

MATH 545 - ASSIGNMENT 5 SOLUTIONS

The data in the file `USCoal-2016.csv` contains monthly observations from January 2001 to December 2016 (192 observations) of the amount (in tons) of coal consumed in energy production in each of five US states (California, Florida, Illinois, North Carolina and Pennsylvania).

Write a report on the analysis of these data sets, commenting on the following aspects

- (a) the deterministic and stochastic time series components for the states; 8 Marks
- (b) similarities and/or differences between the states; 4 Marks
- (c) monthly forecasts for each state for year 2017; 4 Marks
- (d) any other structure or features you deem relevant, or suggested further analyses. 4 Marks

Limit your report to twenty sides of paper including figures and tables. Supply your analysis code as an appendix (of unlimited length).

```
USCoal<-read.csv('USCoal-2016.csv',header=TRUE)
Y.ca<-ts(USCoal[,3],start=c(2001,1),frequency=12)
Y.fl<-ts(USCoal[,4],start=c(2001,1),frequency=12)
Y.il<-ts(USCoal[,5],start=c(2001,1),frequency=12)
Y.nc<-ts(USCoal[,6],start=c(2001,1),frequency=12)
Y.pa<-ts(USCoal[,7],start=c(2001,1),frequency=12)
```

Solutions: There are many possible analyses, and the one included in the following pages is one possible. For full marks, the following key points need to be addressed:

- (a) For each state, you need to note the time-varying mean and seasonality of 12 months, where seasonality could be deterministic or stochastic, and handle it in a suitable way. For the stochastic part, the ARMA structure should be addressed. You should note that California is very different from the other states.

In R, decomposition methods (`decompose`, `stl` etc.) can be used if the results are clearly explained. To identify ARMA model structure `arima` and `auto.arima` can be used, but `auto.arima` results should be explained also. 8 Marks

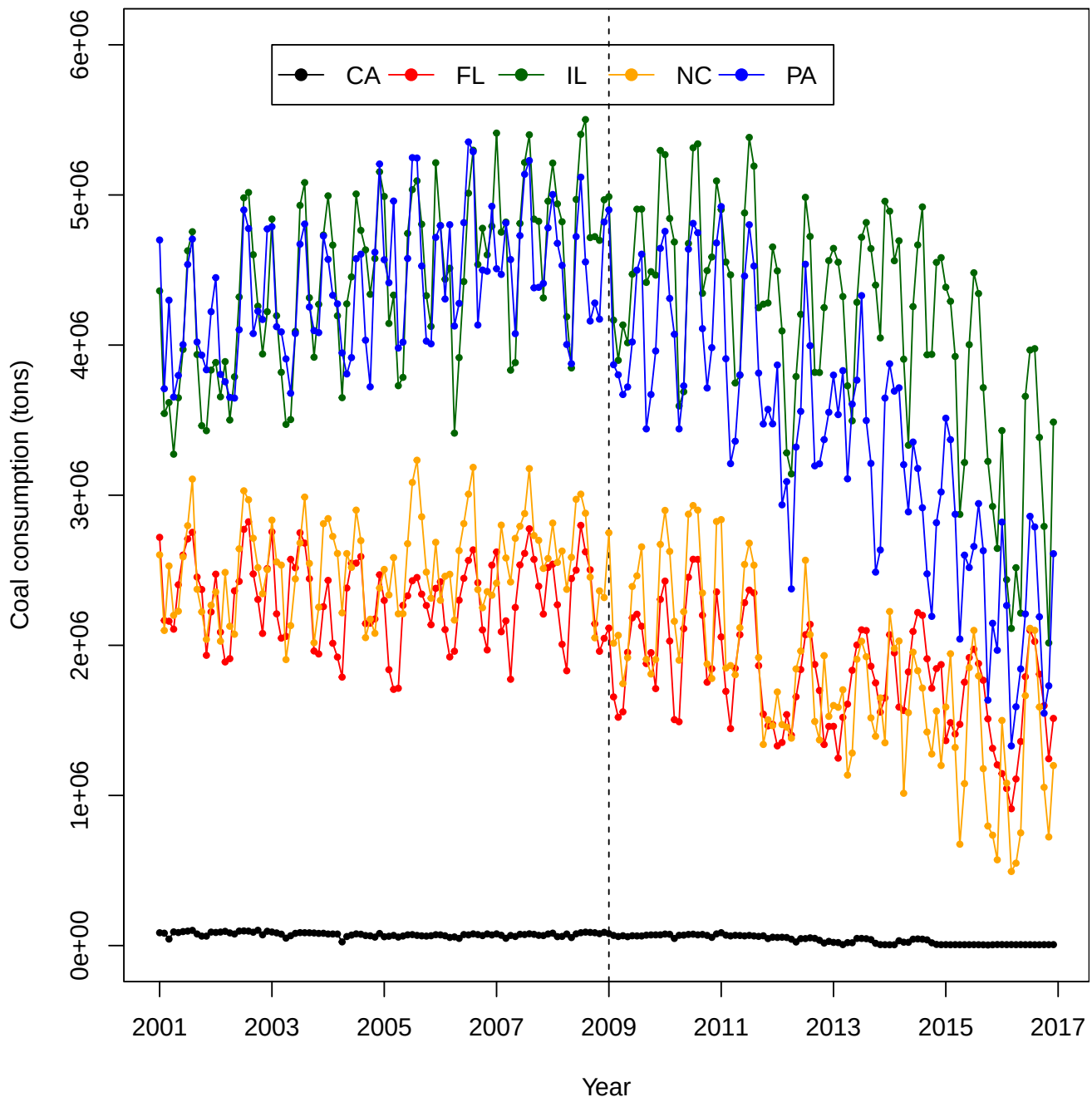
- (b) Should note that California is different at the end of the series, but that otherwise the 12 month seasonality is common to all states. Should note that differencing is usually helpful 4 Marks
- (c) Can use the `forecast` capability in R. 4 Marks
- (d) This depends on what else has been done, but for example you could try stochastic seasonal structures and state space models. 4 Marks

As appropriate, there can be reassignment of marks across the parts.

Detailed Solutions:

A plot of the raw data reveals some common elements:

```
par(mar=c(4,4,1,0));yl<-'Coal consumption (tons)'  
plot(Y.ca,pch=19,xaxt='n',ylim=range(0,6e6),  
      xlab='Year',ylab=yl)  
points(Y.ca,pch=19,cex=0.5);abline(v=2009,lty=2)  
axis(1, xaxp = c(2001, 2017, 8))  
lines(Y.fl,pch=19,col='red');points(Y.fl,pch=19,col='red',cex=0.5)  
lines(Y.il,pch=19,col='darkgreen');points(Y.il,pch=19,col='darkgreen',cex=0.5)  
lines(Y.nc,pch=19,col='orange');points(Y.nc,pch=19,col='orange',cex=0.5)  
lines(Y.pa,pch=19,col='blue');points(Y.pa,pch=19,col='blue',cex=0.5)  
legend(2003,6e6,c('CA','FL','IL','NC','PA'),pch=19,  
       col=c('black','red','darkgreen','orange','blue'),lty=1,horiz=TRUE)
```

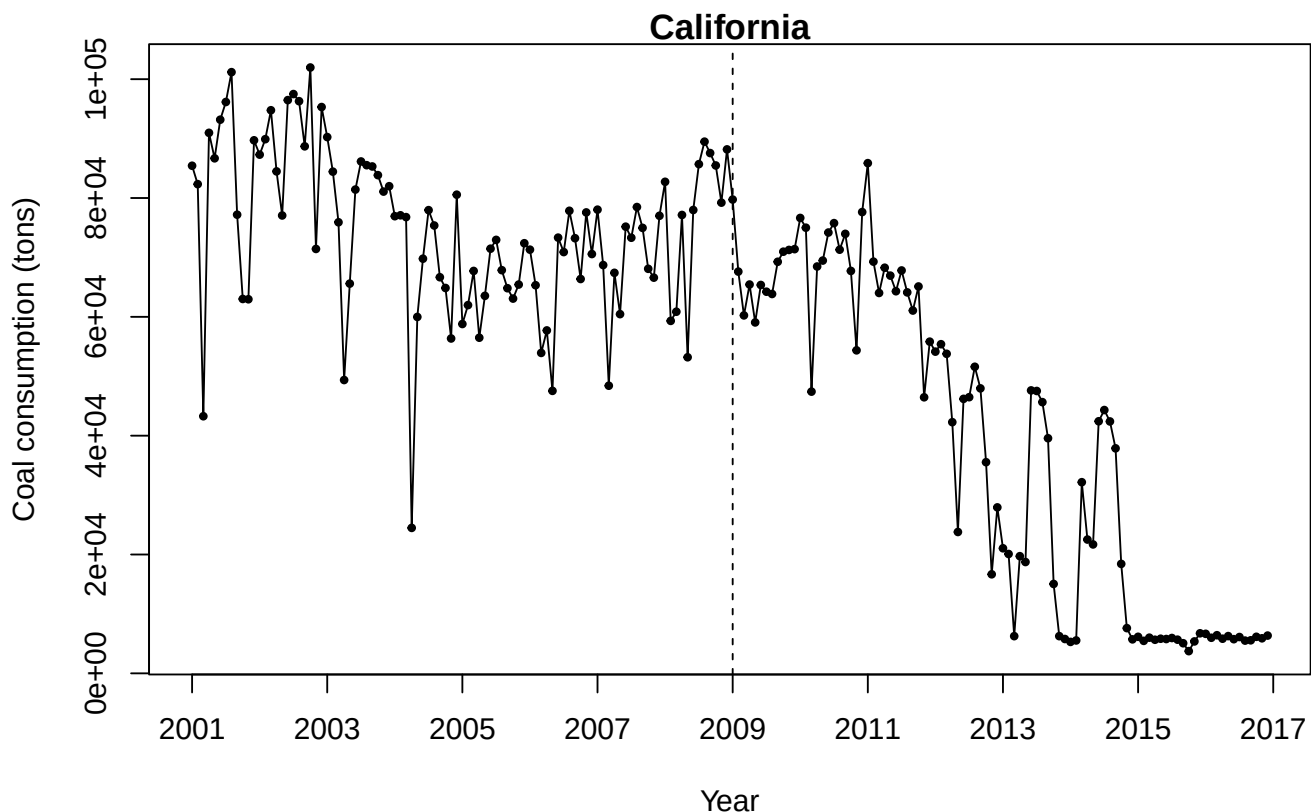


It is evident that

- There is a similarity, non linear change in mean level of production in all five series; there is a decline in the 2009-2017 period, after consumption at a reasonably constant level 2001-2008;
- There is seasonality in the data;
- Illinois and Pennsylvania have similar consumption levels, as to Florida and North Carolina;
- The level of consumption for California is considerably lower than for the other four states, and the general pattern is much less predictable.

