Time series analysis on unemployment data in Yolo County

Qianhai Zhang 912528049
Hong Fan 912524085

Abstract

Unemployment rate is a very important measure of economic conditions. During economic recessions, unemployment rates are generally very high. Using time series analysis to make predictions based on the previous data, government can make use of it in the process of designing new policies. The predictions can also serve as a reference for people who are actively seeking jobs. The goal of our analysis for using the unemployment data in Yolo County is to get some hints about the employment conditions in our local area.

1. Introduction

In this project, we studied unemployment data in Yolo County from 1990 to 2014. The reason for choosing unemployment data is that it reflects economic conditions to some extent and corresponding government policies. Predictions not only play an important role in policy-designing processes but also can offer assistance to people who are seeking jobs. We used time series approach to analyze the unemployment data from three parts: seasonality component, trend component and residual component. For the first part, we used moving average to separate the seasonality. For the second part, we tried both polynomial approach and one-sided moving average method to fit the trend. After deseasonalizing and detrending the data, we fit ARMA models to the stationary residuals. Finally, we made predictions based on the models we selected.

2. Data description

The unemployment data we used came from Federal Economic Data (*Economic Research*). Unemployment rate in Yolo County was recorded monthly and the data we used were collected from 1990 January 1st to 2014 December 31th (*Economic Research*). There are no missing values in our data. Each year we have complete 12 records and thus 300 records in total. By simply looking at the data, we can see that unemployment data in each year have very similar patterns: it went down first, went back slightly, decreased again and then kept increasing until the end of the year. The overall data show the unemployment rate is highest in Yolo County in the beginning of 2011 and lowest in the end of 1999.

3. Data Analysis

3.1 Initial data transformation

The data is splitted into two parts: the first part includes the first 14 years data and they are used to conduct time series analysis, the second part containing the last year's data are used to check our predictions.

Before analyzing the data, we use lambda = 0.067 (suggested by Boxcox) to stable the variability throughout the years (Figure A.1 in appendix).

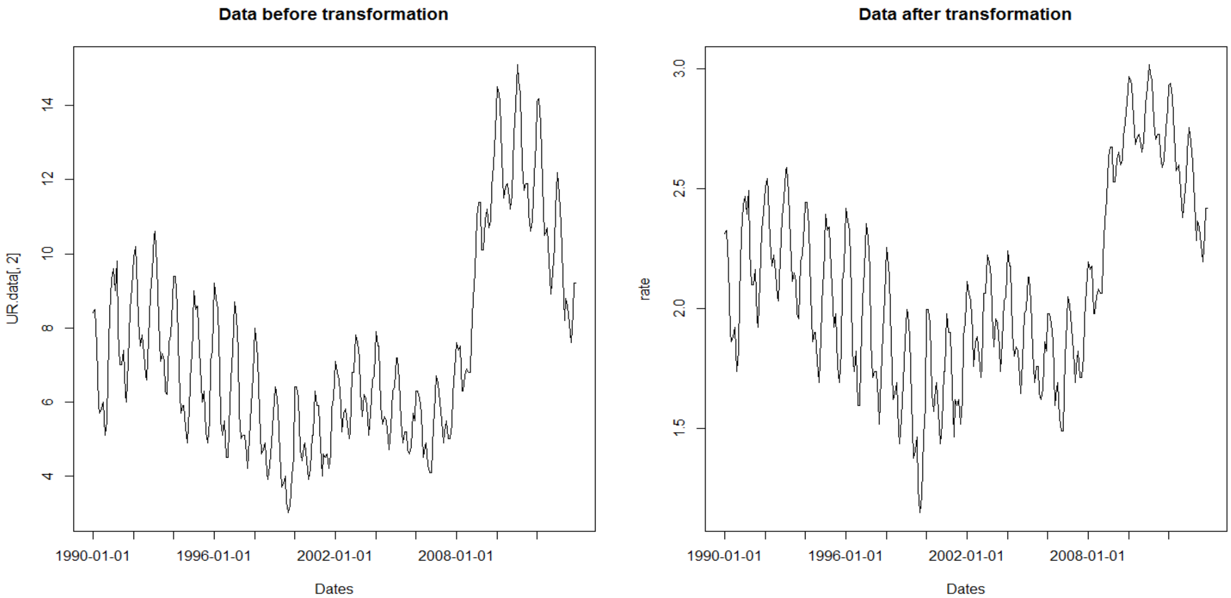

Figure 3.1: Time series plots of the unemployment rate in Yolo County from January 1990 to December 2014 (left) and its transformation (right).

The classical approach taken in time series analysis is based on the postulation that the stochastic process can be analyzed by three components: seasonal part, trend part and a centered random part. The corresponding model:

$$X_t = m_t + s_t + Y_t, \qquad t \in T$$

where mt denotes the trend function, st denotes the seasonality and Yt denotes a zero mean stochastic process.

3.2 Eliminating Seasonal Component:

We use moving average estimation to separate seasonality and analyze it. In the first step, two sided moving average is used to estimate trend. The unemployment data we used has period d = 12, so in this procedure q = 6, n =24.

Two-sided moving average is revised when d is even:

$$\hat{m}_t = \frac{1}{d}(.5x_{t-q} + x_{t-q+1} + \ldots + x_{t+q-1} + .5x_{t+q}), \qquad t = q+1, \ldots, n-q.$$

In the second step, we estimate the seasonality from the detrended data and and then we get deseasonalized data which are used to reestimate the trend in next step.

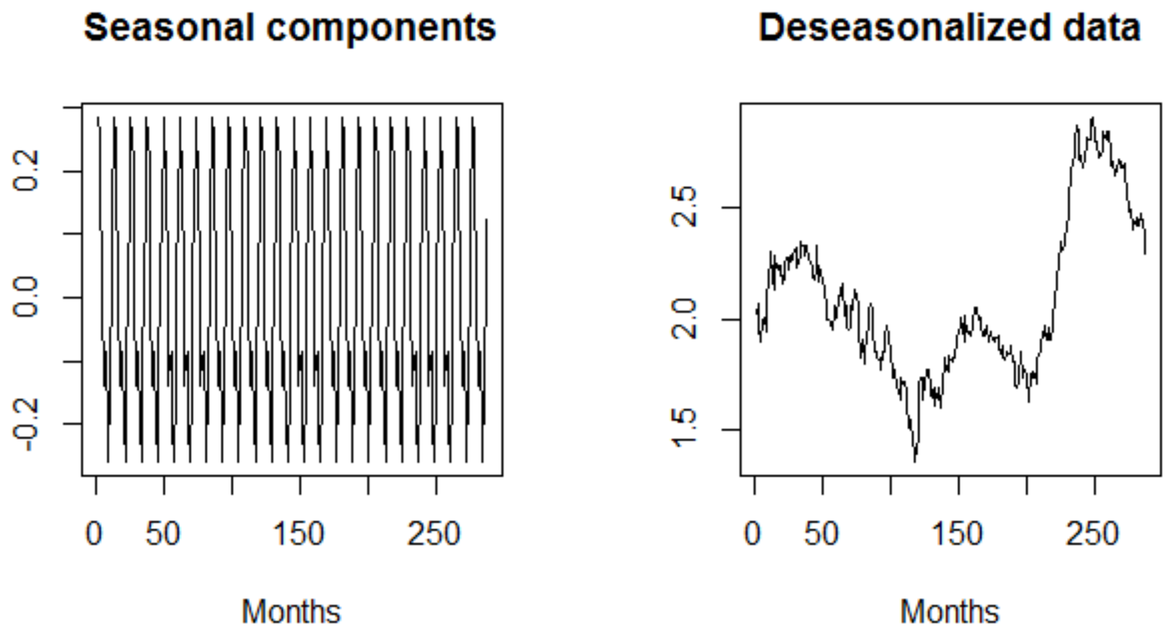

Figure 3.2: seasonality (left) and deseasonalized data (right).

3.3 Eliminating Trend Component

In this part, we use two ways to estimate the trend.

I. Polynomial Estimation

In this process, we fitted polynomials to the deseasonalized data and used adjusted R squared, AIC and BIC as criterion of model selection. Figure A.2 and A.3 clearly show that after polynomial power p = 11 adjusted R squared, AIC and BIC becomes very stable. In order to get better predictions and avoid overfitting, we plotted deseasonalized data and different polynomial fittings as shown below in Figure 1.3.

