Statistics 5350/7110 Fall 2023

P. Shaman

Assignment 2 solution

The file Enplanements.txt gives monthly U.S. enplanements for domestic flights and international flights, in thousands, for the time span 2000(1) through 2023(5). It also includes the logs of these variables and various dummy variables. [After you input the data, be sure to change the class of the variable *Time* from integer to numeric, and redefine *Month* as a factor variable. If you redefine *Month* to be *fMonth*, for example, be sure to add it to the data frame.]

> flights<-read.csv("F:/Stat711023Fall/Enplanements.txt")

> head(flights)

Date Year Month Time TotalEnps logTotalEnps DomEnps logDomEnps IntlEnps

1 2000-01-01 2000 1 1 46492 10.74704 41540 10.63441 4952

2 2000-02-01 2000 2 2 48526 10.78986 43714 10.68542 4812

3 2000-03-01 2000 3 3 58764 10.98128 52977 10.87761 5787

4 2000-04-01 2000 4 4 56032 10.93368 50344 10.82663 5688

5 2000-05-01 2000 5 5 58201 10.97166 52309 10.86492 5892

6 2000-06-01 2000 6 6 61073 11.01983 54673 10.90913 6400

logIntlEnps obs21 obs22 obs23 c348 s348 c432 s432

1 8.507547 0 0 0 -0.57757270 0.8163393 -0.9101060 0.4143756

2 8.478868 0 0 0 -0.33281954 -0.9429905 0.6565858 -0.7542514

3 8.663369 0 0 0 0.96202767 0.2729519 -0.2850193 0.9585218

4 8.646114 0 0 0 -0.77846230 0.6276914 -0.1377903 -0.9904614

5 8.681351 0 0 0 -0.06279052 -0.9980267 0.5358268 0.8443279

6 8.764053 0 0 0 0.85099448 0.5251746 -0.8375280 -0.5463943

> attach(flights)

> Time<-as.numeric(Time)

> fMonth<-as.factor(Month)

> flights<-data.frame(flights,fMonth)

1. Make two separate time series plots, for (i) domestic enplanements and (ii) the log of domestic enplanements.

(a) On each mark the periods of economic contraction, as defined by the Business Cycle Dating Committee. Discuss and compare the two plots in detail. Comment on trend structure, seasonality, outliers (if any), economic downturns, and volatility.

(i)

> DomEnpscontraction<-c(rep(NA,15),DomEnps[16:23],rep(NA,73),DomEnps[97:114],rep(NA,128),DomEnps[243:244],rep(NA,37))

> plot(ts(DomEnps,start=c(2000,1),freq=12),ylab="Emplanements",main="Enplanements of Monthly Domestic Flights, in Thousands",col="green",lwd=2)

> lines(ts(DomEnpscontraction,start=c(2000,1),freq=12),col="red",lwd=2)

> legend("topleft",legend=c("expansion","contraction"),col=c("green","red"),lty=1,cex=0.8)

A graph showing a red and green line

Description automatically generated

**According to the Business Cycle Dating Committee, economic downturns occurred during April 2001 to November 2001, January 2008 to June 2009, and March 2020 to April 2020. Monthly domestic enplanements are level in 2000 and early 2001, and then drop abruptly during the first recession. There is an upward trend with decreasing derivative from 2002 to 2007. During the second downturn, the Great Recession, domestic enplanements drop noticeably, and then remain level for the next five years. There is a subsequent upward trend with increasing slope from 2014 to 2019. The third economic downturn coincides with the onset of COVID in the U.S., which led to a dramatic drop, with air travel essentially coming to a halt. The enplanement numbers do recover in 2021, 2022, and early 2023, but don’t reach the level achieved in 2019. Apart from the effects of the economic recessions, one can argue that domestic air travel has increased since 2000. This can be attributed to an increase in population and a rise in income for many persons.**

**Apart from behavior during the recessions and some subsequent months, there is a strong seasonal pattern, with domestic air travel peaking during the summer months. Additionally, volatility seems relatively stable from 2000 to 2019, except for brief episodes during the first two economic downturns. In contrast, volatility is quite high after the arrival of COVID.**

**Outlier behavior is evident during the economic downturns, especially during the onset of COVID.**

(ii)

> logDomEnpscontraction<-c(rep(NA,15),logDomEnps[16:23],rep(NA,73),logDomEnps[97:114],rep(NA,128),logDomEnps[243:244],rep(NA,37))

> plot(ts(logDomEnps,start=c(2000,1),freq=12),ylab="log of Emplanements",main="log Enplanements of Monthly Domestic Flights, in Thousands",ylim=c(8,12),col="green",lwd=2)

> lines(ts(logDomEnpscontraction,start=c(2000,1),freq=12),col="red",lwd=2)

> legend("topleft",legend=c("expansion","contraction"),col=c("green","red"),lty=1,cex=0.8)

A graph with green and red lines

Description automatically generated

**Trend features for the log enplanement data are similar to those for enplanement data, but are visually less obvious on the log scale. The seasonal structure is clearly visible, and volatility is quite constant from 2000 to 2019.**

**To build models for the data, we’ll exclude the years 2020 to 2023, and the two plots suggest a multiplicative decomposition model is a somewhat better choice than an additive decomposition formulation for model construction. Additionally, the multiplicative approach will permit a convenient interpretation of the statistical results.**

(b) If you identify outliers, explain what is responsible for their behavior.

**There are outliers visible during the first and third recessions. These include September to December in 2001, and the March and April readings in 2020. There are also likely outliers during the period of the Great Recession. As noted, we’ll exclude the post-2019 data in building models.**

*For the remainder of this assignment (except for the initial plots in part 7) restrict calculations to data for the years from 2000 through 2019.*

2. Construct a spectral plot for the domestic log enplanement series. Mark the frequencies which determine seasonal structure and calendar structure. Discuss carefully what the plots reveal, what guidance they provide for model construction. [To resolve spectral peaks at close frequencies (such as 0.333 and 0.348), choose a sufficiently small span. To perform the calculation restricted to the 2000–2019 data, use a command of the form spectrum(logDomEnps[1:240],span=x).]

