AAA-Washington-Case

Yuxin Yao

2024-12-15

AAA Washington Case

In 1993, AAA Washington was one of the two regional automobile clubs affiliated with the American Automobile Association (AAA) operating in Washington State.

Club research had consistently shown that the emergency road service benefit was the primary reason that people join AAA. Providing emergency road service was also the club's single largest operating expense. It was projected that delivering emergency road service would cost $9.5 million, 37% of the club's annual operating budget, in the next fiscal year.

Michael DeCoria objective is to find a way to predict emergency road service call volume for future years.The data on emergency road service call volume is given in **AAAdata.csv**.

Previous analysis addressed the effect of average daily temperature on emergency road service call volume. We've found that the temperature's effect is significant and could explain about half of calls volume variability. We have also discovered that temperature alone does not account for the autocorrelations withing the calls volume series. We have recommended Michael to model these patterns using ARIMA models. He has extended your contract as a consultant to help him apply Box-Jenkins models.

(1.) Visualise the data

Use the monthly data of the calls volume, *Calls*, is recorded from May, 1988 till April, 1993. Produce a time plot of the data, label the axis nicely.

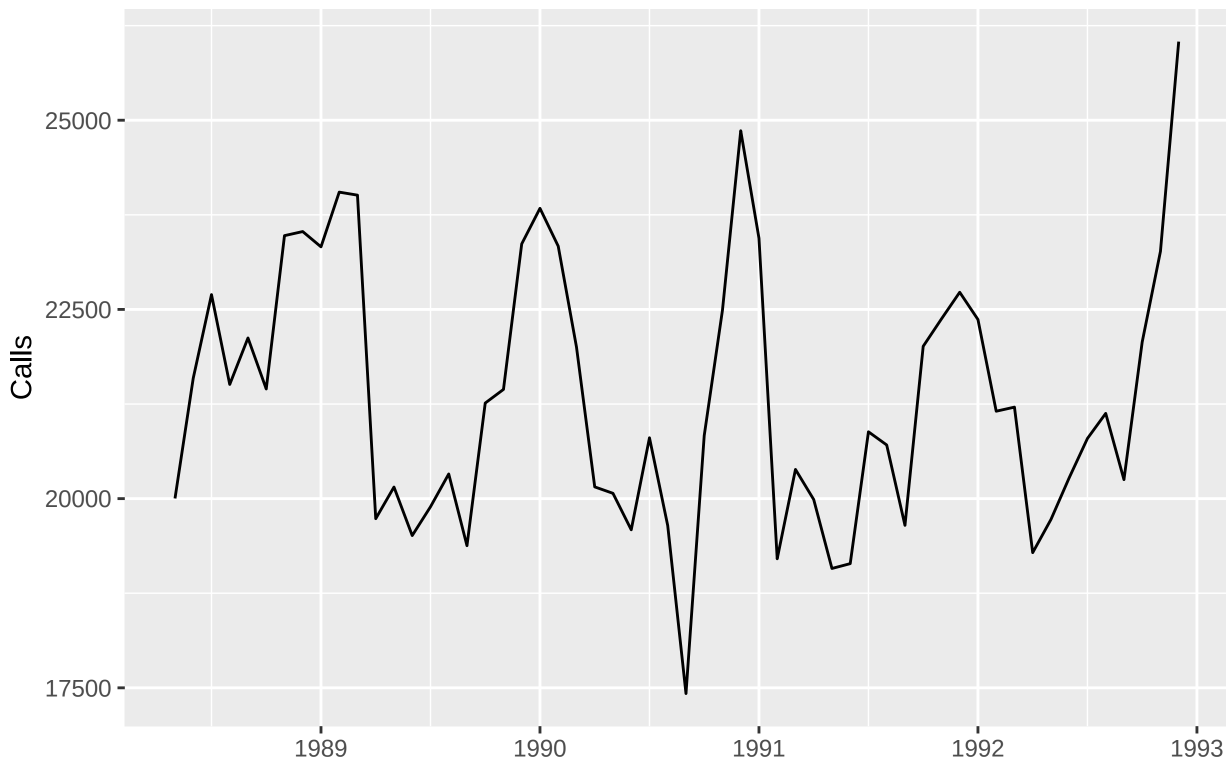
```
data <- read.csv("AAA.csv")
head(data)
```

```
##   Year       Mon Calls Temp
## 1   NA       May 20002 55.1
## 2   NA      June 21591 59.0
## 3   NA      July 22696 63.8
## 4   NA    August 21509 63.8
## 5   NA September 22123 59.1
## 6   NA   October 21449 54.6
```

```
data.ts <- ts(data$Calls, start = c(1988, 5), frequency = 12)
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
autoplot(data.ts, xlab='', ylab='Calls')
```

