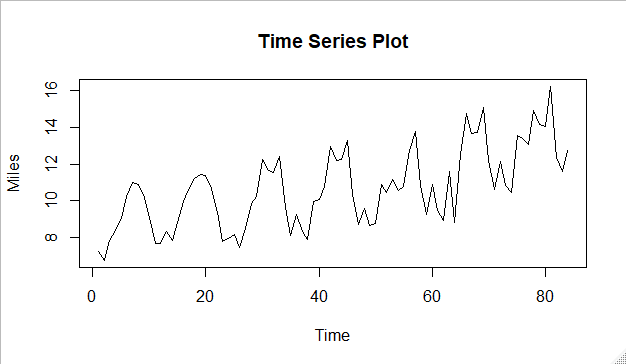
1. Forecast the last 10 observations and calculate the forecast errors. Does the forecasting procedure seem to be working satisfactorily?

|  |  |  |
| --- | --- | --- |
|  | predict | residuals |
| 1 | 46.0000 | 0.0000 |
| 2 | 46.0000 | -1.0000 |
| 3 | 45.4879 | -41.4879 |
| 4 | 24.2438 | 23.7562 |
| 5 | 36.4083 | 10.5917 |
| 6 | 41.8318 | -40.8318 |
| 7 | 20.9237 | -18.9237 |
| 8 | 11.2337 | -8.2337 |
| 9 | 7.0176 | -2.0176 |
| 10 | 5.9845 | 0.0155 |

Q6

Table B. 10 contains 7 years of monthly data on the number of airline miles flown in the United Kingdom. This is seasonal data .

1. Make a time series plot of the data and verify that it is seasonal.



Yes, we can the data is seasonal , and it is going up.

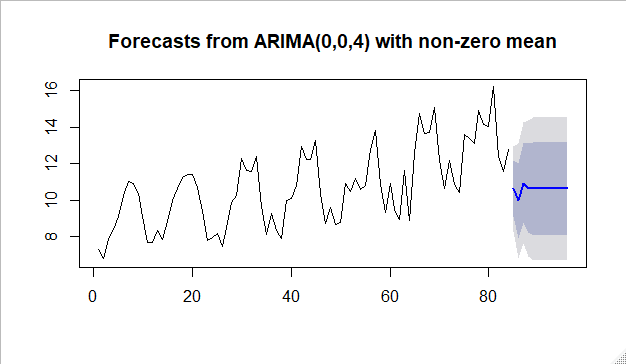
1. Should you use Multiplicative seasonal model or Additive seasonal model to analyze the data? Why?

Yes , if the data shows that it has a strong seasonal sign, we should using the relevance model to fit .

1. Use an appropriate model for the first 6 years to develop a forecasting method for this data. How well does this smoothing procedure work?

|  |
| --- |
| ARIMA(0,0,4) with non-zero mean  Coefficients:  ma1 ma2 ma3 ma4 mean  0.8907 0.4653 0.7370 0.4950 10.6181  s.e. 0.1036 0.1201 0.0948 0.0892 0.4457  sigma^2 estimated as 1.432: log likelihood=-132.79  AIC=277.57 AICc=278.66 BIC=292.16 |
|  |
| |  | | --- | |  | |

1. Make one-step-ahead forecasts of the last 12 months. Determine the forecast errors. Evaluate the forecasting performance based on MPE and MAPE



Q7

Explain the following terms.

1. Strictly stationary

如果一個time series是strictly stationary ，那在其中任取一段時間或空間裡的聯合機率分布，與將這段期間任意平移後的新期間之聯合機率分布相同。期望值、變異數不隨時間或位置變化。例如白噪音，

1. Weakly(Covariance) stationary

Weakly stationary 只需要一階、二階的mgf是隨時間平移不變的。

Q8.

An article in Quality Engineering(The Catapult Problem: The Enhanced Engineering Modeling Using Experimental Design, Vol. 4, 1992)conducted an experiment with a catapult to determine the effects of hook(x1), arm length(x2), start angle(x3)and stop angle(x4) on the distance that the catapult throws a ball. They threw the ball three times for each setting of the factors. Table E3.7 summarizes the experimental results.

