

# Homework 4

Christian Steinmetz  
MATH 8090-Spring 2018

April 18, 2018

## Problem 6.

Consider the data set of global marine temperatures (yearly data) in `globaltmpmarine.txt` (on blackboard). Find the best fitting time series model using diagnostics, parsimony and selection criteria (like BIC). Consider models with and without trend parameters. Decide if the trend is statistically significant. Use your fitted model to forecast the next three values. Write a short report describing your model fitting experience. Be sure to include

- (a) An overview of your model fitting process
- (b) Any plots you used to decide on a model.
- (c) A list of candidate models considered.
- (d) A summary of results for the various models fit to the data (like a table containing BIC values for example)
- (e) A complete description of your final model.
- (f) A plot showing the time series and three forecasts with approximate 95% prediction intervals.

## Solution

A plot of the original data is shown below in Figure 1. We can argue that there is an upward trend in the data that may be fit with a linear or quadratic trend.

First we fit a linear trend and then plot the ACF and PACF of the residuals which are shown in Figure 2 and Figure 3. From the ACF we see significant autocorrelation at lag 1 and lag 4. The PACF shows no significant autocorrelation for  $h > 1$ . This suggests that an AR(1) or AR(4) with linear trend may be an appropriate model.

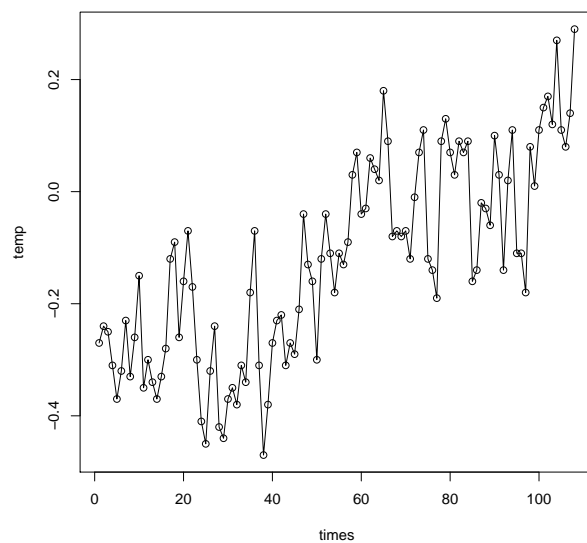


Figure 1: Global martime temperature data.

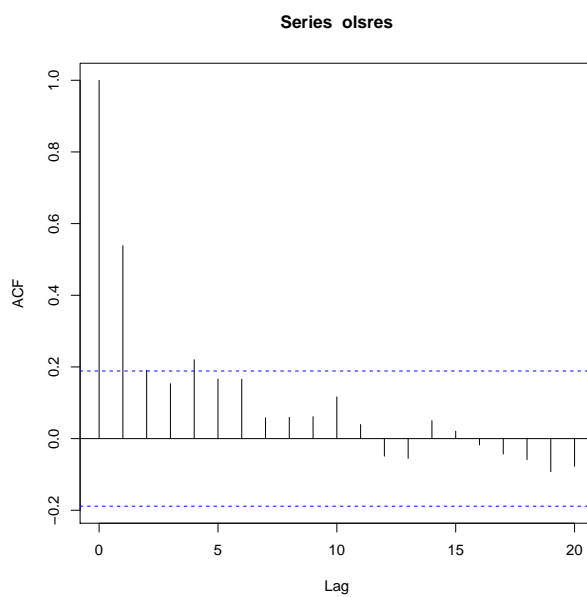


Figure 2: ACF of residuals from linear fit.

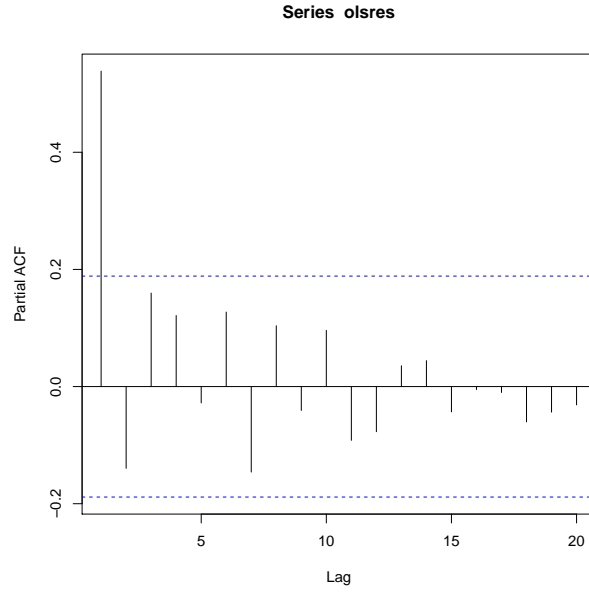


Figure 3: PACF of residuals from linear fit.