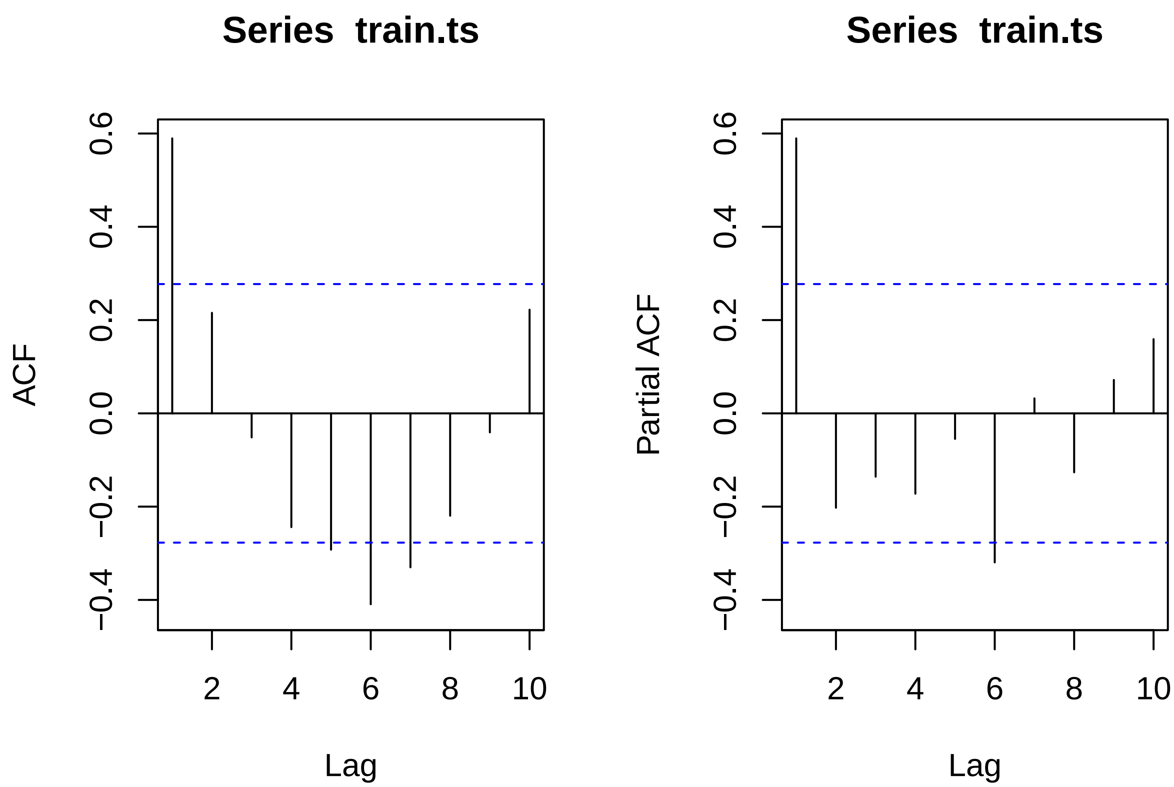


we can see seasonality, trend

(2.) Split the time series into training and validation sets. Leave the last 6 months for the testing set. Use the training set to get ACF and PACF plots.

```r
train.ts <- head(data.ts, (length(data.ts)- 6))
valid.ts <- tail(data.ts, 6)
```

```r
par(mfrow=c(1,2))
Acf(train.ts, 10)
Pacf(train.ts, 10)
```

