

Chapter 6 Solutions

6.2.3. Exercises

Exercise 6.1

- (a) Since $d=1$, R will regular difference x^* once; since $D=1$, then R will seasonally difference once the regular difference. Let's call the differenced time series y_t . After that, an MA(1) model will be fitted to the regular part of y_t , the lower lags (not the lag indicating the cycle when the time series has been made R's ts()) and an MA(1) model will be fitted to the seasonal part (lag=12 and nearby, which if ts object, will be labeled as 1 in the horizontal axis). The model in polynomial form will be

$$(1 - B^{12})(1 - B)x_t^* = (1 + \beta_{12}B^{12})(1 + \beta_1B)w_t$$

. The final forecasting equation will be:

We multiply the polynomials:

$$(1 - B - B^{12} + B^{13})x_t^* = (1 + \beta_1B + \beta_{12}B^{12} + \beta_{12}\beta_1B^{13})w_t,$$

and then just solve for x_t^* .

$$x_t^* = x_{t-1}^* + x_{t-12}^* + x_{t-13}^* + w_t - \hat{\beta}_1w_{t-1} + \hat{\beta}_{12}w_{t-12} + \hat{\beta}_1\hat{\beta}_{12}w_{t-13}$$

and the final forecasting equation for x_t is

$$x_t = e^{x_t^*}$$

- (b) Follow the steps followed in part (a). But the model, in polynomial form, will be:

$$(1 - B^{12})(1 - B)x_t^* = (1 + \beta_{12}B^{12} + \beta_{24}B^{24})(1 + \beta_1B + \beta_2B^2)w_t$$

Multiply the polynomials on the left and on the right, and then do a forecasting equation like the one above.

□

Exercise 6.3

Since x_t is stationary, let y_t denote the raw time series that will be differenced. And let's suppose that, since only the same month of last year affects the value this month, either a seasonal AR or a seasonal MA term is fitted. Let's assume both are fitted. The polynomial form is

$$(1 - \alpha_{12}B^{12})(1 - B)y_t = (1 + \beta_{12}B^{12})w_t.$$

Upon expanding both sides, we get:

$$y_t - y_{t-1} - \alpha_{12}y_{t-12} - \alpha_{12}y_{t-13} = w_t + \beta_{12}w_{t-12}.$$

The forecasting equation expression is then

$$y_t = y_{t-1} + \alpha_{12}y_{t-12} + \alpha_{12}y_{t-13} + w_t + \beta_{12}w_{t-12}.$$

□

Exercise 6.5

ARIMA(3, 1, 1)(0, 0, 0)₄

$$(1 - \alpha_1 B - \alpha_2 B^2 - \alpha_3 B^3)(1 - B)x_t^* = (1 + \beta_1 B)x_t$$

$$x_t^* = x_{t-1}^* + \alpha_1 x_{t-1}^* - \alpha_1 x_{t-2}^* + \alpha_2 x_{t-2} - \alpha_2 x_{t-3} + \alpha_3 x_{t-3} - \alpha_3 x_{t-4} + w_t + \beta_1 w_{t-1}$$

□

Exercise 6.7

(a) The model is ARIMA(1,1,1)(0,1,1)

(b) We expand the polynomial:

$$(1 - B^{12} - 0.219B + 0.219B^{13} - B + B^{13} + 0.219B^2 - 0.219B^{14})z_t =$$

$$(1 - 0.8245 - 0.3745B^{12} + 0.8245 * 0.3745B^{13})w_t.$$

Translate polynomial terms:

$$z_t - 1.219z_{t-1} + 0.219z_{t-2} - z_{t-12} + 1.219z_{t-13} - 0.219z_{t-14} =$$

$$w_t - 0.8245w_{t-1} - 0.3745w_{t-12} + 0.3087753w_{t-13}.$$

So the forecasting equation for z_t is

$$z_t = 1.219z_{t-1} - 0.219z_{t-2} + z_{t-12} - 1.219z_{t-13} + 0.219z_{t-14}$$

$$-0.8245w_{t-1} - 0.3745w_{t-12} + 0.3087753w_{t-13} + w_t$$

and we would predict z_{t+1} as follows:

$$z_{t+1} = 1.219z_t - 0.219z_{t-1} + z_{t-11} - 1.219z_{t-12} + 0.219z_{t-13}$$

$$-0.8245w_t - 0.3745w_{t-11} + 0.3087753w_{t-12} + w_{t+1}$$

(c) $x_{t+1} = e^{z_{t+1}}$

□

Exercise 6.9

We are trying to see whether we need to make the variance stationary.

