Stat479
Dr. Douglas Wiens
Time Series Analysis on the Relationship Between Oil Price and Alberta's Unemployment Rate
Ming Wan

1.Introduction:

As we all know, petroleum industry is the lifeblood of Alberta's economy, according to Alberta government, "the energy sector (oil and gas/mining) accounted for over 22 per cent of Alberta's GDP in 2012". More importantly, " approximately 121,500 people were employed in Alberta's upstream energy sector, which includes oil sands, conventional oil and gas, and mining." We therefore expect a strong correlation between the price of oil and unemployment rate. Here we consider two series of interest, real oil price and the unemployment rate of Alberta, from Jan 1998 to Feb 2015.

Since all of the most often used oil price indices such as the West Texas Intermediate (WTI) and the Brent Crude are not adjusted for inflation, those indices have a tendency to "shrink" the prices in the past, considering that inflation rates are not consistent, it is best that we adjust the prices for inflation before analyzing them. Therefore I decided to find both the monthly Consumer Price Index of U.S. and the WTI, compute the inflation adjusted WTI by dividing the nominal WTI of a given month by the ratio of the CPI in that month to the CPI of Feb. 2015. In this way we can retrieve the real WTI with Feb. 2015 as the basis month.

On the other hand, unemployment rates of Alberta are easily retrieved from Statistics Canada. The series downloaded is not seasonally adjusted, as I intend to fit a seasonal model myself in part 2.

Figure 1.1 shows the plots of these two series with a 6-point moving average superimposed. One can observe a strong negative correlation between oil price and unemployment rate. Intuitively, as oil price rises, petroleum companies would provide more job positions and boost Alberta's economy, therefore stimulating other industries (for instance, service industry) as well, all would lead to more job positions and reduce the unemployment rate. However, as much as we would like to, unemployment rates can never be zero, due to the fact that frictional and structural unemployment will always be present.

One can also notice a significant pike between the years of 2007-2009 for both series. Due to the 2008 oil crisis (08OC), WTI peaked in June 2008 with a record high $144.8/barrel, but a swift plummet ensues and by Dec. 2008, WTI was back to only $ 45.7/barrel. Consequently, in about 10 months the unemployment rate in

Alberta soared from 3% to 7.2%. Such abnormality will prove to be problematic for our model building.

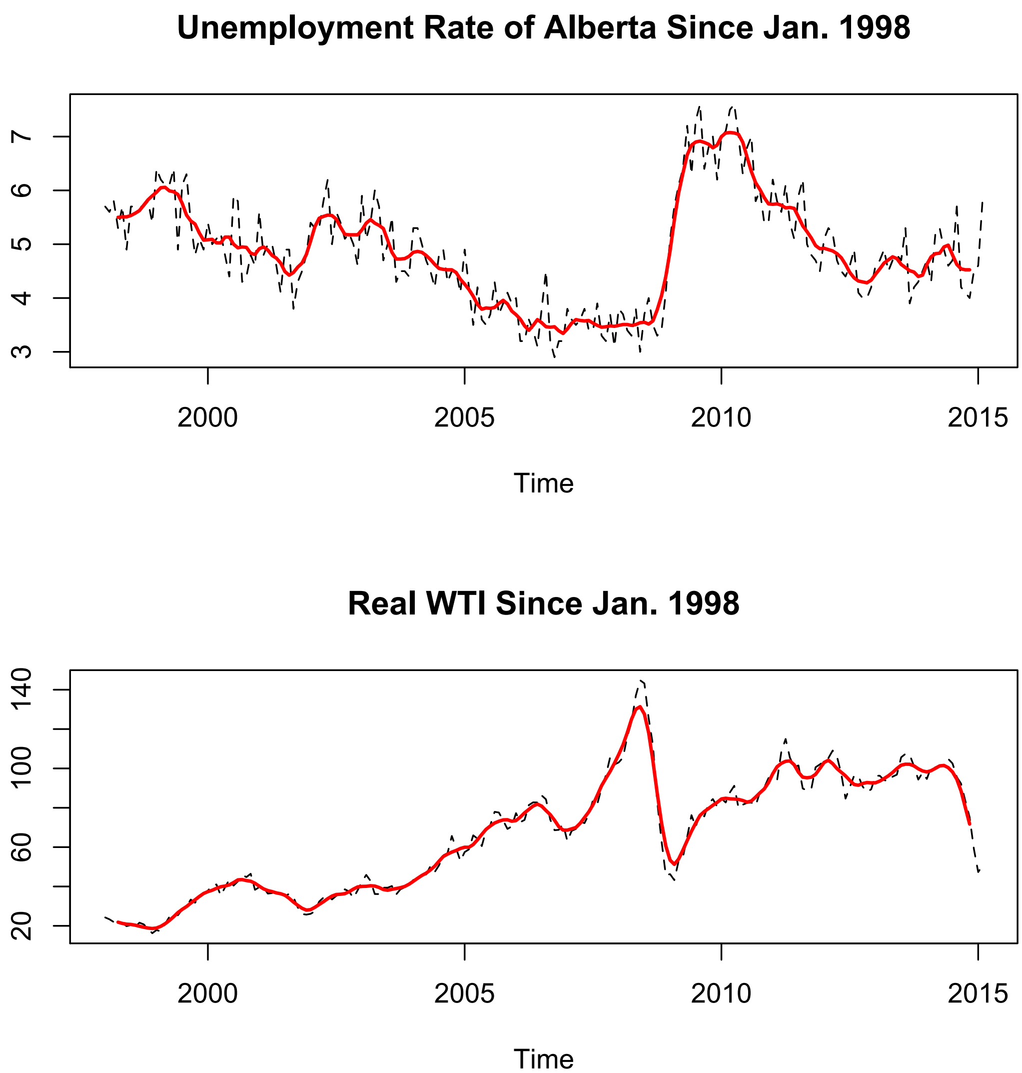

Figure 1.1 Unemployment Rate and Real Oil Price Plotted

Due to 08OC and nonlinear trends, by no means would we consider the two series stationary. Indeed, the ACF and PACF plots below back up our claim. The ACF plot for unemployment decreases slowly and PACF for it is large at 1; for exactly the same reasons, oil price is not stationary either.

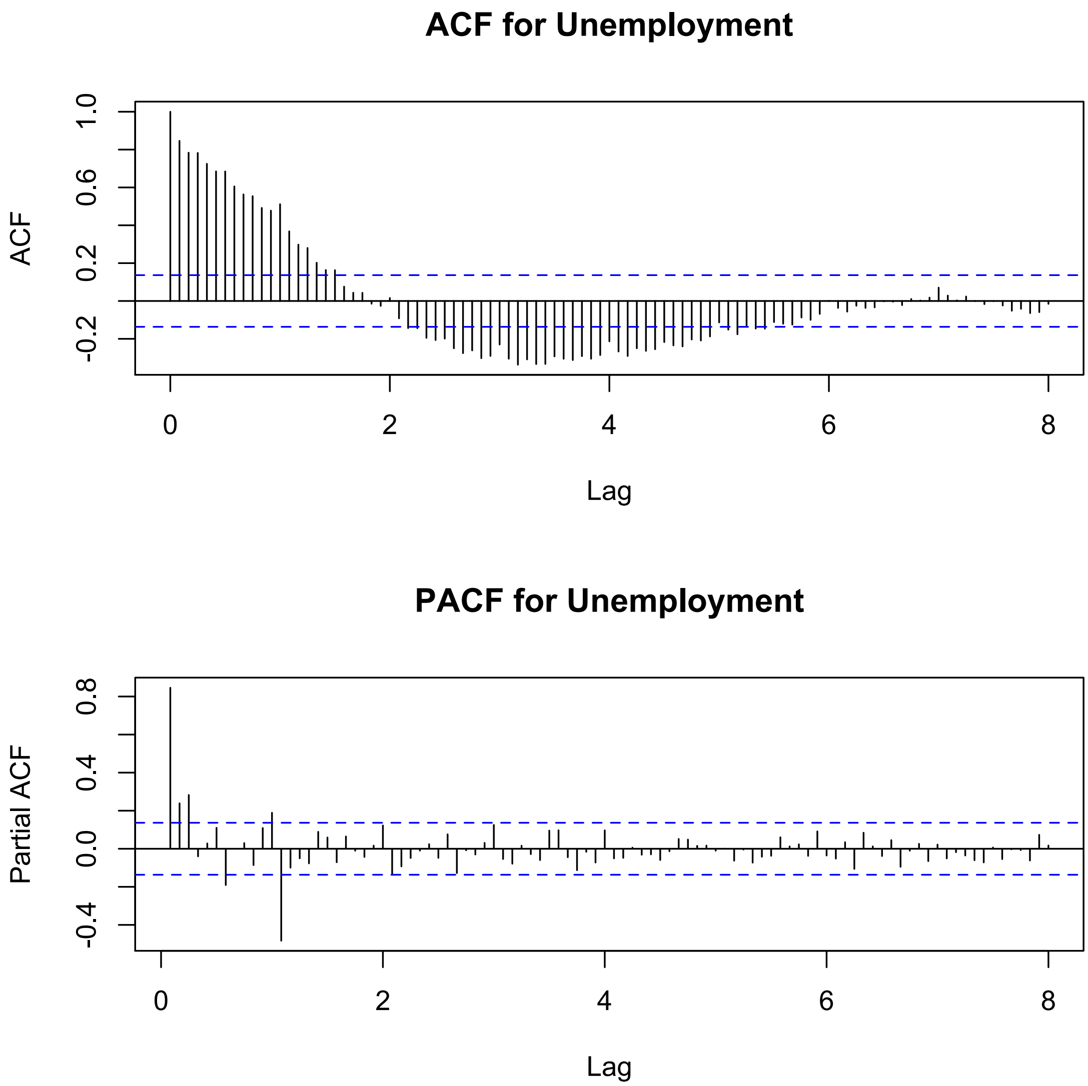

Figure 1.2 a) ACF and PACF for Unemployment Rate

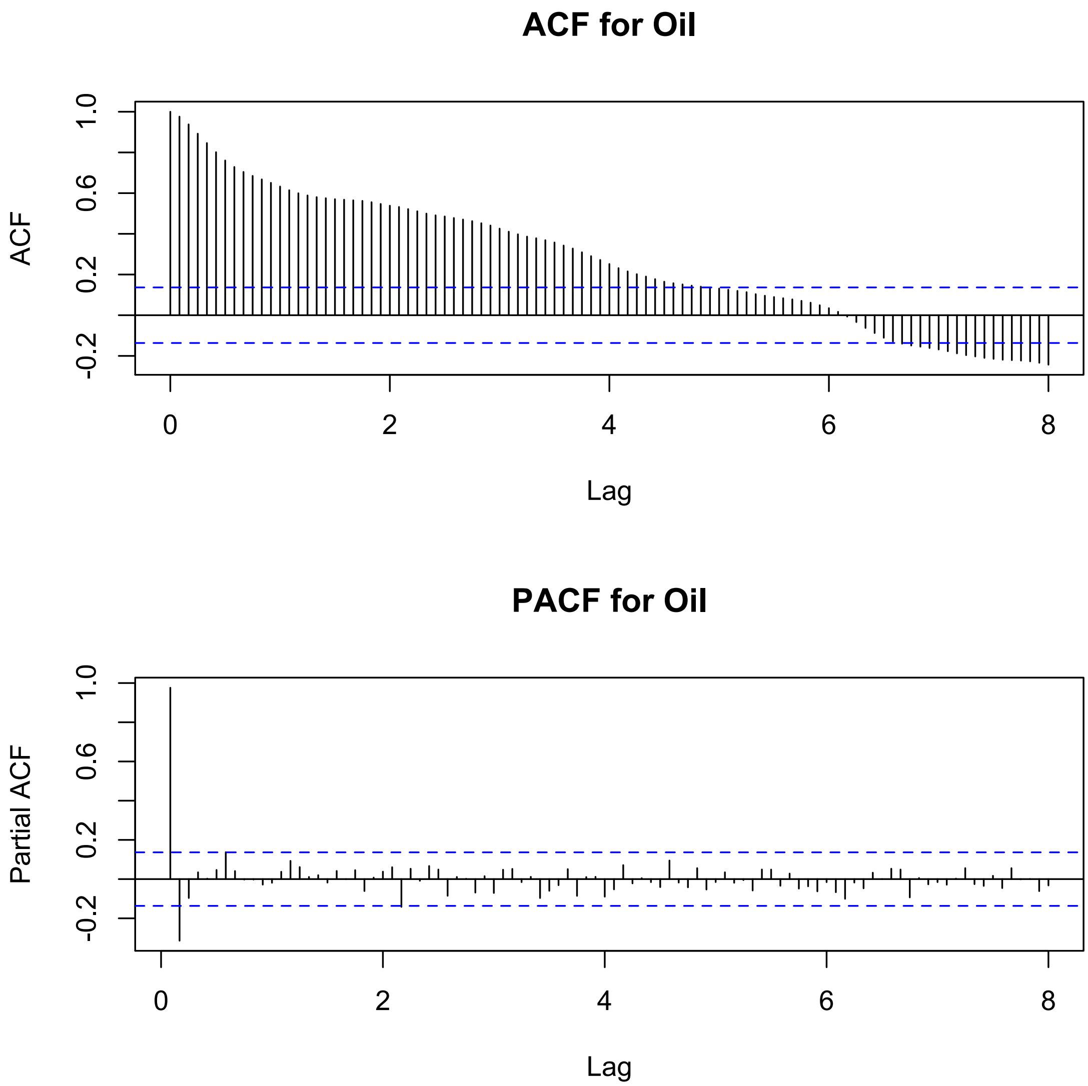

Figure 1.2 b) ACF and PACF for Oil Price

In order to transform our data into stationary series, first we can try to detrend them. If it does not help, differencing might do the trick. Figure 1.3 a) shows the information for detrended unemployment rate series, it does not show much improvement, as ACF and PACF still looks very similar to those of the original series. Figure 1.3 b) indicates some improvement for detrending the oil price (despite high variation during 08OC). Although ACF at lag between 70 and 80 are still significant (which we will just ignore due to the fact that these lags are fairly large so might not be representative if we were to obtain more data), ACF drops at a reasonable rate, its values are still large for lags of 1,2,3,4,5,6, indicating a possible MA(q), q=1,2,3,4,5,6. PACF at lag 1 and 2 are very large so this indicates a potential AR(p), p=1,2. We will still look at the first difference of Oil to see if it will show more improvement, if not, we will just use the detrended oil price series.