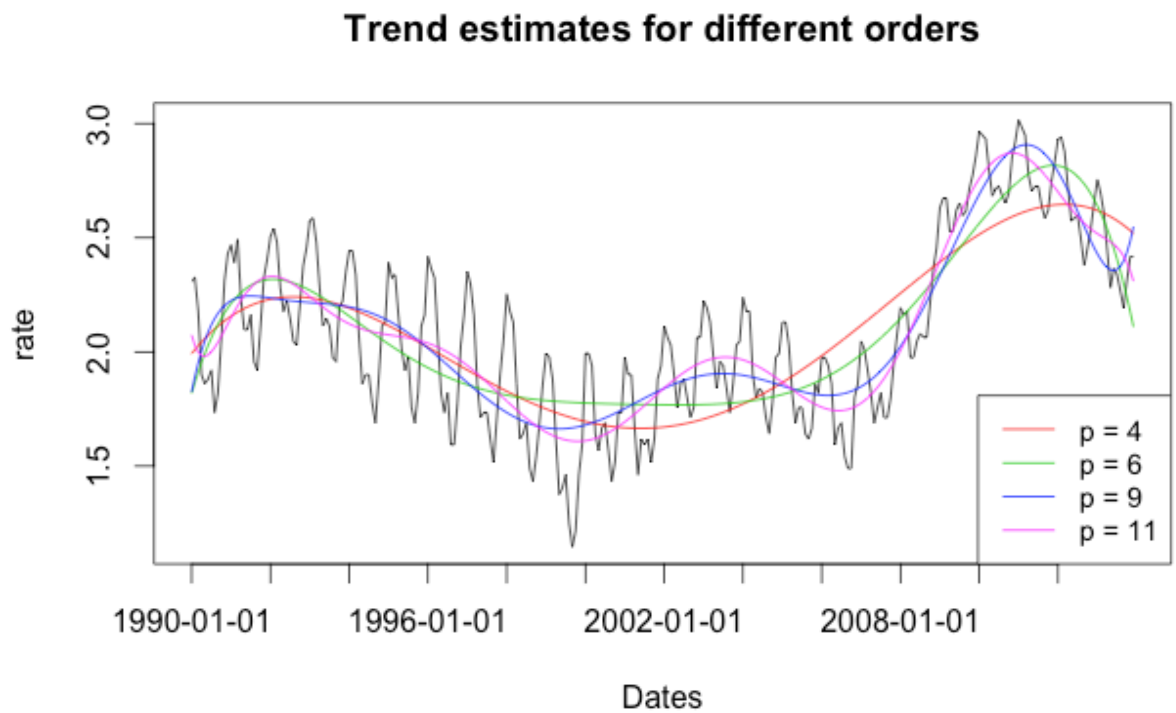

Figure 3.3 Polynomial trend estimates when p = 4, 6, 9, 11.

It is clear to see that in the end of the deseasonalized data, there is a sharp increase followed by some leveling off. After plotting different polynomials, the fourth-order polynomial look reasonable towards the end of our observation period. Since the increase of adjusted R square after order 3 or 4 are very minimal, we chose order 4 for the purpose of forecast.

Using 4th order to detrend our deseasonalized data, the variance of residuals are not stable (Figure A.4), thus we used one-sided moving average to detrend.
For the one-sided moving average procedure, which is defined:

$$\hat{m}_1 = X_1, \qquad \hat{m}_t = aX_t + (1-a)\hat{m}_{t-1}, \quad t = 2, \ldots, n.$$

Figure A.5 shows 9 residual plots in which a ranges from a = 0.1 to 0.9. We chose a = 0.5 as the weight parameter in one-sided moving average since the residuals look stationary. The residuals and corresponding plots are discussed in Assessing Residuals part below.

II. One-sided Moving Average

Our second approach is to apply one-sided moving average directly. Setting a = 0.1, 0.2, 0.3,..., 0.9, we compared corresponding residual plots to find a. Starting from a = 0.1, the residuals become relatively stationary when a = 0.5. It can be seen that when a = 0.5, there is no obvious trend and residuals roughly spread around zero across years (Figure A.6).

3.4 Assessing Residuals

After deseasonalizing and detrending our data, we got residuals from two approaches and we made a comparison between these two:

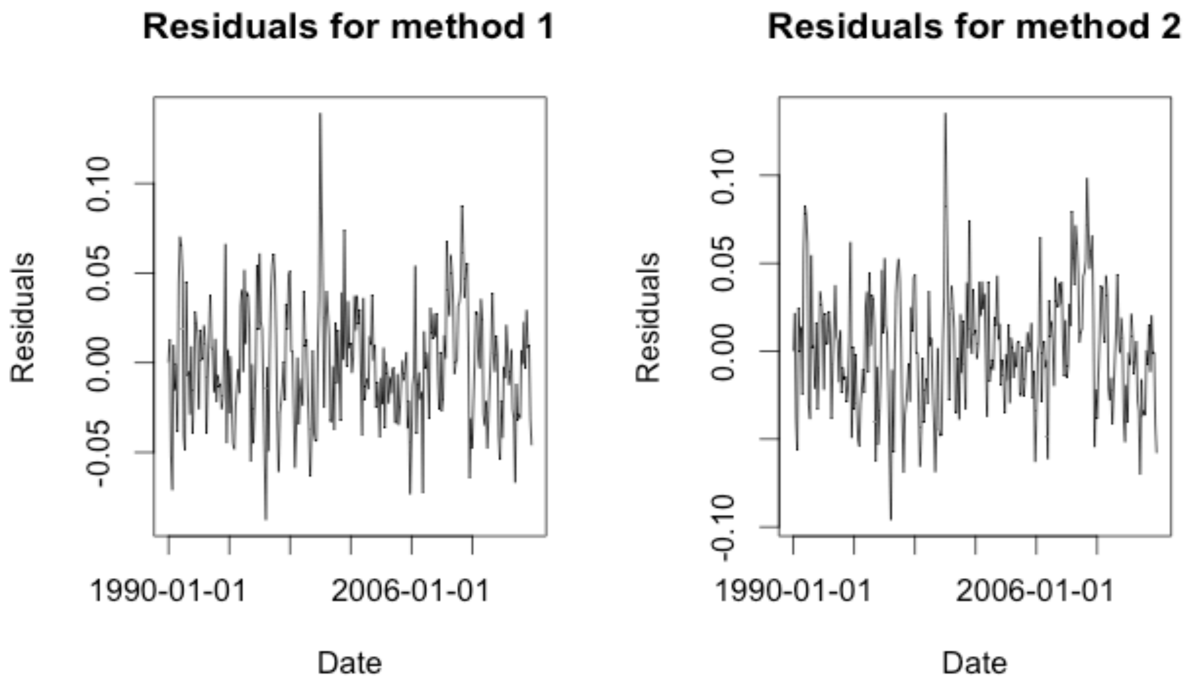

Figure 3.4 Residuals from Polynomial estimation(left) and One-sided moving average estimation(right).

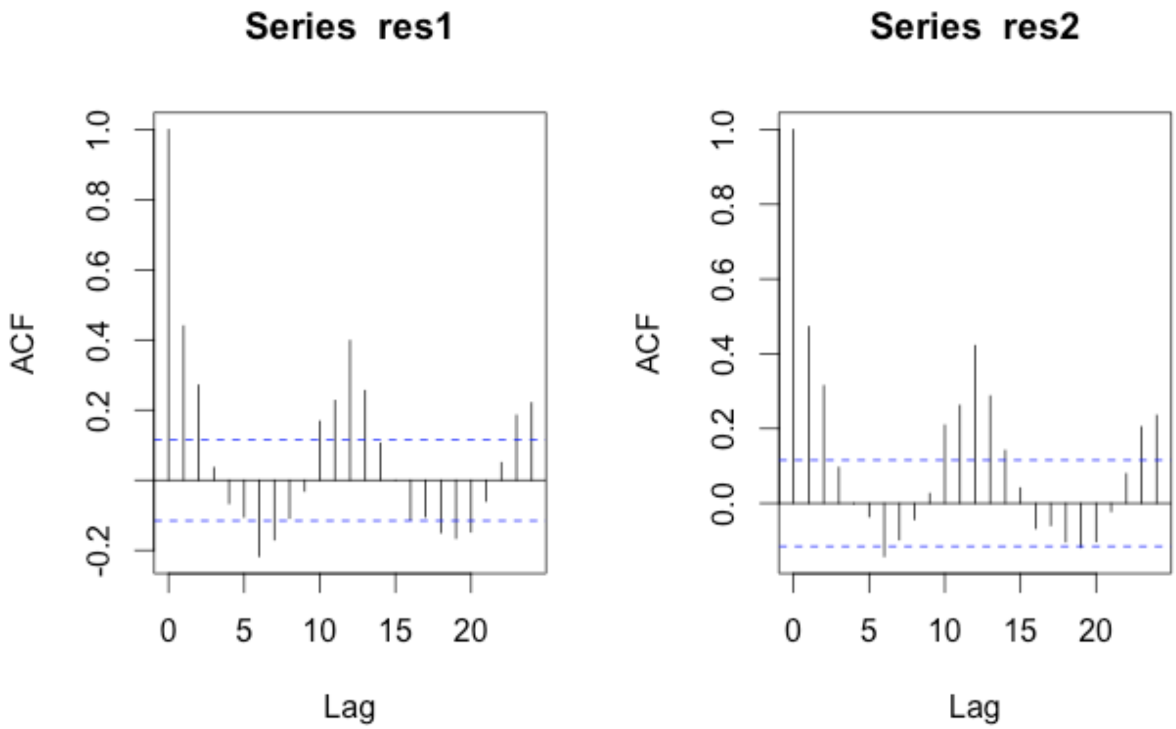

Figure 3.5 The ACFs of Residuals from Polynomial estimation(left) and One-sided moving average estimation(right).