The time plot "Electricity time series" shows an increasing trend and a seasonal apparently proportional to the trend. So we might have suspected heteroscedasticity, nonconstant variance (variance nonstationarity).

The plot of the first difference, viewed in plot "First regular difference of electricity", confirms our suspicion. The variance nonstationarity is present in this plot.

The plot titled "First reg diff of log electricity" indicates that we are trying to correct the variance nonstationarity by taking log of electricity, and that plot shows that we succeeded at that by taking the log. The prevalence of the recurrent waves in the plot though suggests that regular difference alone has not succeeded in accounting for seasonality.

Certainly more inspection of the data with the tools learned is needed.

□

Exercise 6.13

- (a) Taking log was a good idea. It makes the variance more homogeneous across windows of the time series.
- (b) An $ARIMA(0,1,1)(0,0,0)$ model fitted to log electricity was not a good idea. There was clearly some random seasonal term in the time series that should have been accounted for. See Exercise 6.10. The identification done by the report is not correct. The ACF dies away slowly due to the regular seasonal. The PACF dies away as well.
- (c) According to the program china.R, based on what the report says in (a) and (b), the following is the model that they would have fitted:

$$\hat{z}_t = z_{t-1} + w_t - 0.5089w_{t-1}, \text{ where } z_t = \log(\text{electricity})$$

So the model reported in the exercise is model 1, consistent with what the report said in part (a) and (b), but still not the correct model.

The final prediction should certainly be calculated as

$e^{\hat{z}_t}$. but this step should not be done until the residuals of the model fitted are white noise. Certainly it should not be done with model 1.

- (d) The model residuals indicate that we did NOT do a good job with

$$\hat{z}_t = z_{t-1} + w_t - 0.5089w_{t-1}$$

When we run Program china.R, we see that the residuals are not those of white noise. So the model 1 fitted by the report (model 1 in china.R program) is not the correct one.

As we illustrate in Program china.R, it is possible to obtain a better model.

Two other models were considered, $ARIMA(0, 1, 1)(0, 0, 1)_{12}$ and $ARIMA(0, 1, 1)(1, 0, 1)_{12}$. Only the latter results in residuals of white noise, indicating that the model fits well. The AIC is also lower for the second model. An $ARIMA(1, 1, 1)(1, 0, 1)_{12}$ was also considered but the model has higher AIC than the other two models and the residuals were not the residuals of white noise.

The model we suggest is, in polynomial form:

$(1 - 0.985B^{12})(1 - B)z_t = (1 - 0.7390B^{12})(1 - 0.6113B)w_t$ The reader can follow the examples in earlier exercises to expand both sides and write the model in forecasting equation form.

When modeling time series, both the regular part and the seasonal part, after making both stationary, must be modeled. Otherwise, the residuals of the fitted model will not be white noise. The correlogram of the residuals will indicate that there are still autocorrelations to model.

□

Exercise 6.15

The reader should use Program `monthlysunsplotexercise.R` to do this exercise.

- (a) See program
- (b) See program and plots it produces. There does not seem to be an increasing variability as time goes on. Seasonality is not present, as the seasonal box plot displays very similar median number of sunspots every month. There is less variability in October and December, but overall, the variabilities of the sunspots across months is very similar. No outliers are present.

We should difference, because the sample ACF shows that there are significant autocorrelations up to 4 years of month lags, the typical pattern of the ACF of a nonstationary time series. Very possibly, the nonstationarity is caused by those 10 year cycles that we observe in the time plot.

After doing first order differencing, we observe that the time series fluctuates around 0, but there are periods of outbursts, so the variance is not homogeneous. Perhaps we should have logged. The reader can try the exercise starting with the logged series to see if there is any difference. If we continue with what we have instead, we obtain what follows.

- (c) The ACF cuts off after r_3 and the PACF dies away. This is a pattern of an MA(3). We can start with that model: $ARIMA(0, 1, 3)(0, 0, 0)_{12}$.
- (d) The MA(3) model leaves a residual that is white noise according to the residuals ACF and according to the Ljung-Box test (p-value 0.9813, lag=6). The fitted model to the differenced training data is

$$\hat{y}_t^* = -0.4117w_{t-1} - 0.1638W_{t-2} + 0.0751W_{t-3}$$

All the model coefficients are significantly different from 0.