The data for California exhibit particular behaviour:

```
par(mar=c(4,4,1,0))
plot(Y.ca,pch=19,xaxt='n',xlab='Year',ylab='Coal consumption (tons)')
points(Y.ca,pch=19,cex=0.5);abline(v=2009,lty=2)
axis(1,xaxp=c(2001,2017,8))
title('California')
```



The general level for California is much lower than for the other states, and also in the post 2015 period the level of consumption is constant at around 6000 tons per month.

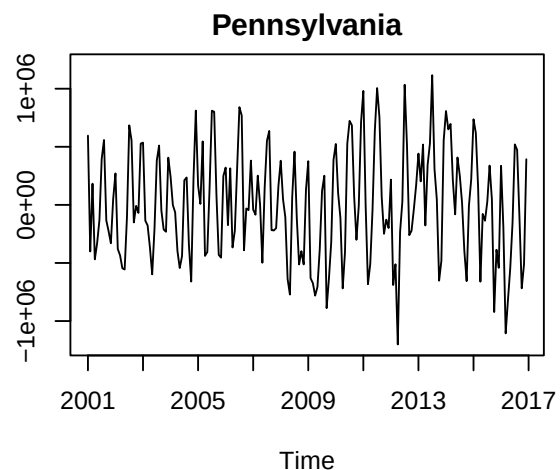
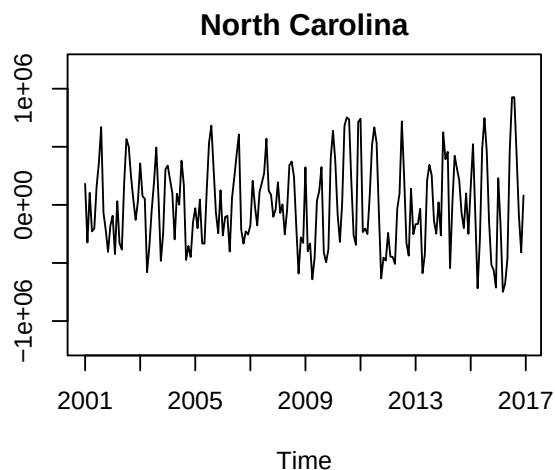
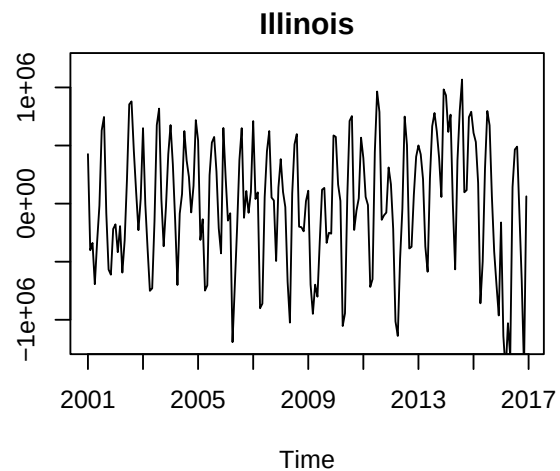
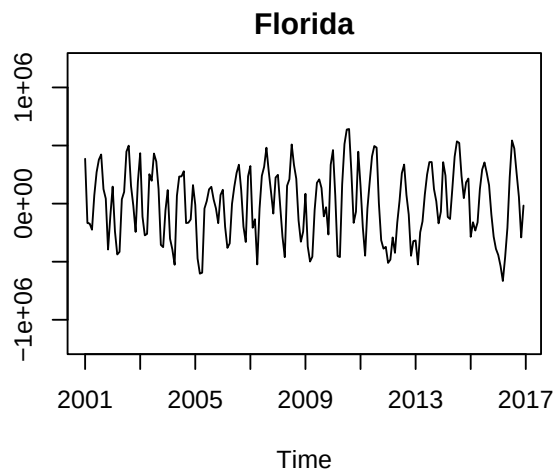
```
Y.ca[97:192]
```

```
+ [1] 79743 67612 60243 65438 59059 65364 64231 63848 69269 70967 71261 71394
+ [13] 76645 75000 47414 68471 69471 74188 75773 71302 73965 67727 54362 77635
+ [25] 85867 69288 64018 68266 66937 64309 67805 64114 61059 65109 46459 55821
+ [37] 54142 55392 53777 42274 23793 46194 46476 51596 47959 35536 16668 27931
+ [49] 21047 20087 6259 19716 18738 47617 47520 45645 39568 15046 6259 5780
+ [61] 5289 5533 32170 22518 21697 42424 44316 42396 37869 18420 7609 5722
+ [73] 6151 5467 6003 5653 5826 5773 5947 5665 5066 3739 5369 6741
+ [85] 6649 5990 6416 5828 6266 5748 6113 5510 5560 6150 5899 6354
```

Therefore, for California, a sensible forecast would be the mean of observations in 2015 and 2016, that is, 6042.36.

For the other four states, we can perform a simple de-meaning by fitting a piecewise linear mean to the data with a breakpoint at the start of 2009 (after 96 observations). This can be easily done using the `lm` function in R.

