Test Exercise 6

Jason

13/02/2018

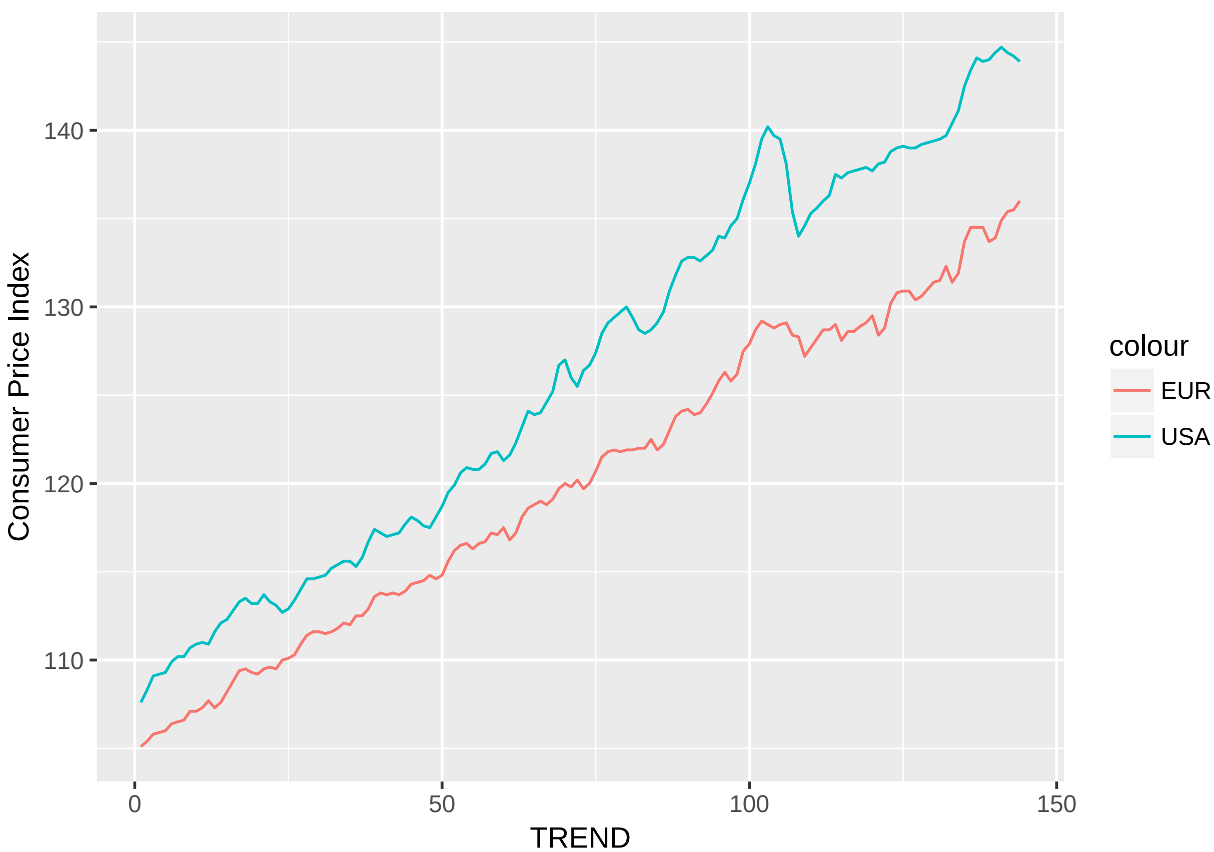
Questions

This test exercise uses data that are available in the data file TestExer6. The question of interest is to model monthly inflation in the Euro area and to investigate whether inflation in the United States of America has predictive power for inflation in the Euro area. Monthly data on the consumer price index (CPI) for the Euro area and the USA are available from January 2000 until December 2011. The data for January 2000 until December 2010 are used for specification and estimation of models, and the data for 2011 are left out for forecast evaluation purposes.