Our linear trend with AR(1) errors is shown in Figure 4 and is followed by our linear trend with AR(4) errors shown in Figure 5.

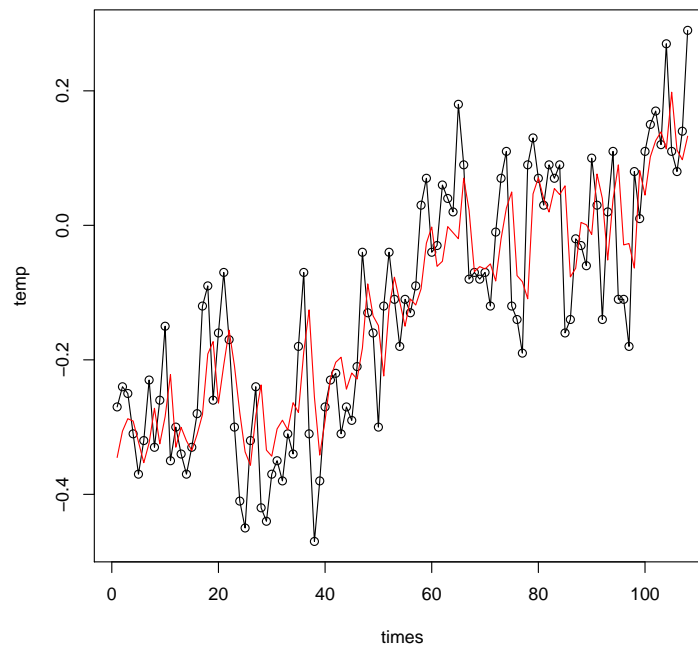


Figure 4: AR(1) fit with linear trend.

```

Coefficients:
      ar1  intercept    times
      0.550      -0.3645  0.0045
s.e.   0.081       0.0391  0.0006

sigma^2 estimated as 0.008728:  log likelihood = 102.6,  aic = -197.2
[BIC] -186.4743

Box-Ljung test

data:  ar1fit$resid
X-squared = 16.053, df = 15, p-value = 0.3785

```

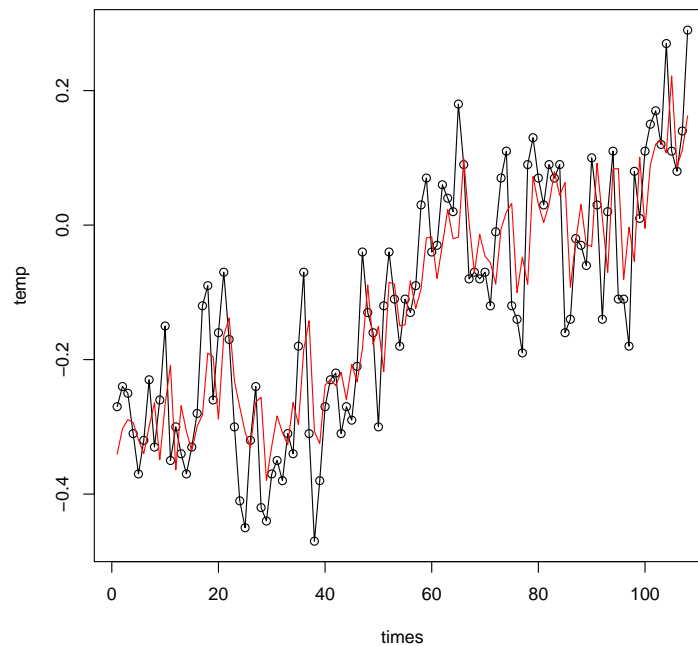


Figure 5: AR(4) fit with linear trend.

```

Coefficients:
      ar1      ar2      ar3      ar4  intercept    times
      0.6342  -0.2199  0.0861  0.1223     -0.3619  0.0044
s.e.   0.0963   0.1134  0.1133  0.0956      0.0440  0.0007

sigma^2 estimated as 0.008174:  log likelihood = 106.05,  aic = -198.1
[BIC] -179.3265

Box-Ljung test

data:  ar4fit$resid
X-squared = 9.2165, df = 15, p-value = 0.8659

```

From the p-values of the Box-Ljung tests we can show that both models are adequate. We also attempt to fit a model without a linear trend using differencing. The results for the ARIMA(1,1,0), ARIMA(4,1,0), and ARIMA(4,1,1) are shown below.