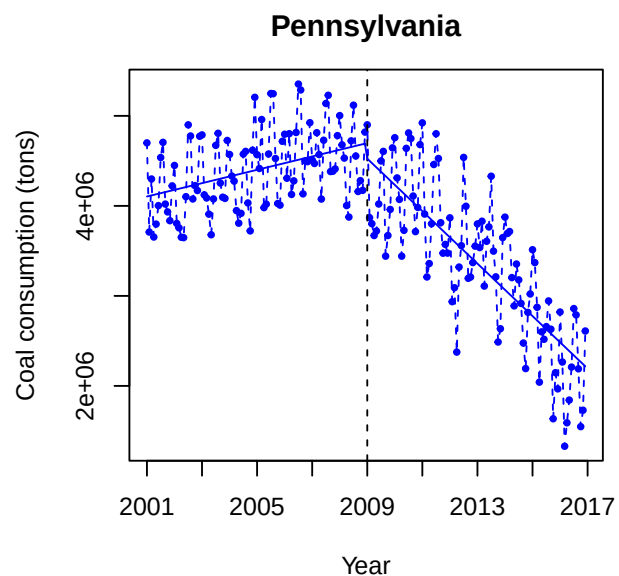
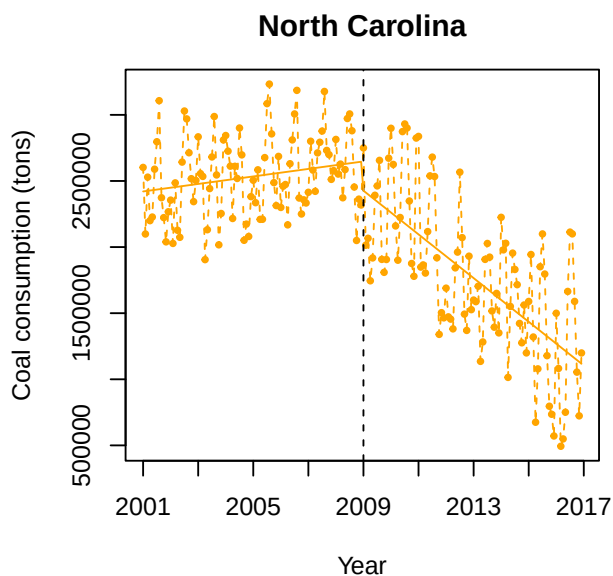
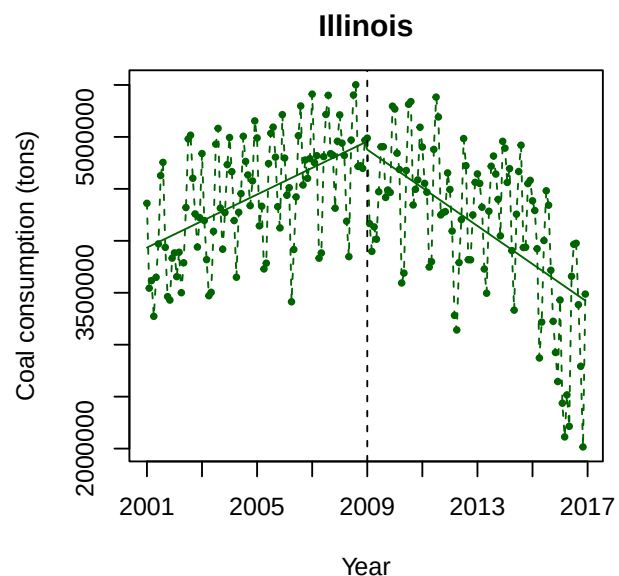
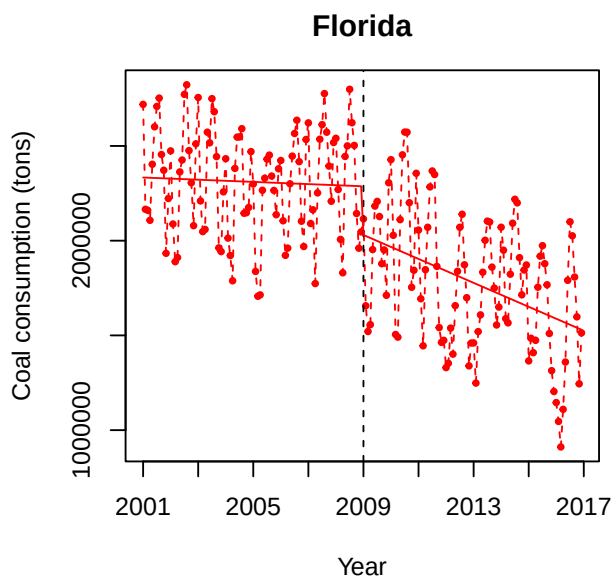
```
Time<-c(1:192); chp<-96
fit.fl<-lm(Y.fl~Time + I(Time>chp) + I((Time-chp)*(Time>chp)))
res.fl<-ts(as.numeric(residuals(fit.fl)), start=c(2001,1), frequency=12)
fit.il<-lm(Y.il~Time + I(Time>chp) + I((Time-chp)*(Time>chp)))
res.il<-ts(residuals(fit.il), start=c(2001,1), frequency=12)
fit.nc<-lm(Y.nc~Time + I(Time>chp) + I((Time-chp)*(Time>chp)))
res.nc<-ts(residuals(fit.nc), start=c(2001,1), frequency=12)
fit.pa<-lm(Y.pa~Time + I(Time>chp) + I((Time-chp)*(Time>chp)))
res.pa<-ts(residuals(fit.pa), start=c(2001,1), frequency=12)
par(mar=c(6,4,2,2), mfrow=c(2,2))
plot(res.fl, ylim=range(-1.2e6, 1.2e6), ylab='', xaxt='n')
axis(1, xaxp = c(2001, 2017, 8)); title('Florida')
plot(res.il, ylim=range(-1.2e6, 1.2e6), ylab='', xaxt='n')
axis(1, xaxp = c(2001, 2017, 8)); title('Illinois')
plot(res.nc, ylim=range(-1.2e6, 1.2e6), ylab='', xaxt='n')
axis(1, xaxp = c(2001, 2017, 8)); title('North Carolina')
plot(res.pa, ylim=range(-1.2e6, 1.2e6), ylab='', xaxt='n')
axis(1, xaxp = c(2001, 2017, 8)); title('Pennsylvania')
```



```

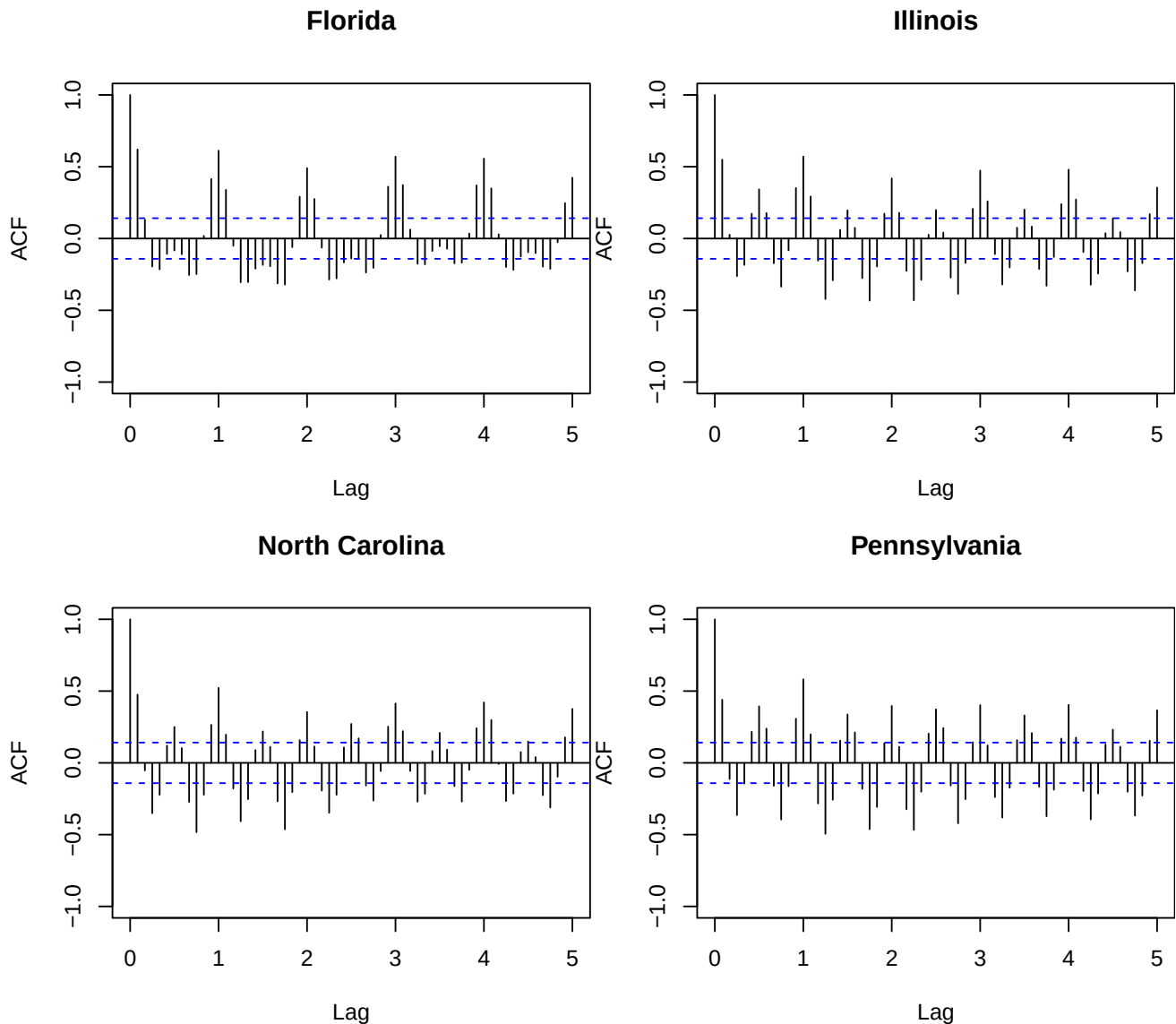
par(mar=c(5,4,3,2),mfrow=c(2,2))
plot(Y.fl,pch=19,col='red',xaxt='n',xlab='Year',ylab=yl,lty=2)
points(Y.fl,pch=19,cex=0.5,col='red');abline(v=2009,lty=2)
lines(ts(fitted(fit.fl),start=c(2001,1),frequency=12),col='red')
axis(1,xaxp=c(2001,2017,8));title('Florida')
plot(Y.il,pch=19,col='darkgreen',xaxt='n',xlab='Year',ylab=yl,lty=2)
points(Y.il,pch=19,cex=0.5,col='darkgreen');abline(v=2009,lty=2)
lines(ts(fitted(fit.il),start=c(2001,1),frequency=12),col='darkgreen')
axis(1,xaxp=c(2001,2017,8));title('Illinois')
plot(Y.nc,pch=19,col='orange',xaxt='n',xlab='Year',ylab=yl,lty=2)
points(Y.nc,pch=19,cex=0.5,col='orange');abline(v=2009,lty=2)
lines(ts(fitted(fit.nc),start=c(2001,1),frequency=12),col='orange')
axis(1,xaxp=c(2001,2017,8));title('North Carolina')
plot(Y.pa,pch=19,col='blue',xaxt='n',xlab='Year',ylab=yl,lty=2)
points(Y.pa,pch=19,cex=0.5,col='blue');abline(v=2009,lty=2)
lines(ts(fitted(fit.pa),start=c(2001,1),frequency=12),col='blue')
axis(1,xaxp=c(2001,2017,8));title('Pennsylvania')

```



The fit of the switching straight lines model is quite good. For the acfs for the residual series, we have the following (here plotted in units of 12 month lags):

```
par(mar=c(5, 4, 3, 0), mfrow=c(2, 2))
acf(res.fl, lag.max=60, ylim=range(-1, 1), main='Florida')
acf(res.il, lag.max=60, ylim=range(-1, 1), main='Illinois')
acf(res.nc, lag.max=60, ylim=range(-1, 1), main='North Carolina')
acf(res.pa, lag.max=60, ylim=range(-1, 1), main='Pennsylvania')
```



For Illinois, North Carolina and Pennsylvania, there seems to be a strong seasonality, although it is not clear whether this is stationary or non-stationary. For Florida there seems to be a 12 month seasonality.

Note that there may well be other valid transformation methods with respect to removing the non-constant mean beyond the linear model approach.