Both Figure 1.4 and Figure 1.5 show residuals getting from these two methods are not very different. Therefore, we applied goodness-of-fit tests to analyze the residuals and conduct forecast if they are stationary.

The sample ACF:
Setting lag = 20, 9/19 of the sample ACFs from the first approach are within the bounds, 10/19 of the sample ACFs from the second approach are within the bounds. The results show that the residuals are not independently, identically distributed.

The Portmanteau test:
All the residuals do not pass Portmanteau test and we have evidence to accept that the residuals are not independent, identically distributed. Thus we can fit stationary ARMA model to our residuals to conduct forecast.

Rank test:
The residuals do not pass rank test, indicating no linear trends exist in the two sets of residuals.

The residual plot shows stationary features and the tests above support that our residuals are not white noise. Residual lag plots (Figure A.6 and Figure A.7), ACF (Figure 1.5) and PACF plots (Figure A.8) do not clearly suggest which ARMA model is appropriate. Thus we fitted different ARMA models using Maximum Likelihood method and compared their AIC values.

We first used ARMA model to fit residual from Polynomial estimation:

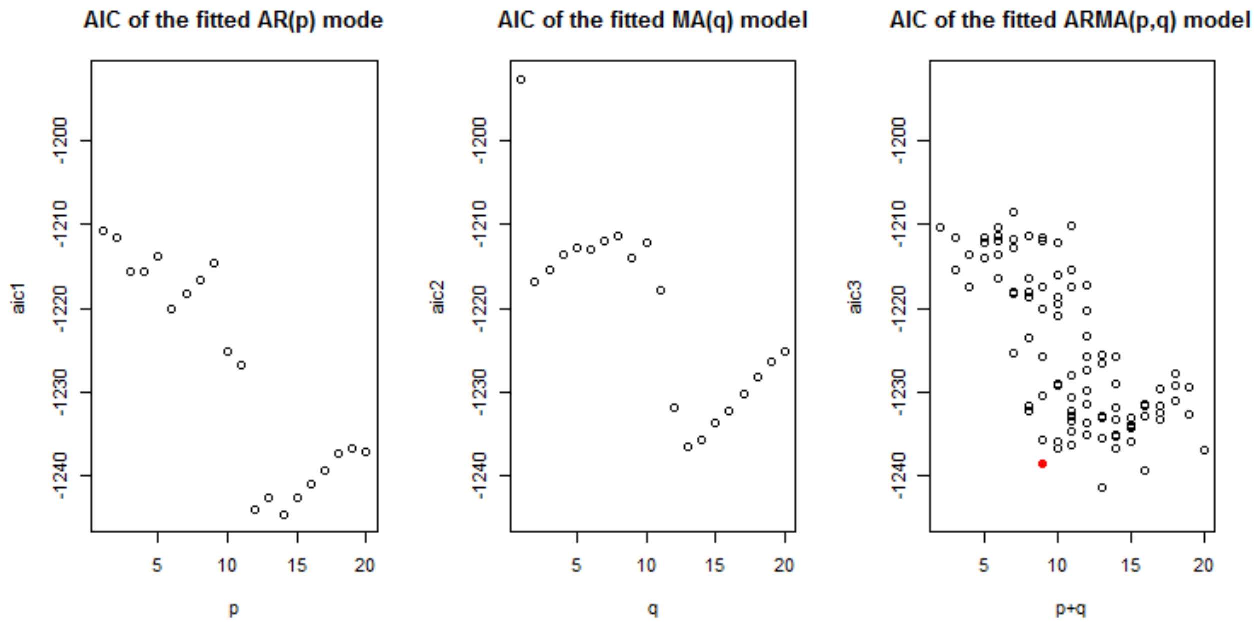

Figure 3.6 AR(left), MA(middle) and ARMA(right) models for from Polynomial estimation.

Since AR suggests p= 12 and MA models suggest q =13 which are very high, we selected ARMA(5,4) (red dot in Figure 3.6) to further conduct residual tests.

We then fit models to residual from One-sided moving average estimation

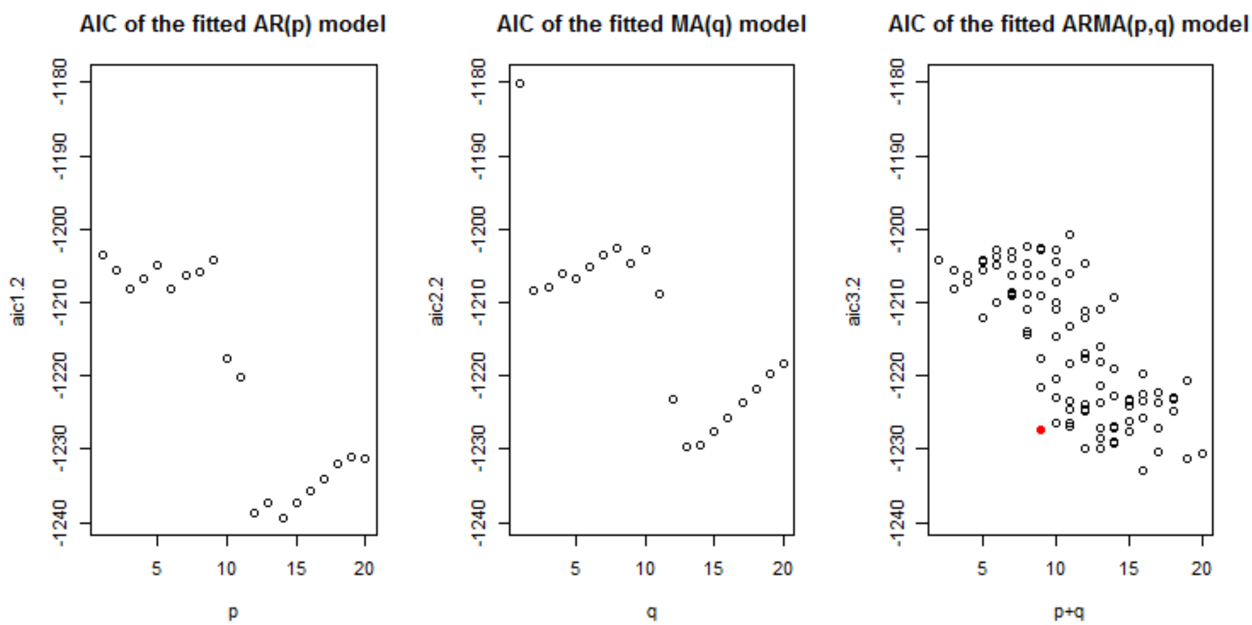

Figure 3.7 AR(left), MA(middle) and ARMA(right) models for residual from One-sided moving average estimation.

Similar to the previous case, we selected ARMA(5,4) (red dot in Figure 1.7) to fit residual 2.