Coefficients:

```
      ar1
      -0.1069
s.e.    0.0966
```

```
sigma^2 estimated as 0.01113:  log likelihood = 88.84,  aic = -173.67
[BIC] -168.3287
```

Box-Ljung test

```
X-squared = 34.807, df = 23, p-value = 0.05436
```

Coefficients:

```
      ar1      ar2      ar3      ar4
      -0.251  -0.4202  -0.2560  -0.0756
s.e.    0.097   0.0965   0.0963   0.0968
```

```
sigma^2 estimated as 0.0091:  log likelihood = 99.36,  aic = -188.71
[BIC] -175.3473
```

Box-Ljung test

```
X-squared = 16.85, df = 23, p-value = 0.8165
```

Coefficients:

```
      ar1      ar2      ar3      ar4      ma1
      0.4958  -0.2587   0.0309   0.0679  -0.7844
s.e.    0.2081   0.1211   0.1329   0.1252   0.1836
```

```
sigma^2 estimated as 0.008849:  log likelihood = 100.77,  aic = -189.54
[BIC] -173.5042
```

Box-Ljung test

```
X-squared = 13.941, df = 23, p-value = 0.9286
```

All four models were found to be adequate and based upon our AIC and BIC values shown in the table below we select the AR(1) model since it had the smallest BIC even though some of the larger models have a higher AIC, although the AIC tends to overfit and does not take into account the number of model parameters.

Model	AIC	BIC
AR(1)	-197.2	-186.4743
AR(4)	-198.1	-179.3265
ARIMA(4,0,1)	-200.8	-179.3456
ARIMA(1,1,0)	-173.67	-168.3287
ARIMA(4,1,0)	-188.71	-175.3473
ARIMA(4,1,1)	-189.54	-173.5042

The AR(1) form is as follows

$$X_t = \mu + \phi_1 X_{t-1} + Z_t,$$

$$\{Z_t\} \sim \text{WN}(0, \sigma^2).$$

The fitted AR(1) with linear regression term is given by

$$X_t = -0.3645 + 0.0045t + 0.550X_{t-1} + Z_t,$$

$$\{Z_t\} \sim \text{WN}(0, \sigma^2).$$

Predicting three time steps ahead produces the following results with the associated 95% confidence intervals shown in Figure 6.

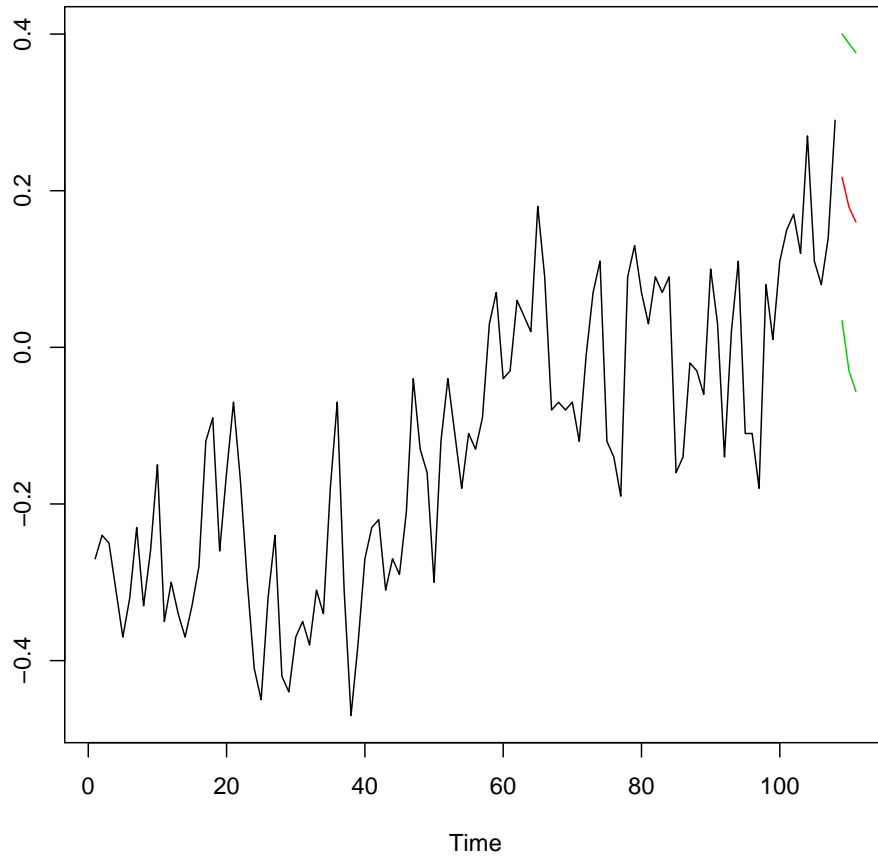


Figure 6: Three step ahead prediction using linear fit with AR(1) errors.