Question (a)

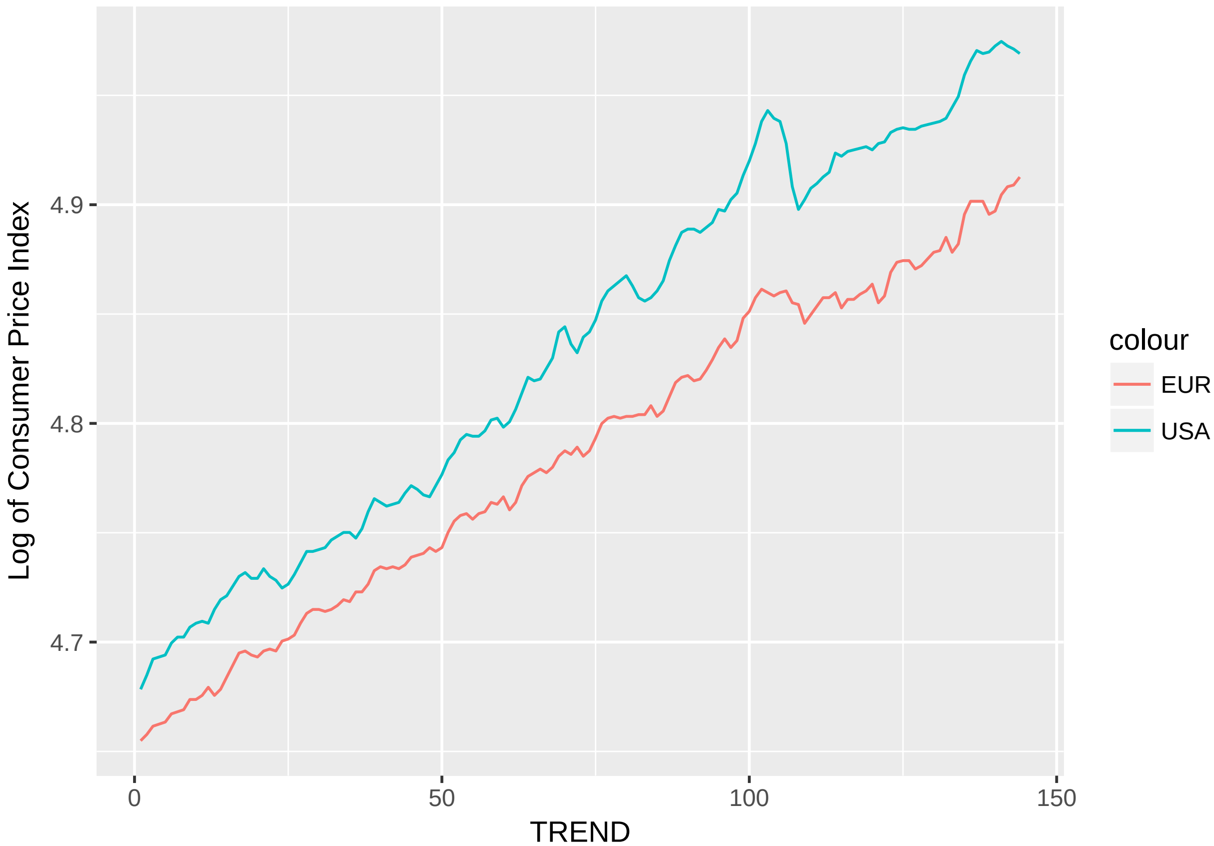
Make time series plots of the CPI of the Euro area and the USA, and also of their logarithm log(CPI) and of the two monthly inflation series DP = log(CPI). What conclusions do you draw from these plots?

```
ggplot(testData) + geom_line(aes(TREND, CPI_EUR, colour='EUR')) + geom_line(aes(TREND, CPI_USA, colour=
```



It can be seen from the graph above that there is a positive trend, where the consumer price index has increased over time, with the exception of some noticable drops. It can also be seen that the USA CPI and Euro CPI are correlated with each other and the USA CPI has always been greater then the EUR CPI value.

```r
ggplot(testData) + geom_line(aes(TREND, LOGPEUR, colour='EUR')) + geom_line(aes(TREND, LOGPUSA, colour=
```