After fitting ARMA(5,4), we had new residual acf plot:

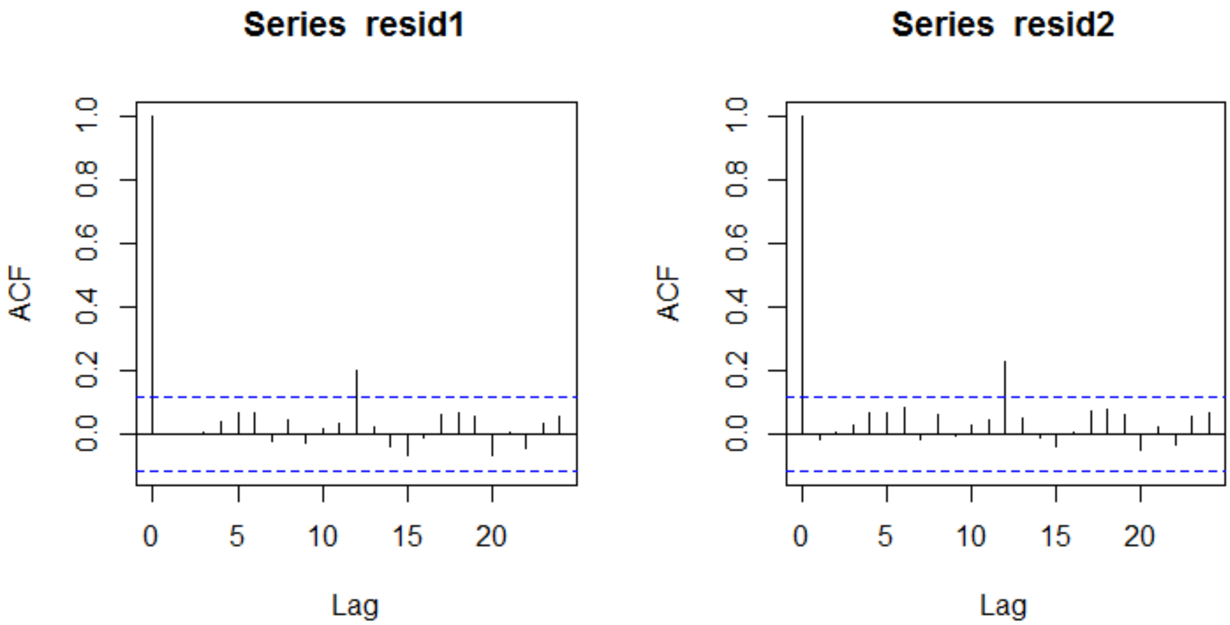

Figure 3.8 Residuals after fitting ARMA(5,4) from residual from Polynomial estimation(left) and One-sided moving average estimation(right).

The Figure above show that almost all the ACF values are within the bounds. Portmanteau Test is used to further check whether the two sets of new residuals for the ARMA model are independently, identically distributed. This test shows our hypothesis that residuals are independent, identically distributed is accepted at level 0.05. This test supports our ARMA(5,4) model for residual 1 and residual 2. The Rank test shows no more trends in the two sets of residuals.

3.5 Prediction

The prediction of the time series is given by the sum of predictions of the trend, seasonality component and the residual component.

3.5.1 Seasonality

In the time series model, we have assumptions on the seasonal component:

$$s_{t+d} = s_t, \qquad \sum_{j=1}^{d} s_j = 0$$

From our estimation of the seasonality, each year we have

```
[1]  0.284191445  0.254641349  0.207832864  0.004650456 -0.140492671 -0.095147173
[7] -0.084009931 -0.206255189 -0.258174180 -0.141000319  0.049960672  0.123802676
```

3.5.2 Trend

Since the trend in our data is not stable (Figure 1.9), we decided to use data in 2013 (12 values) to predict unemployment rate on January in 2014, then used previous 12 points to make January's prediction, and every prediction is based on the previous points. The method we used here is fitting polynomials to the previous 12 points and used adjusted R squared to select the best model.

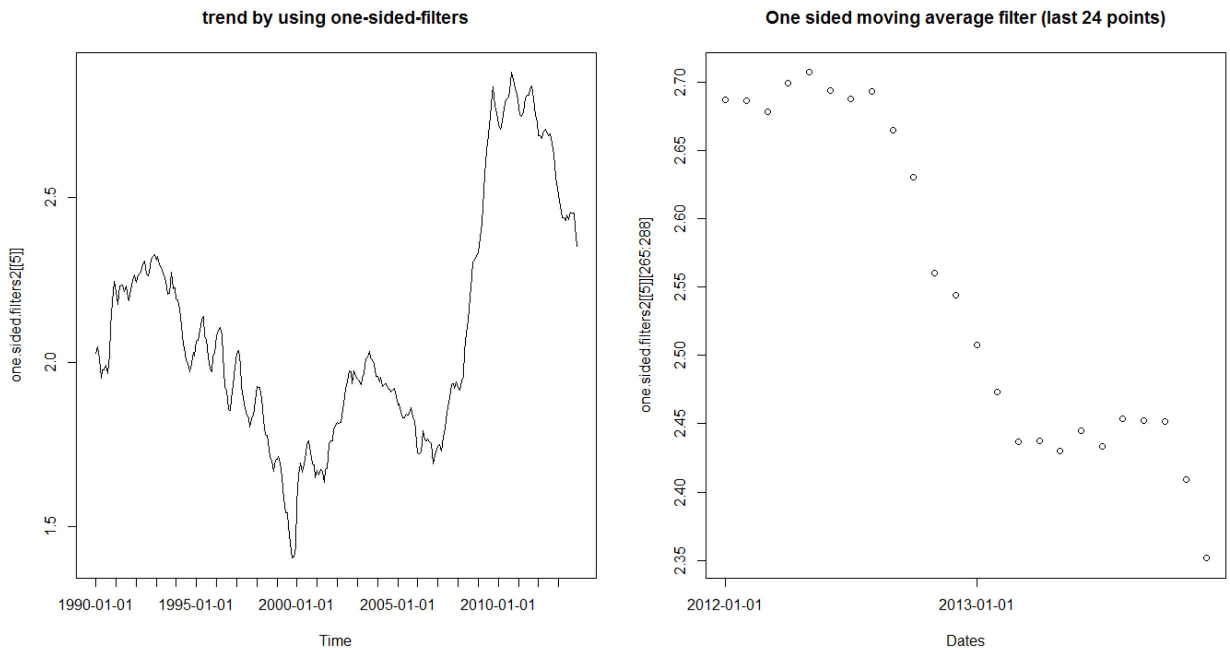

Figure 3.9 Overall trend (left) and trends in 2012 & 2013 years (right).

According to adjusted R squared (Figure A.10) , our first three predictions are based on the 3-order polynomials fitted to their previous 12 points, the rest nine predictions are based on the 2-order polynomials fitted to their previous 12 points. Figure 3.10 shows the predictions of the trends in 2014.

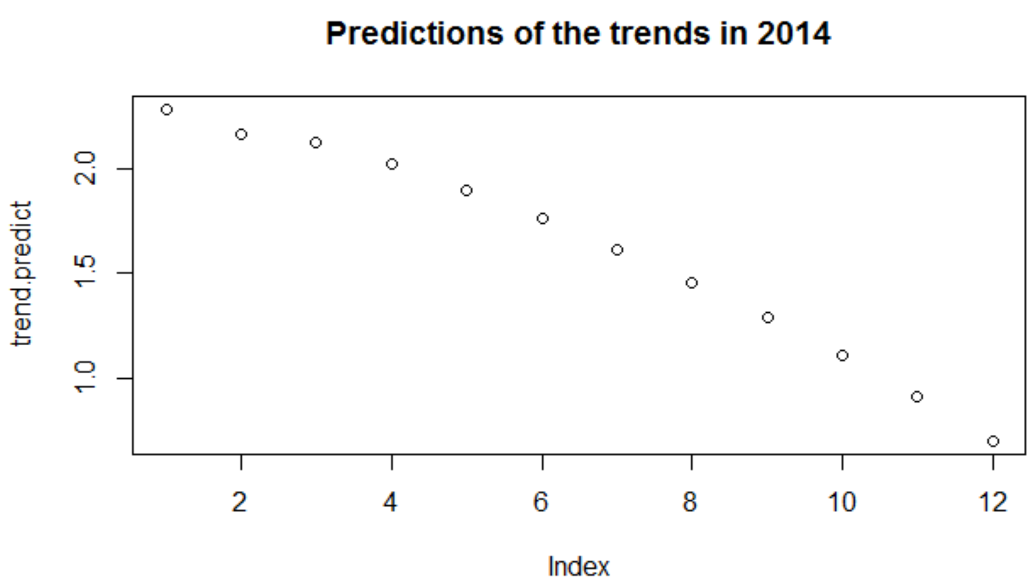

Figure 3.10 Prediction of the trends in 2014

3.5.3 Residuals

Residuals are predicted by ARMA(5,4) model.