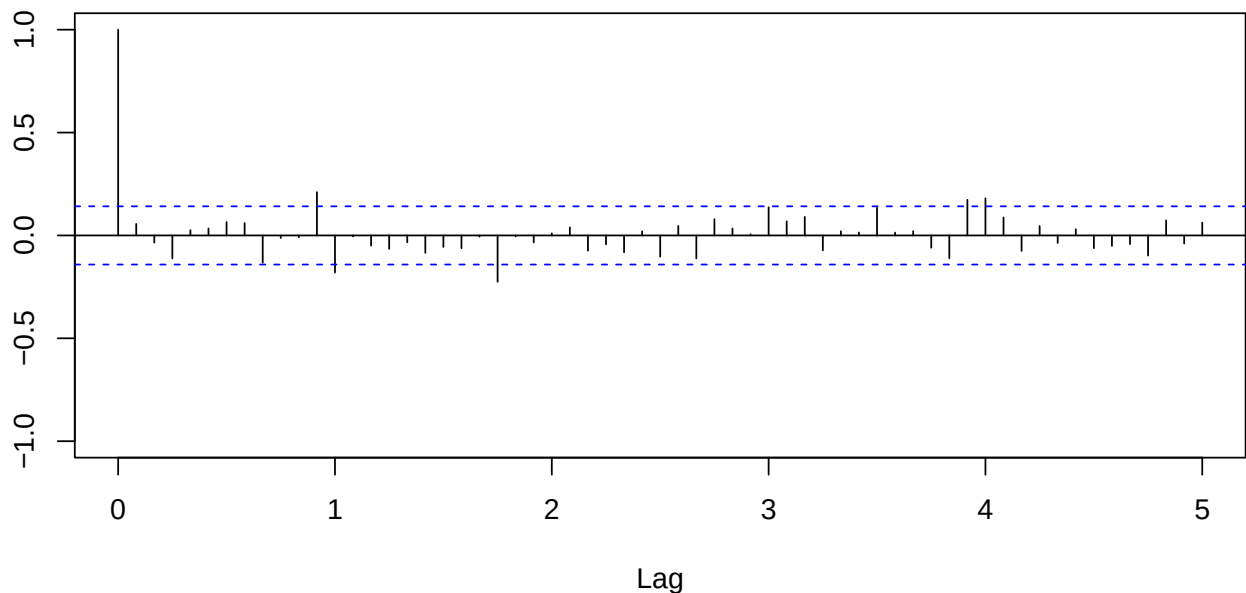
- **Florida:** For the Florida series, we first fit a stationary SARIMA model with seasonal period 12 months, and with one AR term and one seasonal AR term.

```
fit0<-arima(res.fl,order=c(1,0,0),
            seasonal=list(order=c(1,0,0),period=12),include.mean=FALSE)
coef(fit0)

+      ar1      sar1
+ 0.6192894 0.6603626

par(mar=c(4,2,3,0));tv<-'ACF of residuals from SARIMA(1,0,0)x(1,0,0)[12]'
acf(residuals(fit0),ylim=range(-1,1),lag.max=60,main=tv)
```

ACF of residuals from SARIMA(1,0,0)x(1,0,0)[12]



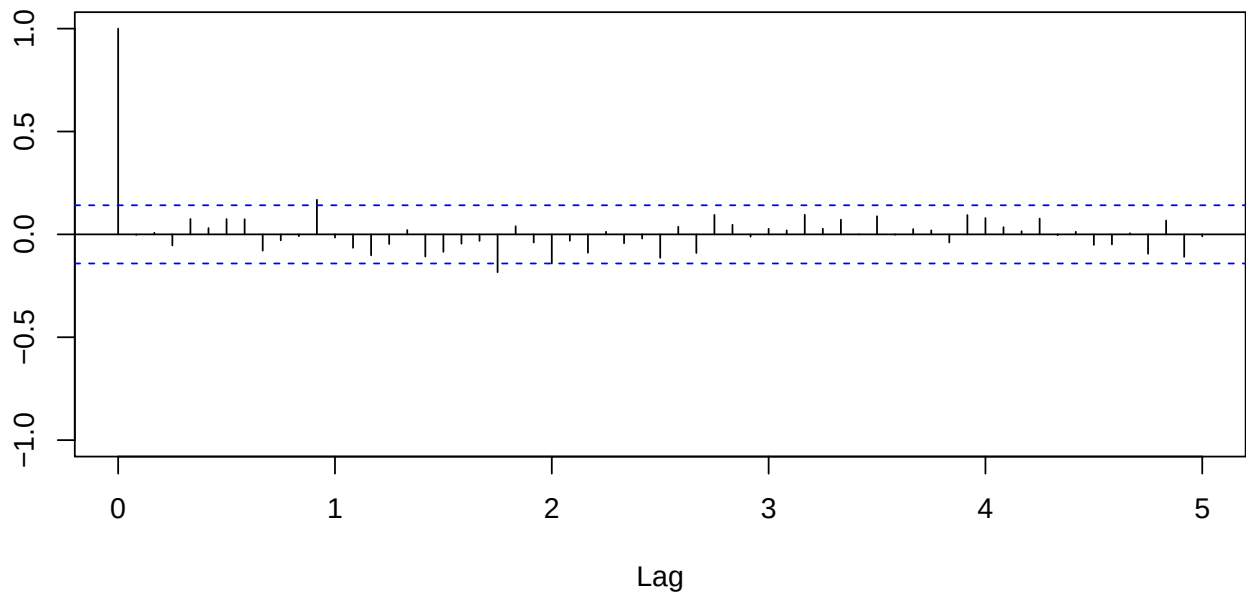
This fit is quite good, although perhaps there is evidence that MA terms are warranted.

```
fit1<-arima(res.fl,order=c(1,0,1),
            seasonal=list(order=c(1,0,1),period=12),include.mean=FALSE)
coef(fit1)

+      ar1      ma1      sar1      sma1
+ 0.5305338 0.1337823 0.9847755 -0.7873842

par(mar=c(4,2,3,0));tv<-'ACF of residuals from SARIMA(1,0,1)x(1,0,1)[12]'
acf(residuals(fit1),ylim=range(-1,1),lag.max=60,main=tv)
```

ACF of residuals from SARIMA(1,0,1)x(1,0,1)[12]



```
AIC(fit0, fit1); sigZsq <- fit1$sigma2
```

```
+      df      AIC
+ fit0  3 5201.181
+ fit1  5 5153.934
```

Using the AIC comparison, it is clear that the $\text{SARIMA}(1, 0, 1) \times (1, 0, 1)_{12}$ model fit is preferable. The estimated model can be written

$$(1 - 0.5305B)(1 - 0.9848B^{12})X_t = (1 - 0.1338B)(1 + 0.7874B^{12})Z_t$$

where $\sigma_Z^2 = 2.2992589 \times 10^{10}$. We use the `forecast` package to forecast the next 12 months

```
library(forecast)
```

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

```
(fit1.pred <- forecast(fit1, h=12))
```

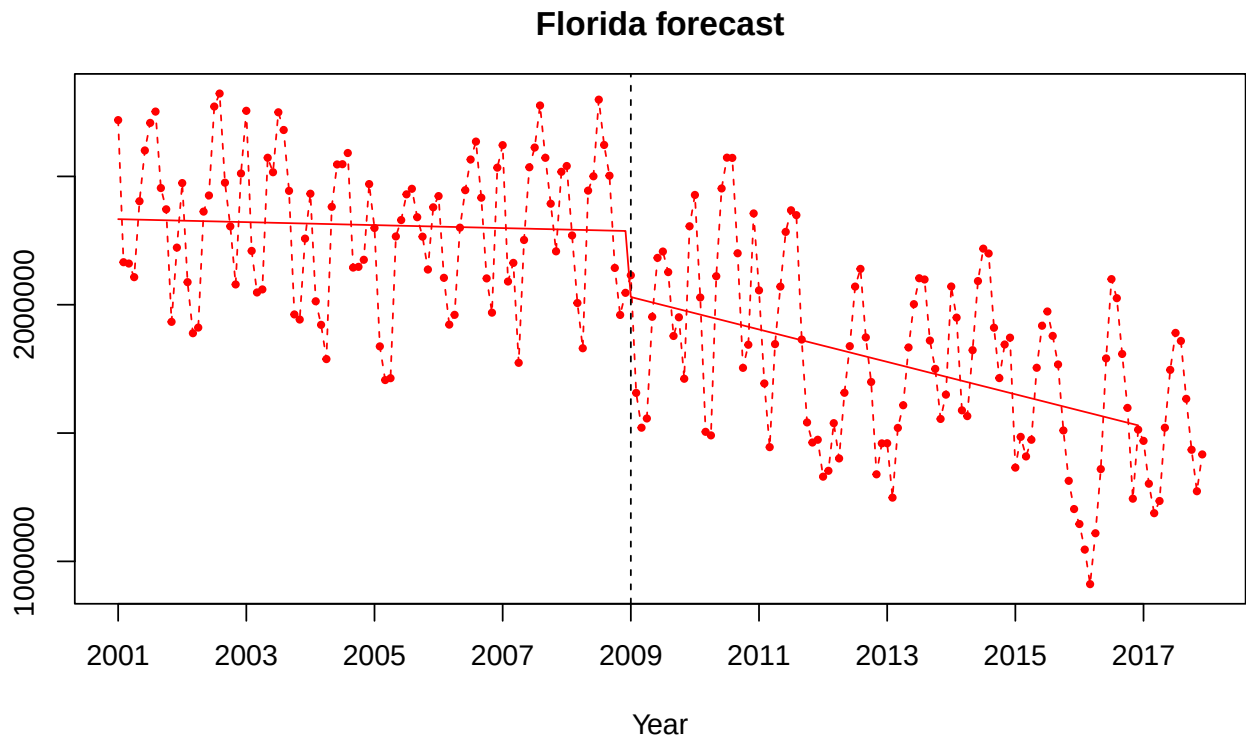
	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
+ Jan 2017	-55188.48	-249530.8172	139153.86	-352409.436	242032.47
+ Feb 2017	-216848.11	-450162.3963	16466.18	-573671.519	139975.30
+ Mar 2017	-326439.83	-569599.5364	-83280.12	-698320.509	45440.85
+ Apr 2017	-273584.87	-519444.6385	-27725.11	-649594.937	102425.19
+ May 2017	17155.49	-229458.9243	263769.89	-360008.707	394319.68
+ Jun 2017	247693.18	866.7832	494519.58	-129795.220	625181.58
+ Jul 2017	397142.58	150256.5447	644028.61	19562.974	774722.18
+ Aug 2017	370876.38	123973.5733	617779.18	-6728.875	748481.63
+ Sep 2017	150691.76	-96215.7224	397599.25	-226920.650	528304.18
+ Oct 2017	-42456.64	-289365.3145	204452.03	-420070.871	335157.59
+ Nov 2017	-198702.16	-445610.6987	48206.39	-576316.186	178911.87
+ Dec 2017	-50090.10	-296996.9530	196816.75	-427701.545	327521.34