**Problem 8.**

Consider the airline data modeled in class (see R code on Canvas). We will assess the predictive ability of time series models for this data as follows. Use the first ten years of data:  $Y_1, \dots, Y_{120}$  to

- (a) Find the best fitting time series regression model you can find for the  $\log(y)$  values.
- (b) Find the best fitting SARIMA model for the first 10 years of data.
- (c) Use each model to forecast the last year of data and compare your total forecast error for the two models. Which model has smaller error? Which model has smaller AICC?
- (d) Use the fitted model to show how the forecast at time 121 is calculated using each of your model fits.

**Solution**

- (a) We start with the AR(1) model and monthly dummy variables to account for the periodic trend. In addition we fit an AR(2) with the same monthly dummy variables since the ACF plot suggests there may be correlation beyond lag 1. It was found that adding further autoregression parameters did not improve the AIC or BIC values of the model. Using the Box-Ljung test we find no evidence of significant autocorrelation in the residuals of the fitted models so we can say both the AR(1) and AR(2) are adequate. We actually find that the AR(2) fit has a slightly greater AIC and BIC value. There the AR(1) model was chosen since it had both the smallest AIC and BIC values and also employed the least amount of parameters.

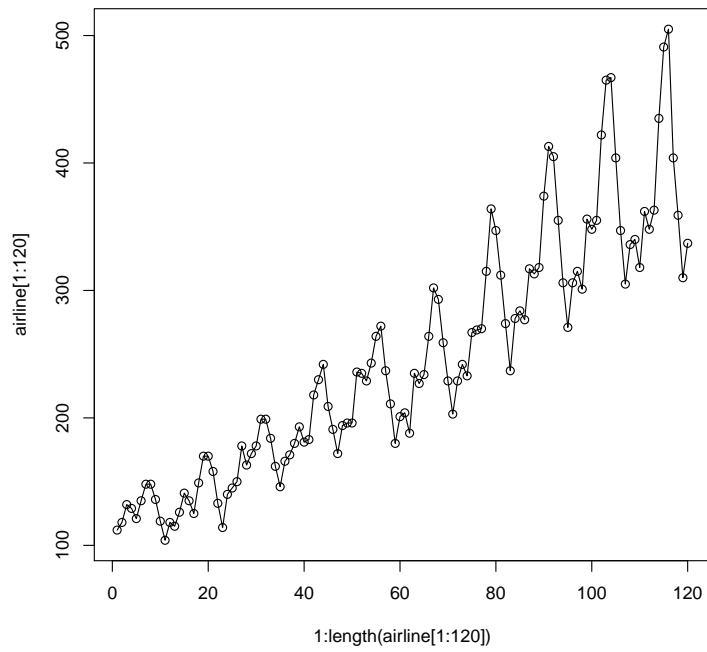


Figure 7: First ten years of monthly airline data.