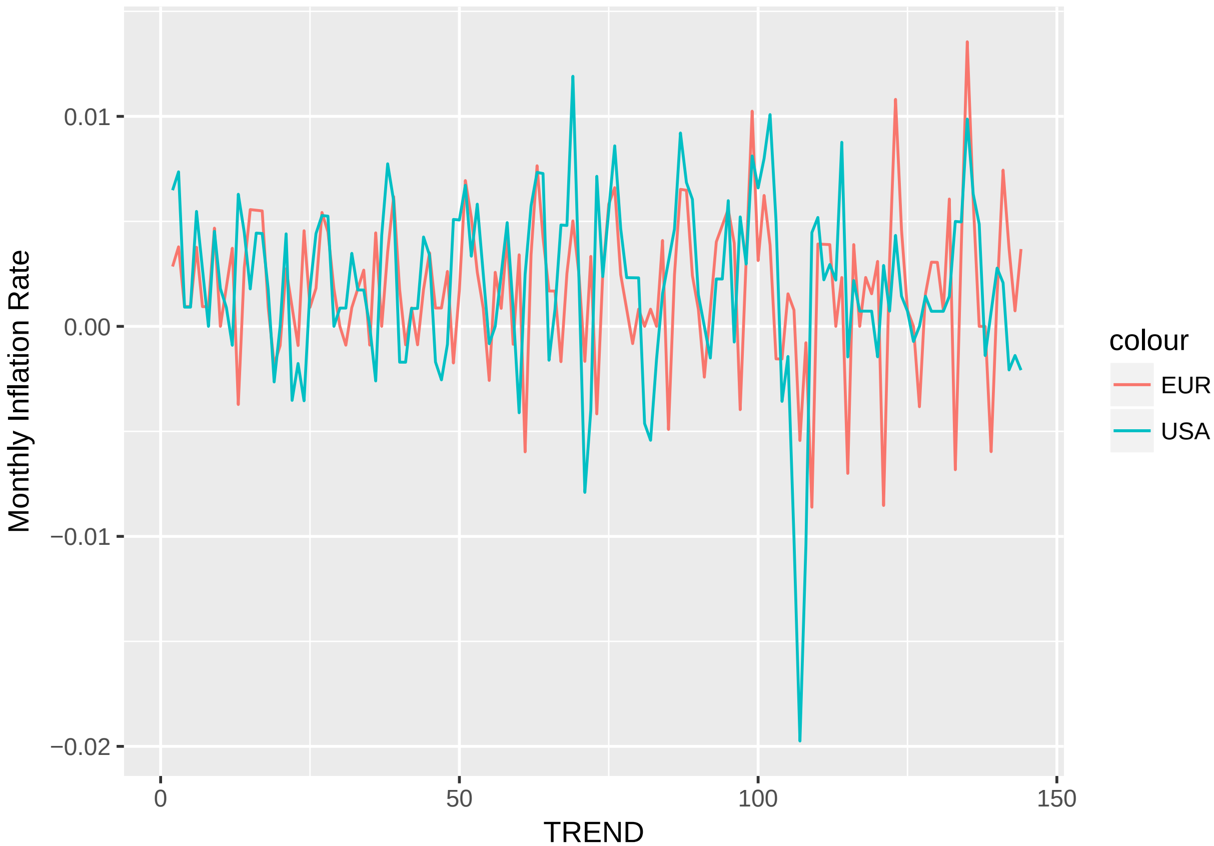


From the graph above, it seems that the second graph exhibits the same charateristics as the first graph.

```r
ggplot(testData) + geom_line(aes(TREND, DPEUR, colour='EUR')) + geom_line(aes(TREND, DPUSA, colour='USA
```

```
## Warning: Removed 1 rows containing missing values (geom_path).

## Warning: Removed 1 rows containing missing values (geom_path).
```



From the graph above, it seems that the inflation rate for both USA and EUR are stationary and correlated. It also looks like the USA inflation rate is slightly more volatile than the EUR inflation rate.