for the residual series, and then add in the linear mean obtained from the earlier fit:

```

be0<-coef(fit.fl)[1]+coef(fit.fl)[3];be1<-coef(fit.fl)[2]+coef(fit.fl)[4]
mfit<-be0+be1*(192+c(1:12))-coef(fit.fl)[4]*chp
Y.fl.pred<-ts(c(Y.fl,mfit+fit1.pred$mean),start=c(2001,1),frequency=12)
par(mar=c(4,2,3,0))
plot(Y.fl.pred,pch=19,col='red',xaxt='n',xlab='Year',ylab=yl,lty=2)
points(Y.fl.pred,pch=19,cex=0.5,col='red');abline(v=2009,lty=2)
lines(ts(fitted(fit.fl),start=c(2001,1),frequency=12),col='red')
axis(1,xaxp=c(2001,2017,8));title('Florida forecast')

```

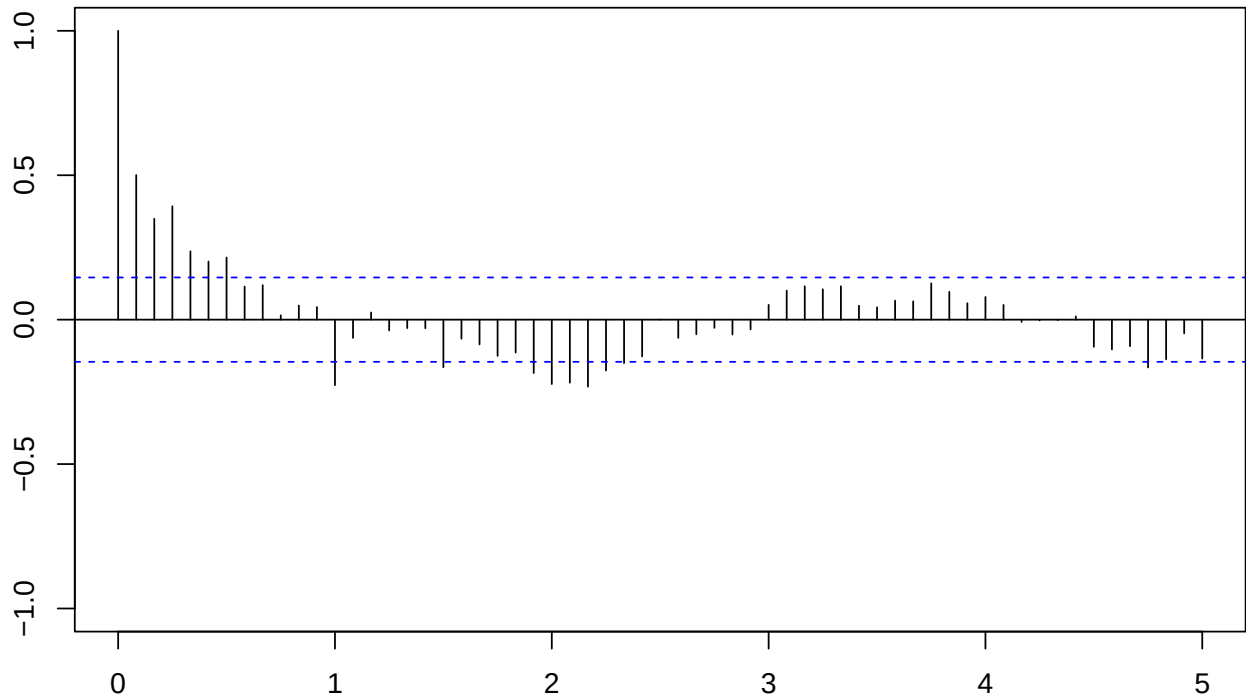


```
ts(Y.fl.pred[192+c(1:12)],start=c(2017,1),frequency=12)
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
2017	1469528	1302605	1187750	1235341	1520819	1746093	1890279	1858750	1633302
	Oct	Nov	Dec						
2017	1434890	1273381	1416730						

- **Illinois:** For Illinois, we try the same strategy. Here there is more evidence of non-stationarity, so we also assess the impact of seasonal differencing. First, we examine the impact of differencing at lag 12

```
par(mar=c(2,2,1,0))
acf(diff(res.il,12),ylim=range(-1,1),main='ACF of differenced series',lag.max=60)
```



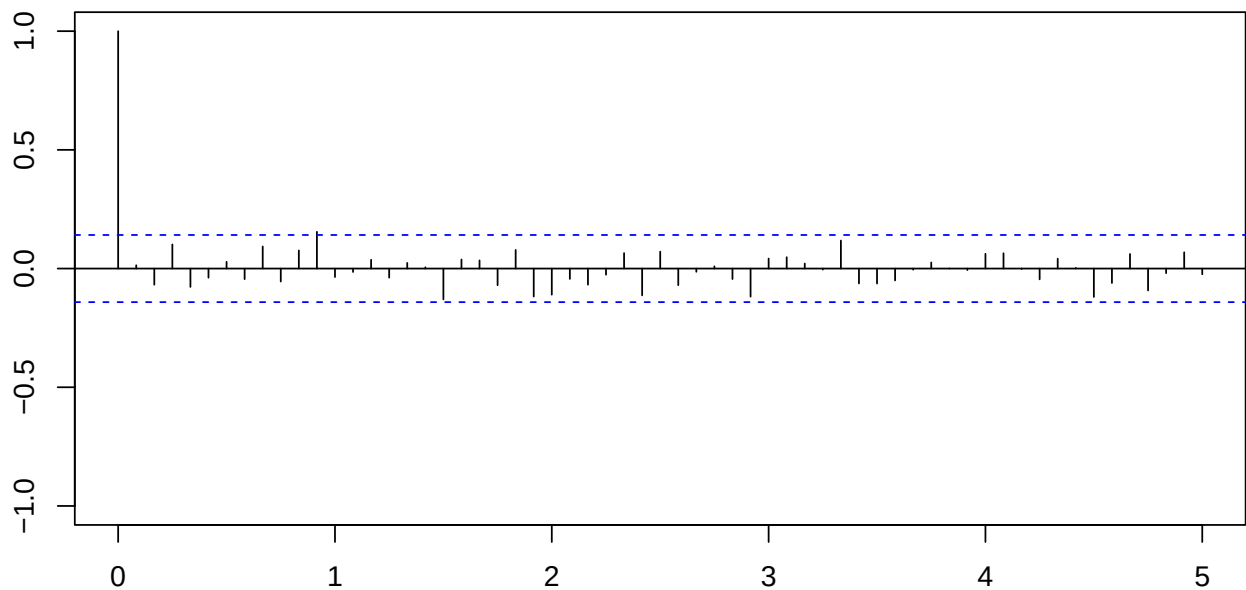
```
fit0<-arima(res.il,order=c(1,0,0),seasonal=list(order=c(0,1,0),period=12))
fit1<-arima(res.il,order=c(1,0,1),seasonal=list(order=c(0,1,0),period=12))
fit2<-arima(res.il,order=c(1,0,1),seasonal=list(order=c(0,1,1),period=12))
fit3<-arima(res.il,order=c(1,0,1),seasonal=list(order=c(1,1,1),period=12))
AIC(fit0,fit1,fit2,fit3)
```

```
+      df      AIC
+ fit0  2 5152.515
+ fit1  3 5148.723
+ fit2  4 5066.848
+ fit3  5 5068.834
```

From this analysis, the SARIMA(1,0,1) × (0,1,1)₁₂ is preferred.

```
par(mar=c(2,2,3,0))
acf(residuals(fit2),ylim=range(-1,1),
    main='ACF of Residuals from SARIMA(1,0,1)(0,1,1)[12]',lag.max=60)
```

ACF of Residuals from SARIMA(1,0,1)(0,1,1)[12]