> spectrum(logDomEnps[1:240],span=5)

> abline(v=c(1/12,2/12,3/12,4/12,5/12,6/12),col="red",lty=2)

> abline(v=c(0.220,0.348,0.432),col="blue",lty=2)

A graph of a graph

Description automatically generated

**In the spectral plot the frequencies for seasonal structure are marked in red, and those for calendar variation are in blue. The peak at low frequency is indicative of trend structure in the data, the peaks at frequencies 1/12, 2/12, 3/12, 4/12, 5/12, and 6/12 show the presence of prominent seasonal structure. The calendar component with frequency 0.432 may be a significant addition in the construction of a model for the data.**

3. (a) Despite the use of the log series in part 2, fit an additive decomposition model to the domestic enplanement data. Include in your model trend and static seasonal components, calendar trigonometric pairs if needed, and dummies for outliers if needed. [To restrict calculation to the 2000–2019 data, use a command of the form lm(DomEnps~x1+x2,data=flights[1:240,]), where flights is the name of the data frame.]

Give a brief description of your model.

**Using the plots in part 1 and the spectral plot in part 2, we construct an additive decomposition model with trend, static seasonality, outliers for the months September to November 2001, and the 0.432 trigonometric pair.**

> model1<-lm(DomEnps~poly(Time,5)+fMonth+obs21+obs22+obs23+c432+s432,data=flights[1:240,]);summary(model1)

Call:

lm(formula = DomEnps ~ poly(Time, 5) + fMonth + obs21 + obs22 +

obs23 + c432 + s432, data = flights[1:240, ])

Residuals:

Min 1Q Median 3Q Max

-5051.4 -929.0 246.2 1117.8 3794.8

Coefficients:

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

(Intercept) 48197.808 389.390 123.778 < 2e-16 \*\*\*

poly(Time, 5)1 69407.746 1765.918 39.304 < 2e-16 \*\*\*

poly(Time, 5)2 21872.614 1753.447 12.474 < 2e-16 \*\*\*

poly(Time, 5)3 25361.149 1745.258 14.531 < 2e-16 \*\*\*

poly(Time, 5)4 6891.889 1740.578 3.960 0.000102 \*\*\*

poly(Time, 5)5 -20743.955 1760.236 -11.785 < 2e-16 \*\*\*

fMonth2 -1358.047 549.918 -2.470 0.014297 \*

fMonth3 9531.574 549.537 17.345 < 2e-16 \*\*\*

fMonth4 6934.303 549.891 12.610 < 2e-16 \*\*\*

fMonth5 9124.425 549.970 16.591 < 2e-16 \*\*\*

fMonth6 11234.268 549.928 20.429 < 2e-16 \*\*\*

fMonth7 13467.724 550.500 24.465 < 2e-16 \*\*\*

fMonth8 11489.347 550.197 20.882 < 2e-16 \*\*\*

fMonth9 2746.826 558.125 4.922 1.69e-06 \*\*\*

fMonth10 7648.501 558.108 13.704 < 2e-16 \*\*\*

fMonth11 4648.169 558.659 8.320 9.56e-15 \*\*\*

fMonth12 5200.523 551.697 9.426 < 2e-16 \*\*\*

obs21 -13010.462 1811.296 -7.183 1.07e-11 \*\*\*

obs22 -8360.064 1810.616 -4.617 6.64e-06 \*\*\*

obs23 -5105.158 1811.146 -2.819 0.005265 \*\*

c432 -4.432 161.392 -0.027 0.978119

s432 -382.242 158.678 -2.409 0.016831 \*

---

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

Residual standard error: 1737 on 218 degrees of freedom

Multiple R-squared: 0.9481, Adjusted R-squared: 0.9431

F-statistic: 189.7 on 21 and 218 DF, p-value: < 2.2e-16

**The calendar pair with frequency 0.348 is not significant and is omitted in model 1. Let’s perform an *F-*test for the pair with frequency 0.432.**

> model1a<-lm(DomEnps~poly(Time,5)+fMonth+obs21+obs22+obs23,data=flights[1:240,])

> anova(model1a,model1)

Analysis of Variance Table

Model 1: DomEnps ~ poly(Time, 5) + fMonth + obs21 + obs22 + obs23

Model 2: DomEnps ~ poly(Time, 5) + fMonth + obs21 + obs22 + obs23 + c432 +

s432

Res.Df RSS Df Sum of Sq F Pr(>F)

1 220 675442555

2 218 657928964 2 17513591 2.9015 0.05707 .

---

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

**The 0.432 frequency pair is marginally significant, and we elect to keep it in the model.**

**There are also a fifth-degree polynomial in time for trend estimation, monthly dummies for seasonality, and three outlier dummies; the last are all significant. *R-*square for the model is 0.9481, and the residual standard error is 1737.**

(b) Perform a thorough residual analysis, including a residual spectral plot, and discuss the results.

> residcontraction1<-c(rep(NA,15),resid(model1)[16:23],rep(NA,73),resid(model1)[97:114],rep(NA,126))

> plot(ts(resid(model1),start=c(2000,1),freq=12),xlab="Time",ylab="Residual",main="Residuals of Model 1",col="green",lwd=2)

> lines(ts(residcontraction1,start=c(2000,1),freq=12),col="red",lwd=2)

A graph showing a green and red line

Description automatically generated

**The residual plot shows inadequate trend estimation, especially after the 2001 and 2008–09 recessions. There is also some changing volatility.**

A graph of a normal q-q plot

Description automatically generated

> shapiro.test(resid(model1))

Shapiro-Wilk normality test

data: resid(model1)

W = 0.97607, p-value = 0.000442

**The residual distribution is not normal, as indicated by the normal quantile plot and the Shapiro–Wilk test. Quite a few observations are overpredicted by the model, as the lower tail of the quantile plot shows.**

A graph with lines and numbers

Description automatically generated

> spectrum(resid(model1),span=4)

> abline(v=c(1/12,2/12,3/12,4/12,5/12,6/12),col="red",lty=2)

> abline(v=c(0.220,0.348,0.432),col="blue",lty=2)

A graph of a graph

Description automatically generated

**The residual acf plot shows failure of the model to reduce to white noise, primarily due to inadequate trend estimation, and there is indication at lags 6 and 12 that there is some remaining dynamic seasonal structure. The low frequency peak in the spectral confirms the trend assertion, and the spectral peaks at frequencies 1/12 and 2/12 show the remaining dynamic seasonal structure.**

(c) Draw a plot to examine if an additive model fit is acceptable, and discuss carefully.