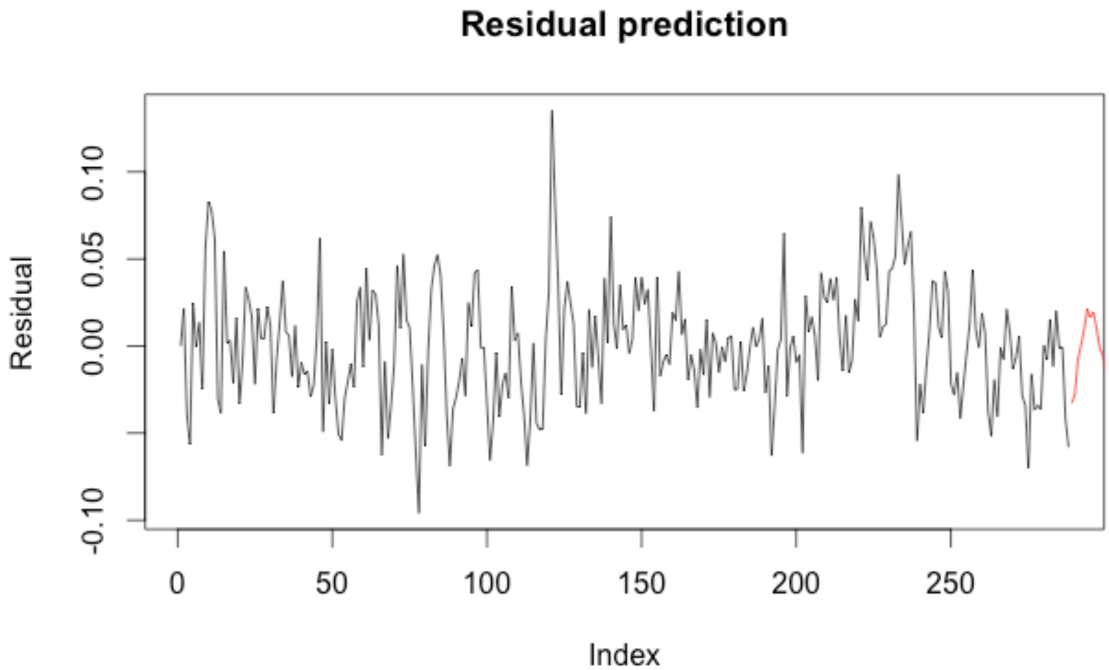

Figure3.11: Prediction of residuals in 2014(the red line) using ARMA(5,4)

The final predictions in 2014 includes three components: trend prediction, seasonality prediction and residual prediction. After transforming our predictions to the previous scale, we plotted our predictions and made comparisons with the actual unemployment rates in 2014.

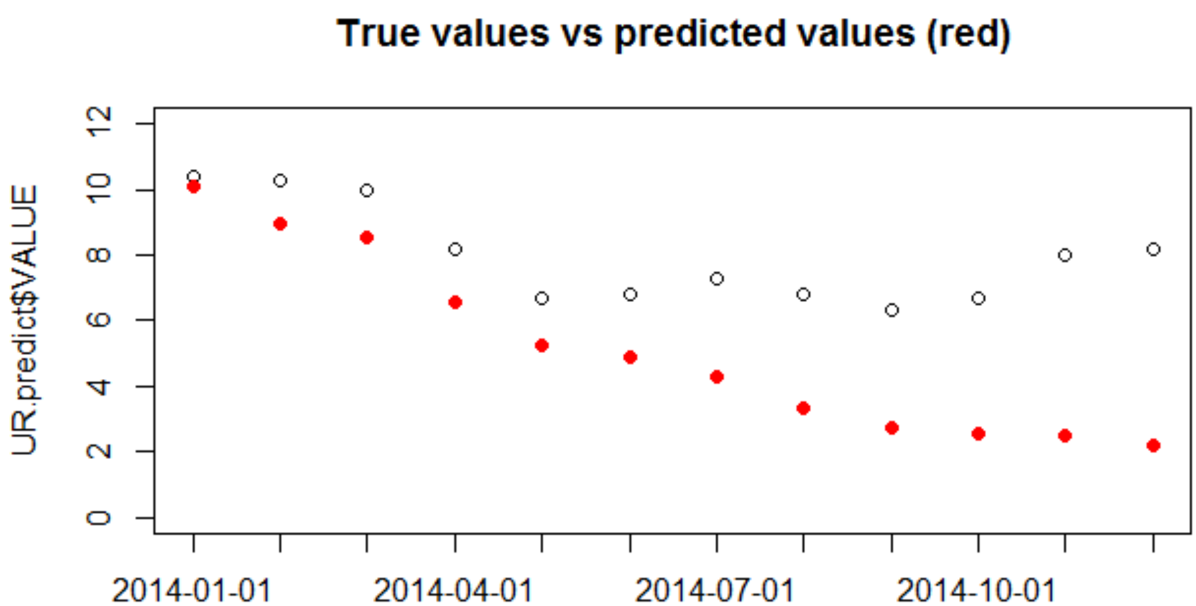

Figure 3.12 The true values VS the predicted values(red points) in 2014.

4. Discussion

We analyzed the data from three components: seasonality, trend and residuals. The seasonality and residuals parts are relatively easy to analyze, compared to the trend part, since the overall trend is not stable and it fluctuated a lot from 1990 to 2013. We used one-sided moving average to detrend and in the prediction process, we tried to use the last three points in 2013 to make a prediction of the unemployment rate on January in 2014, but the last three points lied on a very steep downward line, which indicated that our predictions would go down quickly if we used this approach. Then we checked the last few points in 2013 in the original data which show quickly increase and then leveling off. In order to capture that trend, we fitted polynomial models to the previous 12 points every time we made predictions. We think piecewise regression may predict the trend better.

5. Conclusion

In this project, we applied time series approaches to analyze unemployment data in Yolo County. We used moving average method to deseasonalized our data first and then used two approaches to capture the trend. In the first approach, we used 4th-order polynomial to fit the deseasonalized data and then used one-sided moving average to detrend the data again. In the second approach, we applied one-sided moving average method to the deseasonalized data directly. It turned out residuals are very similar. After using three goodness of fit tests to conduct residual testing, ARMA model is used to fit the stationary residuals. In the prediction process, polynomial method is used to predict the future trends. The final prediction is the sum of predicted seasonality, trend and residual part.

References

Economic Research. Federal Reserve Bank of St. Louis, Updated: 2015. Web. 1 March. 2015. http://research.stlouisfed.org/fred2/series/CAYOLO3URN

Appendix

Plots:

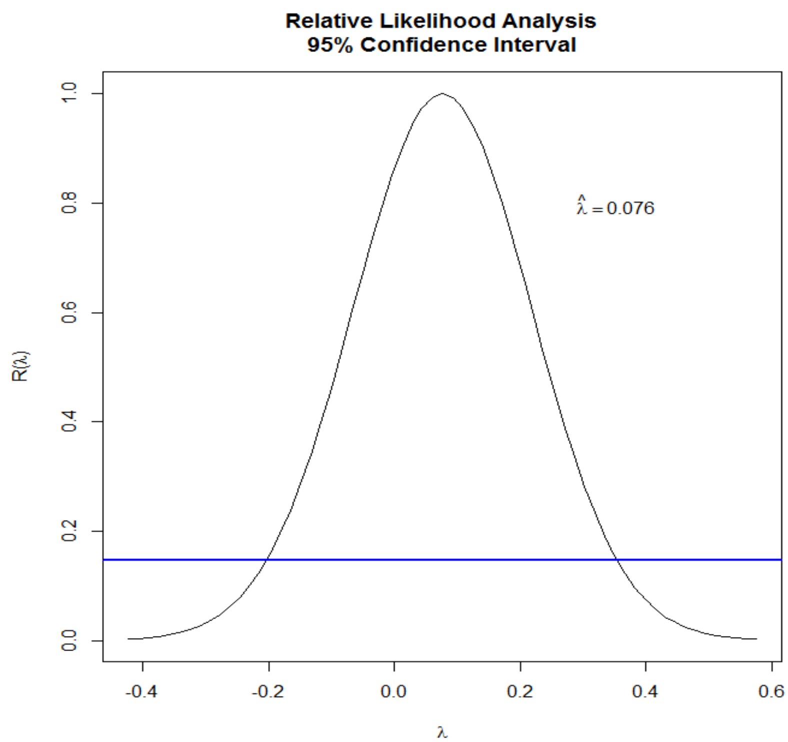

Figure A.1: Boxcox plot on the raw data.