Question (b)

Perform the Augmented Dickey-Fuller (ADF) test for the two log(CPI) series. In the ADF test equation, include a constant (), a deterministic trend term (t), three lags of DP = log(CPI) and, of course, the variable of interest log(CPIt−1). Report the coefficient of log(CPIt−1) and its standard error and t-value, and draw your conclusion.

```
adfEURCPI <- dynlm(DPEUR ~ lag(TREND,0) + lag(DPEUR,1) + lag(DPEUR,2) + lag(DPEUR,3) + lag(LOGPEUR,1), d
summary(adfEURCPI)
```

```
##
## Time series regression with "numeric" data:
## Start = 1, End = 140
##
## Call:
## dynlm(formula = DPEUR ~ lag(TREND, 0) + lag(DPEUR, 1) + lag(DPEUR,
##     2) + lag(DPEUR, 3) + lag(LOGPEUR, 1), data = testData)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0112018 -0.0015085  0.0002827  0.0020131  0.0096450
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)      6.420e-01  2.263e-01   2.837  0.00526 **
## lag(TREND, 0)    2.374e-04  8.496e-05   2.795  0.00596 **
## lag(DPEUR, 1)    1.442e-01  8.665e-02   1.665  0.09833 .
## lag(DPEUR, 2)   -9.022e-02  8.521e-02  -1.059  0.29160
## lag(DPEUR, 3)   -1.128e-01  8.565e-02  -1.317  0.19002
## lag(LOGPEUR, 1) -1.374e-01  4.860e-02  -2.826  0.00543 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00336 on 134 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.1202, Adjusted R-squared:  0.0874
## F-statistic: 3.663 on 5 and 134 DF,  p-value: 0.003875
```

```
adfUSACPI <- dynlm(DPUSA ~ lag(TREND,0) + lag(DPUSA,1) + lag(DPUSA,2) + lag(DPUSA,3) + lag(LOGPUSA,1), 
summary(adfUSACPI)
```

```
##
## Time series regression with "numeric" data:
## Start = 1, End = 140
##
## Call:
## dynlm(formula = DPUSA ~ lag(TREND, 0) + lag(DPUSA, 1) + lag(DPUSA,
##     2) + lag(DPUSA, 3) + lag(LOGPUSA, 1), data = testData)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0131466 -0.0018596 -0.0001258  0.0019564  0.0088758
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.494e-01  1.272e-01   2.747  0.00684 **
## lag(TREND, 0)    1.514e-04  5.723e-05   2.645  0.00914 **
## lag(DPUSA, 1)    6.091e-01  8.404e-02   7.248 3.03e-11 ***
## lag(DPUSA, 2)   -1.513e-01  9.650e-02  -1.567  0.11936
## lag(DPUSA, 3)   -6.450e-03  8.623e-02  -0.075  0.94048
## lag(LOGPUSA, 1) -7.434e-02  2.719e-02  -2.735  0.00709 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.003506 on 134 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.3261, Adjusted R-squared:  0.3009
## F-statistic: 12.97 on 5 and 134 DF,  p-value: 2.721e-10
```

Variable	Coefficient	Std. Error	t-value	p-value
LOGPUSA	-7.434e-02	2.719e-02	-2.735	0.00709
LOGPEUR	-1.374e-01	4.860e-02	-2.826	0.00543

Since t-value for the log(CPI) (-2.826 for EUR and -2.735 for USA) are greater than the critical value which is -3.5, log(CPI) are not stationary.