> plot(predict(model1),resid(model1),xlab="Predicted Enplanements",ylab=" Residual Enplanements",main="Residual by Predicted Plot")

A graph of residuals

Description automatically generated

**The plot does not show evidence of heteroscedasticity, and thus the additive decomposition model is acceptable. Despite this, we turn to a multiplicative decomposition model in part 4, to provide preferable interpretation.**

4. (a) Fit a multiplicative decomposition model to the domestic enplanement data, including variables to estimate trend, static seasonality, calendar structure, and add significant outlier dummies. Give a brief description of your model.

> model2<-lm(logDomEnps~poly(Time,5)+fMonth+obs21+obs22+obs23+c432+s432,data=flights[1:240,]);summary(model2)

Call:

lm(formula = logDomEnps ~ poly(Time, 5) + fMonth + obs21 + obs22 +

obs23 + c432 + s432, data = flights[1:240, ])

Residuals:

Min 1Q Median 3Q Max

-0.120687 -0.017869 0.003565 0.020709 0.076717

Coefficients:

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

(Intercept) 10.776813 0.007686 1402.181 < 2e-16 \*\*\*

poly(Time, 5)1 1.242997 0.034855 35.661 < 2e-16 \*\*\*

poly(Time, 5)2 0.319234 0.034609 9.224 < 2e-16 \*\*\*

poly(Time, 5)3 0.423679 0.034448 12.299 < 2e-16 \*\*\*

poly(Time, 5)4 0.110502 0.034355 3.216 0.001495 \*\*

poly(Time, 5)5 -0.407641 0.034743 -11.733 < 2e-16 \*\*\*

fMonth2 -0.027805 0.010854 -2.562 0.011093 \*

fMonth3 0.182511 0.010847 16.826 < 2e-16 \*\*\*

fMonth4 0.136062 0.010854 12.536 < 2e-16 \*\*\*

fMonth5 0.174406 0.010855 16.066 < 2e-16 \*\*\*

fMonth6 0.211279 0.010854 19.465 < 2e-16 \*\*\*

fMonth7 0.248679 0.010866 22.887 < 2e-16 \*\*\*

fMonth8 0.216423 0.010860 19.929 < 2e-16 \*\*\*

fMonth9 0.056582 0.011016 5.136 6.21e-07 \*\*\*

fMonth10 0.148668 0.011016 13.496 < 2e-16 \*\*\*

fMonth11 0.093895 0.011027 8.515 2.72e-15 \*\*\*

fMonth12 0.103307 0.010889 9.487 < 2e-16 \*\*\*

obs21 -0.368146 0.035751 -10.297 < 2e-16 \*\*\*

obs22 -0.185641 0.035738 -5.195 4.70e-07 \*\*\*

obs23 -0.124573 0.035748 -3.485 0.000595 \*\*\*

c432 0.000398 0.003185 0.125 0.900686

s432 -0.007113 0.003132 -2.271 0.024127 \*

---

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

Residual standard error: 0.03429 on 218 degrees of freedom

Multiple R-squared: 0.9414, Adjusted R-squared: 0.9358

F-statistic: 166.9 on 21 and 218 DF, p-value: < 2.2e-16

**We use a partial *F*-statistic to test significance for the frequency 0.432 calendar pair.**

> model2a<-lm(logDomEnps~poly(Time,5)+fMonth+obs21+obs22+obs23,data=flights[1:240,])

> anova(model2a,model2)

Analysis of Variance Table

Model 1: logDomEnps ~ poly(Time, 5) + fMonth + obs21 + obs22 + obs23

Model 2: logDomEnps ~ poly(Time, 5) + fMonth + obs21 + obs22 + obs23 +

c432 + s432

Res.Df RSS Df Sum of Sq F Pr(>F)

1 220 0.26241

2 218 0.25632 2 0.0060868 2.5884 0.07745 .

---

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

**The trigonometric pair with frequency 0.432 is marginally significant, and we choose to retain it and use model 2. The explanatory variables in model 2 are the same as those in the additive decomposition model fit in part 3. Here *R-*square is 0.9414, and the residual standard error is 0.0343.**

(b) Estimate the seasonal indices, display them with a table and a plot, and interpret your findings.

> b1<-coef(model2)[1]

> b2<-coef(model2)[7:17]+b1

> b3<-c(b1,b2)

> seas4<-exp(b3-mean(b3))

> seas4

(Intercept) fMonth2 fMonth3 fMonth4 fMonth5 fMonth6

0.8792664 0.8551555 1.0553195 1.0074225 1.0468004 1.0861202

fMonth7 fMonth8 fMonth9 fMonth10 fMonth11 fMonth12

1.1275099 1.0917218 0.9304515 1.0202020 0.9658252 0.9749587

> cbind(seas4)

seas4

January 0.8793

February 0.8552

March 1.0553

April 1.0074

May 1.0468

June 1.0861

July 1.1275

August 1.0917

September 0.9305

October 1.0202

November 0.9658

December 0.9750

A graph with numbers and lines

Description automatically generated

**Domestic air travel is estimated to peak in June, July, and August, and there is also a local peak in March. The months with lowest estimated domestic travel are January and February, and the estimates are also below trend level in September, November, and December. In July enplanements are estimated to be 12.8 percent above trend level, and the estimates for March, June, and August are 5.5 percent, 8.6 percent, and 9.2 percent, respectively, above trend. In January, February, and September the estimates are 12.1 percent, 14.5 percent, and 7.0 percent, respectively, below trend level. These results confirm, as one expects, that domestic air travel peaks in the summer months when people take vacations and their children are not attending school. The low level of such travel in January and February is certainly influenced by winter weather and school attendance by children. Perhaps the small estimated peak in March stems from school spring vacations and desire to take a vacation after experiencing winter weather.**

(c) Give a thorough residual analysis of your fitted model. Discuss what structure in the data the model has failed to capture.