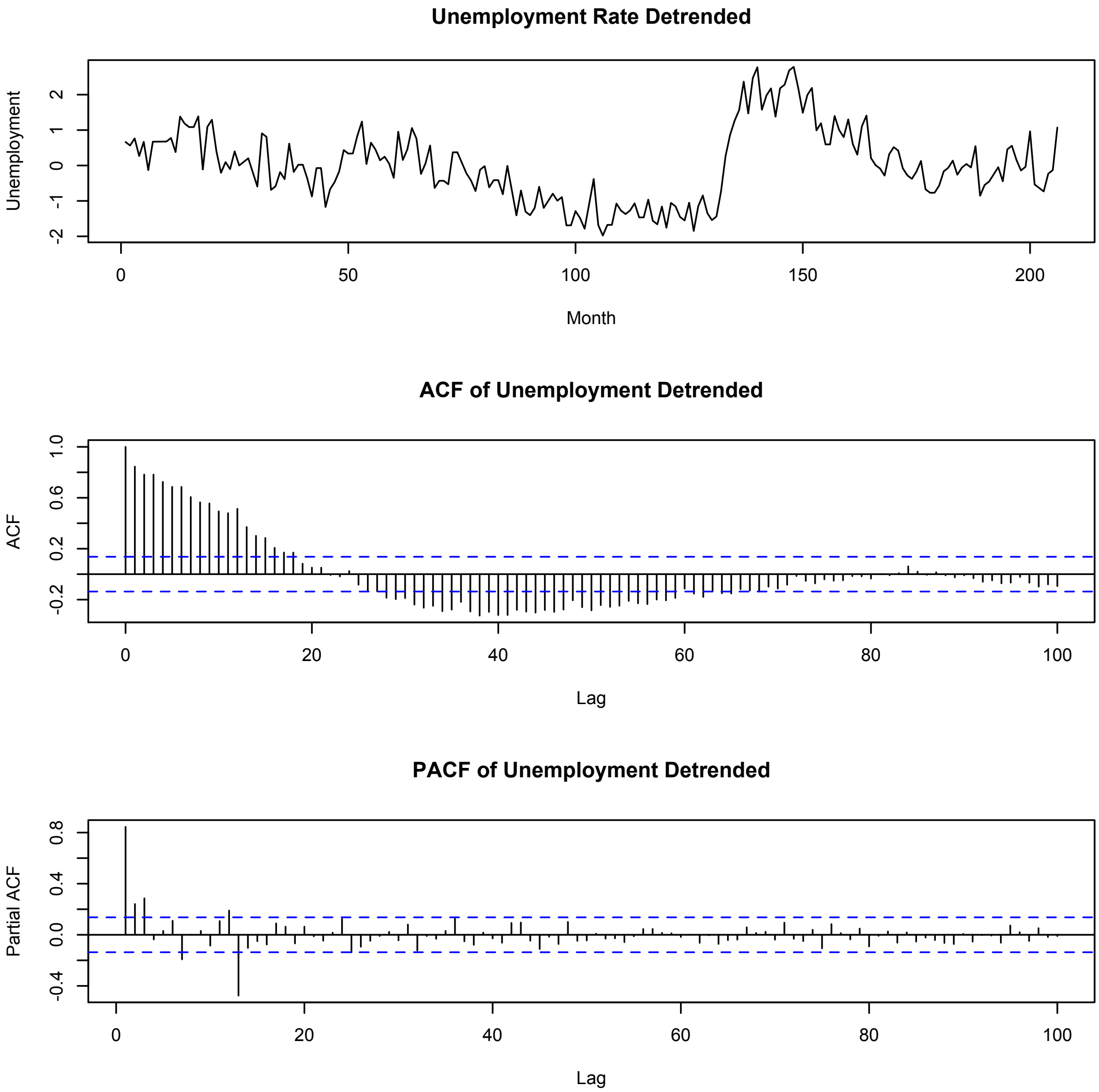

Figure 1.3 a) Plot, ACF and PACF of the detrended unemployment rate

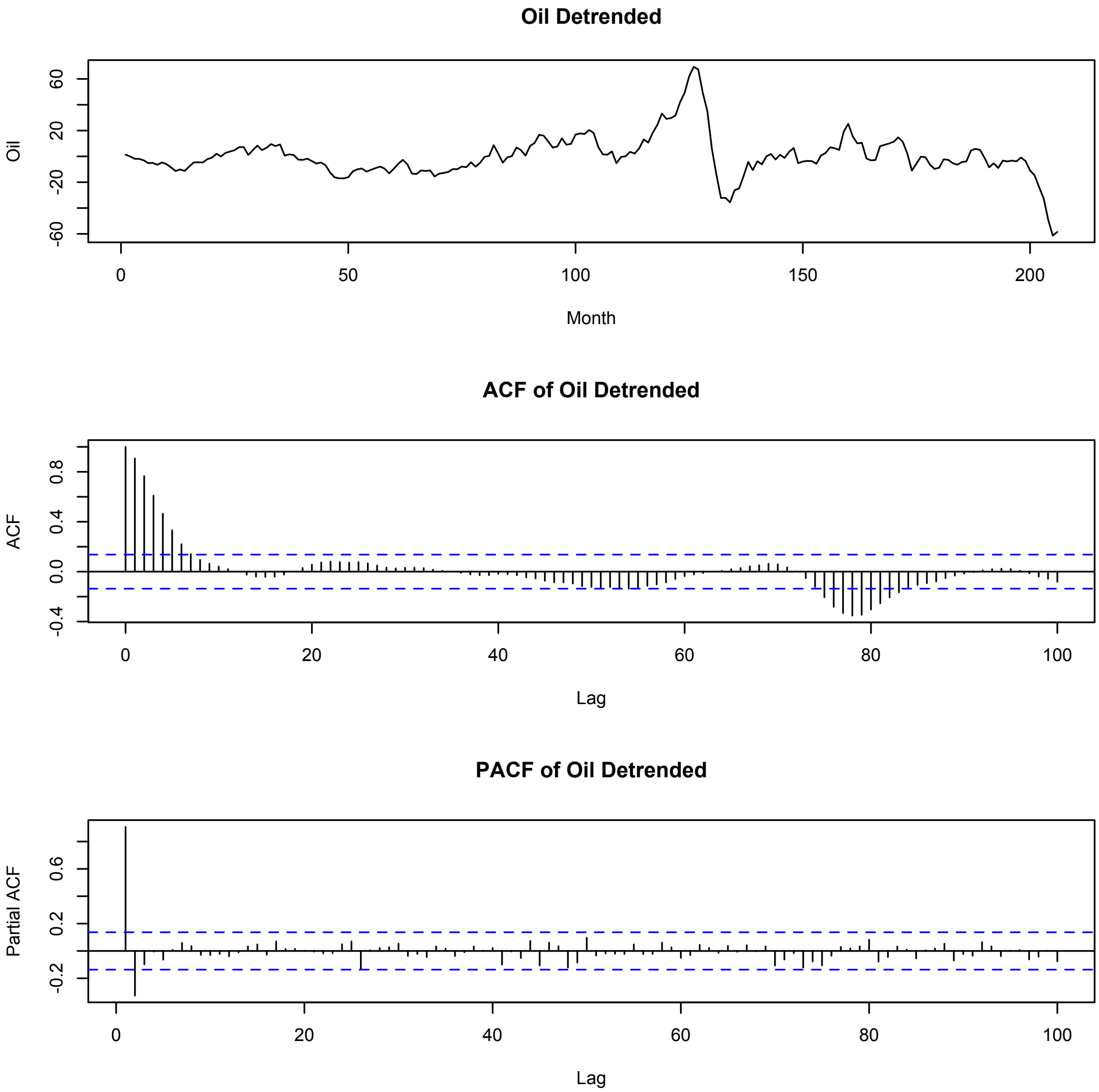

Figure 1.3 b) Plot, ACF and PACF of the detrended oil price

Figure 1.4 a) and Figure 1.4 b) are the plots of the differenced unemployment rate, a) being the first difference and b) being the second difference. We can observe that taking the first difference already transformed the series into a fairly stationary one, but the second difference did not make too much improvement. Therefore we can just use the first difference of unemployment. Since its ACF at s = 12,24,36 are very large and decreasing slowly and PACF is large at s=12, this is a sign of seasonal nonstationarity. Therefore we will observe the behaviour of the annual difference of Diff(unemp) next, as seen in Figure 1.4 c).

The ACF for annual difference of Diff(unemp) at small lag s are all close to 0, when q=1,2 the ACF is barely significant, hence it indicates q for MA(q) might be 0 (we will still investigate cases for q=1 or 2); ACF peaks at s=12 and only at s=12, this is a sign for seasonal MA(Q), Q=1; PACF is barely significant for lag s=1,2 and is close to zero for all lags s>3 except s=12,24,36, implying a possible AR(p), p=0,1,2 and seasonal AR(P), P=1,2,3. And of course we also have d=1, D=1 and s=12. Having speculated all the coefficients for seasonal ARIMA model, we will later find the most suitable one.

Figure 1.4 d) shows the first difference of the oil series. The plot of it looks fairly stationary (again excluding the part corresponding to 08OC) with a slight increase in variance over time, but ACF drops very fast and PACF is only large at very few points. We can use this information to fit an ARIMA model for oil later as well, right now we speculate it to be ARIMA(p,d,q),p=1,d=1,q=1,2.

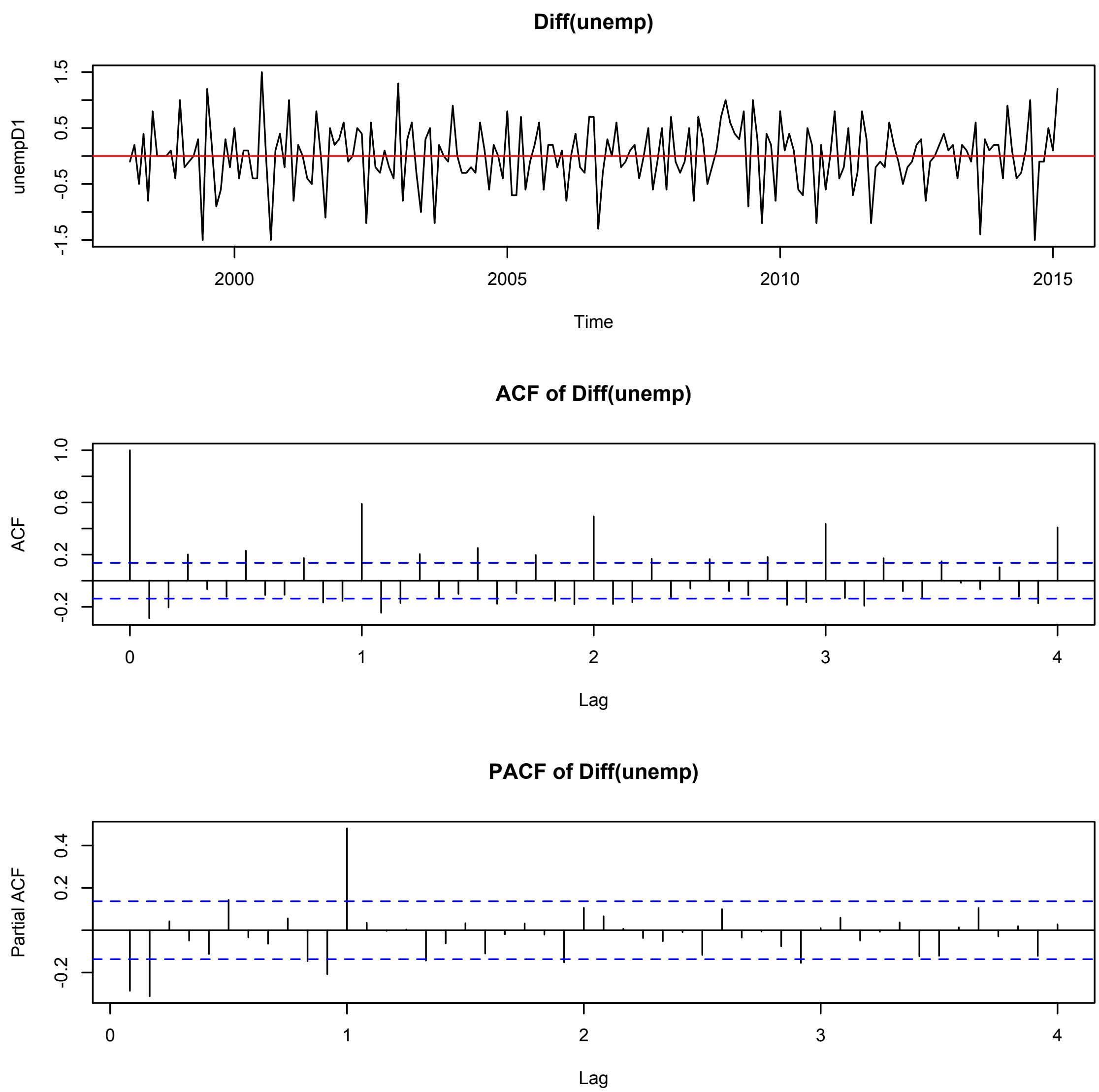

Figure 1.4 a) Plot, ACF and PACF of Diff(unemp)

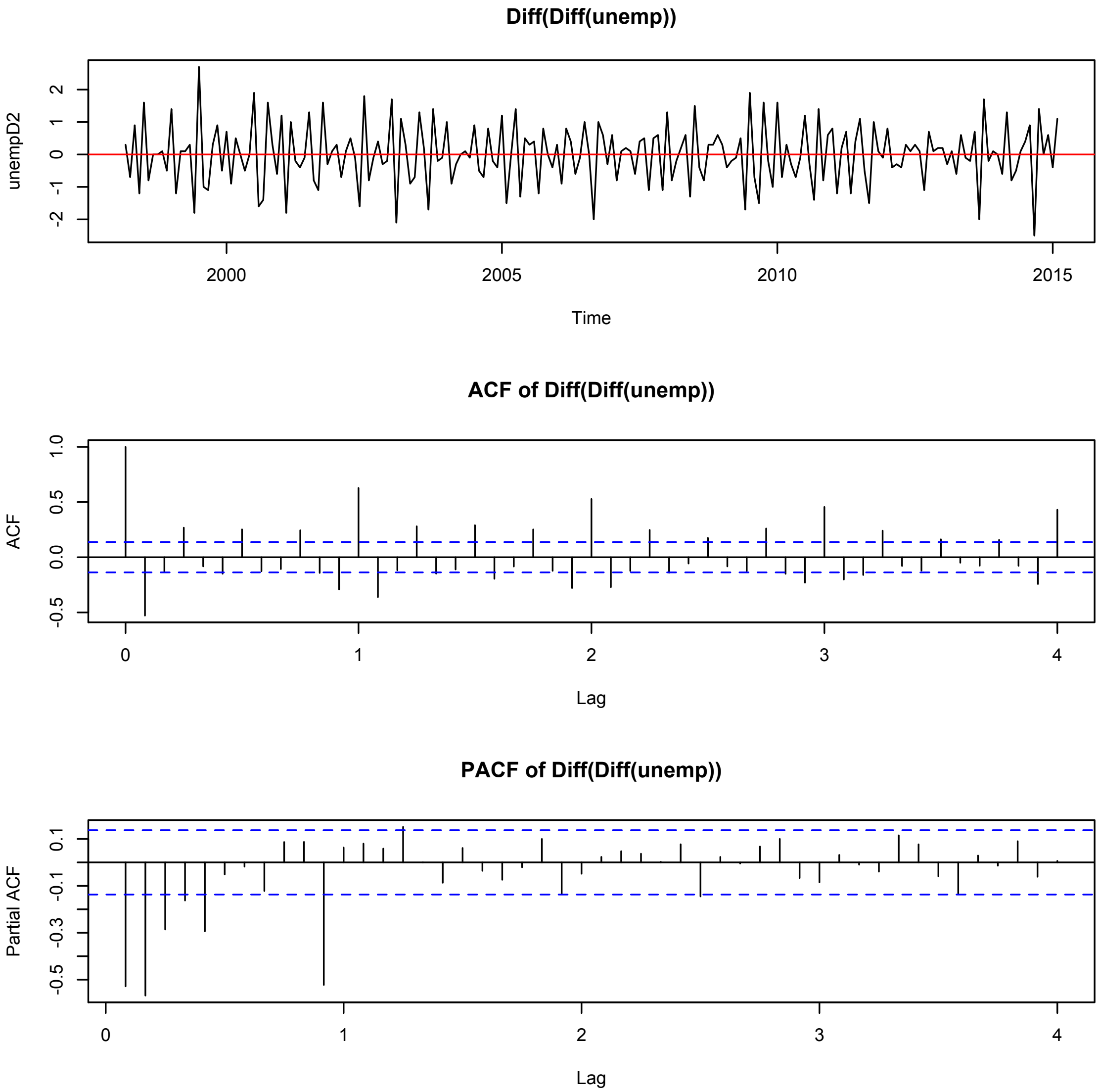

Figure 1.4 b) Plot, ACF and PACF of Diff(Diff(unemp))