Question (c)

As the two series of log(CPI) are not cointegrated (you need not check this), we continue by modelling the monthly inflation series DPEUR = log(CPIEUR) for the Euro area. Determine the sample autocorrelations and the sample partial autocorrelations of this series to motivate the use of the following AR model: DPEURt = +1DPEURt−6 +2DPEURt−12 +t. Estimate the parameters of this model (sample Jan 2000 - Dec 2010).

```
arModel <- ar.ols(testData$DPEUR[2:131], order.max = 12)
arModel
```

```
##
## Call:
## ar.ols(x = testData$DPEUR[2:131], order.max = 12)
##
## Coefficients:
##       1        2        3        4        5        6        7        8
##  0.0590   0.0014  -0.0972   0.0082  -0.1393   0.1943  -0.0567  -0.1271
##       9       10       11       12
## -0.0431  -0.1074   0.0602   0.5168
##
## Intercept: -2.316e-05 (0.000223)
##
## Order selected 12  sigma^2 estimated as  5.857e-06
```

The AR Model is shown below:

$$DPEUR_t = 0.000223 + 0.1943 DPEUR_{t-6} + 0.5168 DPEUR_{t-12}$$

Question (d)

Extend the AR model of part (c) by adding lagged values of monthly inflation in the USA at lags 1, 6, and 12. Check that the coefficient at lag 6 is not significant, and estimate the ADL model DPEURt = + 1DPEURt−6 + 2DPEURt−12 + 1DPUSAt−1 + 2DPUSAt−12 + t (sample Jan 2000 - Dec 2010).

```
testData2 <- testData[2:132,]
modelADL <- dynlm(DPEUR ~ lag(DPEUR, 6) + lag(DPEUR, 12) + lag(DPUSA, 1) + lag(DPUSA, 6) + lag(DPUSA, 1
summary(modelADL)
```

```
##
## Time series regression with "numeric" data:
## Start = 1, End = 119
##
## Call:
## dynlm(formula = DPEUR ~ lag(DPEUR, 6) + lag(DPEUR, 12) + lag(DPUSA,
##     1) + lag(DPUSA, 6) + lag(DPUSA, 12), data = testData2)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0065866 -0.0016535 -0.0000118  0.0012630  0.0082682
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.0004407  0.0002853   1.545    0.125
## lag(DPEUR, 6)   0.2029891  0.0785520   2.584    0.011 *
```

```
## lag(DPEUR, 12)  0.6367464  0.0874766   7.279 4.78e-11 ***
## lag(DPUSA, 1)   0.2264287  0.0511286   4.429 2.20e-05 ***
## lag(DPUSA, 6)  -0.0560565  0.0547645  -1.024    0.308
## lag(DPUSA, 12) -0.2300418  0.0541695  -4.247 4.47e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002272 on 113 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.5602, Adjusted R-squared:  0.5408
## F-statistic: 28.79 on 5 and 113 DF,  p-value: < 2.2e-16
```

```r
modelADL2 <- dynlm(DPEUR ~ lag(DPEUR, 6) + lag(DPEUR, 12) + lag(DPUSA, 1) + lag(DPUSA, 12), data = testD
summary(modelADL2)
```