A graph showing a graph

Description automatically generated with medium confidence

**The model has failed to fully capture trend structure in the data, primarily during and after the recessions in 2001 and 2008–2009. There is also some remaining trend in the last few years prior to the arrival of COVID. The plot is similar to the residual from model 1 in part 3, but now on the log scale.**

A graph of a normal q-q plot

Description automatically generated

> shapiro.test(resid(model2))

Shapiro-Wilk normality test

data: resid(model2)

W = 0.97171, p-value = 0.0001021

**The residual distribution does not conform to normality. The lower tail is long relative to normality, with overprediction of many observations by the model. The parameter estimates given by the model are valid, but *p*-values given by the statistical tests are somewhat questionable.**

A graph with lines and numbers

Description automatically generated

A graph of a graph

Description automatically generated

**The residual acf plot shows failure of the model to achieve reduction to white noise, with failure to estimate trend structure fully the main culprit. This conclusion is confirmed by the residual spectral plot, which shows a peak at low frequency. The smaller peaks at frequencies 1/12, 3/12, and 6/12 provide evidence of some remaining dynamic seasonality.**

5. (a) Calculate the lag 1 and lag 2 residuals from the model in part 4 and add these two new variables to the model you fit in part 4. Be sure to include the calendar trigonometric pairs and outlier dummies if they turn out to be significant in this fitting. [Be careful in calculating the lag 1 and 2 residuals. For example, use the following code for the lag 1 residual:

lag1resid<-c(NA,resid(model)[1:239],rep(0,41))

Add this variable and the lag 2 residual to the data frame.]

**We also form a dummy for an outlier in December 2001, observation 24.**

> lag1resid<-c(NA,resid(model2)[1:239],rep(0,41))

> lag2resid<-c(NA,NA,resid(model2)[1:238],rep(0,41))

> obs24<-ifelse(Time==24,1,0)

> flights<-data.frame(flights,lag1resid,lag2resid,obs24)

> model3<-lm(logDomEnps~poly(Time,5)+fMonth+obs21+obs22+obs23+obs24+c348+s348+c432+s432+lag1resid+lag2resid,data=flights[1:240,]);summary(model3)

Call:

lm(formula = logDomEnps ~ poly(Time, 5) + fMonth + obs21 + obs22 +

obs23 + obs24 + c348 + s348 + c432 + s432 + lag1resid + lag2resid,

data = flights[1:240, ])

Residuals:

Min 1Q Median 3Q Max

-0.075002 -0.012370 0.000667 0.012783 0.068918

Coefficients:

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

(Intercept) 10.7839627 0.0046754 2306.532 < 2e-16 \*\*\*

poly(Time, 5)1 1.2241528 0.0211163 57.972 < 2e-16 \*\*\*

poly(Time, 5)2 0.3530788 0.0211685 16.679 < 2e-16 \*\*\*

poly(Time, 5)3 0.4219399 0.0211575 19.943 < 2e-16 \*\*\*

poly(Time, 5)4 0.1264085 0.0212798 5.940 1.16e-08 \*\*\*

poly(Time, 5)5 -0.3845411 0.0216398 -17.770 < 2e-16 \*\*\*

fMonth2 -0.0362343 0.0066330 -5.463 1.31e-07 \*\*\*

fMonth3 0.1756220 0.0065282 26.902 < 2e-16 \*\*\*

fMonth4 0.1292887 0.0065194 19.831 < 2e-16 \*\*\*

fMonth5 0.1666722 0.0065268 25.537 < 2e-16 \*\*\*

fMonth6 0.2046637 0.0065204 31.388 < 2e-16 \*\*\*

fMonth7 0.2414734 0.0065221 37.024 < 2e-16 \*\*\*

fMonth8 0.2089268 0.0065209 32.040 < 2e-16 \*\*\*

fMonth9 0.0527322 0.0066038 7.985 8.95e-14 \*\*\*

fMonth10 0.1429392 0.0066094 21.627 < 2e-16 \*\*\*

fMonth11 0.0867327 0.0066098 13.122 < 2e-16 \*\*\*

fMonth12 0.1028375 0.0066097 15.559 < 2e-16 \*\*\*

obs21 -0.4274077 0.0215293 -19.852 < 2e-16 \*\*\*

obs22 -0.2168958 0.0217853 -9.956 < 2e-16 \*\*\*

obs23 -0.1320465 0.0212604 -6.211 2.76e-09 \*\*\*

obs24 -0.1297062 0.0212652 -6.099 5.01e-09 \*\*\*

c348 0.0038619 0.0018962 2.037 0.0429 \*

s348 -0.0033796 0.0019292 -1.752 0.0813 .

c432 -0.0007134 0.0019012 -0.375 0.7079

s432 -0.0053261 0.0018688 -2.850 0.0048 \*\*

lag1resid 0.5434709 0.0615296 8.833 4.01e-16 \*\*\*

lag2resid 0.2938549 0.0610970 4.810 2.87e-06 \*\*\*

---

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

Residual standard error: 0.02031 on 211 degrees of freedom

(2 observations deleted due to missingness)

Multiple R-squared: 0.9796, Adjusted R-squared: 0.977

F-statistic: 388.7 on 26 and 211 DF, p-value: < 2.2e-16

**Both lag residual variables are highly significant, and we retain both calendar trigonometric pairs. Also, note that the dummy variable for December 2001 is very significant. *R*-square for the model is 0.9796, and the residual standard error is 0.0203.**

(b) Give a residual analysis of this model, and compare your findings to those for the model in part 4. If there is remaining structure the model has not captured, discuss what it is.

A graph showing a red and green line

Description automatically generated

**The residual plot shows that model 3 has captured most of the trend structure in the data. The exception is the period of the 2008–09 recession, although the trend estimation shortfall there is rather modest. The residuals include some spikes, but, apart from these, volatility is relatively constant. Clearly addition of the two lag residual variables has produced substantial improvement, relative to the performance of model 2.**

A graph with a line

Description automatically generated

> shapiro.test(resid(model3))

Shapiro-Wilk normality test

data: resid(model3)

W = 0.98594, p-value = 0.01915

**The normal quantile plot of the residuals displays a somewhat long upper tail, with underestimation of data points by the model, and there is an outlier, where the data point for July 2002 is overestimated. We conclude that the residual distribution is not normal, but still closer to normality than the plot for model 2.**

A graph with lines and numbers

Description automatically generated

**There are significant residual autocorrelations at lags 4, 12, 16, 19, and 34. Thus, there is evidence of some remaining dynamic seasonality, and overall we conclude that model 3 is a bit shy of achieving reduction to white noise. The residual spectral plot which follows has modest peaks at frequencies associated with seasonality, 3/12 and 6/12. We are almost at reduction to white noise, but not quite there.**

A graph of a graph

Description automatically generated

6. (a) Fit the model in part 4 (that is, with the same variables you selected in part 4) to the data spanning the years 2000 to 2018. That is, withhold the 2019 data. Use the resulting model to forecast data for the year 2019. [To do so, use the command forecast<-predict(model,newdata=flights[229:240,] Here flights is the name of the data frame.] Compare the year 2019 data and the forecast with a table and a plot. Discuss the result.