```
coef(fit2);sigZsq<-fit2$sigma2
```

```
+      ar1      ma1      sma1
+ 0.8189121 -0.3644542 -0.9999967
```

The residual series seems adequate. The fitted model is

$$(1 - 0.8189B)(1 - B^{12})X_t = (1 + 0.3645B)(1 + B^{12})Z_t$$

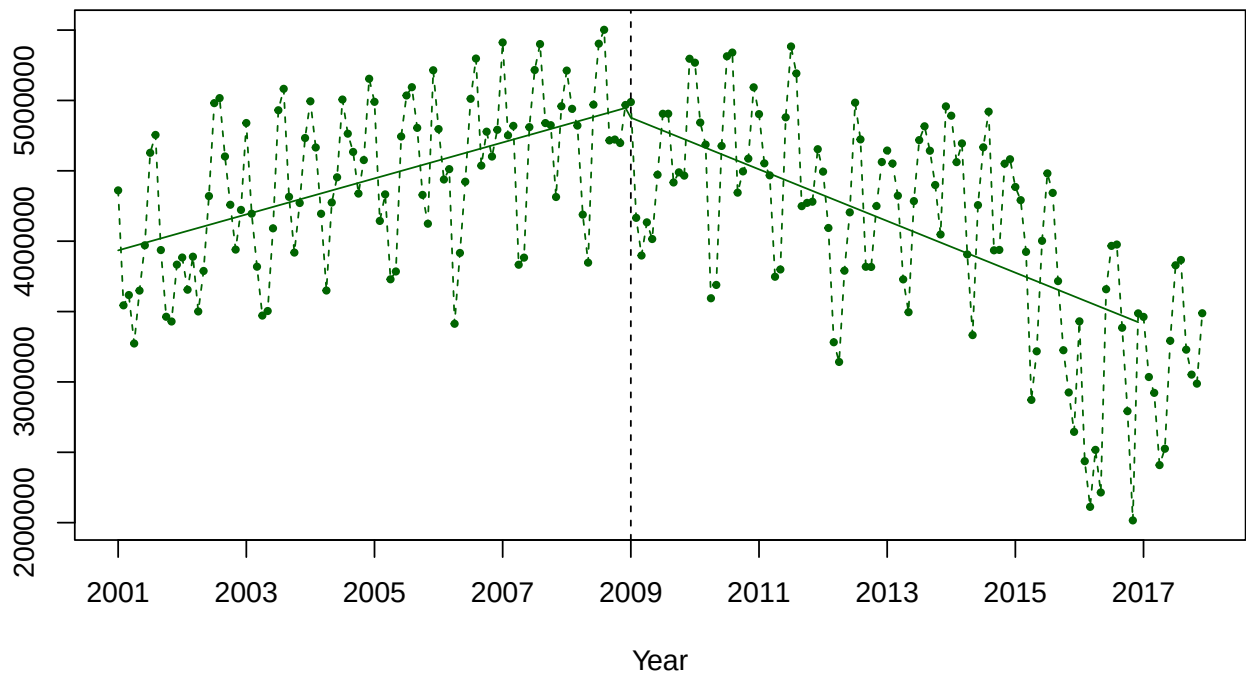
where $\sigma_Z^2 = 7.8000026 \times 10^{10}$. For the forecast for the next 12 months in the residual series, we use the forecast function add then in the appropriate time-varying linear mean obtained from the earlier fit:

```
(fit2.pred<-forecast(fit2,h=12));yfit<-fit2.pred$mean
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
+ Jan 2017	52982.84	-316167.55	422133.23	-511584.0	617549.6
+ Feb 2017	-359034.18	-764380.78	46312.43	-978958.3	260890.0
+ Mar 2017	-456287.76	-884194.41	-28381.11	-1110714.5	198139.0
+ Apr 2017	-954162.25	-1396551.12	-511773.39	-1630737.6	-277586.9
+ May 2017	-822634.19	-1274470.81	-370797.56	-1513658.7	-131609.7
+ Jun 2017	-40322.94	-498379.91	417734.04	-740860.6	660214.8
+ Jul 2017	511455.08	49282.90	973627.27	-195376.3	1218286.4
+ Aug 2017	564122.58	99224.96	1029020.20	-146877.0	1275122.1
+ Sep 2017	-57607.89	-524303.69	409087.92	-771357.5	656141.7
+ Oct 2017	-219565.88	-687433.02	248301.25	-935106.9	495975.2
+ Nov 2017	-268471.67	-737077.00	200133.66	-985141.7	448198.3
+ Dec 2017	246954.77	-222076.90	715986.44	-470367.3	964276.8

```
be0<-coef(fit.il)[1]+coef(fit.il)[3]
be1<-coef(fit.il)[2]+coef(fit.il)[4]
mfit<-be0+be1*(192+c(1:12))-coef(fit.il)[4]*chp
Y.il.pred<-ts(c(Y.il,yfit+mfit),start=c(2001,1),frequency=12)
par(mar=c(4,2,3,0))
plot(Y.il.pred,pch=19,col='darkgreen',xaxt='n',xlab='Year',ylab=yl,lty=2)
points(Y.il.pred,pch=19,cex=0.5,col='darkgreen');abline(v=2009,lty=2)
lines(ts(fitted(fit.il),start=c(2001,1),frequency=12),col='darkgreen')
axis(1,xaxp=c(2001,2017,8));title('Illinois forecast')
```

Illinois forecast

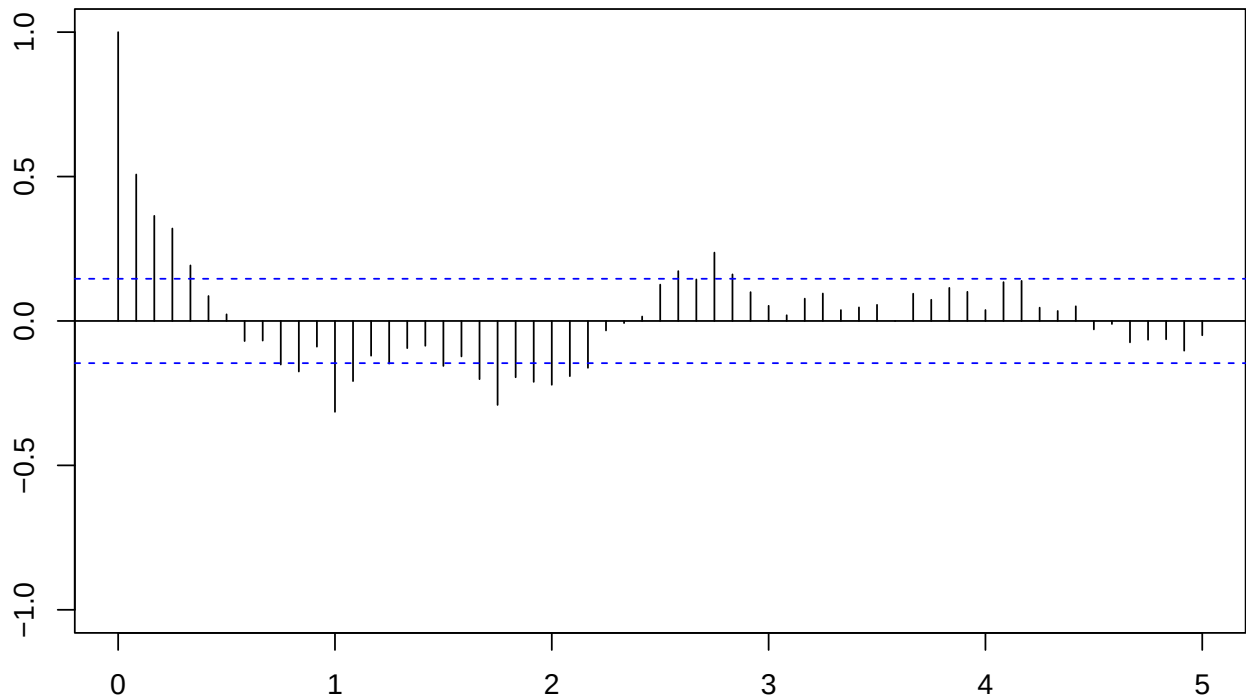


```
ts(Y.il.pred[192+c(1:12)], start=c(2017,1), frequency=12)
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
2017	3462153	3034840	2922291	2409120	2525353	3292368	3828850	3866222	3229196
	Oct	Nov	Dec						
2017	3051942	2987741	3487871						

- **North Carolina:** For North Carolina, we try the same strategy as for Illinois

```
par(mar=c(2,2,1,0))
acf(diff(res.nc,12),ylim=range(-1,1),main='ACF of differenced series',lag.max=60)
```



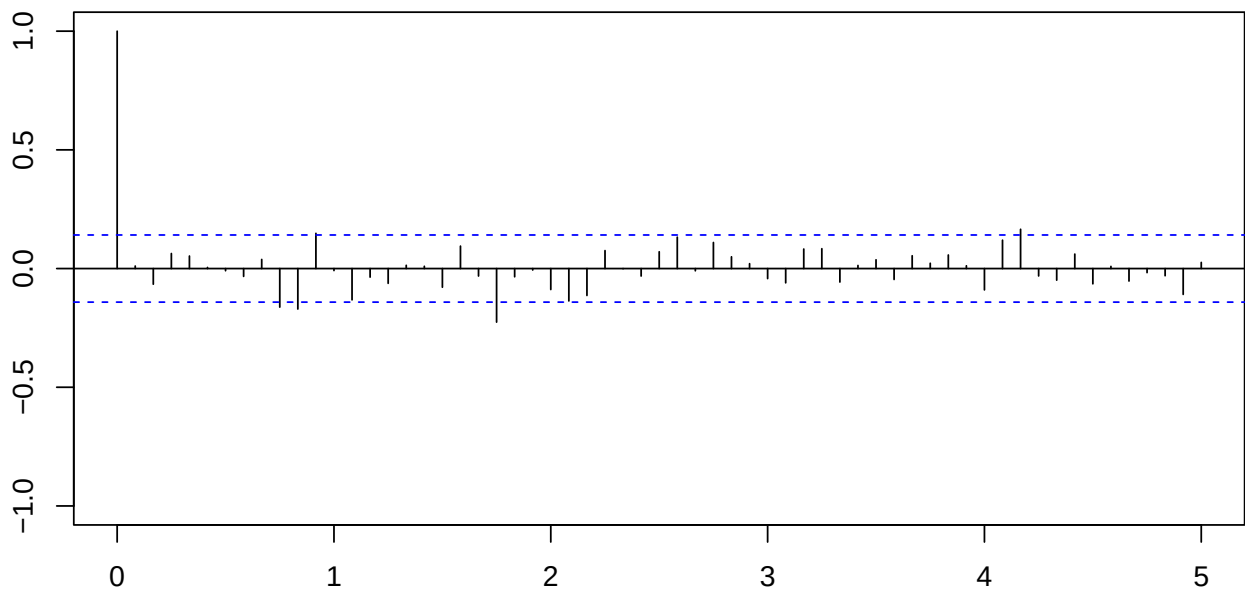
```
fit0<-arima(res.nc,order=c(1,0,0),seasonal=list(order=c(0,1,0),period=12))
fit1<-arima(res.nc,order=c(1,0,1),seasonal=list(order=c(0,1,0),period=12))
fit2<-arima(res.nc,order=c(1,0,1),seasonal=list(order=c(0,1,1),period=12))
fit3<-arima(res.nc,order=c(1,0,1),seasonal=list(order=c(1,1,1),period=12))
fit4<-arima(res.nc,order=c(1,0,1),seasonal=list(order=c(1,1,2),period=12))
fit5<-arima(res.nc,order=c(1,0,2),seasonal=list(order=c(1,1,1),period=12))
fit6<-arima(res.nc,order=c(1,0,1),seasonal=list(order=c(2,1,1),period=12))
fit7<-arima(res.nc,order=c(2,0,1),seasonal=list(order=c(1,1,1),period=12))
AIC(fit0,fit1,fit2,fit3,fit4,fit5,fit6,fit7)
```

	df	AIC
+ fit0	2	5043.172
+ fit1	3	5040.221
+ fit2	4	4993.055
+ fit3	5	4990.097
+ fit4	6	4990.762
+ fit5	6	4991.213
+ fit6	6	4990.430
+ fit7	6	4991.674

Here the $SARIMA(1,0,1) \times (1,1,1)_{12}$ is preferred, although there is little to choose between several of the models. It certainly seems that no more complex model is needed.