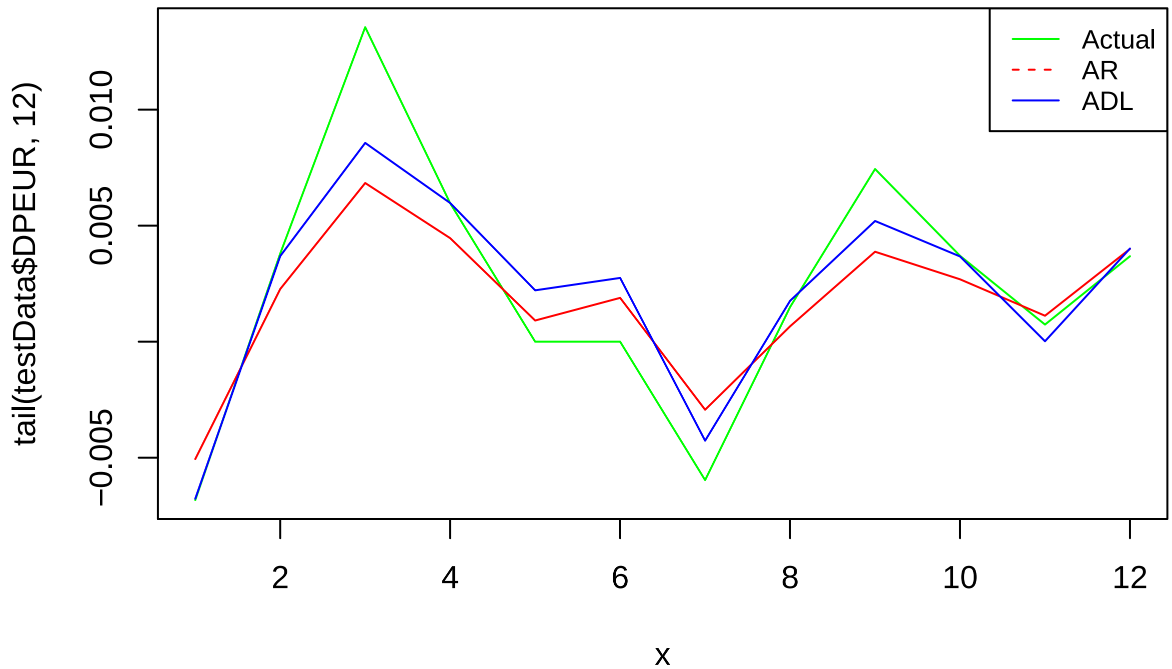
```
##
## Time series regression with "numeric" data:
## Start = 1, End = 119
##
## Call:
## dynlm(formula = DPEUR ~ lag(DPEUR, 6) + lag(DPEUR, 12) + lag(DPUSA,
##     1) + lag(DPUSA, 12), data = testData2)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0067809 -0.0016356  0.0000532  0.0013660  0.0082448
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.0003391  0.0002676   1.267   0.2076
## lag(DPEUR, 6)   0.1687310  0.0710801   2.374   0.0193 *
## lag(DPEUR, 12)  0.6551529  0.0856263   7.651 6.93e-12 ***
## lag(DPUSA, 1)   0.2326460  0.0507772   4.582 1.19e-05 ***
## lag(DPUSA, 12) -0.2264880  0.0540694  -4.189 5.55e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002273 on 114 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.5561, Adjusted R-squared:  0.5406
## F-statistic: 35.71 on 4 and 114 DF,  p-value: < 2.2e-16
```

The ADL Model is shown below:

$$DPEUR_t = 0.0003391 + 0.1687310 DPEUR_{t-6} + 0.6551529 DPEUR_{t-12} + 0.2326460 DPUSA_{t-1} - 0.2264880 DPUSA_{t-12}$$

Question (e)

Use the models of parts (c) and (d) to make two series of 12 monthly inflation forecasts for 2011. At each month, you should use the data that are then available, for example, to forecast inflation for September 2011 you can use the data up to and including August 2011. However, do not re-estimate the model and use the coefficients as obtained in parts (c) and (d). For each of the two forecast series, compute the values of the root mean squared error (RMSE), mean absolute error (MAE), and the sum of the forecast errors (SUM). Finally, give your interpretation of the outcomes.




```r
# Function that returns Root Mean Squared Error
rmse <- function(error)
{
    sqrt(mean(error^2))
}

# Function that returns Mean Absolute Error
mae <- function(error)
{
    mean(abs(error))
}

sumError <- function(error)
{
    sum(error)
}

actual <- tail(testData$DPEUR,12)
predictedAR <- adfPre$pred

errorAR <- actual - predictedAR

rmseAR <- rmse(errorAR)
maeAR <- mae(errorAR)
sumErrorAR <- sumError(errorAR)

predictedADL <- tail(predict(modelADL3, n.ahead=12),12)

errorADL <- actual - predictedADL

rmseADL <- rmse(errorADL)
maeADL <- mae(errorADL)
```

```r
sumErrorADL <- sumError(errorADL)
```

Model	RMSE	MAE	Sum of Error
AR Model	0.002599311	0.001954232	0.00685721
ADL Model	0.001956689	0.00128424	0.000749603

It can be seem from the table above that the ADL model performs much better then the AR model.