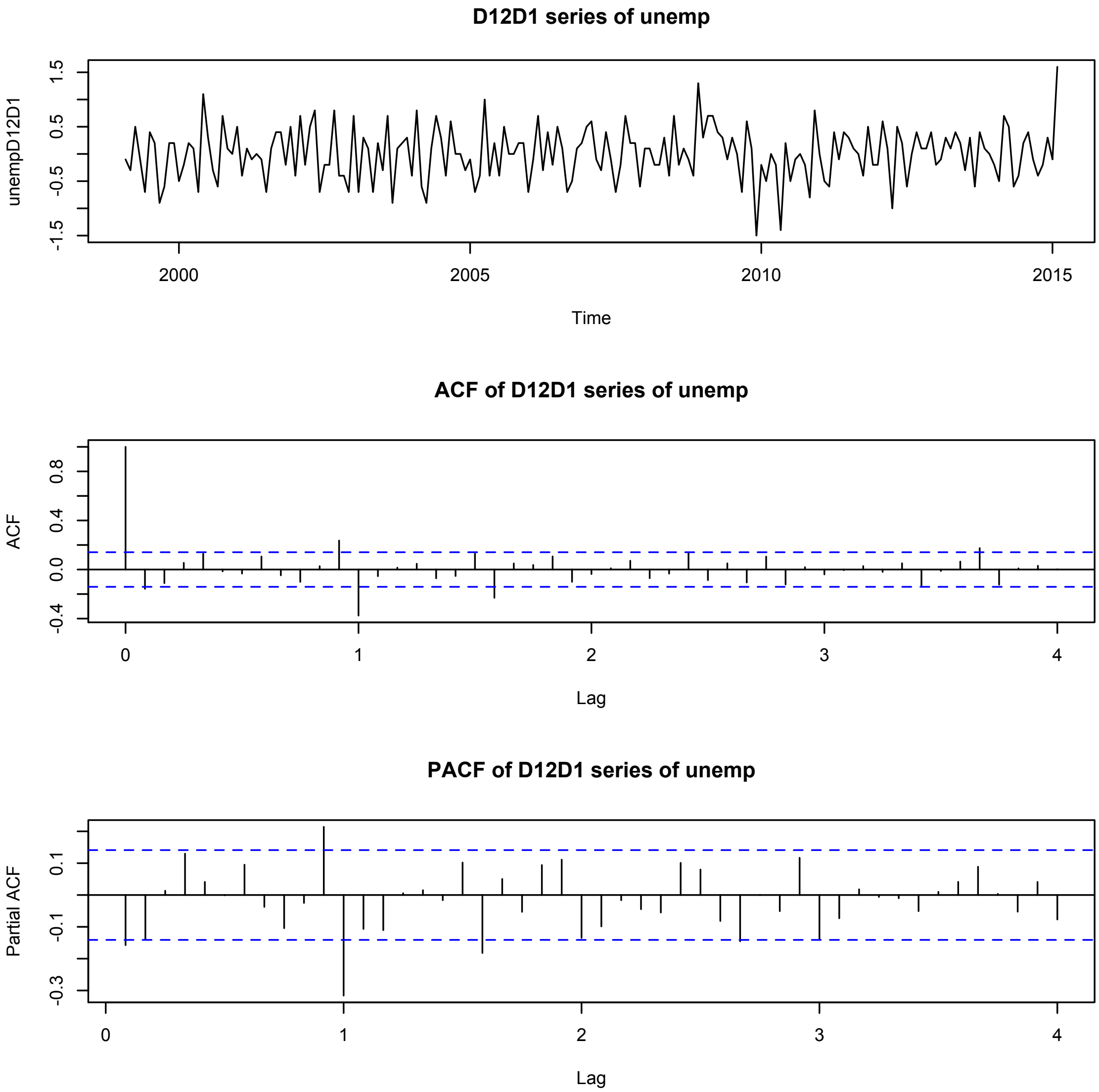

Figure 1.4 c) Plot, ACF and PACF of annual difference(Diff(unemp))

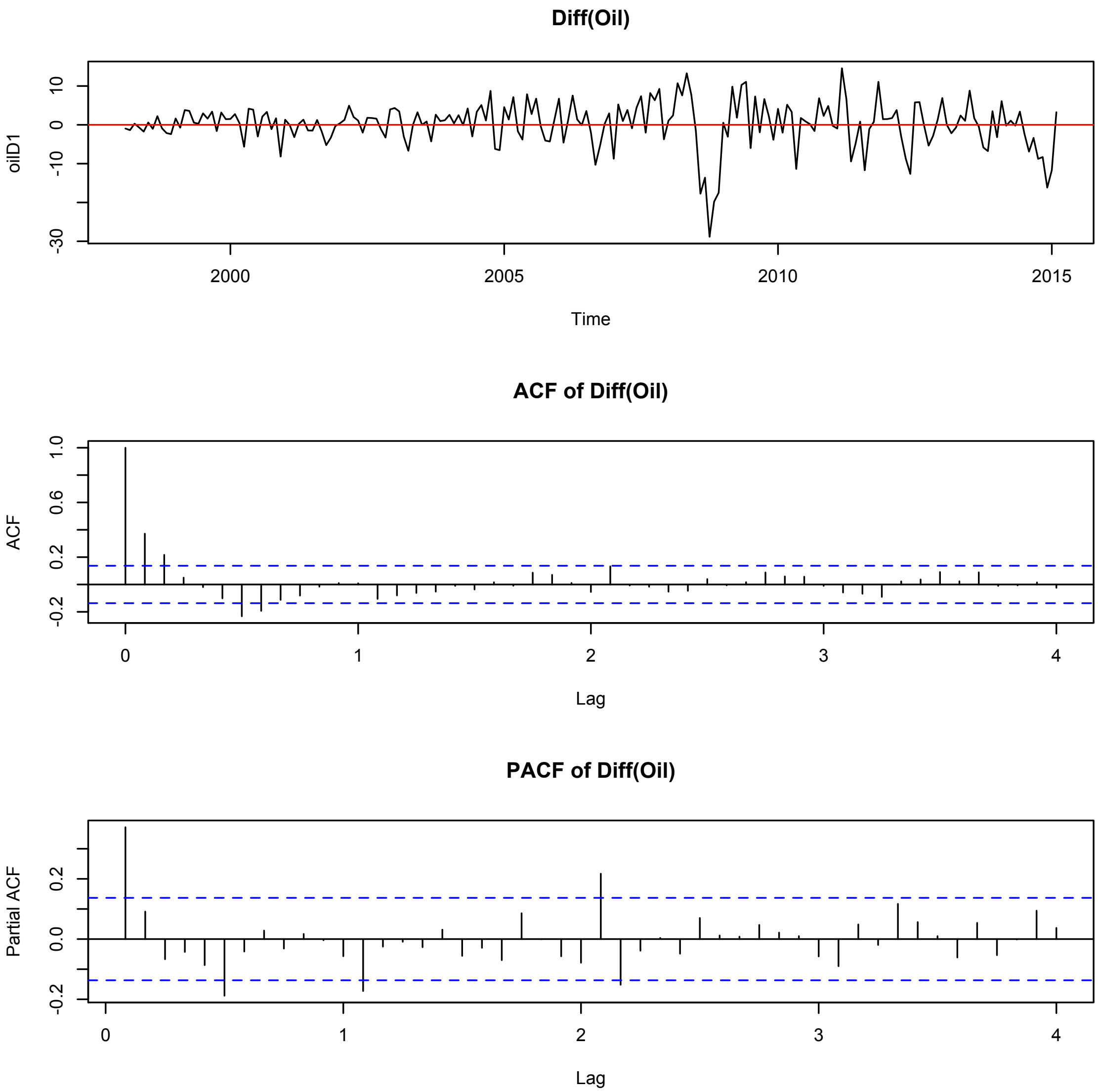

Figure 1.4 d) Plot, ACF and PACF of Diff(Oil)

2. Time Domain Analysis:

In this part we will try to find the best seasonal ARIMA model for unemp and ARIMA model for oil, primarily using AIC, AICc and BIC as guidelines. After obtaining the models, forecasts of the next year for both series will be our main interest.

Table 2.1 shows all the seasonal ARIMA models considered for unemployment alongside with their AIC, AICc and BIC values.

	p	d	q	P	D	Q	S	AIC	AICc	BIC
[1,]	0	1	0	1	1	1	12	-0.8176706	-0.8073851	-1.785361
[2,]	0	1	0	2	1	1	12	-0.8098916	-0.7992168	-1.761427
[3,]	0	1	0	3	1	1	12	-0.8078412	-0.7966761	-1.743222

[4,]	**0**	**1**	**1**	**1**	**1**	**1**	**12**	**-0.8854270**	**-0.8747522**	**-1.836963**
[5,]	0	1	1	2	1	1	12	-0.8787721	-0.8676070	-1.814153
[6,]	0	1	1	3	1	1	12	-0.8782773	-0.8665195	-1.797504
[7,]	0	1	2	1	1	1	12	-0.8757874	-0.8646224	-1.811168
[8,]	0	1	2	2	1	1	12	-0.8697374	-0.8579796	-1.788964
[9,]	0	1	2	3	1	1	12	-0.8689533	-0.8564987	-1.772025
[10,]	1	1	0	1	1	1	12	-0.8798706	-0.8691958	-1.831406
[11,]	1	1	0	2	1	1	12	-0.8741406	-0.8629756	-1.809522
[12,]	1	1	0	3	1	1	12	-0.8722958	-0.8605380	-1.791522
[13,]	1	1	1	1	1	1	12	-0.8760890	-0.8649239	-1.811470
[14,]	1	1	1	2	1	1	12	-0.8697617	-0.8580039	-1.788988
[15,]	1	1	1	3	1	1	12	-0.8716215	-0.8591669	-1.774693
[16,]	1	1	2	1	1	1	12	-0.8652429	-0.8534851	-1.784469
[17,]	1	1	2	2	1	1	12	-0.8594829	-0.8470283	-1.762554
[18,]	**1**	**1**	**2**	**3**	**1**	**1**	**12**	**-0.9266510**	**-0.9133939**	**-1.813568**
[19,]	2	1	0	1	1	1	12	-0.8756541	-0.8644890	-1.811035
[20,]	2	1	0	2	1	1	12	-0.8695373	-0.8577794	-1.788764
[21,]	2	1	0	3	1	1	12	-0.8693214	-0.8568667	-1.772393
[22,]	2	1	1	1	1	1	12	-0.8660669	-0.8543090	-1.785293
[23,]	2	1	1	2	1	1	12	-0.8599007	-0.8474461	-1.762972
[24,]	2	1	1	3	1	1	12	-0.8596841	-0.8464269	-1.746601
[25,]	2	1	2	1	1	1	12	-0.8787773	-0.8663227	-1.781849
[26,]	2	1	2	2	1	1	12	-0.8696031	-0.8563460	-1.756520

Table 2.1 All seasonal ARIMA models considered with AIC, AICc and BIC, Unemp

If we select a model with the smallest AIC and AICc, ARIMA$(1,1,2)\times(3,1,1)_{12}$ will be the most suitable model, however according to BIC, ARIMA$(0,1,1)\times(1,1,1)_{12}$ is the best model available. With principle of parsimony in mind, if ARIMA$(0,1,1)\times(1,1,1)_{12}$ does not terribly violate all the assumptions then we will use this model, because it is way more simple. Figure 2.1 is the residual plot for this model. The standardized residuals do not exhibit any obvious trends. The Ljung-Box test for hypothesis of ACF residual whiteness has a few p-values on the edge of significance at a few lags, but overall there are no critically significant p-values. Histogram and QQ plot of standardized residuals are shown in Figure 2.2, although there is a slight light right tail, the residuals look well normally distributed. Furthermore, p-value of Shapiro-Wilk test is far greater than 0.05, therefore we have great confidence that this model meets all the assumptions.

Shapiro-Wilk normality test

data: stdres
W = 0.9914, p-value = 0.2669

Coefficients:

	ma1	sar1	sma1
	-0.2692	0.1127	-0.8475
s.e.	0.0704	0.0930	0.0720

The final seasonal ARIMA model for unemployment rate is:

$$\mathbf{(1-0.1127*B^{12})*(1-B^{12})*(1-B)*Xt = (1-0.8475*B^{12})*(1-0.2692B)*wt}$$

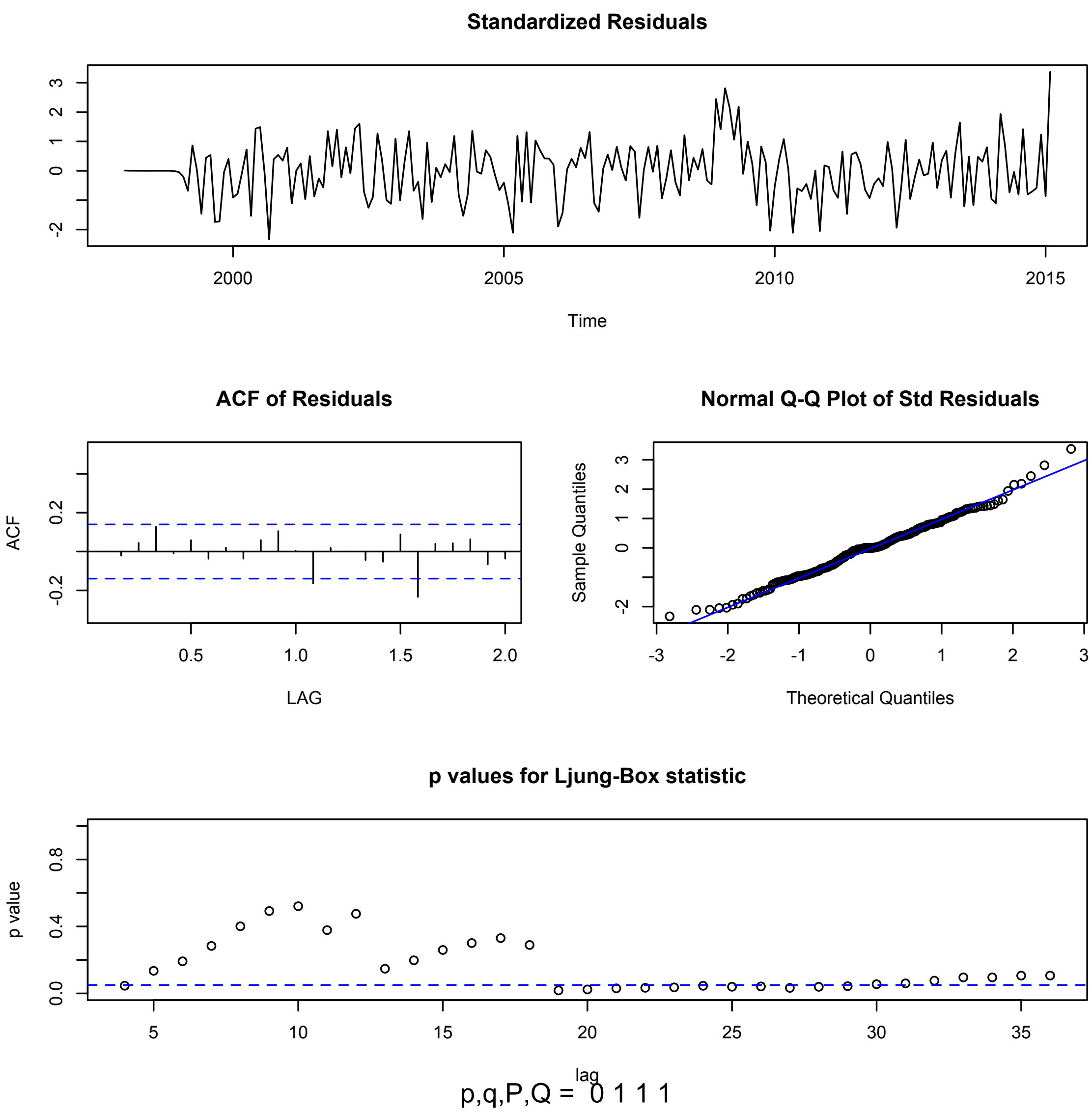

Figure 2.1 Residual plots of the ARIMA$(0,1,1)\times(1,1,1)_{12}$ model for Unemployment