1. Fit a regression model to the data and perform a residual analysis for the model.

Call:

lm(formula = y ~ x1 + x2 + x3 + x4, data = data)

Residuals:

1 2 3 4 5 6 7 8

18.325 1.225 9.675 -29.225 -18.325 -1.225 -9.675 29.225

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 79.150 10.348 7.649 0.00464 \*\*

x1 41.825 10.348 4.042 0.02726 \*

x2 9.850 10.348 0.952 0.41139

x3 26.375 10.348 2.549 0.08403 .

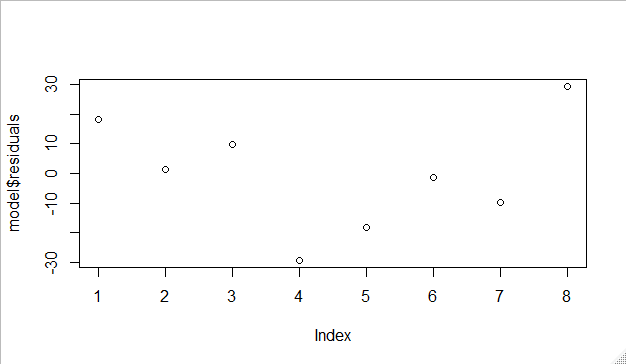
x4 -8.575 10.348 -0.829 0.46807

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 29.27 on 3 degrees of freedom

Multiple R-squared: 0.8906, Adjusted R-squared: 0.7448

F-statistic: 6.106 on 4 and 3 DF, p-value: 0.08451



1. Use the sample variances as the basis for WLS estimation of the original data(not the sample means)

lm(formula = y ~ x1 + x2 + x3 + x4, data = data, weights = data$V8)

Weighted Residuals:

1 2 3 4 5 6 7 8

14.079 -1.786 1.822 -13.189 -13.186 3.430 -4.058 23.736

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 75.226 9.065 8.298 0.00367 \*\*

x1 38.183 8.319 4.590 0.01943 \*

x2 7.099 8.468 0.838 0.46338

x3 20.927 8.412 2.488 0.08867 .

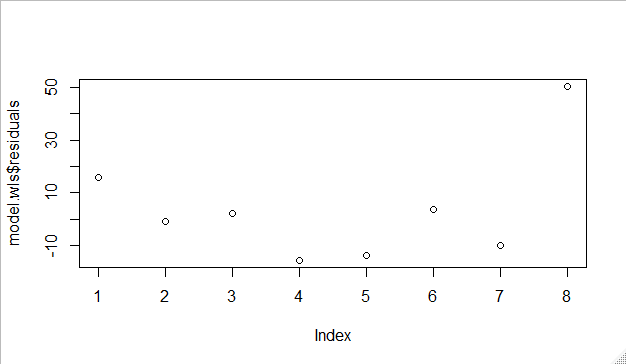
x4 -3.340 8.501 -0.393 0.72069

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19.53 on 3 degrees of freedom

Multiple R-squared: 0.8931, Adjusted R-squared: 0.7506

F-statistic: 6.268 on 4 and 3 DF, p-value: 0.08174



1. Fit an appropriate model to the sample variances. Use this model to develop the appropriate weights and repeat part (b).

lm(formula = samplevariances ~ x1 + x2 + x3 + x4, data = data)

Residuals:

1 2 3 4 5 6 7 8

-0.45917 0.69083 -0.07917 -0.15250 0.45917 -0.69083 0.07917 0.15250

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.9333 0.2445 3.817 0.0316 \*

x1 -0.4025 0.2445 -1.646 0.1983

x2 -0.4533 0.2445 -1.854 0.1608

x3 0.2567 0.2445 1.050 0.3710

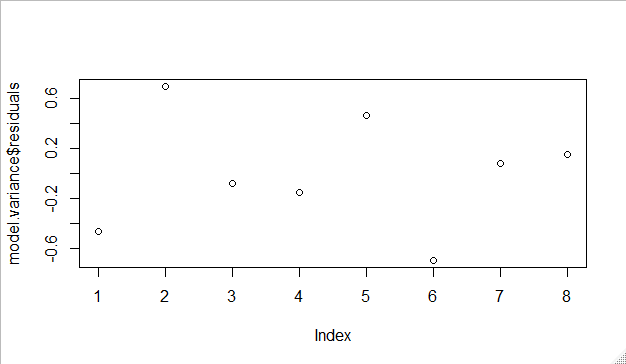
x4 0.2633 0.2445 1.077 0.3604

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6917 on 3 degrees of freedom