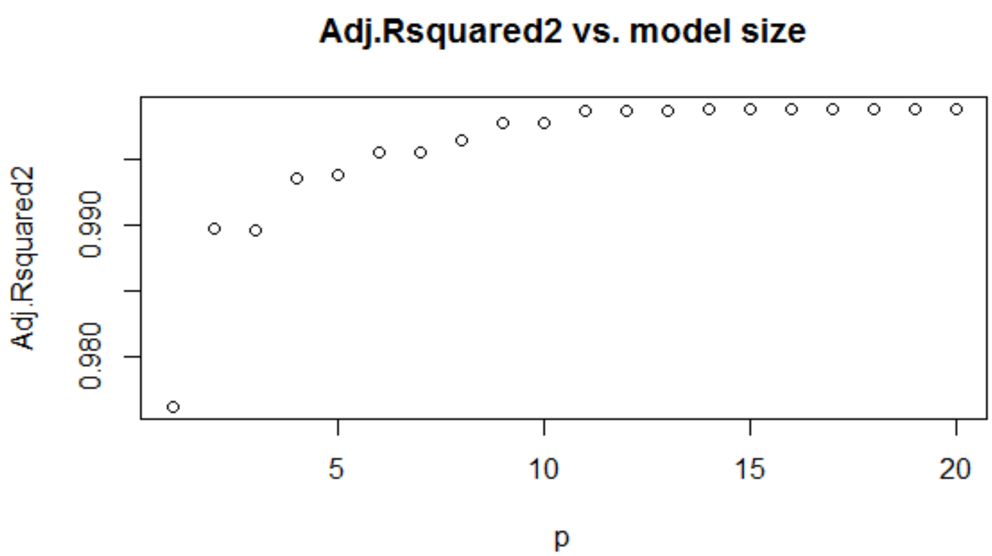

Figure A.2: Adj.Rsquared2 VS model size when fitting trend with polynomial method.

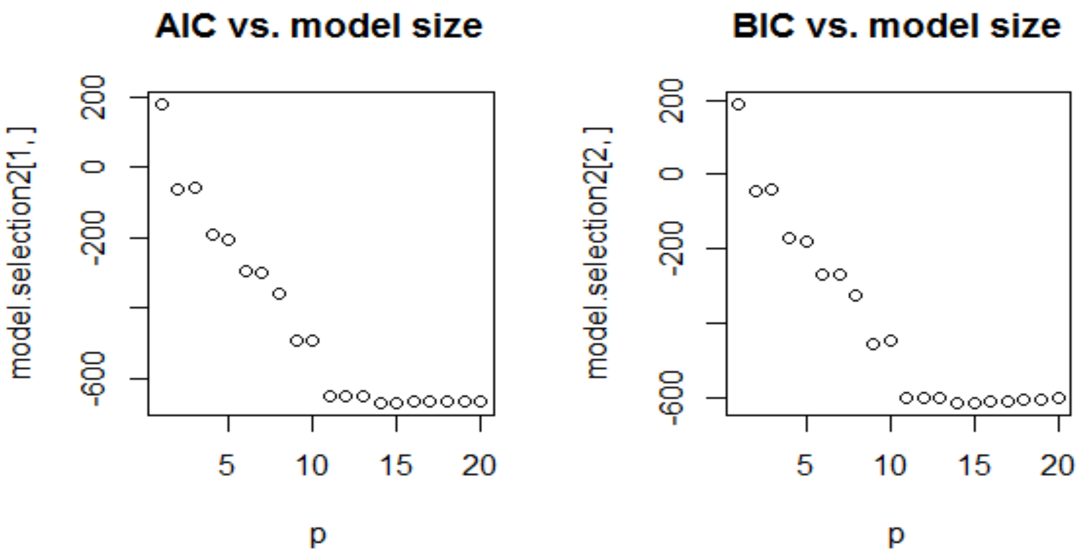

Figure A.3: AIC(BIC) VS model size when fitting trend with polynomial method.

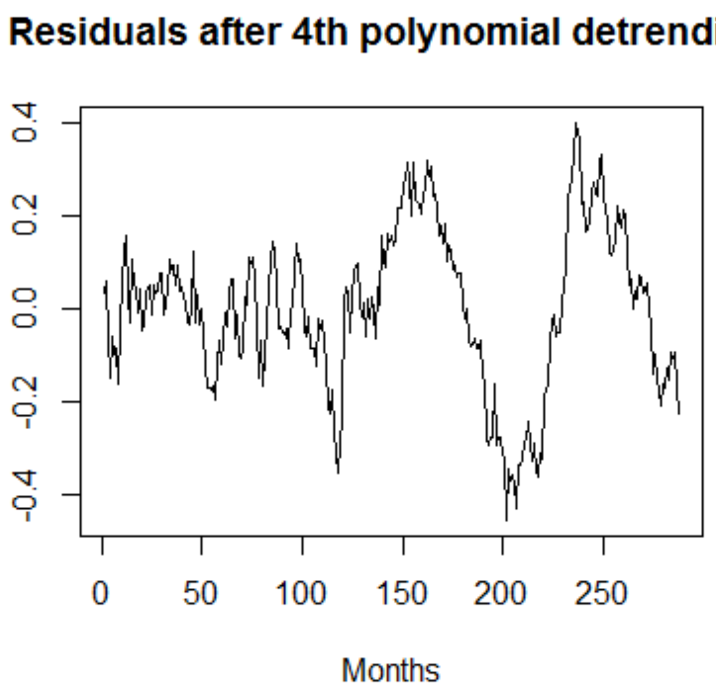

Figure A.4: Residuals after 4th polynomial detrending

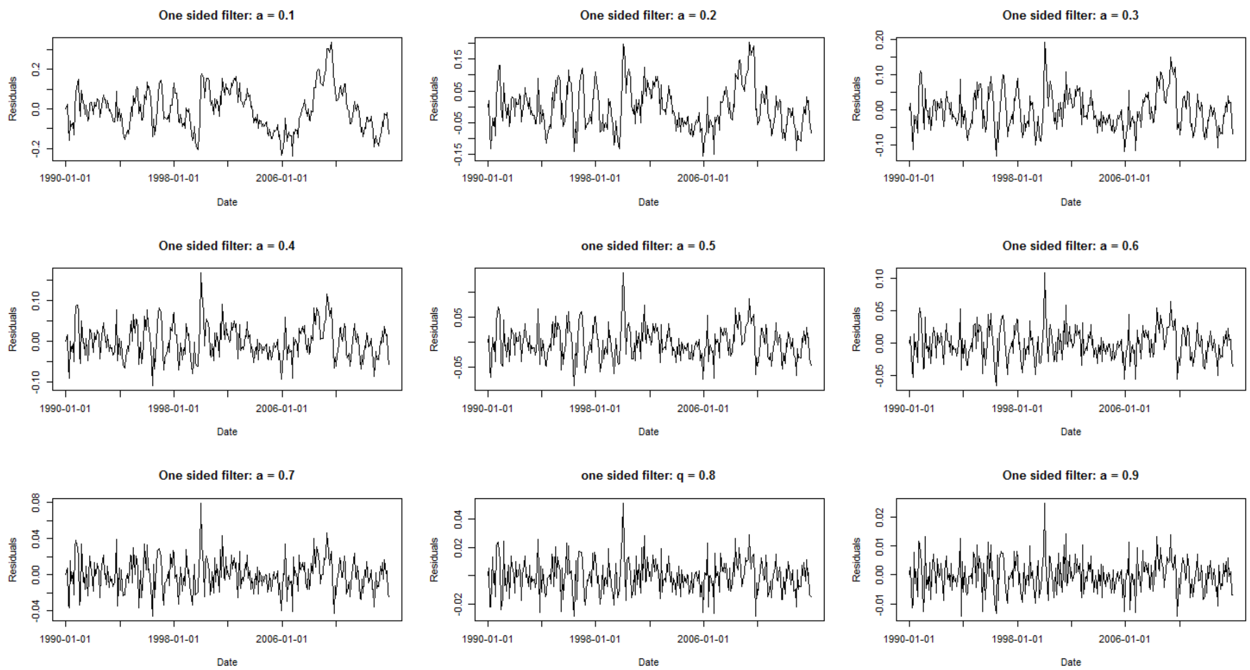

Figure A.5: Residuals after redetrending using moving average with a = 0.1, 0.2, …, 0.9 from polynomial method