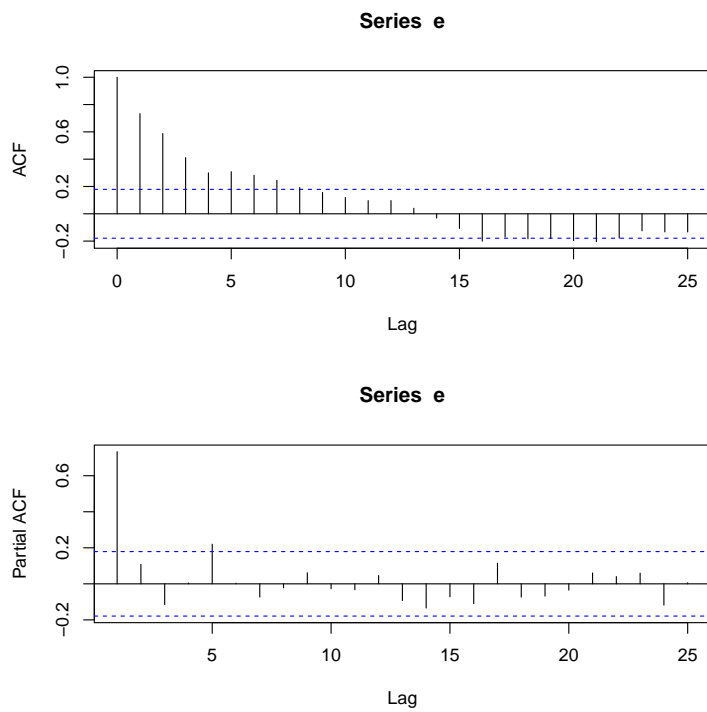


Figure 8: ACF and PACF of regression residuals using monthly dummy variables.



- (b) We improve upon the  $\text{ARIMA}(3, 0, 0) \times (1, 1, 0)_{12}$  model from class by fitting an  $\text{ARIMA}(1, 0, 1) \times (1, 1, 0)_{12}$ . The original model has  $\text{AIC} = -385.57$  and  $\text{BIC} = -369.48$  while the second model has smaller AIC and BIC values of  $\text{AIC} = -387.1$  and  $\text{BIC} = -373.69$ . We also note that with the Box-Ljung test both models produce residuals that appear to have no significant autocorrelation.
- (c) We use both models to predict the next year and then compare these predictions with the actual time series values. The predictions and their associated 95% confidence intervals are shown. To compare the prediction error the mean squared error was calculated for both models. The regression model performed better with an MSE of -633.4 while the SARIMA model had an MSE of 1804.8. This agrees with our AIC and BIC metrics. The regression model had a lower BIC at -397.40 while the SARIMA model had a BIC of -373.7.

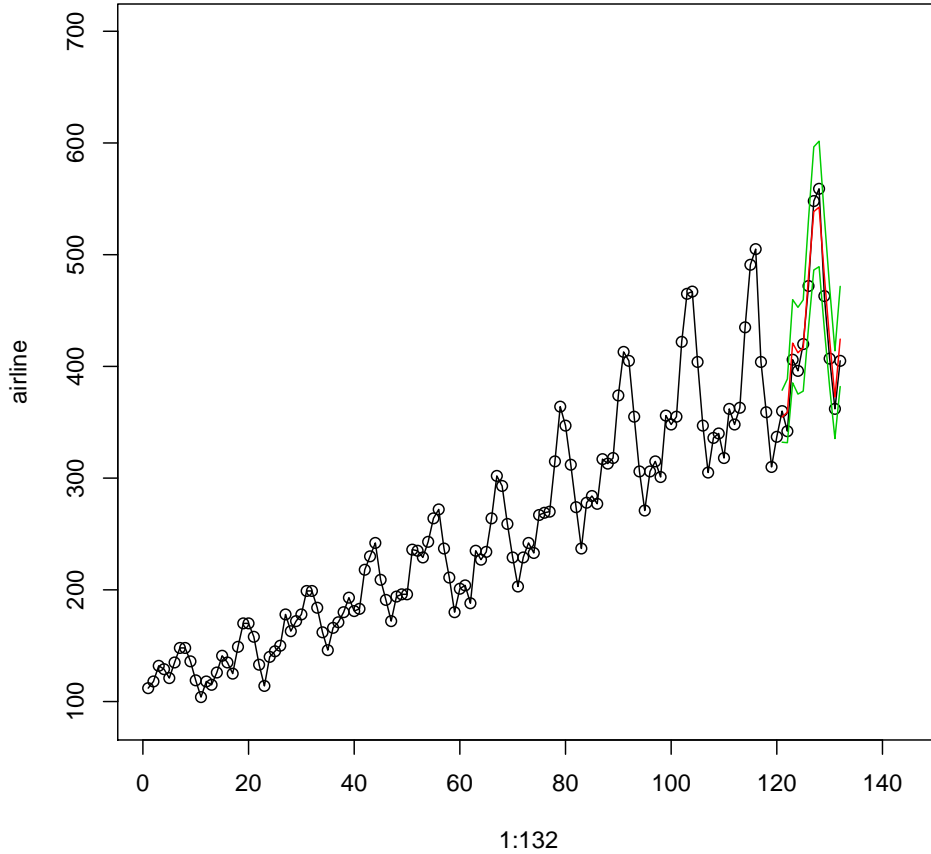


Figure 9: 1 year ahead prediction using the AR(1) regression model.