**Using the model from part 4 for the 2000–2018 data, we find that the calendar pair with frequency 0.432 is not significant. The *p*-value of the partial *F*-test is 0.1109, and we don’t include the pair in model 4, following. The model has a fifth-degree polynomial for trend estimation and includes dummies for outliers in September, October, and November of 2001.**

> model4<-lm(logDomEnps~poly(Time,5)+fMonth+obs21+obs22+obs23,data=flights[1:228,]);summary(model4)

Call:

lm(formula = logDomEnps ~ poly(Time, 5) + fMonth + obs21 + obs22 +

obs23, data = flights[1:228, ])

Residuals:

Min 1Q Median 3Q Max

-0.117710 -0.017712 0.004109 0.021042 0.086282

Coefficients:

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

(Intercept) 10.765884 0.007748 1389.434 < 2e-16 \*\*\*

poly(Time, 5)1 1.036406 0.034283 30.231 < 2e-16 \*\*\*

poly(Time, 5)2 0.193452 0.034003 5.689 4.29e-08 \*\*\*

poly(Time, 5)3 0.388606 0.033854 11.479 < 2e-16 \*\*\*

poly(Time, 5)4 0.186034 0.033796 5.505 1.08e-07 \*\*\*

poly(Time, 5)5 -0.404631 0.034212 -11.827 < 2e-16 \*\*\*

fMonth2 -0.027668 0.010936 -2.530 0.01214 \*

fMonth3 0.182237 0.010937 16.663 < 2e-16 \*\*\*

fMonth4 0.136120 0.010939 12.444 < 2e-16 \*\*\*

fMonth5 0.172892 0.010941 15.802 < 2e-16 \*\*\*

fMonth6 0.211747 0.010945 19.347 < 2e-16 \*\*\*

fMonth7 0.248484 0.010949 22.695 < 2e-16 \*\*\*

fMonth8 0.217693 0.010953 19.874 < 2e-16 \*\*\*

fMonth9 0.055190 0.011113 4.966 1.42e-06 \*\*\*

fMonth10 0.148182 0.011119 13.327 < 2e-16 \*\*\*

fMonth11 0.094401 0.011125 8.485 4.03e-15 \*\*\*

fMonth12 0.100809 0.010980 9.181 < 2e-16 \*\*\*

obs21 -0.363727 0.035074 -10.370 < 2e-16 \*\*\*

obs22 -0.179477 0.035074 -5.117 7.04e-07 \*\*\*

obs23 -0.115962 0.035074 -3.306 0.00111 \*\*

---

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

Residual standard error: 0.0337 on 208 degrees of freedom

Multiple R-squared: 0.9361, Adjusted R-squared: 0.9303

F-statistic: 160.4 on 19 and 208 DF, p-value: < 2.2e-16

**Let’s examine several residual diagnostics before constructing a forecast of the 2019 data.**

A graph of a normal q-q plot

Description automatically generated

> shapiro.test(resid(model4))

Shapiro-Wilk normality test

data: resid(model4)

W = 0.97227, p-value = 0.0001897

**According to the normal quantile plot and the Shapiro–Wilk test, the residual distribution is nonnormal. The following residual acf plot shows that model 4 fails to adequately estimate trend structure. Nonetheless let’s proceed to forecast the year 2019 data.**

A graph with lines and numbers

Description automatically generated

**Tabulation and plot of the forecasts for 2019 follow. A polynomial of degree five has been used for trend estimation.**

> forecast<-predict(model4,newdata=flights[229:240,])

> cbind(DomEnps[229:240],exp(forecast))

Actual Forecast

229 58034 55710.44

230 55679 53995.67

231 70233 66332.17

232 66938 63045.82

233 71365 65062.82

234 72791 67243.41

235 75284 69306.09

236 72718 66723.81

237 63981 56271.87

238 69922 61228.57

239 64817 57485.09

240 69719 57274.95

> plot(ts(DomEnps[229:240]),ylab="Enplanements",xlab="Year 2019",main="Forecast from Model 4 of 2019 Enplanements",ylim=c(50000,75000),col="red",lwd=2)

> lines(ts(exp(forecast)),col="blue",lwd=2)

> legend("topleft",legend=c("2019 data","forecast"),col=c("red","blue"),lty=1,cex=0.8)

A graph of a graph showing the number of individuals

Description automatically generated with medium confidence

**The forecasts underpredict the 2019 data, and the underprediction becomes more severe as the forecast horizon increases.**

(b) Redo the year 2019 forecast by using a polynomial of one less degree from that employed in part (a). Compare the year 2019 data and the forecast with a table and a plot, and discuss the result.

(c) Discuss the different outcomes in parts (a) and (b). What do you think is responsible for the difference? What do you think will happen if additional polynomial degrees to estimate trend are tried?

**Let’s produce forecasts of the 2019 data using the model in part (a) with the polynomial degree ranging from 2 to 8 (part (a) has already performed the calculation for degree 5).**

pflights<-matrix(rep(0,96),ncol=8)

> pflights[,1]<-DomEnps[229:240]

> for(i in 2:8){

+ model<-lm(logDomEnps~poly(Time,i)+fMonth+obs21+obs22+obs23,data=flights[1:228,])

+ pflights[,i]<-exp(predict(model,newdata=flights[229:240,]))

+ }

> pflights

**Here are the forecasts in tabular form. The first column has the actual 2019 data.**

Polynomial Degree

Actual 2 3 4 5 6 7 8

Jan 58034 55020.87 59210.17 61482.55 55710.44 57926.79 60262.24 56622.30

Feb 55679 53599.19 57856.42 60283.61 53995.67 56567.05 59371.72 54642.42

Mar 70233 66220.45 71704.41 74977.30 66332.17 70054.75 74266.15 66757.58

Apr 66938 63337.43 68803.49 72206.41 63045.82 67164.54 72002.52 63014.22

May 71365 65818.58 71734.95 75565.52 65062.82 69962.23 75940.69 64481.01

Jun 72791 68542.52 74956.70 79264.47 67243.41 73033.09 80374.11 65960.04

Jul 75284 71230.79 78166.78 82987.79 69306.09 76083.13 85013.53 67150.51

Aug 72718 69193.12 76200.41 81231.23 66723.81 74091.59 84183.36 63710.76

Sep 63981 59087.33 65263.14 69840.13 56271.87 63198.45 73177.32 52821.66

Oct 69922 64964.47 72027.34 77407.62 61228.57 69669.02 82293.10 56326.87

Nov 64817 61675.83 68646.54 74098.39 57485.09 66326.61 80064.31 51666.18

Dec 69719 62009.93 69322.58 75206.83 57274.95 67129.50 82937.41 50142.56

> plot(ts(pflights[,1]),lwd=2,col="black",ylim=c(40000,100000),main="Year 2019 Forecasts")