```
par(mar=c(2,2,3,0))
acf(residuals(fit3),ylim=range(-1,1),
    main='ACF of Residuals from SARIMA(1,0,1)(1,1,1)[12]',lag.max=60)
```

ACF of Residuals from SARIMA(1,0,1)(1,1,1)[12]



```
coef(fit3);sigZsq<-fit2$sigma2
```

```
+      ar1      ma1      sar1      sma1
+ 0.6051419 -0.1834945 0.2246379 -0.9133418
```

The residual series seems adequate. The fitted model is

$$(1 - 0.6051B)(1 - 0.2246B^{12})(1 - B^{12})X_t = (1 + 0.1835B)(1 + 0.9133B^{12})Z_t$$

where $\sigma_Z^2 = 5.8006302 \times 10^{10}$. For the forecast for the next 12 months in the residual series, we use the forecast function add then in the appropriate time-varying linear mean obtained from the earlier fit:

```
(fit3.pred<-forecast(fit3,h=12));yfit<-fit3.pred$mean
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
+ Jan 2017	346838.16	43711.1320	649965.19	-116754.62	810430.95
+ Feb 2017	24139.52	-304787.2011	353066.25	-478910.49	527189.53
+ Mar 2017	-209486.06	-547368.1127	128395.99	-726232.07	307259.94
+ Apr 2017	-478947.62	-820050.2666	-137844.98	-1000619.10	42723.86
+ May 2017	-227860.80	-570135.2078	114413.61	-751324.33	295602.73
+ Jun 2017	342119.04	-583.3855	684821.47	-181999.09	866237.18
+ Jul 2017	621320.89	278462.0632	964179.73	96963.56	1145678.23
+ Aug 2017	595966.99	253051.4498	938882.53	71522.93	1120411.05
+ Sep 2017	79748.41	-263186.3868	422683.21	-444725.10	604221.93
+ Oct 2017	-249824.02	-592761.7705	93113.72	-774302.05	274654.00
+ Nov 2017	-286468.60	-629396.2159	56459.02	-810931.13	237993.93
+ Dec 2017	-59220.88	-402114.1820	283672.41	-583630.93	465189.16

```
be0<-coef(fit.nc)[1]+coef(fit.nc)[3]
be1<-coef(fit.nc)[2]+coef(fit.nc)[4]
mfit<-be0+be1*(192+c(1:12))-coef(fit.nc)[4]*chp
Y.nc.pred<-ts(c(Y.nc,yfit+mfit),start=c(2001,1),frequency=12)
par(mar=c(4,2,3,0))
plot(Y.nc.pred,pch=19,col='orange',xaxt='n',xlab='Year',ylab=yl,lty=2)
points(Y.nc.pred,pch=19,cex=0.5,col='orange');abline(v=2009,lty=2)
lines(ts(fitted(fit.nc),start=c(2001,1),frequency=12),col='orange')
axis(1,xaxp=c(2001,2017,8));title('North Carolina forecast')
```

North Carolina forecast

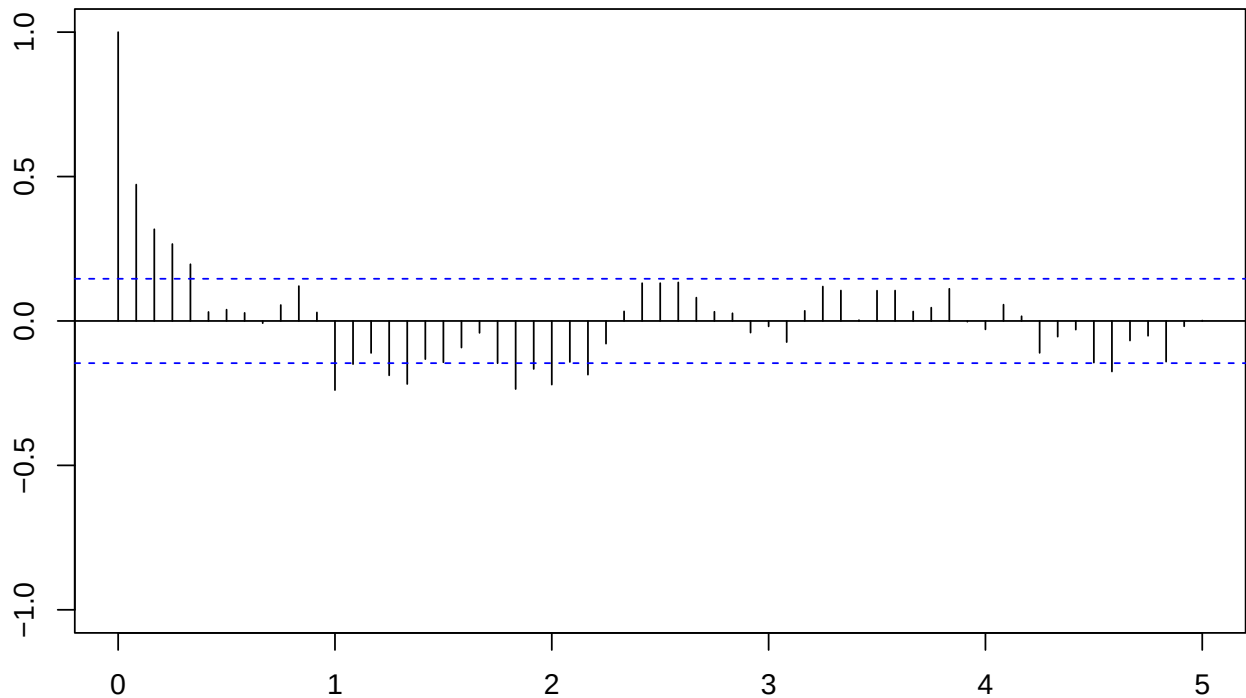


```
ts(Y.nc.pred[192+c(1:12)], start=c(2017,1), frequency=12)
```

		Jan	Feb	Mar	Apr	May	Jun	Jul
+	2017	1451088.6	1114615.7	867215.9	583980.1	821292.7	1377498.3	1642925.9
		Aug	Sep	Oct	Nov	Dec		
+	2017	1603797.8	1073805.0	730458.3	680039.5	893513.0		

- **Pennsylvania:** For Pennsylvania, we try the same strategy as for Illinois and North Carolina.

```
par(mar=c(2,2,1,0))
acf(diff(res.pa,12),ylim=range(-1,1),main='ACF of differenced series',lag.max=60)
```



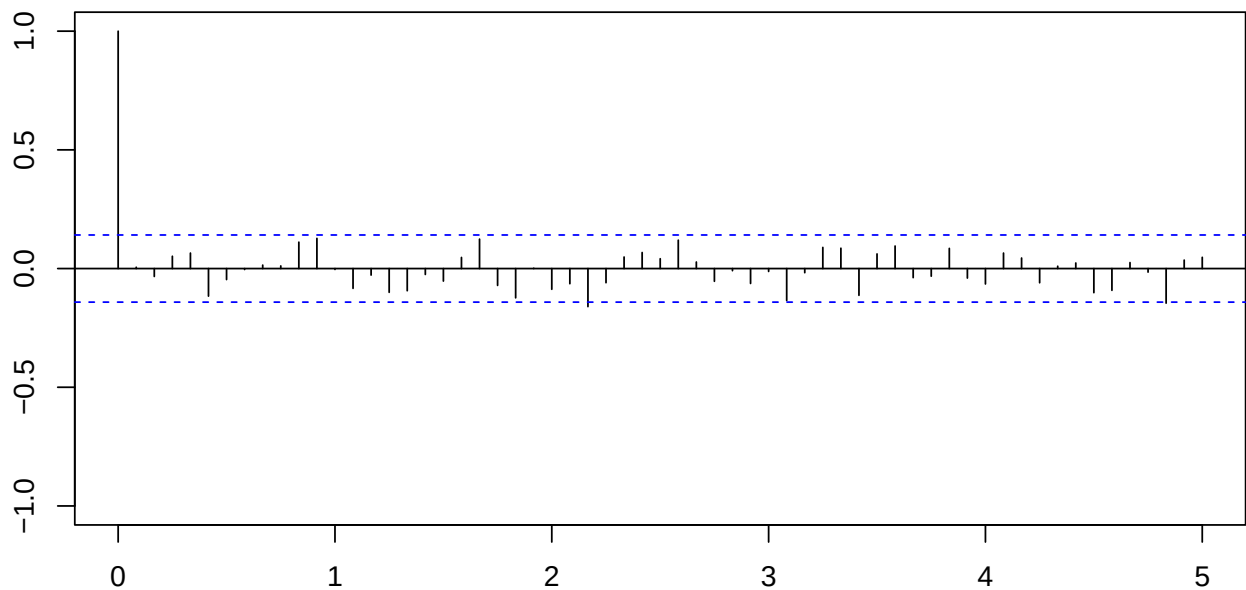
```
fit0<-arima(res.pa,order=c(1,0,0),seasonal=list(order=c(0,1,0),period=12))
fit1<-arima(res.pa,order=c(1,0,1),seasonal=list(order=c(0,1,0),period=12))
fit2<-arima(res.pa,order=c(1,0,1),seasonal=list(order=c(0,1,1),period=12))
fit3<-arima(res.pa,order=c(1,0,1),seasonal=list(order=c(1,1,1),period=12))
fit4<-arima(res.pa,order=c(1,0,1),seasonal=list(order=c(1,1,2),period=12))
fit5<-arima(res.pa,order=c(1,0,2),seasonal=list(order=c(1,1,1),period=12))
fit6<-arima(res.pa,order=c(1,0,1),seasonal=list(order=c(2,1,1),period=12))
fit7<-arima(res.pa,order=c(2,0,1),seasonal=list(order=c(1,1,1),period=12))
AIC(fit0,fit1,fit2,fit3,fit4,fit5,fit6,fit7)
```