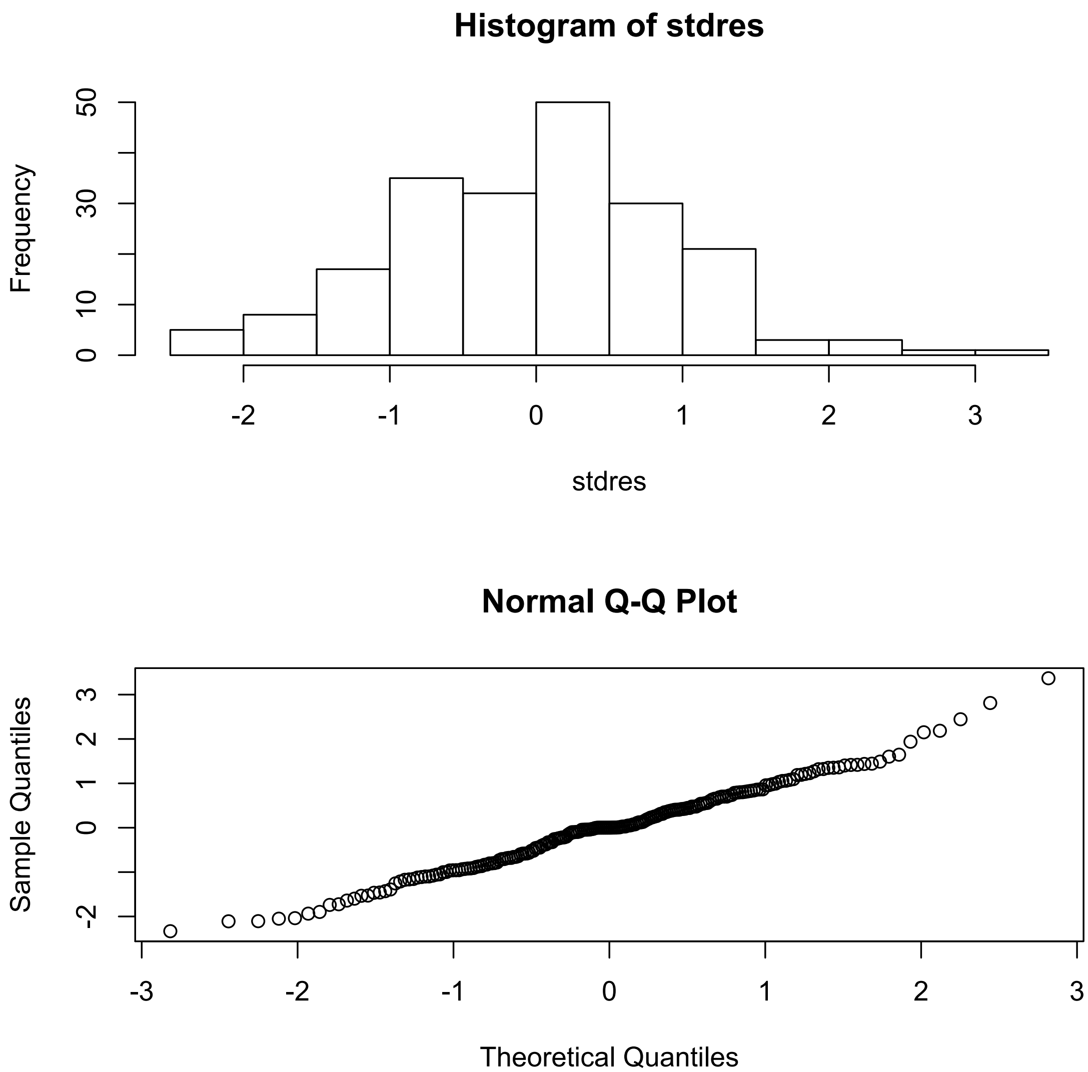

Figure 2.2 Histogram and QQ plot of the standardized residuals for unemp

In a much similar fashion we attempt to fit an ARIMA model for oil series. I decided to range p and q from 1 to 3 for more model options. It says the best model around is ARIMA(3,1,3), however, ACF of residuals are large for lag = 1 and 2, and the p-values for Ljung-Box statistic are critically small at all lags, this means the residual whiteness assumption might be violated. Since ACF residuals are suspiciously large at 1 and 2, perhaps we should have fitted a seasonal ARIMA for oil as well?

	p	d	q	AIC	AICc	BIC
[1,]	1	1	1	-0.2712716	-0.2605968	-1.222807
[2,]	1	1	2	-0.2821115	-0.2709464	-1.217493
[3,]	1	1	3	-0.2806299	-0.2688721	-1.199856
[4,]	2	1	1	-0.2976663	-0.2865013	-1.233047
[5,]	2	1	2	-0.2882597	-0.2765018	-1.207486

[6,] 2 1 3 -0.4462953 -0.4338407 -1.349367
[7,] 3 1 1 -0.3160585 -0.3043007 -1.235285
[8,] 3 1 2 -0.2809646 -0.2685099 -1.184036
[9,] 3 1 3 -0.4991149 -0.4858578 -1.386032

Table 2.2 All ARIMA models considered with AIC, AICc and BIC

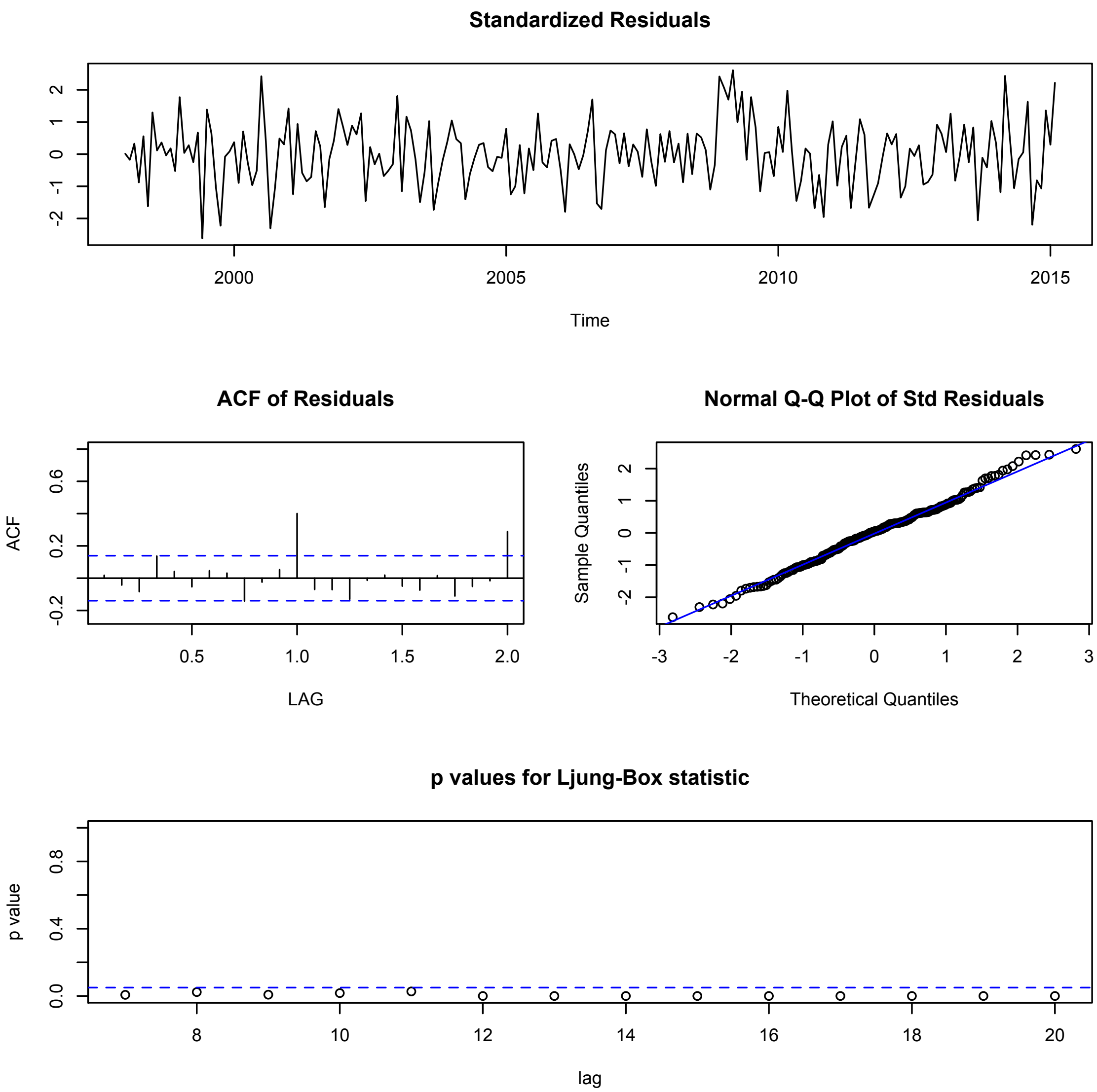

Figure 2.3 Residual plots of the ARIMA(3,1,3) model for Oil

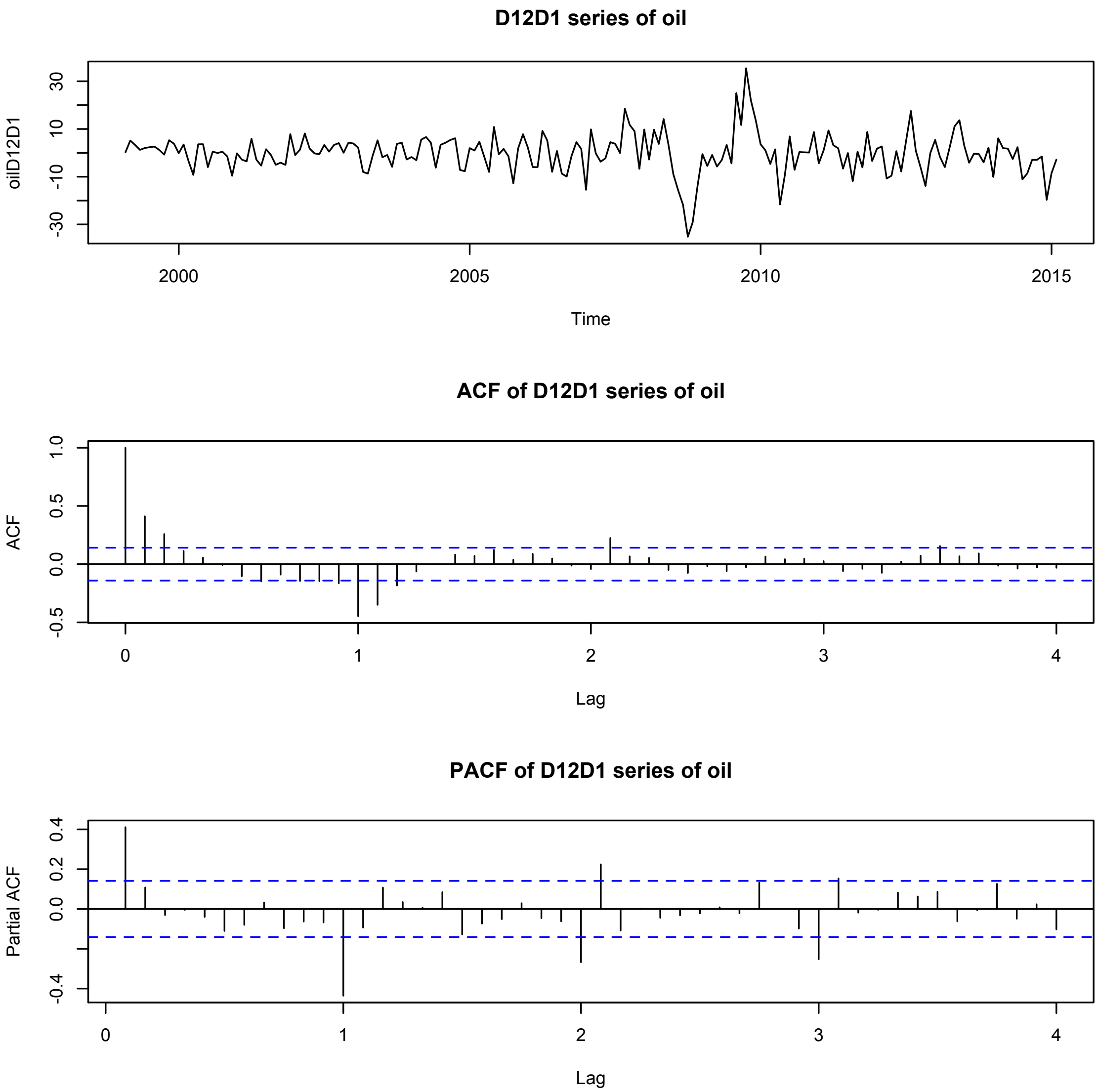

Figure 2.4 Plot, ACF and PACF of annual difference of Diff(oil)

I went back to take the annual difference of Diff(oil), indeed it now shows strong seasonal trends which we did not discover if we were to take only the first difference. Figure 2.4 indicates a potential seasonal ARIMA(p,1,q)×$(P,1,Q)_{12}$,p=1,2, q=1,2,3, P=1,2,3, Q=1 model. Table 2.3 indicates that the most simple model with minimal AIC or BIC is ARIMA(1,1,1)×$(2,1,1)_{12}$, its residual plots, Ljung-Box residual tests and histograms all suggest that we have found the right model this time, although Shapiro-Wilk normality test has highly significant p-value, I decided to mainly use the graphical criterions.

	p	d	q	P	D	Q	S	AIC	AICc	BIC
[1,]	1	1	1	1	1	1	12	4.387195	4.398360	3.451813
[2,]	**1**	**1**	**1**	**2**	**1**	**1**	**12**	**4.351081**	**4.362839**	**3.431855**
[3,]	1	1	1	3	1	1	12	4.356406	4.368861	3.453335

[4,]	1	1	2	1	1	1	12	4.392817	4.404575	3.473590
[5,]	1	1	2	2	1	1	12	4.357776	4.370231	3.454705
[6,]	1	1	2	3	1	1	12	4.363387	4.376644	3.476470
[7,]	1	1	3	1	1	1	12	4.402448	4.414903	3.499377
[8,]	1	1	3	2	1	1	12	4.367426	4.380683	3.480509
[9,]	1	1	3	3	1	1	12	4.373120	4.387287	3.502358
[10,]	2	1	1	1	1	1	12	4.393273	4.405031	3.474047
[11,]	2	1	1	2	1	1	12	4.359194	4.371649	3.456123
[12,]	2	1	1	3	1	1	12	4.364492	4.377749	3.477575
[13,]	2	1	2	1	1	1	12	4.350499	4.362953	3.447427