- (e) See the *Exercise6.15.R* program to obtain what is being asked. We observe in the plot of the forecast and the prediction intervals that the test data falls inside the interval, and the forecast is approximately the same number of sun spots

□

6.4.3 Exercises

Exercise 6.16

The following models were fit to y^{**} , which denotes the seasonal difference of the regular difference of the square root of the `osvisit` time series, using times 1977:1-1994:12 for the model fit (that is the training data). The `sigma2` then is the root mean square of the residuals, a measure of how good the model fit is. The `acf` of y^{**} helped us see that the time series thus transformed was stationary.

□

Model	(p,d,q)(P,D,Q) _F	Res white noise? What is not properly modeled?	Akaike AIC	sigma2 (in sample fit measure)	rmse forecast
Model 7	ARIMA(1,1,1)(1,1,1)	WN	1415.504	58.3783	5637.484
Model 9	ARIMA(1,1,1)(1,1,0)	almost WN- a little seas	1422.092	61.41928	6610.954
Model 8	ARIMA(1,1,1)(0,1,1)	almost WN a little seas	1415.157	58.98177	5618.592
Model 3	ARIMA(1,1,0)(1,1,0)	NOT WN	1450.531	71.61326 regular	7043.702
Model 5	ARIMA(1,1,0)(0,1,1)	NOT WN	1440.316	67.53381 regular	6395.195
Model 1	ARIMA(1,1,0)(0,1,0)	NOT WN Reg and seas	1468.177	79.35731	8507.463
Model 6	ARIMA(0,1,1)(1,1,0)	NOT WN seas	1426.524	63.54052	7511.455
Model 4	ARIMA(0,1,1)(0,1,1)	NOT WN seas	1420.422	61.3432	6644.038
Model 2	ARIMA(0,1,0)(1,1,0)	NOT WN regular	1477.937	82.72045	6542.025

Marked in red are the lowest of each column.

6.9 Problems

Problem 6.2

$$(1 - B^{12})(1 - B)y_t = (1 - 0.5B)(1 - 0.5B^{12})w_t$$

$$ARIMA(0, 1, 1,)(0, 1, 1)_{12}$$

□

Problem 6.4

ACF, Ljung Box, Run test, Durbin Watson. Be able to interpret the R output from these tests and see the limitations of each.

□

Problem 6.9

The ACF of viscosity dies down quickly (at the first few lags) in sinusoidal form, and the PACF cuts off after lag 2. An AR(2) model suffices. After fitting the following model

$$(1 - 0.6821B - 0.4333B^2)y_t = 34.9464 + w_t,$$

the residuals are white noise which indicates the model is a good model. The square of the residuals is also white noise., indicating, together with the white noise residuals, that there is no volatility in the residuals. Garch is not needed. The following quick code to check (not to obtain plots for presentation, which should have titles, and informative labels and the time) will allow confirming the conclusions.

```
data=scan("viscosity.txt")
y=ts(data)
plot(y)

par(mfrow=c(2,1))
acf(y); pacf(y)
model=arima(y,order=c(2,0,0))
acf(model$residuals)
acf(model$residuals^2)
```

□

Problem 6.10

- (a) The following R code will allow the reader to produce results. The code will continue in the following parts, as needed.

```
rooms=scan("rooms.txt")
rooms.ts=ts(rooms,frequency=12,start=c(1977,1)) #make the data a time series
start(rooms.ts); end(rooms.ts)
rooms1=window(rooms.ts,end=c(1989:12)) # training set

# Pre-differencing transformations needed? Check variance stationarity

# The time plot looks like the variability of the data increases over time.
# Transformations to consider (will choose just one) for variance stationarity

par(mfrow=c(3,1))
lnrooms.ts=ts(log(rooms1),frequency=12,start=c(1977,1)) #log transform
sqrrooms.ts=ts(sqrt(rooms1),frequency=12,start=c(1977,1)) # square root
qrrooms.ts=ts(rooms1^(0.25),frequency=12,start=c(1977,1)) # quartic
dev.off()

# We will plot all three transformations now to see which stabilizes
# the variance better.

par(mfrow=c(3,1))
plot.ts(lnrooms.ts,type="l",main="time series plot of ln(rooms) training data")
plot.ts(sqrrooms.ts,type="l",main="time series plot of sqr(rooms)training data")
plot.ts(qrrooms.ts,type="l",main="time series plot of qrt(rooms) training data")
dev.off()

# We will work with the quartic root $y*=y^0.25.$ So let?s
# now look at mean stationarity.
# For that, we must look at the SAC of the quartic root of the data.

acf(qrrooms.ts, main="ACF of training quartic root data (1977:1-1989:12))
```

```
# The SAC pf the quartic root reveals the nonstationarity that
we see both in the time series plot (both trend over time and
seasonality nonstationarity) although the long decaying behavior
of the ACF that would reveal the trend in the time plot over time is
masked by the smooth seasonal correlations.
We must difference the data to remove the mean nonstationarities.

par(mfrow=c(3,1))
diff1=diff(qrtrrooms.ts,lag=1,differences=1) #This does only regular differencing
acf(diff1)
diff12=diff(qrtrrooms.ts,lag=12,differences=1) # only seasonal differencing
acf(diff12)
diff1diff12=diff(diff1,lag=12,differences=1) # seasonal diff of regular diff
acf(diff1diff12)
dev.off()
```