> lines(ts(pflights[,2]),lwd=2,col="red")

> lines(ts(pflights[,3]),lwd=2,col="blue")

> lines(ts(pflights[,4]),lwd=2,col="green")

> lines(ts(pflights[,5]),lwd=2,col="brown")

> lines(ts(pflights[,6]),lwd=2,col="purple")

> lines(ts(pflights[,7]),lwd=2,col="orange")

> lines(ts(pflights[,8]),lwd=2,col="cyan")

> legend("topleft",legend=c("actual","2","3","4","5","6","7","8"),col=c("black","red","blue","green","brown","purple","orange","cyan"),lty=1,cex=0.7)

A graph of a number of colored lines

Description automatically generated with medium confidence

**The actual data are given by the black curve. Degree 6 (purple) produces the most accurate predictions, and forecasts for degree 3 (blue) are the second most accurate, with some overprediction. Degree 2 (red) underpredicts, degree 4 (green) overpredicts, and degree 5 (brown) underpredicts. The worst overprediction is given by degree 7 (orange), the worst underprediction by degree 8 (cyan).**

degree 2 3 4 5 6 7 8

performance     close  

**The results are highly dependent upon the degree of the polynomial used for trend estimation. As the degree of the polynomial becomes large (7 and 8 here), the quality of the forecasts deteriorates. Most of the degrees considered here don’t produce satisfactory predictions. Polynomial trend estimation is designed to track the actual data used to build a model. However, when we extrapolate beyond the range of the data, the polynomial used for estimation increases or decreases very quickly and sharply, and this movement leads to poor forecasting. It is perhaps unusual that in this example degree 6 (purple) produces reasonable forecasts. This, however, does not suggest that degree 6 is a good choice for other data sets.**

**To gauge the closeness of the seven forecasts, we calculate the root mean square error for each.**

> rmse<-rep(0,7)

> for(i in 1:7){

+ rmse[i]<-sqrt(sum((pflights[,1]-pflights[,i+1])^2)/12)

+ }

> rmse

[1] 4453.057 2197.220 6333.492 6610.438 1121.596 9173.781 9735.746

**Via this criterion degree 6 produces forecasts closest to the data, and degree 3 is second best. Degrees 7 and 8 are worst. It’s interesting that degree 2 is third best by the rmse metric. Degree 2 does not track the 2000–2018 data, but it does forecast somewhat well. In conclusion, we can observe that a model which tracks the data well may forecast poorly, and, conversely, a model which tracks the data poorly may forecast well.**

**We should also note that the forecast horizon considered here is one year ahead. As the forecast horizon increases, all polynomial degrees are not able to produce reasonable forecasts.**

7. Repeat parts 1, 2, 4, and 5 for the international enplanement data. In part 5 add just the lag 1 residual to the part 4 model for international enplanements.

A graph showing a line of green and red

Description automatically generated

**The trend behavior in the plot of international enplanements is similar to that for the plot of domestic enplanements. There are two notable differences in the behavior of this international series relative to the domestic series. The seasonal peaks in the summer months are more prominent, and there is a more noticeable increase in volatility as the level of the time series rises.**

A graph with green and red lines

Description automatically generated

**The log transformation has largely stabilized volatility for the years 2000–2019. This indicates that a multiplicative decomposition model should be employed in analyzing the international series.**

**The following plot gives spectral estimation for the log series.**

A graph of a graph

Description automatically generated

**The spectral result indicates the data series has trending and a prominent seasonal structure. In addition, there is possible calendar structure which needs to be explored—the frequencies 0.348 and 0.432 should be investigated.**

**Model 6, following, includes a fifth-degree polynomial to track trend, month dummies to estimate static seasonality, and dummies for outliers in September, October, and November 2021. The calendar trigonometric pairs are not included—they are not significant. The model *R*-square is 0.9559, and the residual standard error is 0.0470.**

> model6<-lm(logIntlEnps~poly(Time,5)+fMonth+obs21+obs22+obs23,data=flights[1:240,]);summary(model6)

Call:

lm(formula = logIntlEnps ~ poly(Time, 5) + fMonth + obs21 + obs22 +

obs23, data = flights[1:240, ])

Residuals:

Min 1Q Median 3Q Max

-0.202315 -0.022966 0.003016 0.026903 0.133365

Coefficients:

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

(Intercept) 8.84958 0.01053 840.602 < 2e-16 \*\*\*

poly(Time, 5)1 2.58088 0.04777 54.032 < 2e-16 \*\*\*

poly(Time, 5)2 -0.29326 0.04742 -6.184 3.01e-09 \*\*\*

poly(Time, 5)3 0.07733 0.04721 1.638 0.102831

poly(Time, 5)4 0.37878 0.04708 8.046 5.30e-14 \*\*\*

poly(Time, 5)5 -0.44172 0.04761 -9.278 < 2e-16 \*\*\*

fMonth2 -0.11175 0.01486 -7.520 1.38e-12 \*\*\*

fMonth3 0.09652 0.01486 6.494 5.48e-10 \*\*\*

fMonth4 0.03646 0.01486 2.453 0.014965 \*

fMonth5 0.06347 0.01487 4.269 2.92e-05 \*\*\*

fMonth6 0.15640 0.01487 10.516 < 2e-16 \*\*\*

fMonth7 0.24829 0.01488 16.689 < 2e-16 \*\*\*

fMonth8 0.20902 0.01488 14.044 < 2e-16 \*\*\*

fMonth9 -0.04339 0.01509 -2.876 0.004425 \*\*

fMonth10 -0.03674 0.01510 -2.434 0.015735 \*

fMonth11 -0.08356 0.01510 -5.532 8.93e-08 \*\*\*

fMonth12 0.02898 0.01492 1.943 0.053 288 .