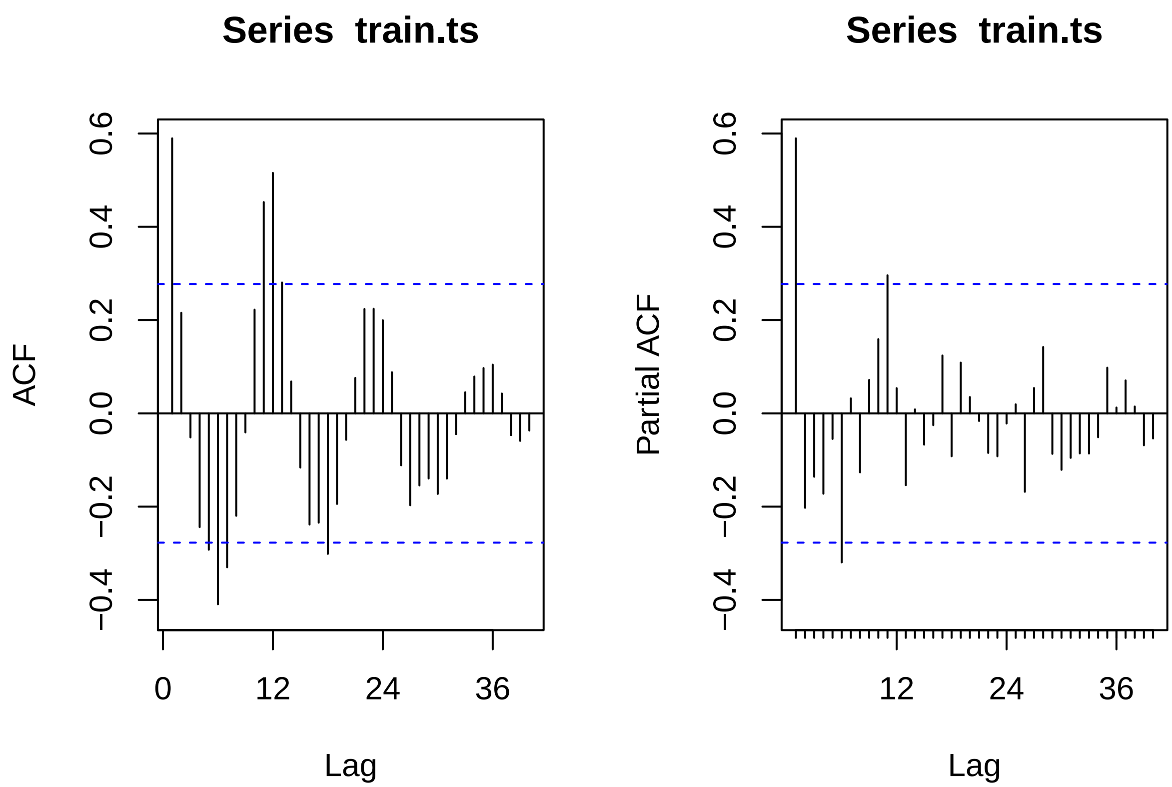



```
par(mfrow=c(1,1))
```

(1,0,0)(,0,) acf seems to tail off, pacf cuts off – AR(1)

both cut off, no pattern on the list. but try arma(1,1)

```
# seasonal
par(mfrow=c(1,2))
Acf(train.ts, 40)
Pacf(train.ts, 40)
```