Table 2.3 All seasonal ARIMA models considered with AIC, AICc and BIC

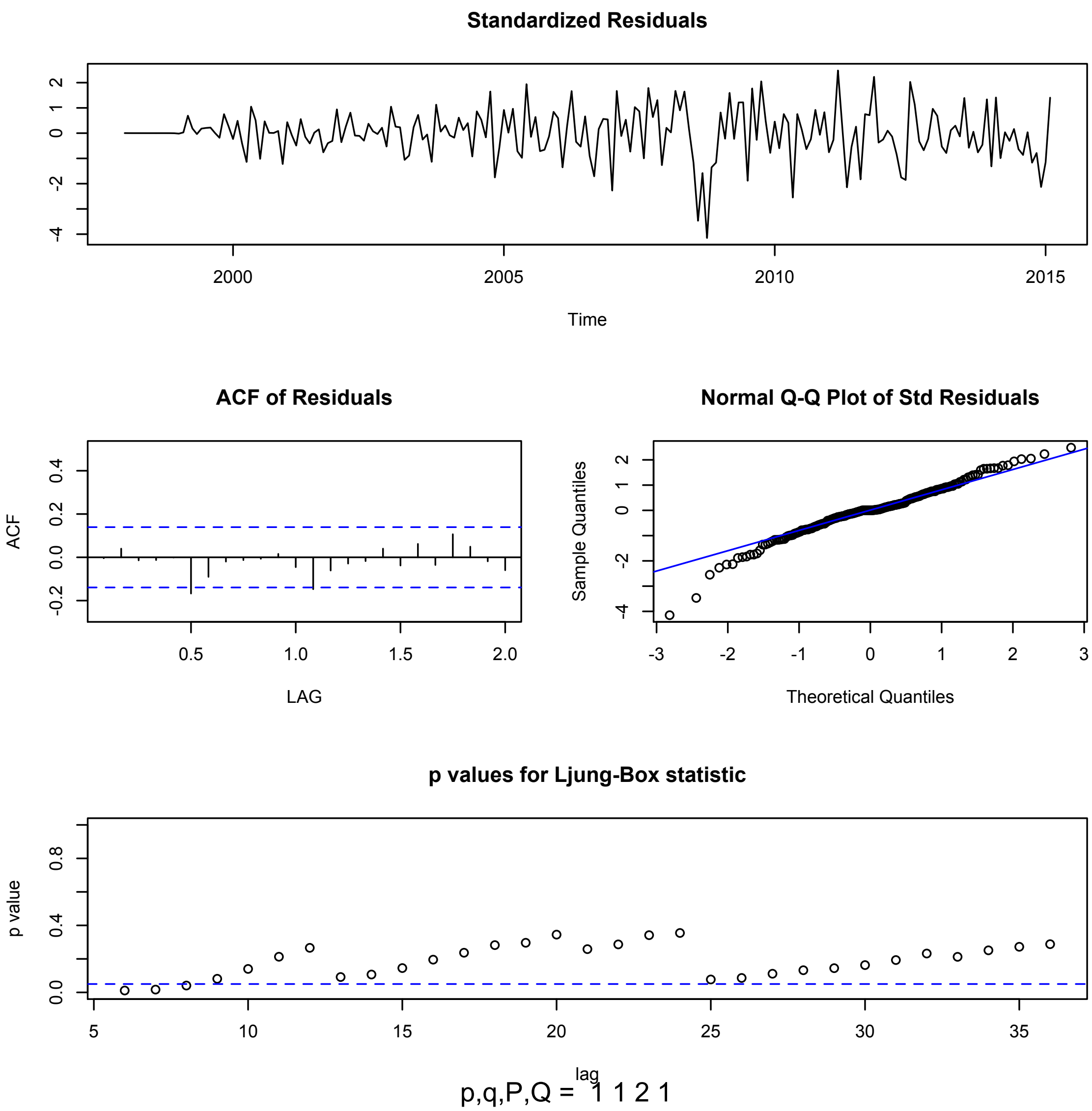

Figure 2.5 Residual plots of the seasonal ARIMA(1,1,1)×(2,1,1)$_{12}$ model for Oil

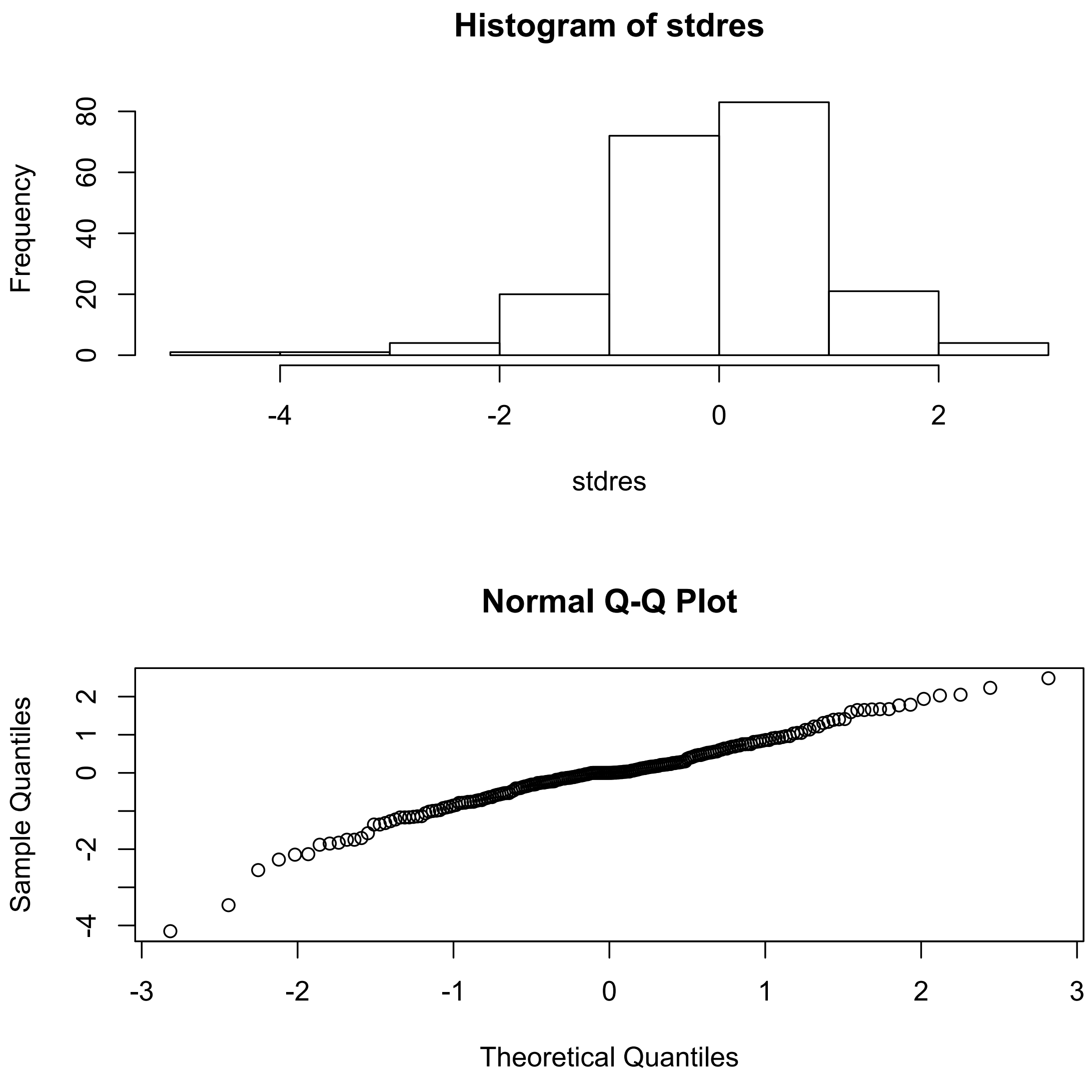

Figure 2.6 Histogram and QQ plot of the standardized residuals for oil

Shapiro-Wilk normality test

data: stdres
W = 0.9711, p-value = 0.0003102

Coefficients:

	ar1	ma1	sar1	sar2	sma1
	0.5310	**-0.1623**	**0.0487**	**-0.1626**	**-1.0000**
s.e.	**0.1335**	**0.1512**	**0.0745**	**0.0741**	**0.1426**

The final seasonal ARIMA model for oil is:

$(1+0.0487*B^{12}-0.1626*B^{24})*(1+0.5310*B)*(1-B^{12})*(1-B)*Xt = (1-B^{12})*(1-0.1623B)*wt$

Once the seasonal ARIMA models are obtained, we are able to forecast both series, which is the main interest in this part, but first it is better to see how well the forecast model would be, if we hold back an entire year's information and use the models to "predict" the values of last year and compare with all the actual values from the last year. If satisfying, we can move on to predicting next year's values.

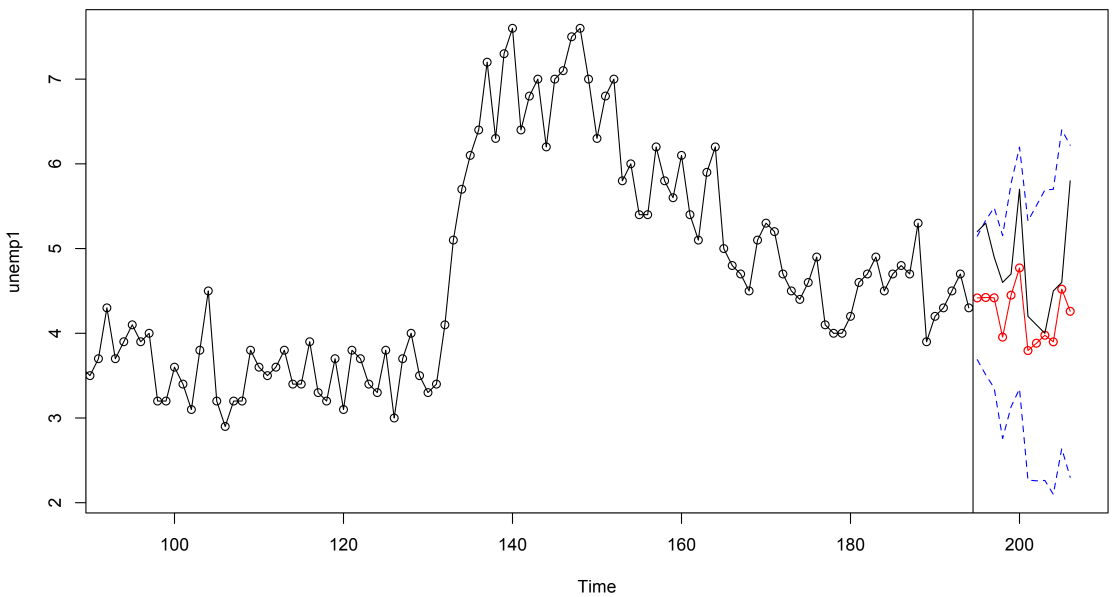

Figure 2.7 Forecast for the held-back-1-year unemp

	predictions	se	actual	error
195	4.417320	0.3644422	5.2	0.78268044
196	4.422482	0.4561202	5.3	0.87751825
197	4.418777	0.5322341	4.9	0.48122290
198	3.954880	0.5987493	4.6	0.64511951
199	4.449626	0.6585804	4.7	0.25037444
200	4.771371	0.7134113	5.7	0.92862914
201	3.796157	0.7643188	4.2	0.40384284
202	3.882486	0.8120411	4.1	0.21751388
203	3.977344	0.8571104	4.0	0.02265621
204	3.899089	0.8999254	4.5	0.60091140
205	4.520926	0.9408081	4.6	0.07907416
206	4.259324	0.9798460	5.8	1.54067625

Table 2.4 Comparison of actual data vs. predicted data for unemployment

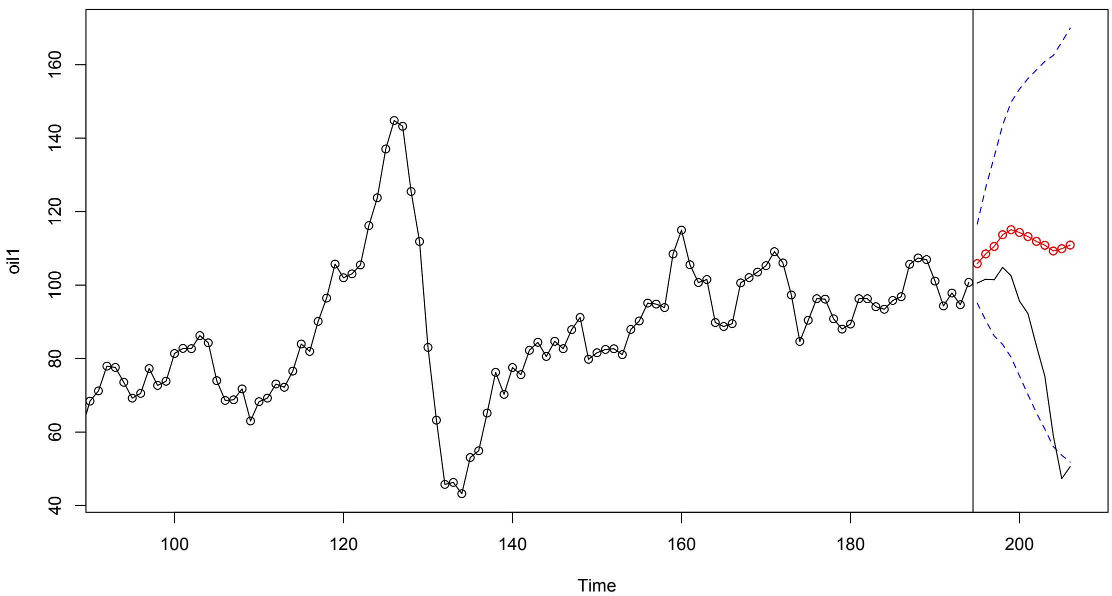

Figure 2.8 Forecast for the held-back-1-year oil

	predictions	se	actual	error
195	105.8477	5.379689	100.54179	-5.305877
196	108.4789	9.025104	101.61461	-6.864287
197	110.5131	12.172823	101.41931	-9.093796
198	113.7003	14.918133	104.82636	-8.873933
199	115.0705	17.345152	102.53926	-12.531271
200	114.3155	19.522715	95.63604	-18.679436
201	113.1986	21.503586	92.25290	-20.945747
202	111.8974	23.327227	83.48874	-28.408668
203	110.8432	25.023028	75.18865	-35.654564
204	109.2752	26.612950	59.01448	-50.260699
205	109.8928	28.113270	47.32242	-62.570395
206	110.8957	29.534149	50.58000	-60.315712

Table 2.5 Comparison of actual data vs. predicted data for oil

The model for unemployment is pretty decent, while the oil predictor did a poor job. Our oil predictor definitely cannot foresee the price plummet for the last several months, so there is really not much to blame about our model. Below are the predictions for the next twelve months, just out of curiosity.