obs21 -0.17839 0.04880 -3.655 0.000321 \*\*\*

obs22 -0.26579 0.04880 -5.447 1.37e-07 \*\*\*

obs23 -0.15715 0.04880 -3.220 0.001475 \*\*

---

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

Residual standard error: 0.04699 on 220 degrees of freedom

Multiple R-squared: 0.9559, Adjusted R-squared: 0.952

F-statistic: 250.7 on 19 and 220 DF, p-value: < 2.2e-16

A graph showing a number of green and red lines

Description automatically generated

**As the residual plot above shows, the model fails to capture some of the trend structure in the data, especially in the vicinity of the 2001 and 2008–09 economic downturns.**

> qq<-qqnorm(resid(model6))

> qqline(resid(model6))

> identify(qq)

[1] 1 40 41

A graph of a normal q-q plot

Description automatically generated

> obs40<-ifelse(Time==40,1,0)

> obs41<-ifelse(Time==41,1,0)

> obs1<-ifelse(Time==1,1,0)

> flights<-data.frame(flights,obs1,obs40,obs41)

**The residual normal quantile plot for model 6 has a long lower tail and departure from normality. The three outliers identified on the plot are for January 2000, and April and May of 2003. Dummies are defined for them and added to the model fit, giving the model 7 result following.**

> model7<-lm(logIntlEnps~poly(Time,5)+fMonth+obs1+obs21+obs22+obs23+obs40+obs41,data=flights[1:240,]);summary(model7)

Call:

lm(formula = logIntlEnps ~ poly(Time, 5) + fMonth + obs1 + obs21 +

obs22 + obs23 + obs40 + obs41, data = flights[1:240, ])

Residuals:

Min 1Q Median 3Q Max

-0.13734 -0.02381 0.00196 0.02563 0.12188

Coefficients:

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

(Intercept) 8.858255 0.009595 923.209 < 2e-16 \*\*\*

poly(Time, 5)1 2.529351 0.043002 58.819 < 2e-16 \*\*\*

poly(Time, 5)2 -0.255718 0.042696 -5.989 8.66e-09 \*\*\*

poly(Time, 5)3 0.066732 0.042737 1.561 0.119878

poly(Time, 5)4 0.377839 0.043055 8.776 5.09e-16 \*\*\*

poly(Time, 5)5 -0.446831 0.043524 -10.266 < 2e-16 \*\*\*

fMonth2 -0.120750 0.013403 -9.009 < 2e-16 \*\*\*

fMonth3 0.087595 0.013398 6.538 4.39e-10 \*\*\*

fMonth4 0.038461 0.013575 2.833 0.005041 \*\*

fMonth5 0.064311 0.013571 4.739 3.89e-06 \*\*\*

fMonth6 0.147696 0.013390 11.031 < 2e-16 \*\*\*

fMonth7 0.239658 0.013389 17.899 < 2e-16 \*\*\*

fMonth8 0.200460 0.013390 14.971 < 2e-16 \*\*\*

fMonth9 -0.051474 0.013563 -3.795 0.000191 \*\*\*

fMonth10 -0.044761 0.013566 -3.299 0.001132 \*\*

fMonth11 -0.091519 0.013570 -6.744 1.38e-10 \*\*\*

fMonth12 0.020696 0.013402 1.544 0.123986

obs1 -0.181574 0.046131 -3.936 0.000112 \*\*\*

obs21 -0.186612 0.043390 -4.301 2.57e-05 \*\*\*

obs22 -0.273848 0.043390 -6.311 1.54e-09 \*\*\*

obs23 -0.165039 0.043390 -3.804 0.000185 \*\*\*

obs40 -0.217085 0.043361 -5.006 1.15e-06 \*\*\*

obs41 -0.192328 0.043342 -4.437 1.45e-05 \*\*\*

---

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

Residual standard error: 0.04177 on 217 degrees of freedom

Multiple R-squared: 0.9656, Adjusted R-squared: 0.9621

F-statistic: 276.9 on 22 and 217 DF, p-value: < 2.2e-16

**The additional dummies are all significant, and we retain them. *R*-square has increased from 0.9559 to 0.9656, and the residual standard error has decreased from 0.0470 to 0.0418.**

A graph showing a red and green line

Description automatically generated

**Three large negative values evident in the residual plot for model 6 have been adjusted in the present plot for model 7. Otherwise there are no changes in the residual plot.**

A graph of a normal q-q plot

Description automatically generated

> shapiro.test(resid(model7))

Shapiro-Wilk normality test

data: resid(model7)

W = 0.97976, p-value = 0.001663

**The residual normal quantile plot for model 7 is improved from that for model 6, but, as the Shapiro–Wilk test shows, the residual distribution is still nonnormal. And the residual acf plot below gives evidence of much remaining trend not captured by model 7.**

A graph with lines and numbers

Description automatically generated

A graph of a graph

Description automatically generated

**The residual spectral plot is very informative. The peak at low frequency is evidence of remaining trend structure, and the smaller peaks at frequencies 1/12, 2/12, and 4/12 show the presence of remaining dynamic seasonality which model 7 cannot capture. Additionally, some calendar structure may be present—see the spectral ordinates at frequencies 0.220 and 0.348.**

**Next, we form static seasonal estimates from model 7.**

> b1<-coef(model7)[1]

> b2<-coef(model7)[7:17]+b1

> b3<-c(b1,b2)

> seas7<-exp(b3-mean(b3))

> seas7

(Intercept) fMonth2 fMonth3 fMonth4 fMonth5 fMonth6

0.9599592 0.8507692 1.0478391 0.9975998 1.0237233 1.1127472

fMonth7 fMonth8 fMonth9 fMonth10 fMonth11 fMonth12

1.2199301 1.1730363 0.9117964 0.9179382 0.8760050 0.9800340

> cbind(seas7)

seas7

January 0.9600

February 0.8508

March 1.0478

April 0.9976

May 1.0237

June 1.1127

July 1.2199

August 1.1730

September 0.9118

October 0.9179

November 0.8760

December 0.9800

A graph with lines and numbers

Description automatically generated

**International air travel is estimated to peak very sharply in July and August, and there is also a small local peak in March. The months with lowest estimated international travel are February, September, October, and November, and the January estimate is also below trend level. In July enplanements are estimated to be 22.0 percent above trend level, and the estimates for August, June, and March are 17.3 percent, 11.3 percent, and 4.8 percent, respectively, above trend. The February, November, September, and October estimates are 14.9 percent, 12.4 percent, 8.8 percent, and 8.2 percent, respectively, below trend level.**