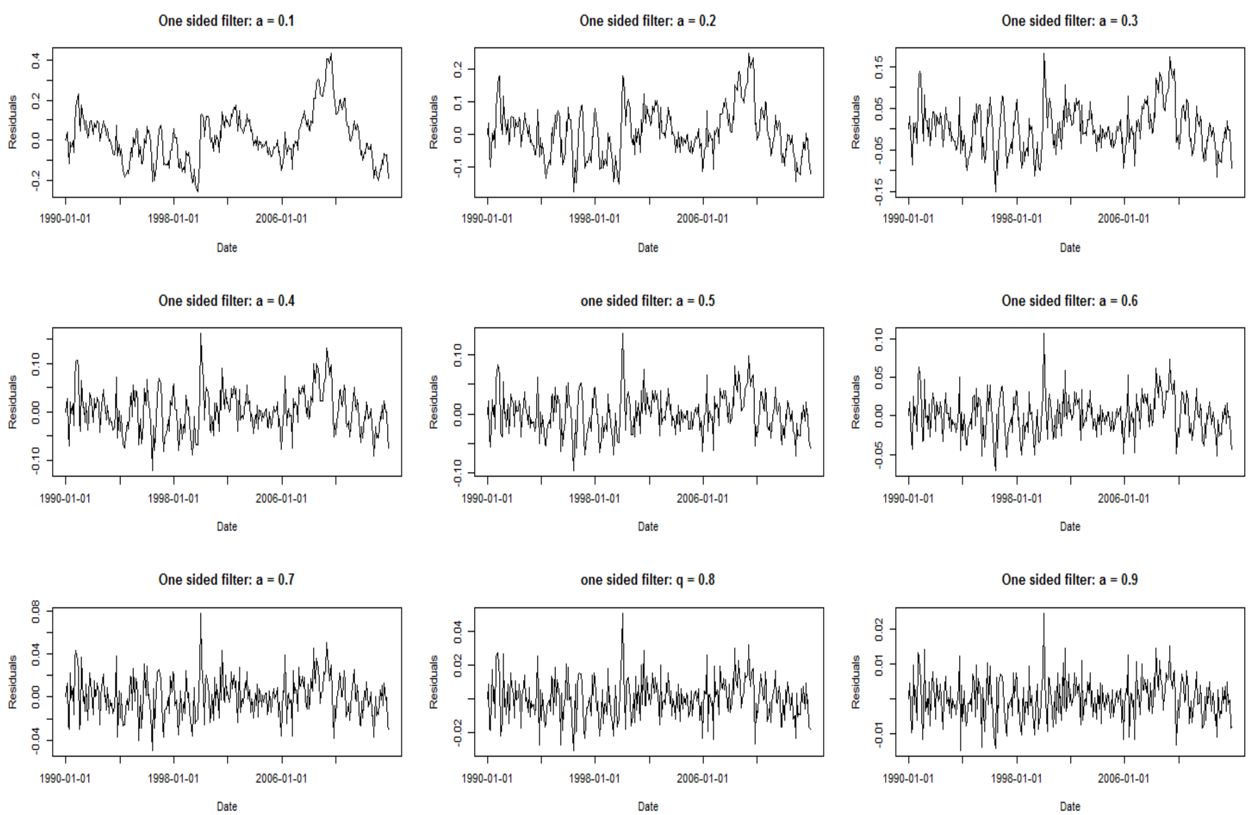

Figure A.6: Residuals after detrending using moving average with a = 0.1, 0.2, …, 0.9 from one-sided moving average method.

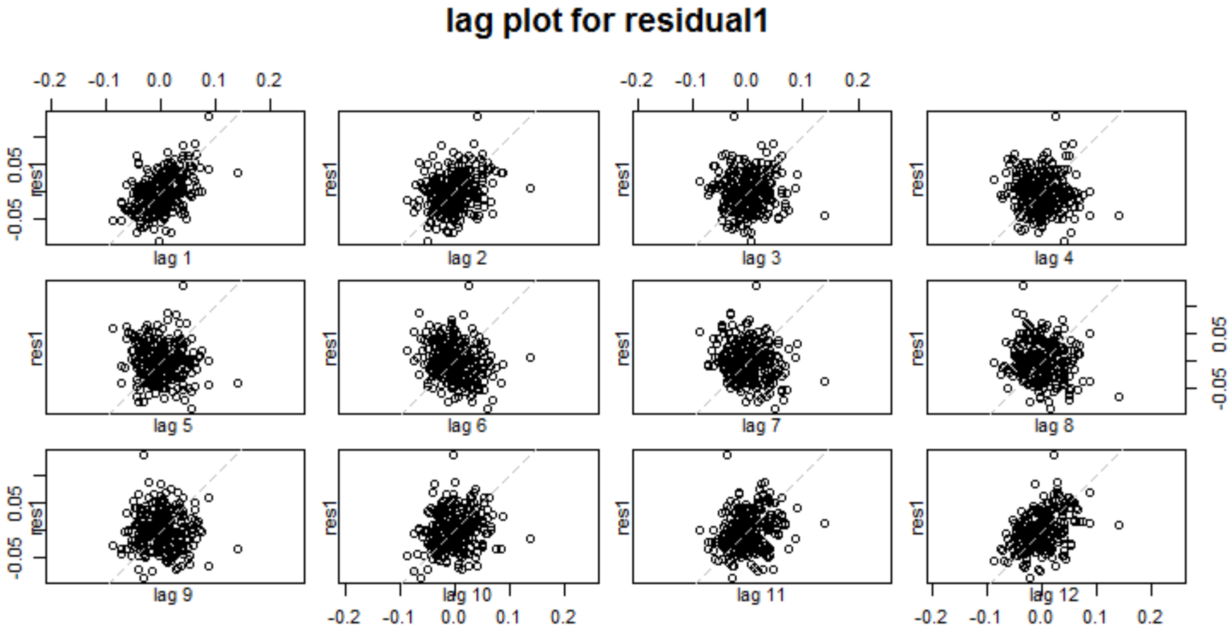

Figure A.7: Scatterplot matrix relating current residuals to past residuals for the lags h = 1,...,12 for method1 polynomial.

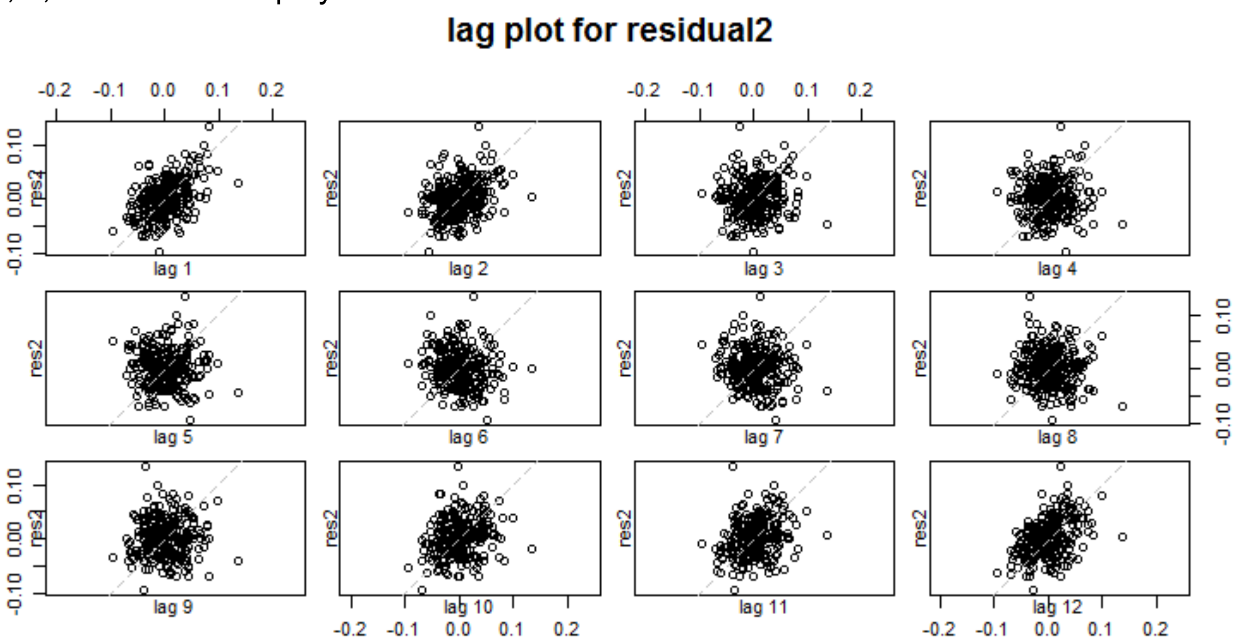

Figure A.8: Scatterplot matrix relating current residuals to past residuals for the lags h = 1,...,12 for method2 one-sided moving average.