- (b) This question is asking to identify a tentative model to start with. We do not have to fit it in this part.

Since we think that the seasonal difference gives us a sample ACF that reveals that the quartic root has been made stationary, we will use for modeling the following variable $z_t = y_t^* - y_{t-12}^*$.

To identify a model we need to look both at the ACF and the PACF. Add the following code. d

```
par(mfrow=c(2,1))
acf(diff12, main="ACF of seasonal diff of training data")
pacf(diff12, main="PACF of seasonal diff of training data")
dev.off()
```

Now looking at the sample ACF and PACF of the seasonal difference of the quartic root of the training data, we can see that at the nonseasonal frequencies, the PACF has significant spikes at lags 1, 3, and cuts off after that, while the ACF dies down quickly, suggesting an AR model with terms at lags 1 to start with. The reader could consider other close interpretations.

Continuing with the identification, and looking at the seasonal frequencies, (the 1 in the horizontal axis corresponding to lag 12) we can see that the ACF cuts off quickly at lag 12 and the PACF dies down quickly, suggesting a seasonal model MA with a term at lag 12, seasonal MA(1). So the model identified is $ARIMA(1, 0, 1)(0, 1, 1)_{12}$ which in polynomial form would be:

$$(1 - \alpha_1)(1 - B^{12})y_t^* = (1 - \beta_{12}B^{12})w_{t-12} + w_t$$

This is the model that would be fitted to start with. It will not be the last model.

- (c) Add the following code to the program.

```
model1 = arima(qrtrrooms.ts, order=c(1,0,0),seas=list(order=c(0,1,1), 12),
include.mean=FALSE)
model1
```

$$(1 - 0.9206)(1 - B^{12})y_t^* = (1 - (-0.4796)B^{12})w_{t-12} + w_t \quad se_{\hat{\alpha}_1} = 0.0487, t - statistic = 18.9; se_{\hat{\beta}_{12}} = 0.1368; t - statistic = -3.5$$

$$AIC = -574.15; \hat{\sigma}^2 = 0.001009$$

The t-test statistics are larger than 2 (the dataset is large enough to use this benchmark) so the model coefficient estimates are statistically significant.

- (d) We check first whether the residuals of the fitted model are white noise. We can look at the ACF of the residuals and the Ljung-Box test. Add the following code:

```
acf(residuals(model1), na.action=na.omit, lag=50)
Box.test(residuals(model1), lag=6, type="Ljung-Box")
```

The ACF of the model residuals indicates that although the seasonal part has been captured by the model (no significant autocorrelation at lag 12 (1 number in the horizontal axis)), the lower autocorrelations are not captured. We need to consider other models.

Keeping it small, the reader can start considering several variations of the modeling for the regular low lags. Could consider and run for example $ARIMA(5, 0, 1)(0, 1, 1)_{12}$. It is a rather large model, so consider smaller order variations for the regular part instead of (5,0,1)

- (e) Once satisfied with the model, use code from program *secondarimaexample.R* to obtain the forecasts, after creating the test set which will be needed to calculate the RMSE.