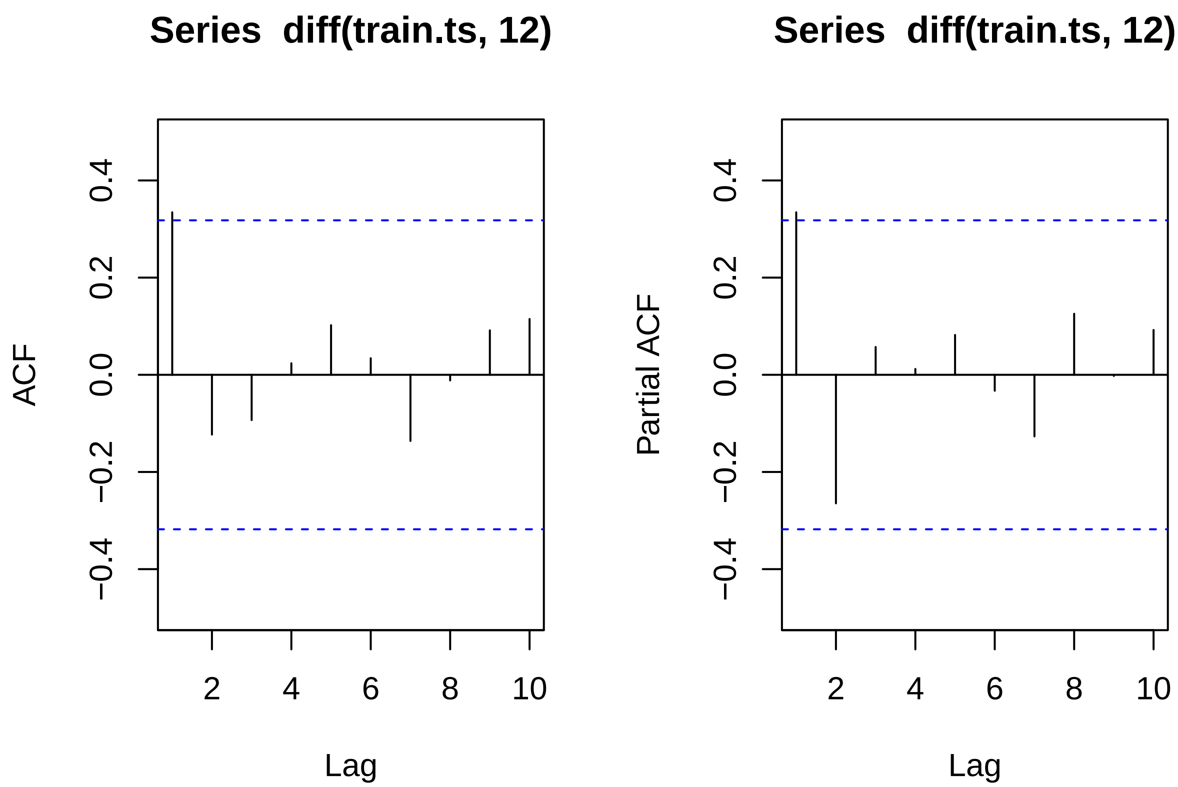



```r
par(mfrow=c(1,1))
```

seasonal: (1,0,0)(1,0,0) acf tails off, pacf cuts off – AR(1) SAR(1)

both cut off, no pattern on the list, but try arima(1,1)(?) (1,0,1)*(1,0,0)12 seasonality will be spike at 12. and there is a trend.

```r
par(mfrow=c(1,2))
Acf(diff(train.ts, 12), 10)
Pacf(diff(train.ts, 12), 10)
```




```r
par(mfrow=c(1,1))
```

we can try (1,0,1) since it is not clear what is going on. (0,0,1) maybe (0,0,0)

(3.) After a quick discussion with Michael I've decided to check the performance of the following models:

(1) SARIMA(1,0,0)x(1,0,0)12,
(2) SARIMA(1,0,0)x(1,1,0)12,
(3) SARIMA(1,0,0)x(0,1,0)12,
(4) SARIMA(1,0,1)x(0,1,0)12,
(5) SARIMA(0,0,1)x(0,1,0)12,
(6) SARIMA(1,1,1)x(0,1,1)12.