**The other months are all close to trend level.**

**These results confirm, as one expects, that international air travel peaks in the summer months when people take vacations and their children are not attending school. The low level of international travel occurs in February and the fall months. The small estimated peak in March is consistent with the seasonal pattern for domestic travel.**

**Next, we add the lag 1 residual to model 7. A lag 2 residual is tried, but is not statistically significant.**

> lag1resid2<-c(NA,resid(model7)[1:239],rep(0,41))

> flights<-data.frame(flights,lag1resid2)

> model8<-lm(logIntlEnps~poly(Time,5)+fMonth+obs21+obs22+obs23+obs40+obs41+c348+s348+c432+s432+lag1resid2,data=flights[1:240,]);summary(model8)

Call:

lm(formula = logIntlEnps ~ poly(Time, 5) + fMonth + obs21 + obs22 +

obs23 + obs40 + obs41 + c348 + s348 + c432 + s432 + lag1resid2,

data = flights[1:240, ])

Residuals:

Min 1Q Median 3Q Max

-0.127614 -0.009857 0.000000 0.015046 0.083219

Coefficients:

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

(Intercept) 8.860900 0.006093 1454.288 < 2e-16 \*\*\*

poly(Time, 5)1 2.532512 0.027264 92.887 < 2e-16 \*\*\*

poly(Time, 5)2 -0.245028 0.027077 -9.049 < 2e-16 \*\*\*

poly(Time, 5)3 0.066759 0.027097 2.464 0.014547 \*

poly(Time, 5)4 0.392432 0.027310 14.369 < 2e-16 \*\*\*

poly(Time, 5)5 -0.438836 0.027604 -15.897 < 2e-16 \*\*\*

fMonth2 -0.122920 0.008512 -14.441 < 2e-16 \*\*\*

fMonth3 0.084213 0.008510 9.896 < 2e-16 \*\*\*

fMonth4 0.031338 0.008619 3.636 0.000348 \*\*\*

fMonth5 0.062812 0.008628 7.280 6.44e-12 \*\*\*

fMonth6 0.143822 0.008500 16.920 < 2e-16 \*\*\*

fMonth7 0.238317 0.008499 28.039 < 2e-16 \*\*\*

fMonth8 0.197256 0.008500 23.207 < 2e-16 \*\*\*

fMonth9 -0.050637 0.008609 -5.882 1.57e-08 \*\*\*

fMonth10 -0.047093 0.008604 -5.473 1.24e-07 \*\*\*

fMonth11 -0.094413 0.008615 -10.959 < 2e-16 \*\*\*

fMonth12 0.017898 0.008512 2.103 0.036668 \*

obs21 -0.265812 0.028023 -9.485 < 2e-16 \*\*\*

obs22 -0.262489 0.027708 -9.473 < 2e-16 \*\*\*

obs23 -0.174910 0.027721 -6.310 1.61e-09 \*\*\*

obs40 -0.117540 0.028255 -4.160 4.62e-05 \*\*\*

obs41 -0.205366 0.027746 -7.402 3.11e-12 \*\*\*

c348 0.004656 0.002465 1.889 0.060245 .

s348 0.008544 0.002468 3.462 0.000649 \*\*\*

c432 0.002775 0.002471 1.123 0.262691

s432 -0.005427 0.002448 -2.217 0.027698 \*

lag1resid2 0.802035 0.044797 17.904 < 2e-16 \*\*\*

---

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

Residual standard error: 0.02648 on 212 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.9863, Adjusted R-squared: 0.9846

F-statistic: 587.3 on 26 and 212 DF, p-value: < 2.2e-16

**The added lag 1 residual variable is highly significant. And both trigonometric calendar pair are significant enough for inclusion in the model. Here *R*-square is 0.9863, an increase from 0.9656 for model 7, and the residual standard error is 0.0265 in comparison to 0.0418 for model 7. Residual analysis of model 8 follows.**

A graph showing a graph of a graph

Description automatically generated with medium confidence

**The added lag 1 residual variable has allowed the model to capture essentially all trend structure in the data. There are several large negative residuals corresponding to time points overestimated by the model. One occurs at the end of the 2001 recession, and a second near the end of the 2008–09 recession. There are a few brief episodes of increased volatility. However, in summary this residual plot is a substantial improvement from the model 7 plot.**

A graph of a normal q-q plot

Description automatically generated

> shapiro.test(resid(model8))

Shapiro-Wilk normality test

data: resid(model8)

W = 0.89807, p-value = 1.195e-11

**The residual distribution is decidedly nonnormal. Three large negative residuals are visible in the quantile plot, and the lower tail of the residual distribution is long relative to normality.**

A graph of a series resid

Description automatically generated

A graph of a graph

Description automatically generated

**There are several very modestly significant residual autocorrelations, including at lags 18 and 36. However, the acf plot suggests that model 8 almost achieves reduction to white noise. The residual spectral plot is more informative. Spectral peaks at frequencies 2/12 and 4/12 indicate there is remaining mild dynamic seasonality. The spectral plot also indicates that model 8 has almost achieved reduction to white noise. We can conclude that model 8 offers effective description of the international enplanement time series.**

8. Compare the estimated static seasonal indices for the domestic and international data, using a table and a plot. Discuss carefully.

> cbind(seas4,seas7)

seas4 seas7

January 0.8793 0.9600

Februry 0.8552 0.8508

March 1.0553 1.0478

April 1.0074 0.9976

May 1.0468 1.0237

June 1.0861 1.1127

July 1.1275 1.2199

August 1.0917 1.1730

September 0.9305 0.9118

October 1.0202 0.9179

November 0.9658 0.8760

December 0.9750 0.9800

> plot(ts(seas4),ylab="Seasonal Index",xlab="Month",main="Estimated Seasonal Indices, Models 2 and 8",ylim=c(0.8,1.25),col="red",lwd=2)

> lines(ts(seas7),col="blue",lwd=2)

> legend("topleft",legend=c("model 2","model 8"),col=c("red","blue"),lty=1,cex=0.8)

A graph with red and blue lines

Description automatically generated

**There are several notable differences between the estimated seasonal indices for domestic flights and the indices for international flights. In comparison to domestic travel, international travel is more prominent in January, July, and August and less prominent in October and November. The differences between the indices for these months are quite large. The other monthly indices are similar. The summer months dominate international enplanements. Both modes of travel show similar peaks relative to trend in March.**