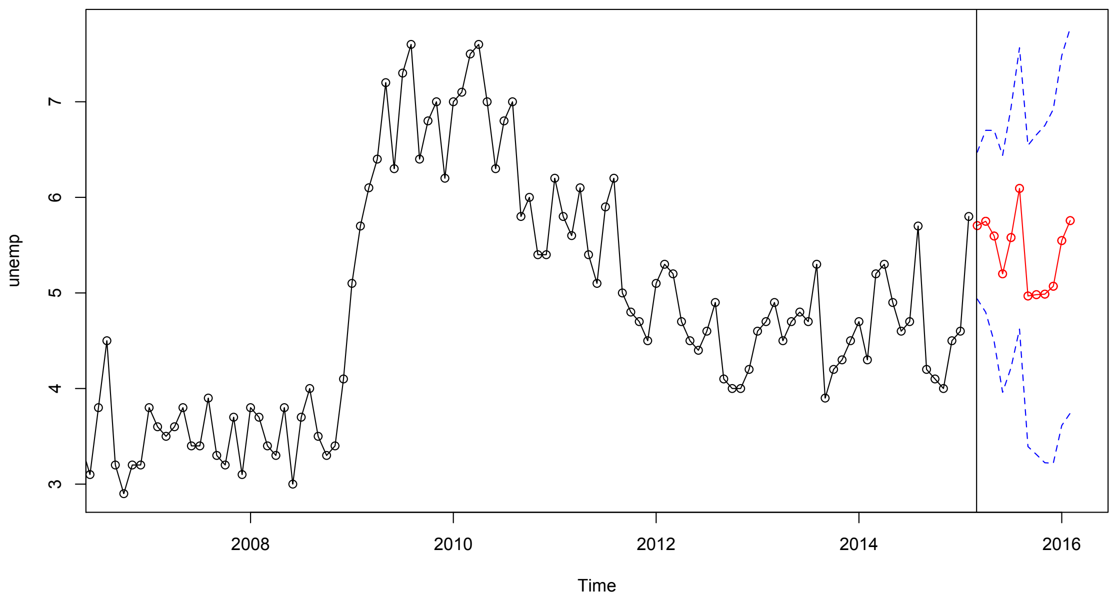

Figure 2.9 Forecast of unemployment rate for the next 12 months

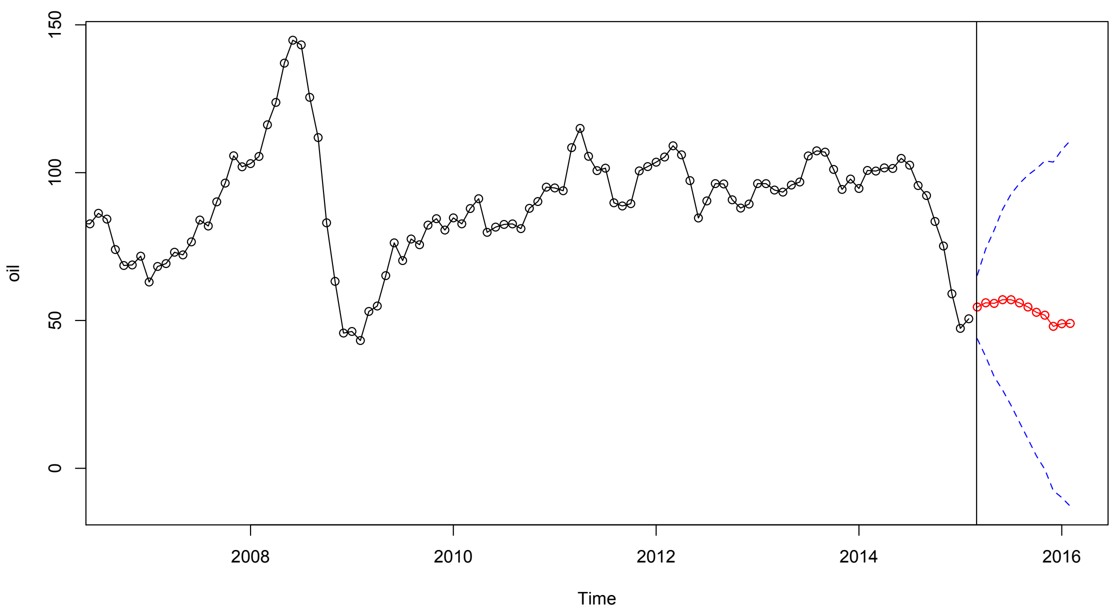

Figure 2.10 Forecast of oil price for the next 12 months

3. Frequency Domain Analysis:

In this part we examine any potential relationships between unemployment rate and oil price. First we will observe the periodograms and cross-periodograms to find interesting frequencies that these two series agree on, so we can apply filters and in the end obtain a lagged regression model with oil price as input and unemployment rate as output.

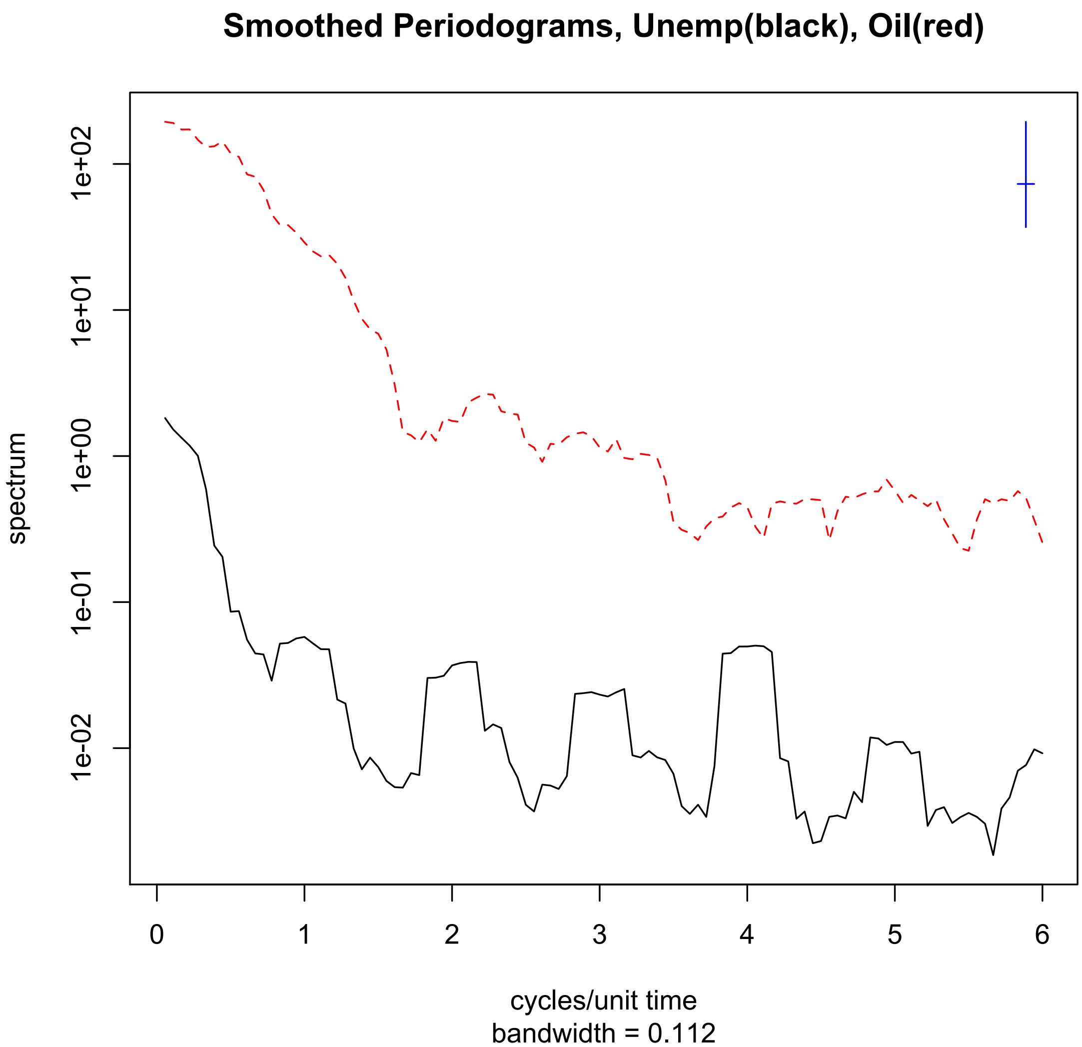

Figure 3.1 Smoothed Periodograms for both series with L=7

Figure 3.1 is the log transformed smoothed periodograms for both series with a smoothing factor L=7, unit time here is in years. As expected the unemployment rate shows strong seasonality with large powers at low frequencies (v<1, which means less than 1 cycle per year) and at annual/seasonal frequencies such as when v is around 1/2/3/4 cycles per year (annual/biannual/4-months/seasonal periods). On the other hand oil prices are less apparent as a seasonal series, since its powers are mainly at low frequencies. From the periodograms we suspect that there might be relationships between the two series in low frequencies.

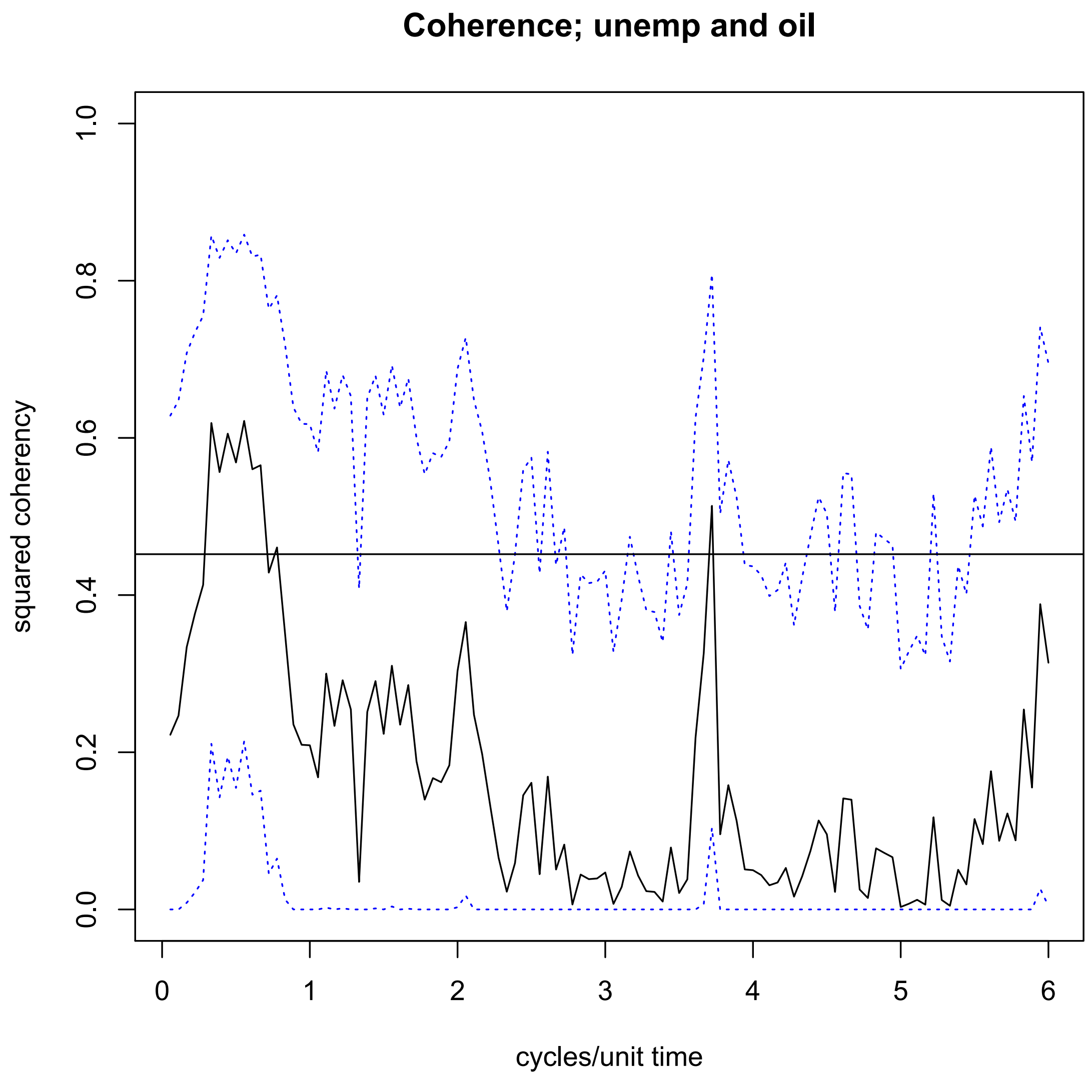

Figure 3.2 Coherence plot between the two series, with alpha=0.05

Figure 3.2 shows the coherence plot with horizontal line indicating significance bound for coherence at alpha = 0.05. Much as expected, coherence between the two series is strongest when the frequencies are less than 1 cycle/year, but there is also a very narrow pike around 4 cycles/year i.e. seasonal coherence. However, we will mainly look at the larger cycles. Attempting to filter out the nuisance frequencies we will look at Table 3.1 to determine the band of frequencies we would like to keep. It seems that v<0.09 would be the general area of focus, but without any signal extraction there can be no precise conclusions.

	freq	UnempPW	OilPW
[1,]	0.004629630	1.819604195	194.3335611
[2,]	0.009259259	1.519136262	190.8056296
[3,]	0.013888889	1.334437631	172.0722377

[4,]	0.018518519	1.178066299	172.3786918
[5,]	0.023148148	1.004315040	146.2163356
[6,]	0.027777778	0.591131107	130.4446976
[7,]	0.032407407	0.243556316	132.1455389
[8,]	0.037037037	0.204028071	142.5836610
[9,]	0.041666667	0.085951976	118.2417450
[10,]	0.046296296	0.086758446	111.9607072
[11,]	0.050925926	0.055206553	84.9191298
[12,]	0.055555556	0.044520593	81.6931264
[13,]	0.060185185	0.043827634	66.4145694
[14,]	0.064814815	0.028953934	45.2916095
[15,]	0.069444444	0.051969940	38.3049083
[16,]	0.074074074	0.052501053	38.1327524
[17,]	0.078703704	0.056359912	33.8187937
[18,]	0.083333333	0.057751291	29.0259598
[19,]	0.087962963	0.052320801	25.2685675
[20,]	0.092592593	0.047605156	23.3172934
[21,]	0.097222222	0.047532499	23.8025017
[22,]	0.101851852	0.021511269	20.7281621
[23,]	0.106481481	0.020205690	16.5968519
[24,]	0.111111111	0.009956351	11.5645611
[25,]	0.115740741	0.007158843	8.7027351
[26,]	0.120370370	0.008610804	7.3788715
[27,]	0.125000000	0.007415470	6.8752750

Table 3.1 Frequencies and powers of the two series, frequency measured in months here

Now it is time to apply linear filters to highlight the peak powers at low frequencies. Both extraction procedures are done with smoothing factor L=7, number of terms used in the lagged regression approximation M=32 and max frequency of 0.09. The results are as in Figure 3.3 and 3.4. The attained frequencies for both series are within a desired range, and after filtering all lower frequencies are highlighted, still quite similar to the original periodograms. In conclusion our linear filters are close to optimal.