	df	AIC
+ fit0	2	5115.622
+ fit1	3	5114.297
+ fit2	4	5064.587
+ fit3	5	5062.002
+ fit4	6	5062.768
+ fit5	6	5063.839
+ fit6	6	5062.723
+ fit7	6	5063.910

As for North Carolina, the SARIMA(1,0,1) × (1,1,1)₁₂ is preferred.

```
par(mar=c(2,2,3,0))
acf(residuals(fit3),ylim=range(-1,1),
    main='ACF of Residuals from SARIMA(1,0,1)(1,1,1)[12]',lag.max=60)
```

ACF of Residuals from SARIMA(1,0,1)(1,1,1)[12]



```
coef(fit3);sigZsq<-fit2$sigma2
```

```
+      ar1      ma1      sar1      sma1
+ 0.6770659 -0.2627954 0.2186547 -0.8879562
```

The residual series seems adequate. The fitted model is

$$(1 - 0.6771B)(1 - 0.2187B^{12})(1 - B^{12})X_t = (1 + 0.2628B)(1 + 0.888B^{12})Z_t$$

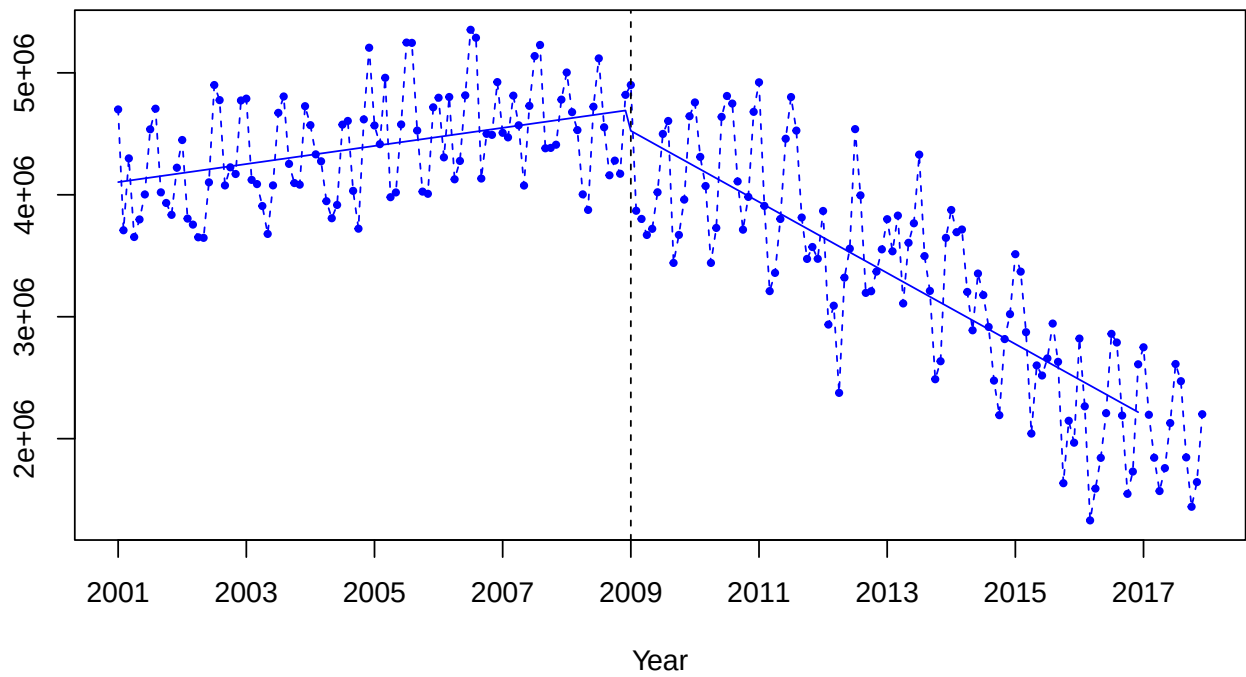
where $\sigma_Z^2 = 8.7334851 \times 10^{10}$. For the forecast for the next 12 months in the residual series, we use the forecast function add then in the appropriate time-varying linear mean obtained from the earlier fit:

```
(fit3.pred<-forecast(fit3,h=12));yfit<-fit3.pred$mean
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
+ Jan 2017	556370.39	184774.13	927966.64	-11937.04	1124677.81
+ Feb 2017	27030.75	-375156.70	429218.21	-588061.88	642123.39
+ Mar 2017	-300770.74	-716229.45	114687.97	-936160.00	334618.53
+ Apr 2017	-549729.11	-971131.87	-128326.34	-1194209.02	94750.81
+ May 2017	-337543.59	-761643.28	86556.11	-986148.10	311060.92
+ Jun 2017	56445.29	-368884.82	481775.40	-594040.98	706931.56
+ Jul 2017	564717.08	138824.52	990609.64	-86629.39	1216063.55
+ Aug 2017	448367.88	22218.61	874517.14	-203371.18	1100106.93
+ Sep 2017	-152267.77	-578532.73	273997.18	-804183.76	499648.22
+ Oct 2017	-532924.73	-959238.49	-106610.97	-1184915.36	119065.90
+ Nov 2017	-306448.09	-732775.01	119878.84	-958458.85	345562.67
+ Dec 2017	274563.51	-151749.36	700876.38	-377425.76	926552.78

```
be0<-coef(fit.pa)[1]+coef(fit.pa)[3]
be1<-coef(fit.pa)[2]+coef(fit.pa)[4]
mfit<-be0+be1*(192+c(1:12))-coef(fit.pa)[4]*chp
Y.pa.pred<-ts(c(Y.pa,yfit+mfit),start=c(2001,1),frequency=12)
par(mar=c(4,2,3,0))
plot(Y.pa.pred,pch=19,col='blue',xaxt='n',xlab='Year',ylab=yl,lty=2)
points(Y.pa.pred,pch=19,cex=0.5,col='blue');abline(v=2009,lty=2)
lines(ts(fitted(fit.pa),start=c(2001,1),frequency=12),col='blue')
axis(1,xaxp=c(2001,2017,8));title('Pennsylvania forecast')
```

Pennsylvania forecast



```
ts(Y.pa.pred[192+c(1:12)], start=c(2017,1), frequency=12)
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
2017	2749457	2195834	1843749	1570507	1758409	2128114	2612102	2471469	1846550
2017	1441610	1643803	2200531						

Summary: The key points of the analysis are

- Four states (Florida, Illinois, North Carolina and Pennsylvania) exhibit qualitatively similar characteristics in terms of the amount of coal used in energy production. The variation can be explained by a low order SARIMA model, with seasonality 12 months.
- The California data is qualitatively very different; since the start of 2015, the coal usage is reasonably constant at a low level.
- For the four similar states, there is evidence of a changepoint in the amount of coal used in energy production, with the changepoint happening around the start of 2009. From 2009 onwards, there is a steady decline in the amount of coal used in energy production.
- Forecasting using the SARIMA model and the linear changepoint model is relatively straightforward. The forecasts assume that extrapolation for 2017 is possible from the data and model for the 2009–2016 period.

```
preds<-cbind(Y.fl.pred,Y.il.pred,Y.nc.pred,Y.pa.pred)
preds2017<-preds[193:204,]
preds2017
```

	Y.fl.pred	Y.il.pred	Y.nc.pred	Y.pa.pred
+ [1,]	1469528	3462153	1451088.6	2749457
+ [2,]	1302605	3034840	1114615.7	2195834
+ [3,]	1187750	2922291	867215.9	1843749
+ [4,]	1235341	2409120	583980.1	1570507
+ [5,]	1520819	2525353	821292.7	1758409
+ [6,]	1746093	3292368	1377498.3	2128114
+ [7,]	1890279	3828850	1642925.9	2612102
+ [8,]	1858750	3866222	1603797.8	2471469
+ [9,]	1633302	3229196	1073805.0	1846550
+ [10,]	1434890	3051942	730458.3	1441610
+ [11,]	1273381	2987741	680039.5	1643803
+ [12,]	1416730	3487871	893513.0	2200531