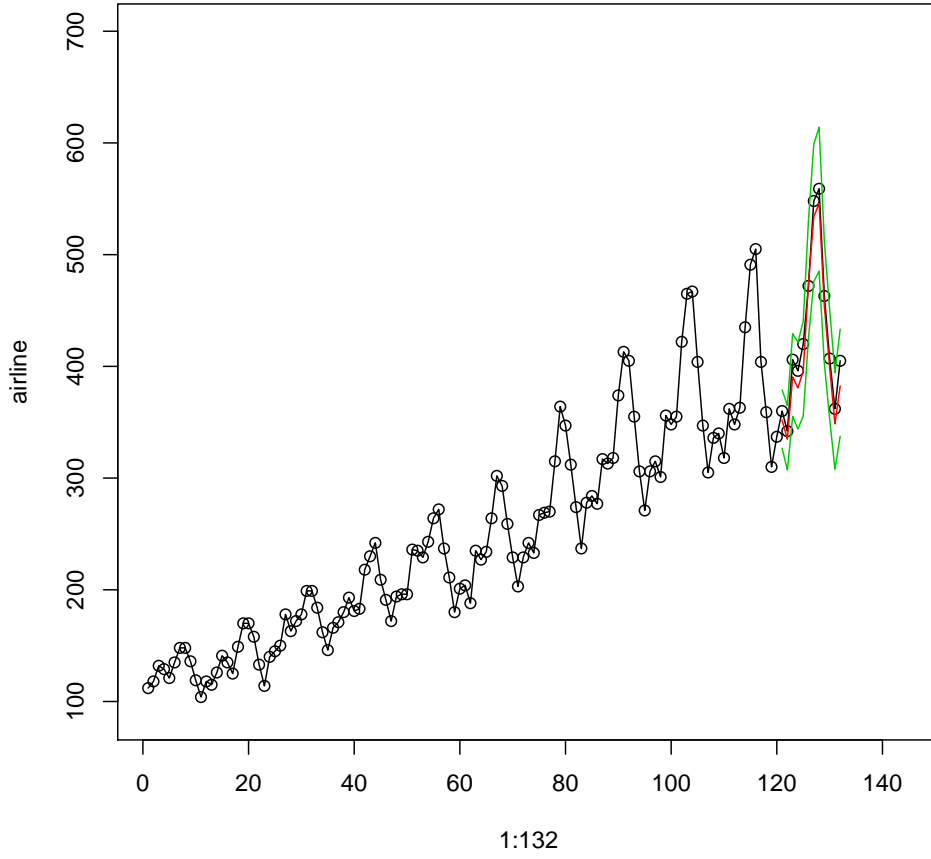


Figure 10: a year ahead prediction using the  $\text{ARIMA}(1, 0, 1) \times (1, 1, 0)_{12}$  model.

(d) Our first model is given by

$$Y_t = \beta_0 + \beta_1 t + \beta_2 M_1 + \beta_3 M_2 + \beta_4 M_3 + \beta_5 M_4 + \beta_6 M_5 + \beta_7 M_6 + \beta_8 M_7 + \beta_9 M_8 + \beta_{10} M_9 + \beta_{11} M_{10} + \epsilon_t.$$

$$\epsilon_t = \phi_1 X_{t-1} + Z_t,$$

$$\{Z_t\} \sim \text{WN}(0, \sigma^2).$$

Using this model to predict  $Y_{121}$  we use the coefficients from the fit below.

```

Coefficients:
      ar1      ar2  intercept    times      M1      M2      M3      M4
0.6825  0.1276    4.6981  0.0103  0.0124  0.0008  0.1363  0.0946
s.e.    0.0918  0.0939    0.0318  0.0004  0.0117  0.0141  0.0158  0.0168

      M5      M6      M7      M8      M9      M10      M11
0.0863  0.2113  0.3085  0.3003  0.1655  0.0245  -0.1174
s.e.    0.0173  0.0175  0.0173  0.0166  0.0155  0.0137  0.0112

sigma^2 estimated as 0.001118:  log likelihood = 237,   aic = -442

```

$$Y_{121} = \beta_0 + \beta_1 121 + \beta_2 M_1 + \beta_3 M_2 + \beta_4 M_3 + \beta_5 M_4 + \beta_6 M_5 + \beta_7 M_6 + \beta_8 M_7 + \beta_9 M_8 + \beta_{10} M_9 + \beta_{11} M_{10} + \phi_1 X_{120}$$