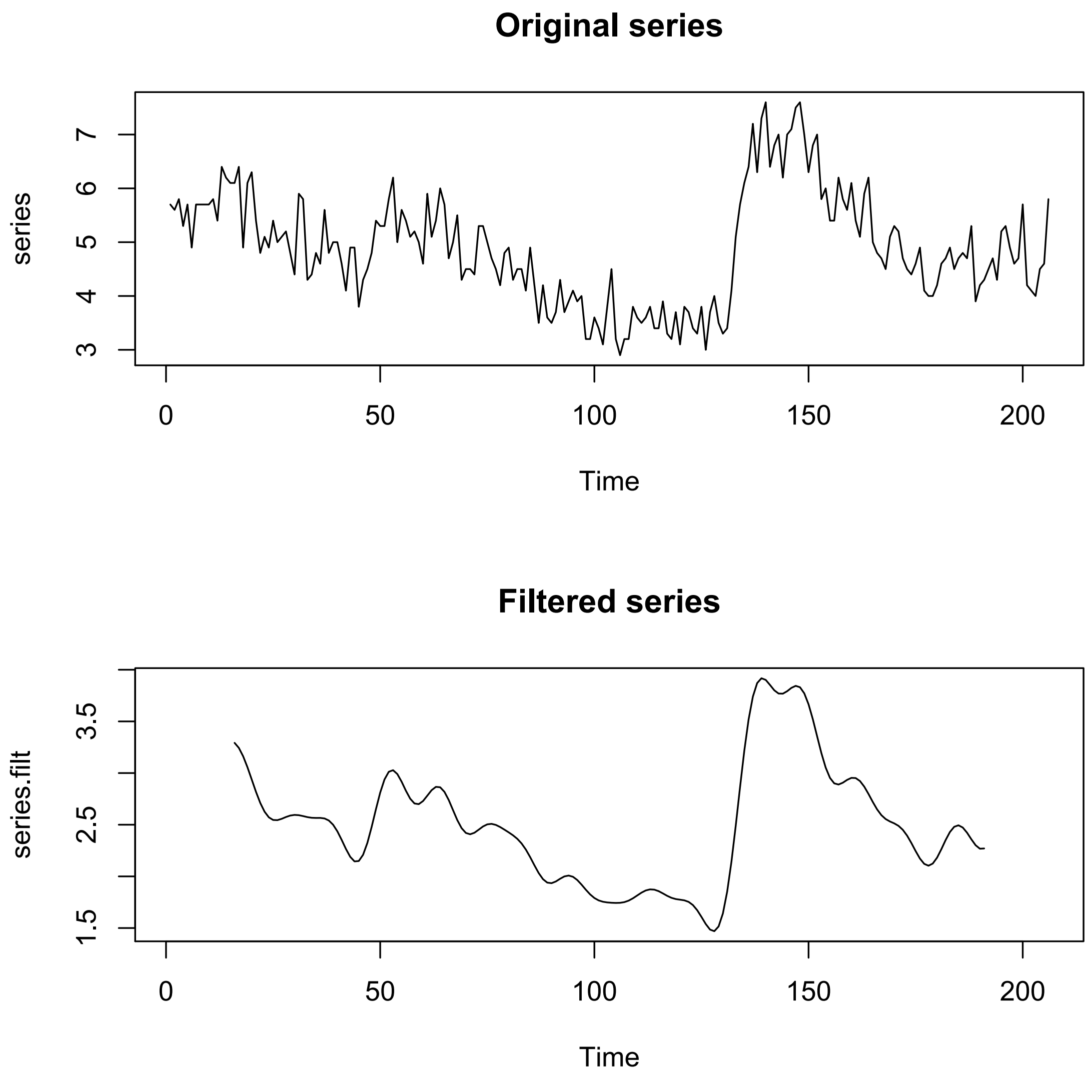

Figure 3.3 a) Original series and filtered series for Unemployment

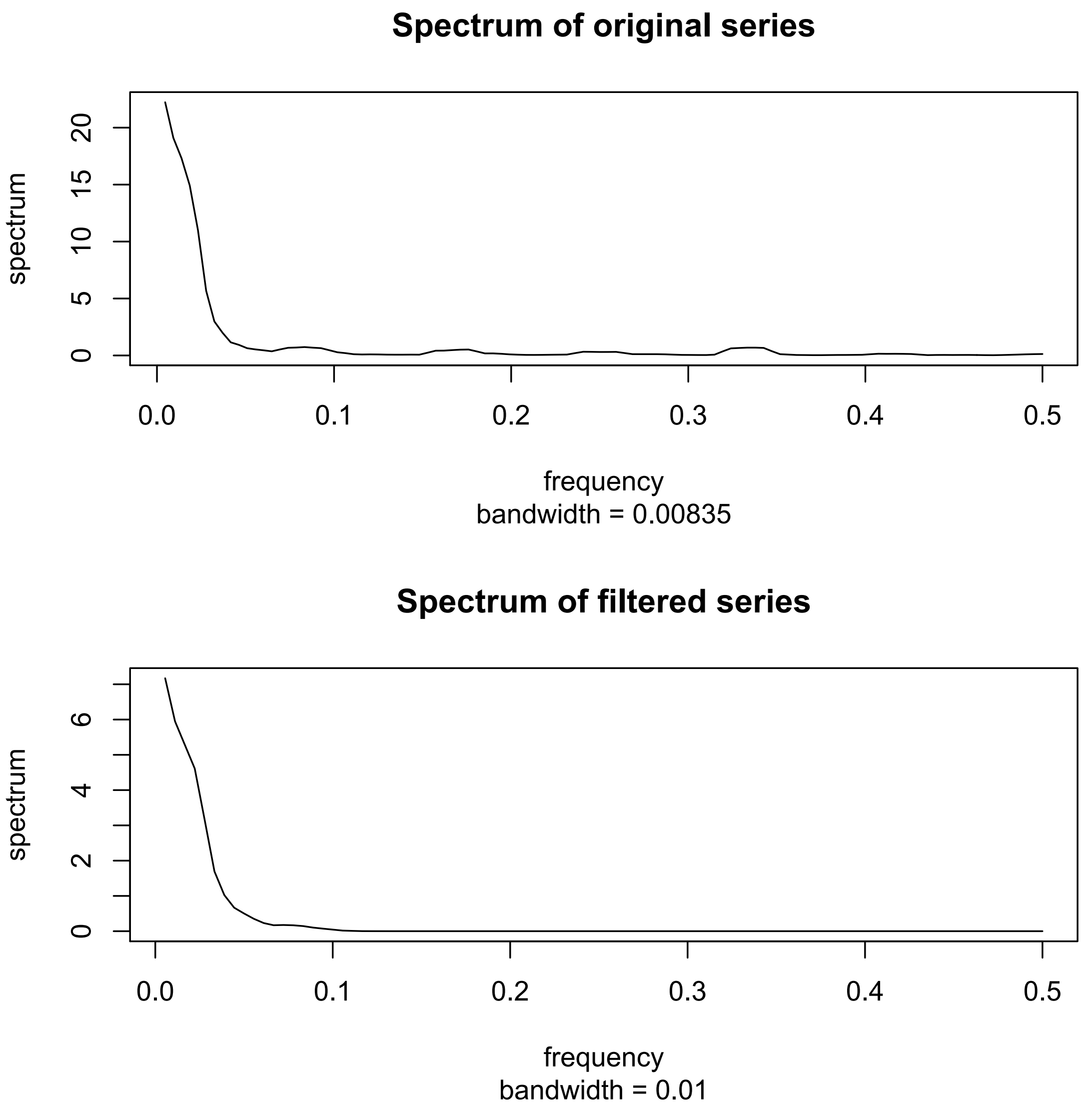

Figure 3.3 b) Spectrum of original and filtered Unemployment

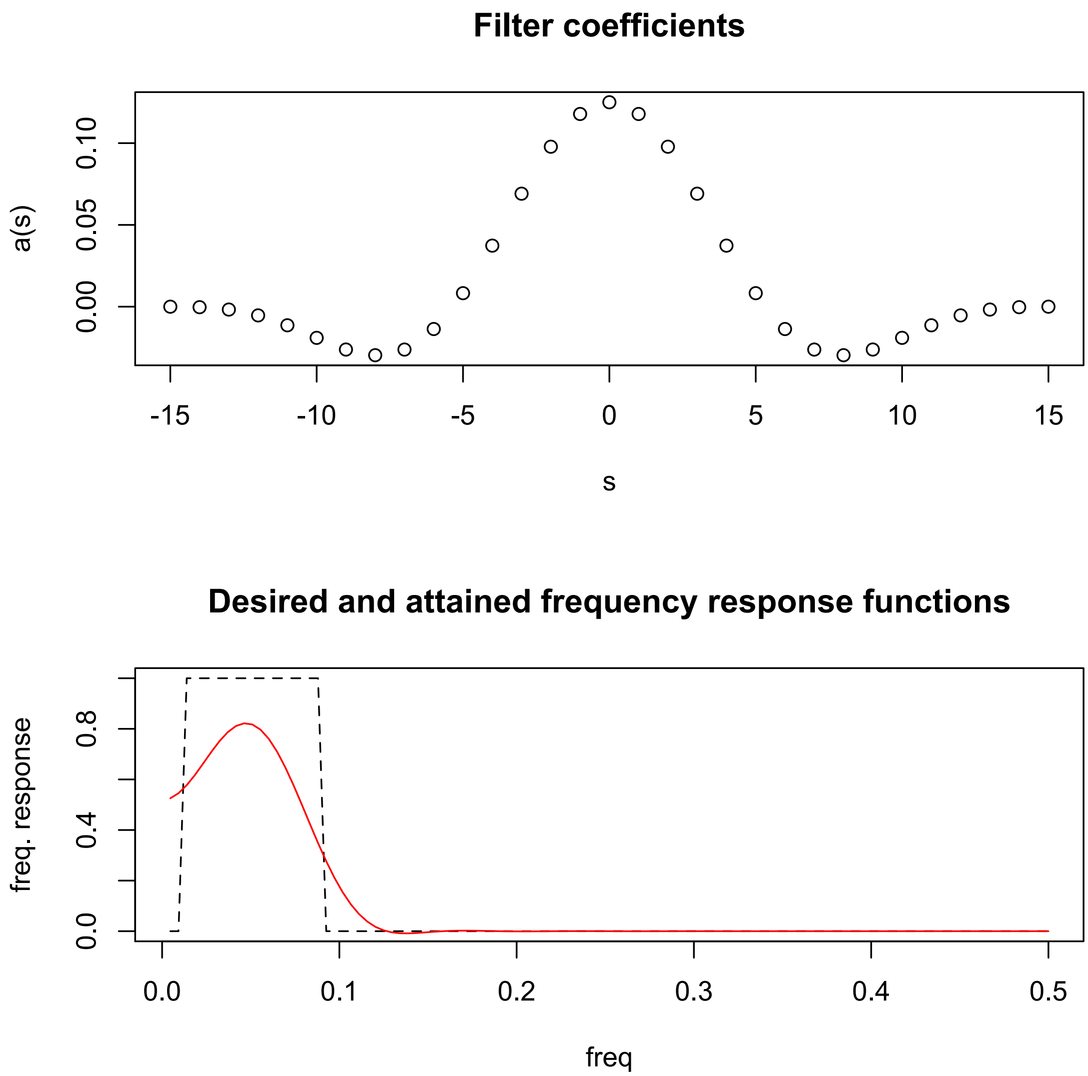

Figure 3.3 c) Filter coefficients and frequency response functions for Unemployment