Fit the models and access their accuracy. Report all supporting material.

```r
# one of the options we might consider in previous question
m = arima(train.ts, order = c(0,0,0),
          seasonal = list(order= c(0,1,0),
                          period=12))
m.p = forecast(m, h=length(valid.ts))
accuracy(m.p, valid.ts)
```

```
##                     ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
## Training set -297.1301 1263.969 856.3499 -1.543848 4.175302 0.7647778 0.3518849
## Test set      865.6667 1432.735 895.0000  3.557380 3.698440 0.7992949 0.1025405
```

```
##                Theil's U
## Training set        NA
## Test set     0.9227046
```

```
m1 = arima(train.ts, order = c(1,0,0),
           seasonal = list(order= c(1,0,0),
                           period=12))
m2 = arima(train.ts, order = c(1,0,0),
           seasonal = list(order= c(1,1,0),
                           period=12))
m3 = arima(train.ts, order = c(1,0,0),
           seasonal = list(order= c(0,1,0),
                           period=12))
m4 = arima(train.ts, order = c(1,0,1),
           seasonal = list(order= c(0,1,0),
                           period=12))
m5 = arima(train.ts, order = c(0,0,1),
           seasonal = list(order= c(0,1,0),
                           period=12))
m6 = arima(train.ts, order = c(1,1,1),
           seasonal = list(order= c(0,1,1),
                           period=12))

m1.predict = forecast(m1, h=length(valid.ts))
m2.predict = forecast(m2, h=length(valid.ts))
m3.predict = forecast(m3, h=length(valid.ts))
m4.predict = forecast(m4, h=length(valid.ts))
m5.predict = forecast(m5, h=length(valid.ts))
m6.predict = forecast(m6, h=length(valid.ts))


accuracy(m1.predict, valid.ts)
```

```
##                      ME     RMSE       MAE        MPE     MAPE      MASE
## Training set -58.41925 1139.870  884.5108 -0.5798217 4.224527 0.7899274
## Test set     876.66610 1739.257 1071.1828  3.4300779 4.375837 0.9566380
##                    ACF1 Theil's U
## Training set 0.08285354        NA
## Test set     0.27680462    1.1158
```

```
accuracy(m2.predict, valid.ts)
```

```
##                      ME     RMSE       MAE       MPE     MAPE      MASE
## Training set -222.6510 1113.257  754.6887 -1.182794 3.689937 0.6739875
## Test set      853.8167 1293.186 1016.0962  3.598587 4.378965 0.9074420
##                    ACF1 Theil's U
## Training set 0.03655911        NA
## Test set     0.02920397 0.8359899
```

```
accuracy(m3.predict, valid.ts)
```

```
##                      ME     RMSE      MAE        MPE     MAPE      MASE
```

```
## Training set -170.7967 1160.563 787.7008 -0.9132158 3.829909 0.7034696
## Test set      745.5311 1434.847 922.1255  2.9830912 3.832307 0.8235198
##                    ACF1 Theil's U
## Training set 0.0749217        NA
## Test set     0.1270180 0.9121691
```

```
accuracy(m4.predict, valid.ts)
```

```
##                     ME     RMSE      MAE        MPE     MAPE      MASE
## Training set -189.4610 1128.471 759.6392 -0.9959035 3.697709 0.6784087
## Test set      805.4565 1442.607 950.5981  3.2679958 3.965960 0.8489478
##                     ACF1 Theil's U
## Training set -0.01991635        NA
## Test set      0.10296213  0.922173
```

```
accuracy(m5.predict, valid.ts)
```

```
##                     ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -194.0847 1128.805 758.0224 -1.018901 3.690655 0.6769648
## Test set      811.3985 1442.204 949.2681  3.296413 3.959407 0.8477600
##                     ACF1 Theil's U
## Training set -0.01066256        NA
## Test set      0.10129766 0.9227046
```

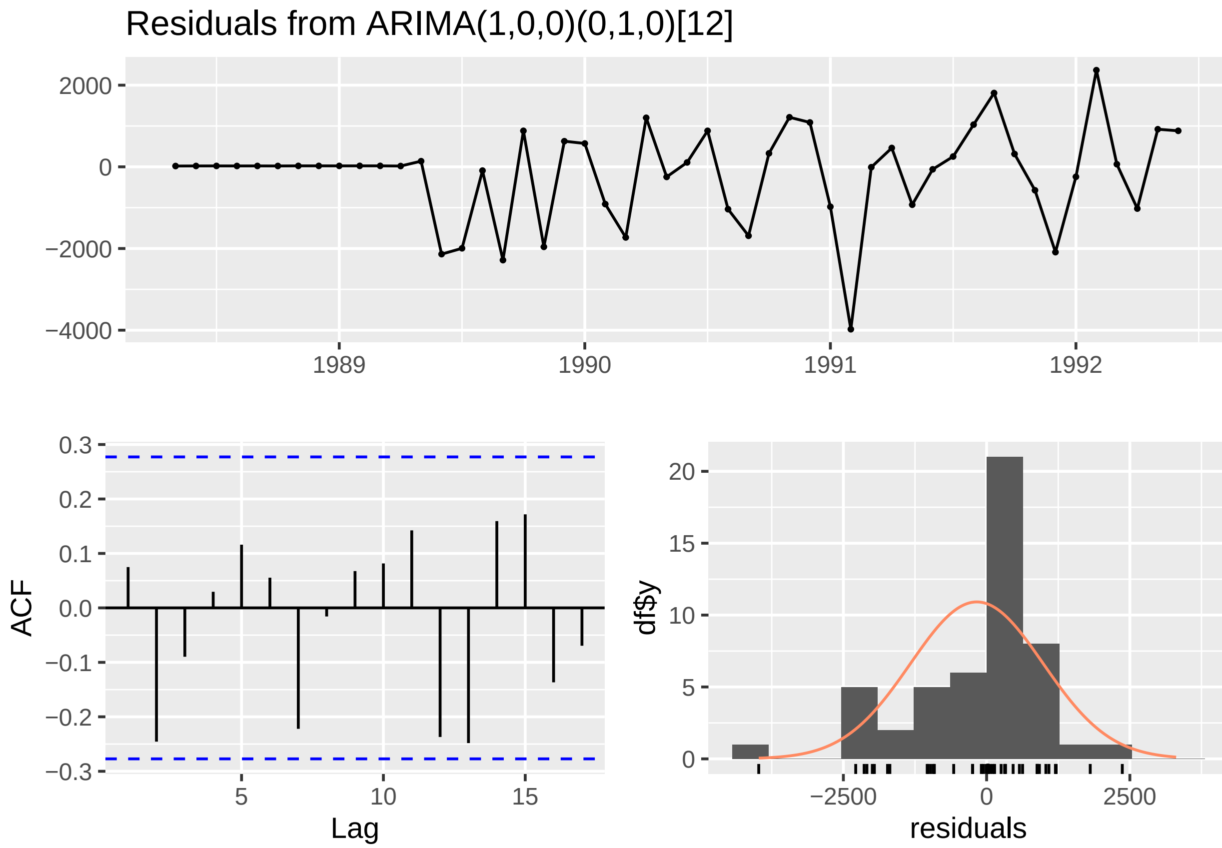
```
accuracy(m6.predict, valid.ts)
```

```
##                     ME      RMSE       MAE       MPE     MAPE      MASE
## Training set  142.5315  958.9937  649.1195 0.5684073 3.150749 0.5797072
## Test set     1555.2995 1801.4593 1555.2995 6.7598073 6.759807 1.3889866
##                   ACF1 Theil's U
## Training set 0.05508401        NA
## Test set     0.11121306  1.203731
```