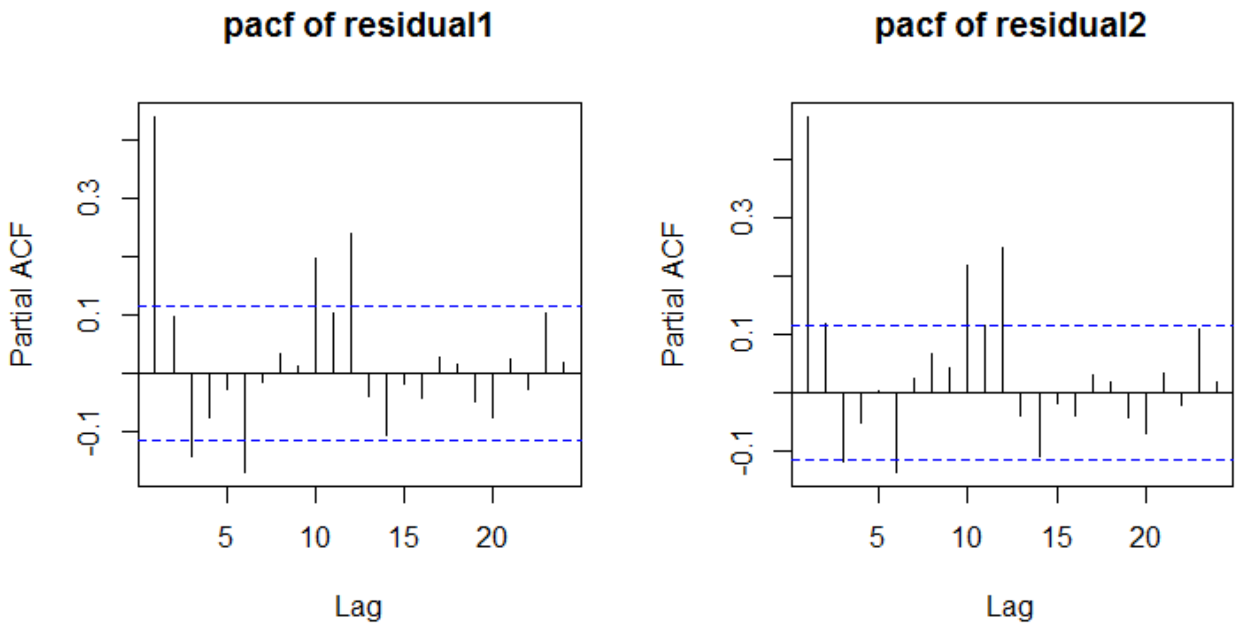

Figure A.9: PACF plots of residuals after method 1 - polynomial (left) and method 2 - one-sided moving average (right).

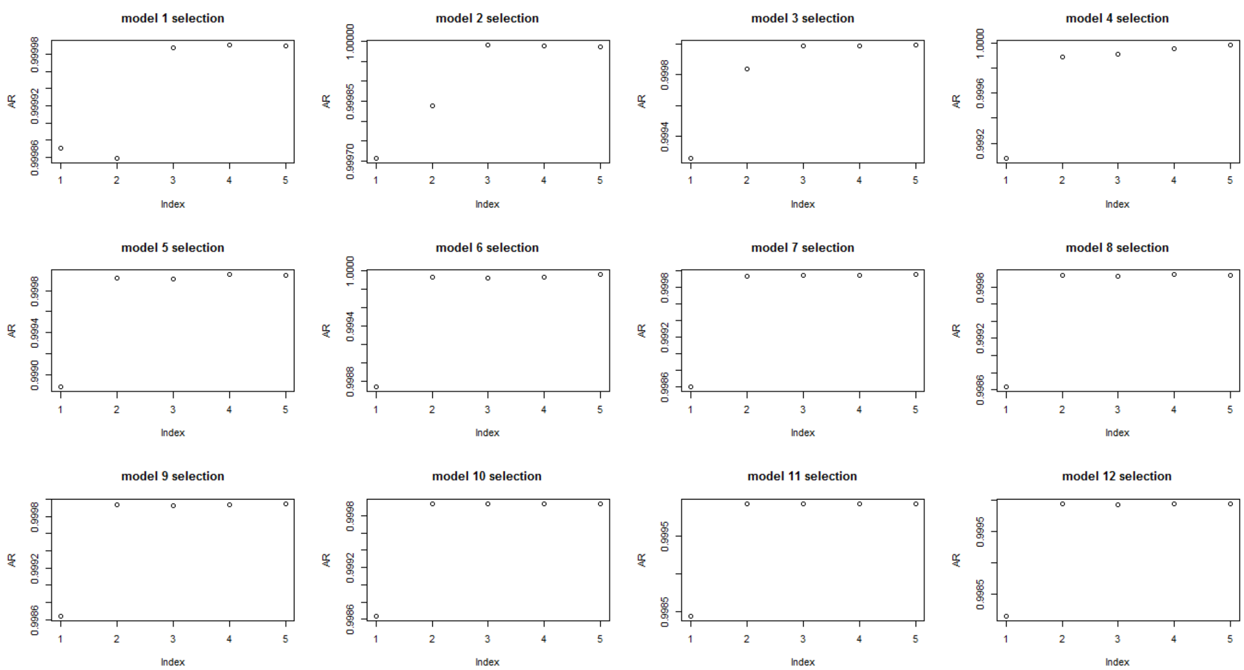

Figure A.10 AICs vs model size p during the trend prediction procedure.

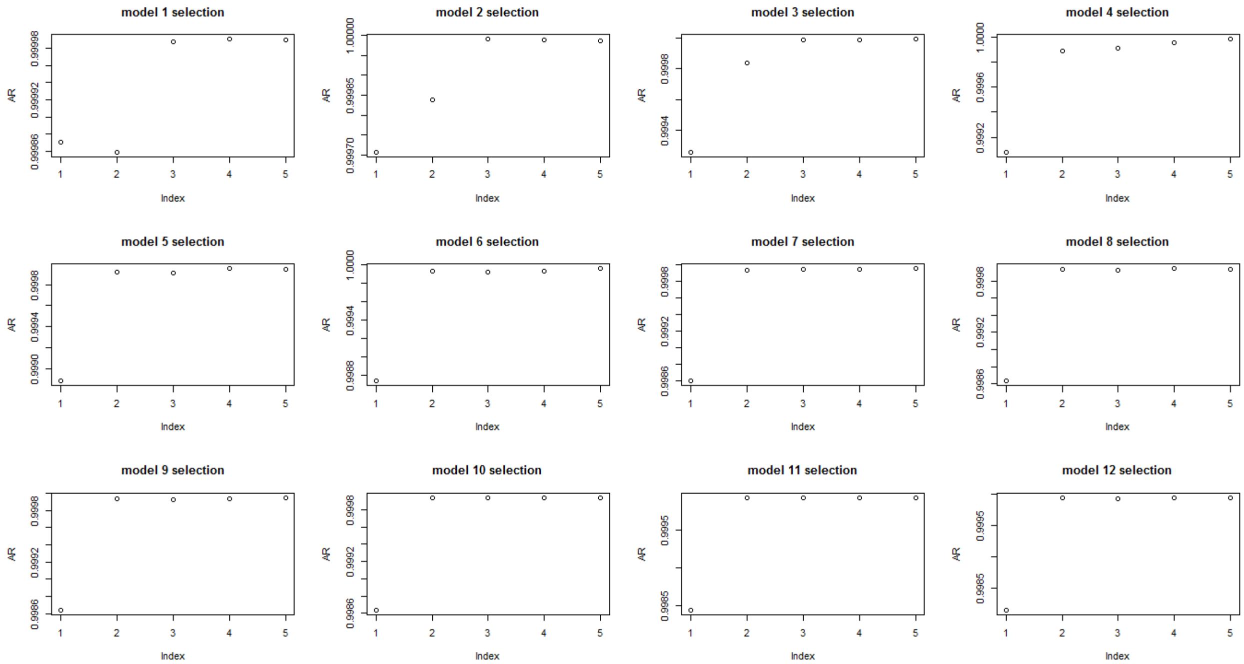

Figure A.10 AICs vs model size p during the trend prediction procedure.