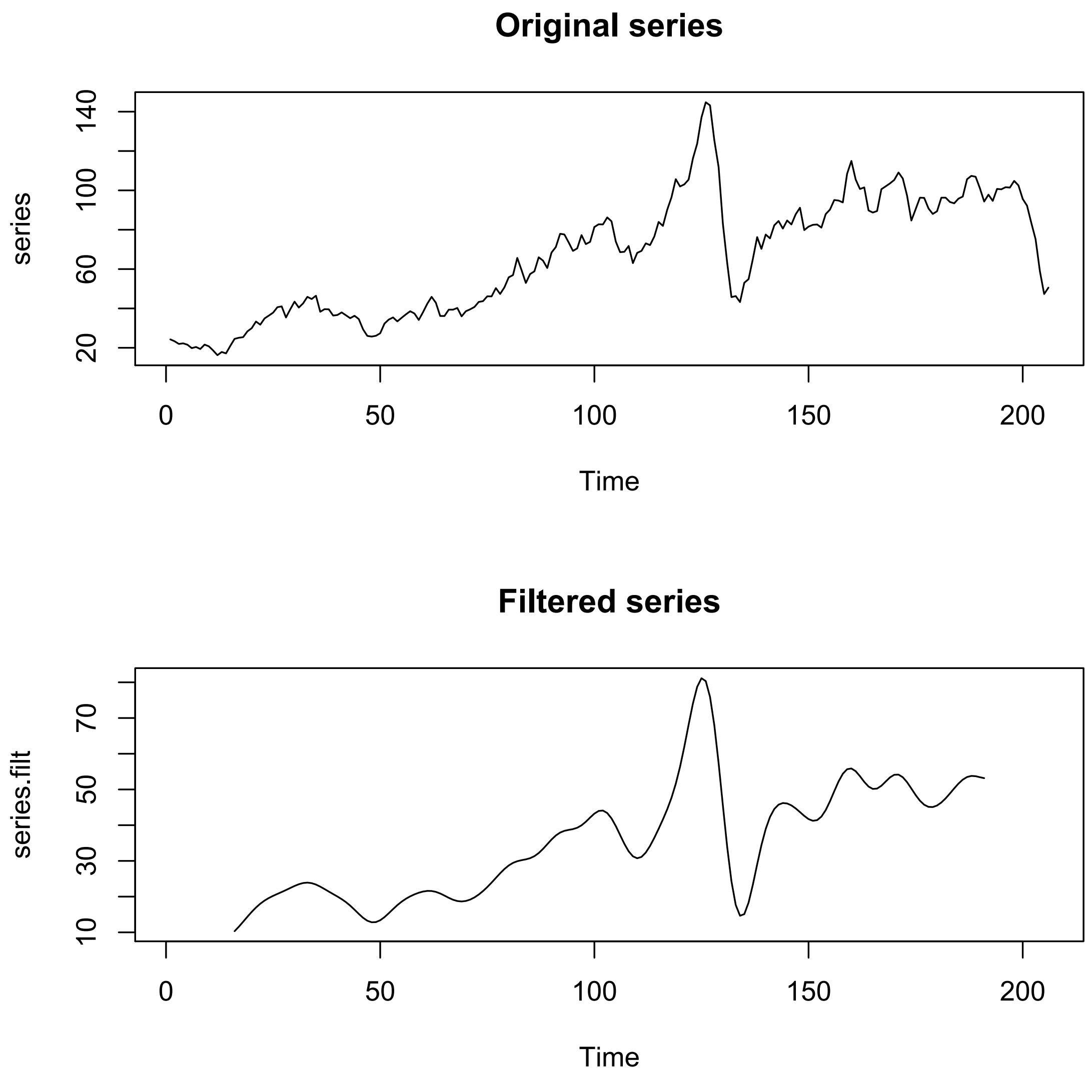

Figure 3.4 a) Original series and filtered series for Oil

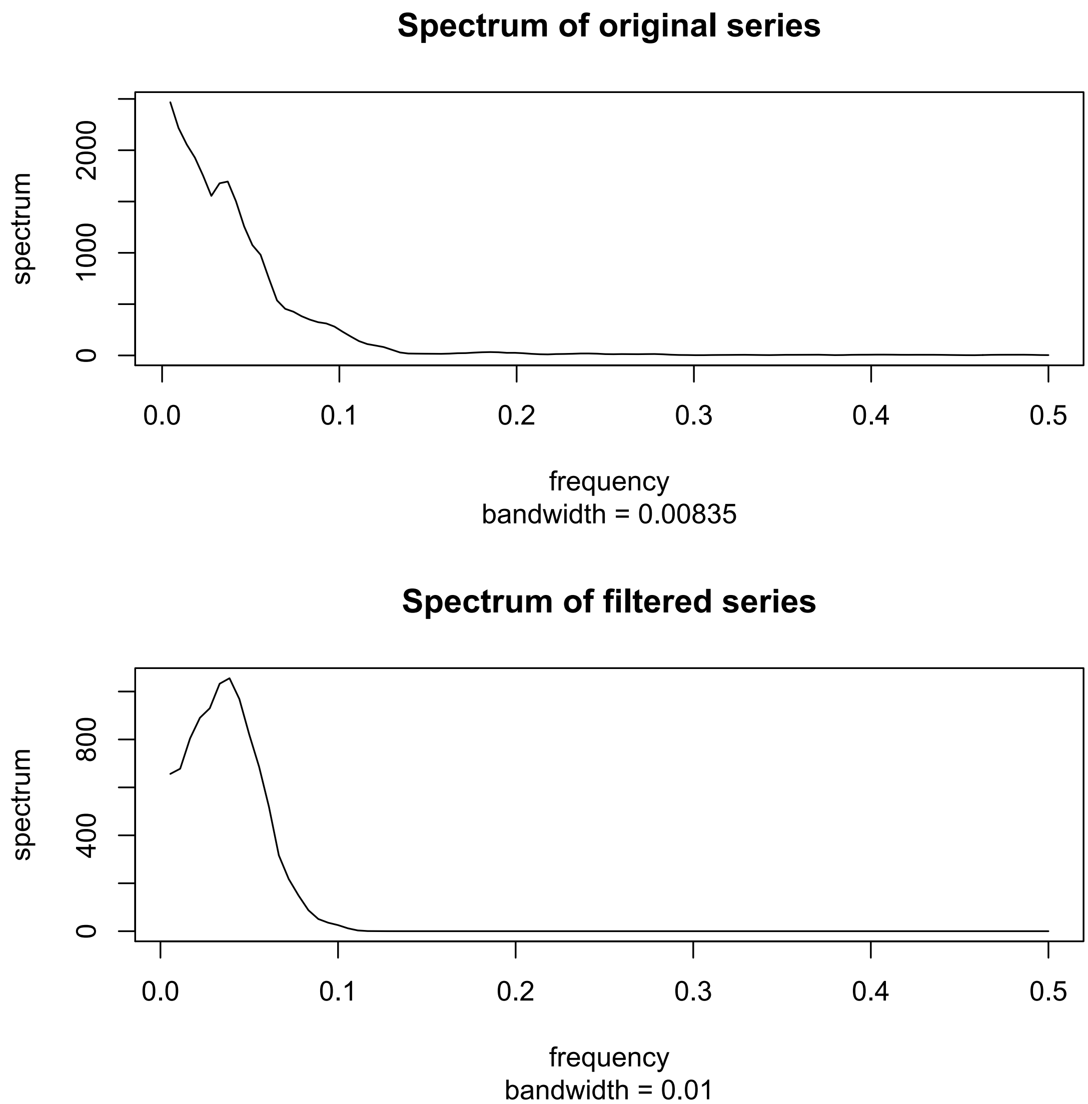

Figure 3.4 b) Spectrum of original and filtered Oil

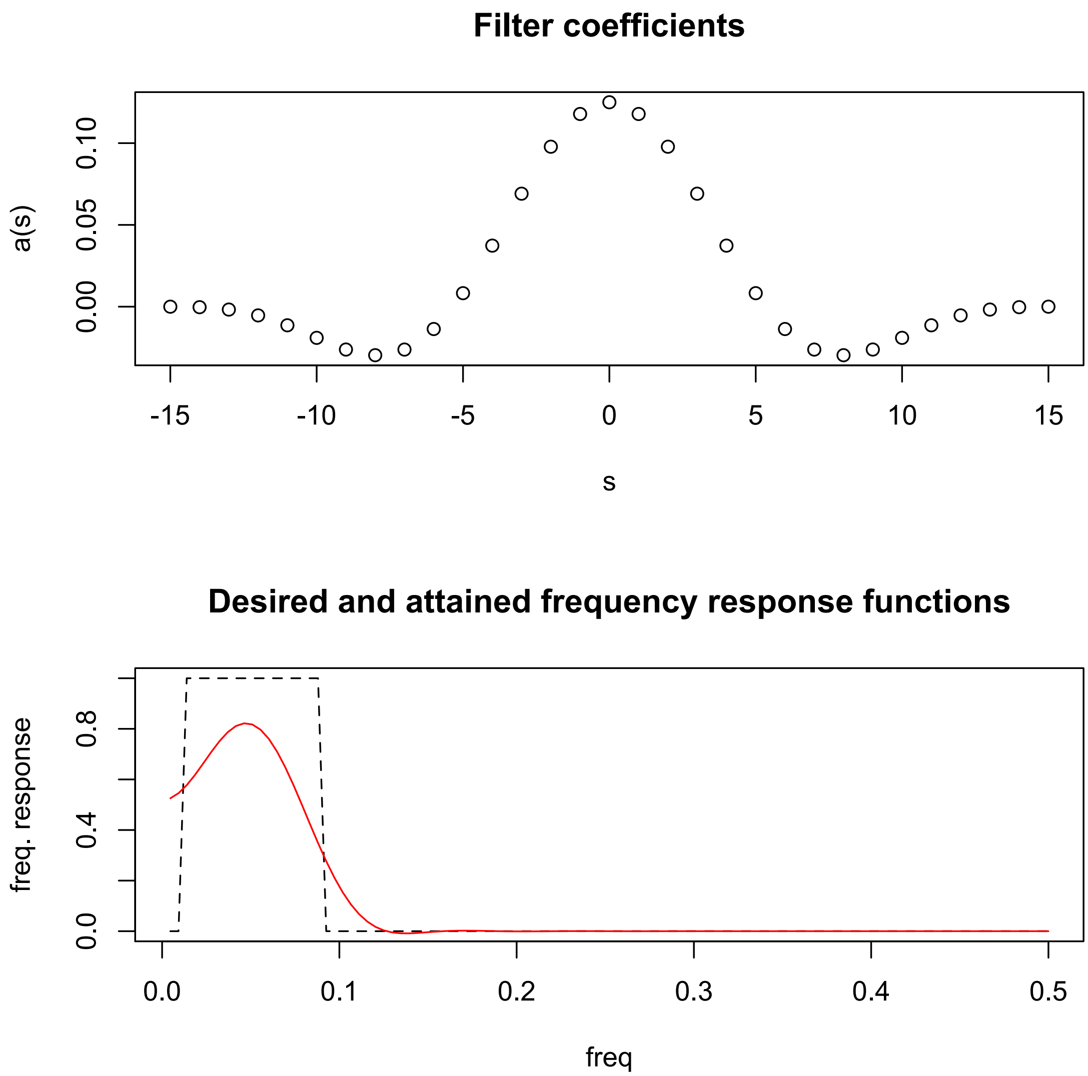

Figure 3.3 c) Filter coefficients and frequency response functions for Oil

Last but not least, lagged regression will be fitted to predict unemployment rate from oil. To obtain a rather precise model for unemployment, I decided to use inversed lagged-regression with unemp as input and oil as output, save two beta's so that the final model, after re-arranging and shifting coefficients, would contain both unemp and oil as inputs at different lags, with current unemp as output.

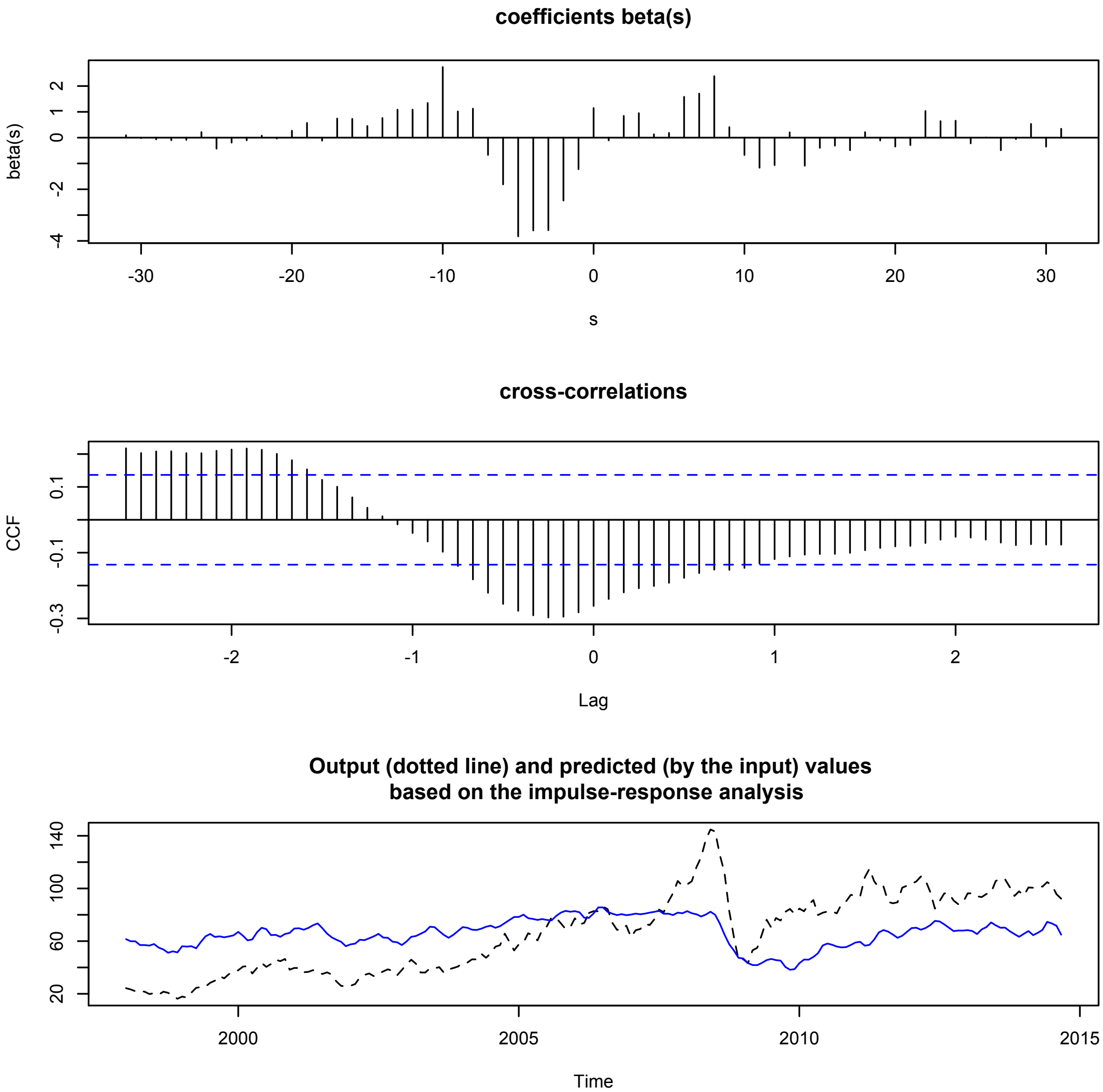

Figure 3.4 Inversed Lagged regression with Output=oil and Input=unemp

	lag s	beta(s)
[1,]	2	-3.586883
[2,]	3	-3.596436
[3,]	4	-3.823426

Table 3.2 large beta's and corresponding lags (The lags are off by 1, so the correct lags should be 3,4 and 5)

If we indeed use re-arranging and shifting coefficients, the final model would be less accurate than that if we re-fit a dynamic regression model, given that all the input variables are already obtained from the previous lagged-regression. Table 3.3 shows the final regression model. Figure 3.5 shows that this model is fairly accurate and predicts unemployment rate well, as the prediction matches most of the real values.

Time series regression with "ts" data:
Start = 1998(6), End = 2015(2)

Call:
dynlm(formula = unemp ~ L(unemp, 1) + L(unemp, 2) + L(oil, 5))

Coefficients:

(Intercept)	L(unemp, 1)	L(unemp, 2)	L(oil, 5)
0.6410374	0.6494843	0.2274589	-0.0006481

Final model: Unemp(t) = 0.6410374+0.6494843*Unemp(t-1)+0.2274589*Unemp(t-2)-0.0006481*Oil(t-5)+w(t)

Table 3.3 Final model and its coefficients

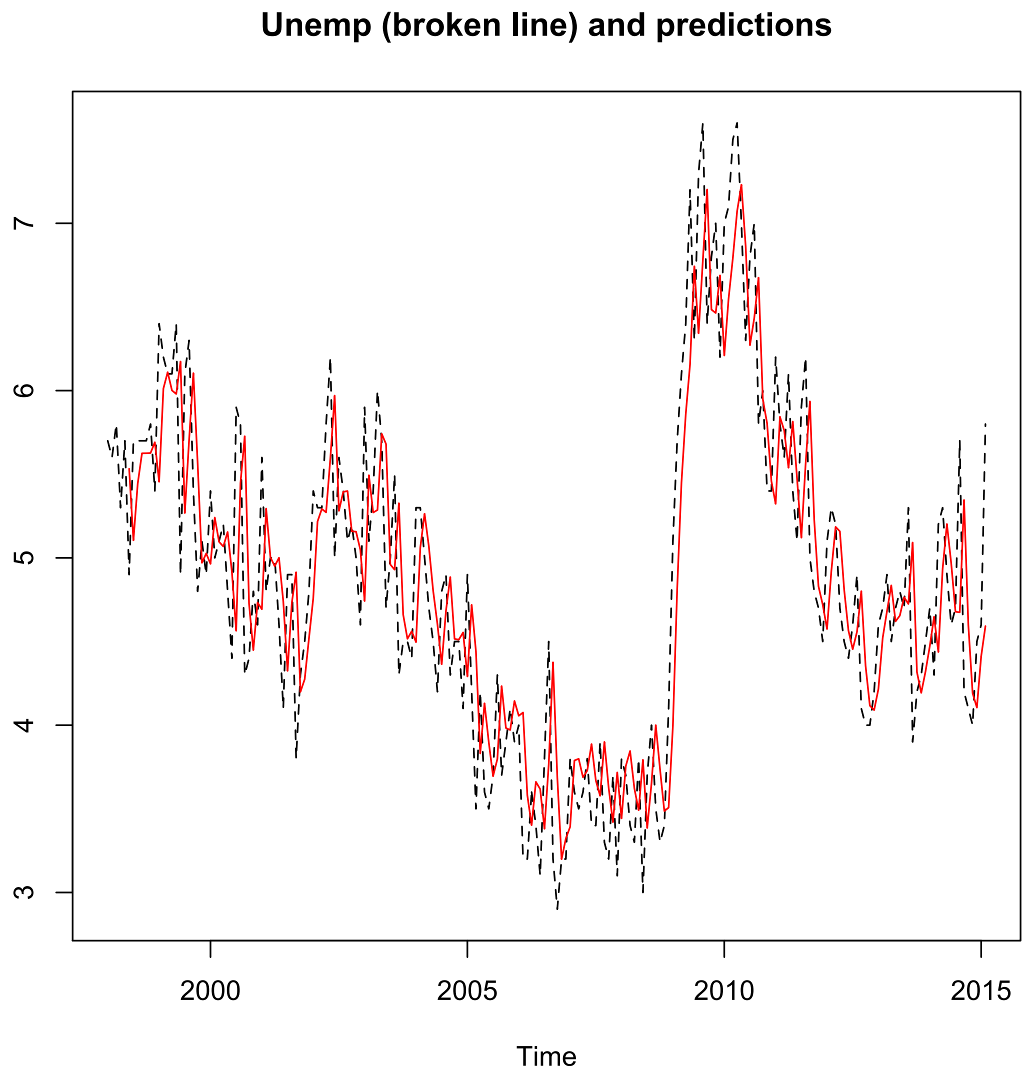

Figure 3.5 Unemployment rate and its predictions shown in red

4. Summary:

Both unemployment rate and oil are non-stationary series; the best way to transform them into stationary series is to take an annual difference of the first difference.

Both series can fit seasonal ARIMA models. The final seasonal ARIMA model for unemployment rate is:

$$(1-0.1127*B^{12})*(1-B^{12})*(1-B)*Xt = (1-0.8475*B^{12})*(1-0.2692B)*wt$$

While the final seasonal ARIMA model for oil is:

$$(1+0.0487*B^{12}-0.1626*B^{24})*(1+0.5310*B)*(1-B^{12})*(1-B)*Xt = (1-B^{12})*(1-0.1623B)*wt$$

We used these two seasonal ARIMA models to predict future values, however since there are other factors that can potentially affect them, the predictors are not very informative.

Coherence plot indicates significant correlation between the two series at low frequencies, the final model to predict unemployment rate from its own past as well as the lagged oil price is:

$$Unemp(t) = 0.6410374+0.6494843*Unemp(t-1)+0.2274589*Unemp(t-2)-0.0006481*Oil(t-5)+w(t)$$

We should also keep in mind that 08OC contributed more errors to all the models we fitted. To summarize, oil price leads unemployment of Alberta by about 5 months.

5.References:

1. http://oilsands.alberta.ca/economicinvestment.html
2. http://www5.statcan.gc.ca/cansim/a47
3. http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=rwtc&f=m