Multiple R-squared: 0.737, Adjusted R-squared: 0.3863

F-statistic: 2.102 on 4 and 3 DF, p-value: 0.284



lm(formula = V8 ~ x1 + x2 + x3 + x4, data = data, weights = data$V8)

Weighted Residuals:

1 2 3 4 5 6 7 8

-0.58441 0.39935 -0.06932 -0.12152 0.54737 -0.76698 0.15437 0.2186

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.9427 0.3261 2.891 0.063 .

x1 -0.5528 0.2992 -1.848 0.162

x2 -0.6189 0.3046 -2.032 0.135

x3 0.3223 0.3026 1.065 0.365

x4 0.3327 0.3058 1.088 0.356

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7024 on 3 degrees of freedom

Multiple R-squared: 0.8374, Adjusted R-squared: 0.6206

F-statistic: 3.862 on 4 and 3 DF, p-value: 0.1479

At first , I use sample variances ~ x1 +x2+x3+x4, but R-squared is bad , so using sample variances ~ x1+x2+x3+x4 by weighting = sample variance, we can find that our R-squared change from 0.3863 to 0.6206.

Q9.

Consider the motor vehicle facilities data in Appendix Table B.25. There are several candidate predictors that could be added to the model.

1. Use the stepwise regression to find an appropriate subset of predictors for the facilities data. Analyze the residuals from the model, including the Durbin-Watson test, and comment on model adequacy.

Step: AIC=719.17

Fatalities ~ Year + Annual.Unemployment.Rate.... + Licensed.Drivers..Thousands. + Resident.Population..Thousands.

Df Sum of Sq RSS AIC

<none> 167914829 719.17

+ Registered.Motor.Vehicles..Thousands. 1 6992274 160922555 719.18

+ Vehicle.Miles.Traveled..Billions. 1 975106 166939724 720.90

I use stepwise and stop here. So , I do the full regression about this model

lm(formula = Fatalities ~ Registered.Motor.Vehicles..Thousands. +

Vehicle.Miles.Traveled..Billions., data = data)

Residuals:

Min 1Q Median 3Q Max

-5420 -2206 -64 2247 5218

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.316e+04 2.146e+03 29.427 < 2e-16 \*\*\*

Registered.Motor.Vehicles..Thousands. -1.605e-01 5.201e-02 -3.085 0.00351 \*\*

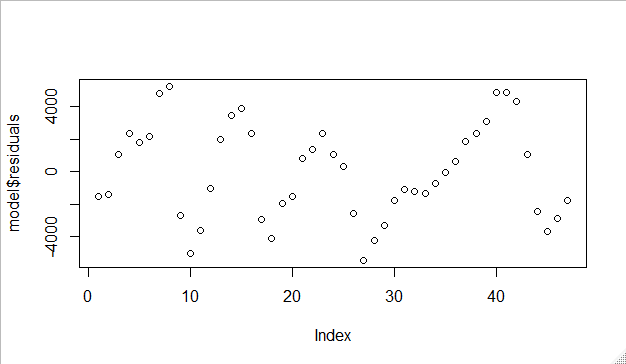
Vehicle.Miles.Traveled..Billions. 4.976e+00 3.781e+00 1.316 0.19503

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2949 on 44 degrees of freedom

Multiple R-squared: 0.7318, Adjusted R-squared: 0.7197

F-statistic: 60.04 on 2 and 44 DF, p-value: 2.659e-13



Durbin-Watson test

data: Fatalities ~ Registered.Motor.Vehicles..Thousands. + Vehicle.Miles.Traveled..Billions.

DW = 0.53134, p-value = 6.272e-11

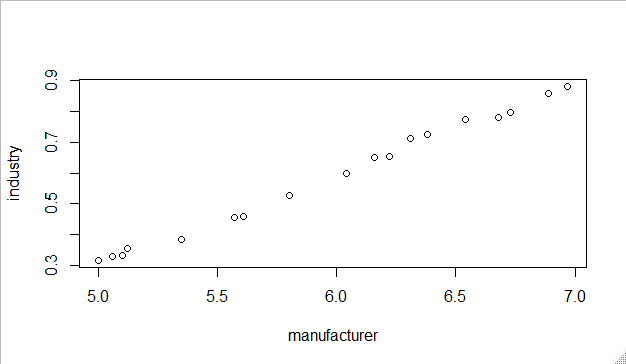
alternative hypothesis: true autocorrelation is greater than 0

So we can see that this model is not bad , and has the smallest AIC, but the residuals plot is too messy.