Propose the best model for the calls volume forecasting. Report its estimated accuracy, comment on its robustness. the best model is SARIMA(1,0,0)x(0,1,0)12, since it has the lowest MAPE.

(4.) Test the best model's assumptions: check whether the residuals are approximately normal and independent. Visual analysis techniques suffice. Include your spript, comments, and all supporting aterial.

```
checkresiduals(m3)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0)(0,1,0)[12]
## Q* = 8.7373, df = 9, p-value = 0.4619
##
## Model df: 1.   Total lags used: 10
```

the residuals are approximately normal and independent.

(5.) Write a short report for Mr. DeCoria summarizing your findings. Provide the best model he should use to forecast the calls volume and the estimated error rate he should expect.

```r
m3f = arima(data.ts, order = c(1,0,0),
           seasonal = list(order= c(0,1,0),
                           period=12))
m3f.predict = forecast(m3f, h=18)
m3f.predict # May 1994 is 19765.59
```

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 1993       23740.26 22039.42 25441.09 21139.06 26341.46
## Feb 1993       21724.39 19883.15 23565.63 18908.46 24540.33
## Mar 1993       21445.09 19580.78 23309.40 18593.87 24296.31
## Apr 1993       19383.89 17515.64 21252.14 16526.65 22241.13
## May 1993       19765.59 17896.66 21634.52 16907.31 22623.87
```

```
## Jun 1993       20292.83 18423.79 22161.87 17434.37 23151.28
## Jul 1993       20801.98 18932.91 22671.04 17943.49 23660.46
## Aug 1993       21128.89 19259.83 22997.96 18270.40 23987.38
## Sep 1993       20252.20 18383.13 22121.27 17393.71 23110.69
## Oct 1993       22069.50 20200.43 23938.57 19211.00 24927.99
## Nov 1993       23268.21 21399.14 25137.27 20409.71 26126.70
## Dec 1993       26039.09 24170.02 27908.15 23180.59 28897.58
## Jan 1994       23740.29 21213.16 26267.42 19875.38 27605.21
## Feb 1994       21724.41 19100.72 24348.10 17711.82 25737.00
## Mar 1994       21445.09 18805.16 24085.03 17407.66 25482.53
## Apr 1994       19383.89 16741.17 22026.61 15342.20 23425.58
## May 1994       19765.59 17122.39 22408.78 15723.17 23808.01
## Jun 1994       20292.83 17649.55 22936.11 16250.28 24335.37
```

The best model is SARIMA(1,0,0)x(0,1,0)12, because it has the lowest error. The forecast for May 1994 is 19765.59.