□

Problem 6.11

This problem is about how to explain the findings in a narrative. If the regular differenced U_t is stationary and an AR(1) then it is the U_t that is $ARIMA(1,1,0)(0,0,0)$, not the differenced U_t . The differenced U_t is $ARMA(1,0,0)$. The 1 in the middle of $ARIMA(1,1,0)$ is indicating that the variable we are talking about is being differenced and that is unemployment. If we fit a model to the differenced data using the *arima* function, then we should not use `order=c(1,1,0)` but `order=c(1,0,0)`. The program that helps us see what is the right thing to do if we are interested in forecasting U_t is *ch6unemploycanada*, listed next.

```
install.packages("vars")
library(vars)
data(Canada)
head(Canada)
class(Canada)
start(Canada); end(Canada); frequency(Canada)

#time plot for Canada' unemployment

plot.ts(Canada[,4],
main="Unemployment in Canada (1980:1-2000:4)",
ylab="Unemployment rate")

boxplot(Canada[,4]~cycle(Canada[,4]),
main="Seasonal boxplot for Canada's unemployment",
ylab="Unemployment rate")

acf(Canada[,4], lag=50, main="Unemployment is not stationary")
```



```
#training and test data#
Canada.training=window(Canada[,4],start=c(1980,1),end=c(1998,4))
Canada.test=window(Canada[,4],start=c(1999,1), end=c(2000,4))

#Regular differenced data#
diff.Canada=diff(Canada.training,lag=1,diff=1)

par(mfrow=c(2,1))
plot.ts(diff.Canada,
main="First regular difference of unemployment\n (1980:1-1998:4)")
acf(diff.Canada,
main="ACF of first regular difference of unemployment\n (1980:1-1998:4)")
dev.off()

## Identify a model

par(mfrow=c(2,1))
acf(diff.Canada, lag=50,main="acf of reg differenced Unemployment ")
pacf(diff.Canada,lag=50,main="pacf of reg differenced Unemployment")
dev.off()

# Fitting an ARIMA model to the training data#
model=arima(Canada.training,order=c(1,1,0), seas=list(order=c(0,0,0),4))
model
#notice that we fit the model to Canada.training, because we are specifying
# order=c(1,1,0) indicating that the Canada.training data will be differenced
# by the program.

acf(residuals(model),main="ACF of residuals of model1")
Box.test(residuals(model),lag=12,type="Ljung-Box")

forecast=predict(model,8)$pred
forecast
start(forecast)
end(forecast)
forecast.se=predict(model,8)$se
#CI for forecast1#
ci.low= ts(forecast-1.96*forecast.se, start=c(1999,1),end=c(2000,4),freq=4)
ci.high=ts(forecast+1.96*forecast.se,start=c(1999,1),end=c(2000,4),freq=4)

# Plot the training data and forecast
ts.plot(cbind(Canada.training,forecast,ci.low, ci.high),
lty=c(1,2,3,3), col=c("black", "red","blue","blue"),
main="Predicted 8 quarterly values of unemployment",ylab="diff unemployment")
lines(Canada.test, col="purple",lty=8)
abline(v=1999.1)

rmse= function(x,y){
```

```
a=sum((x-y)^2)
b=length(x)
rmse=sqrt(a/b)
rmse
}
rmse(Canada.test,forecast )
```

Of course, if our interest is in forecasting the change in unemployment, then the program should be rewritten.

□

Problem 6.18

$$\begin{aligned}y_t &= y_{t-1} + y_{t-12} - y_{t-13} + 0.2191y_{t-1} - 0.2191y_{t-2} \\ &- 0.2191y_{t-13} + 0.2191y_{t-14} - 0.3745w_{t-12} - 0.8245w_{t-1} \\ &+ (0.8245)(0.3745)w_{t-13} + w_t\end{aligned}$$

$$\begin{aligned}y_{t+1}^{\hat{}} &= y_t + y_{t-11} - y_{t-12} + 0.2191y_t - 0.2191y_{t-1} \\ &- 0.2191y_{t-12} + 0.2191y_{t-13} - 0.3745w_{t-11} - 0.8245w_t \\ &+ (0.8245)(0.3745)w_{t-12} + w_t\end{aligned}$$

and

$$x_{t+1}^{\hat{}} = e^{y_{t+1}^{\hat{}}}$$

□

6.10 Quiz

Some questions may have more than one answer.

Question 6.3

FALSE

□

Question 6.4

need to see both the sample ACF and the sample PACF of the observed univariate time series

□

Question 6.5

$$(1 - B^{12} - 0.219B + 0.219B^{13} - B + B^{13} + 0.219B^2 - 0.219B^{14})y_t = (1 - 0.8245 - 0.3745B^{12} + 0.8245 * 0.3745B^{13})w_t$$

$$y_t - 1.219y_{t-1} + 0.219y_{t-2} - y_{t-12} + 1.219y_{t-13} - 0.219y_{t-14} = w_t - 0.8245w_{t-1} - 0.3745w_{t-12} + 0.3087753w_{t-13}$$

The coefficient estimate is 1.219 but when you solve for y_t the sign will change.

□

Question 6.6

Nonstationary

Differenced regular and seasonally before maximum likelihood was used to estimate the model.

□

Question 6.9

3.5

□

Question 6.11

that the variance of the process behaves like an AR(1) process. See example 6.7.

□

Question 6.13

(d) and (e)

□