1. Use the all-possible-model approach to find an appropriate subset of predictors for the facilities data. Analyze the residuals from the model, including the Durbin-Watson test, and comment on model adequacy.

Q10.(20pts)The data in Table E3.6 give the monthly sales for a cosmetics manufacturer(yt) and the corresponding monthly sales for the entire industry(xt). The units of both variables are millions of dollars.

1. Build a simple linear regression model relating company sales to industry sales. Plot the residuals against time. Is there any indication of autocorrelation?



lm(formula = industry ~ manufacturer)

Residuals:

Min 1Q Median 3Q Max

-0.0222738 -0.0130087 0.0004658 0.0126908 0.0260620

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.157963 0.033688 -34.37 <2e-16 \*\*\*

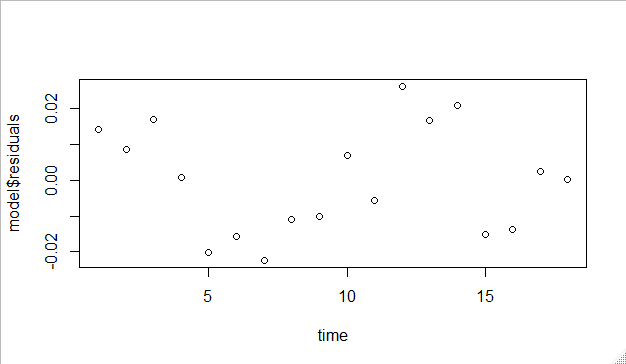
manufacturer 0.292377 0.005606 52.15 <2e-16 \*\*\*

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01545 on 16 degrees of freedom

Multiple R-squared: 0.9942, Adjusted R-squared: 0.9938

F-statistic: 2720 on 1 and 16 DF, p-value: < 2.2e-16



The data has a too strong correlation about industry and manufacturer. 而且還是一個很明顯穩定遞增的序列，這裡面應該有很強的auto-correlation.

1. Use the Durbin-Watson test to determine if there is positive autocorrelation in the errors. What are your conclusions?

Durbin-Watson test

data: industry ~ manufacturer

DW = 1.0801, p-value = 0.007689

alternative hypothesis: true autocorrelation is greater than 0

Yes we reject null hypothesis , so there is positive autocorrelation in the errors.

1. Use one iteration of the Cochrane-Orcutt procedure to estimate the model parameters. Compare the standard error of these regression coefficients with the standard errors of the least squares estimates.

Cochrane-orcutt estimation for first order autocorrelation

Call:

lm(formula = industry ~ manufacturer)

number of interaction: 7

rho 0.43638

Durbin-Watson statistic

(original): 1.08012 , p-value: 7.689e-03

(transformed): 2.06581 , p-value: 4.444e-01

coefficients:

(Intercept) manufacturer

-1.184780 0.296519

1. Test for positive autocorrelation following the first iteration. Has the procedure been successful?

Durbin-Watson test

data: model.diff

DW = 2.5427, p-value = 0.8737

alternative hypothesis: true autocorrelation is greater than 0

By testing positive autocorrelation following the first difference, we can’t reject the null hypothesis. So , it is negative.

1. Define a new set of transformation variables as the first difference of the original variables, denoted as yt’ and xt’. Regress yt’ on xt’ through the origin. Compare the estimate of the slope from this first difference approach with the estimate obtained from the iterative method in (c).

Call:

lm(formula = yt ~ xt, data = diff)

Residuals:

Min 1Q Median 3Q Max

-0.035502 -0.007898 -0.000366 0.010188 0.032929

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.002706 0.007352 -0.368 0.718

xt 0.308633 0.053212 5.800 3.5e-05 \*\*\*

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01651 on 15 degrees of freedom

Multiple R-squared: 0.6916, Adjusted R-squared: 0.6711

F-statistic: 33.64 on 1 and 15 DF, p-value: 3.5e-05

In (c) the slope = 0.296519 and in (e) the slope = 0.308633 , 兩個差不多