Homework 2

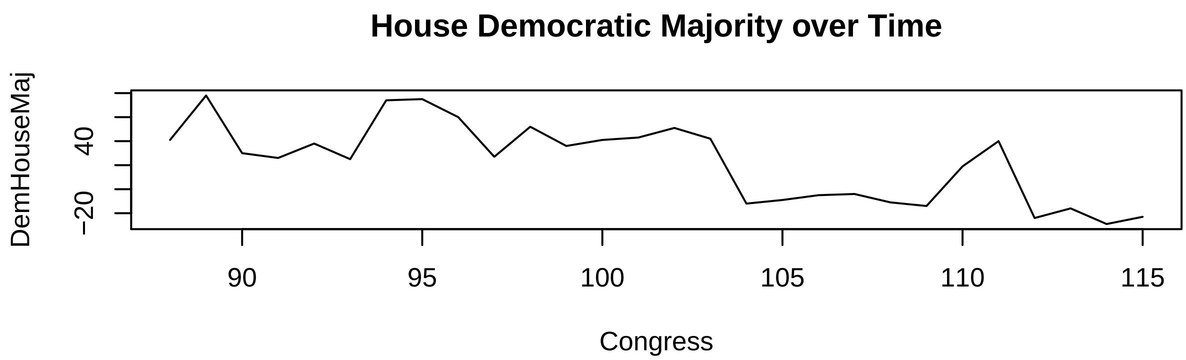
Austin Sell

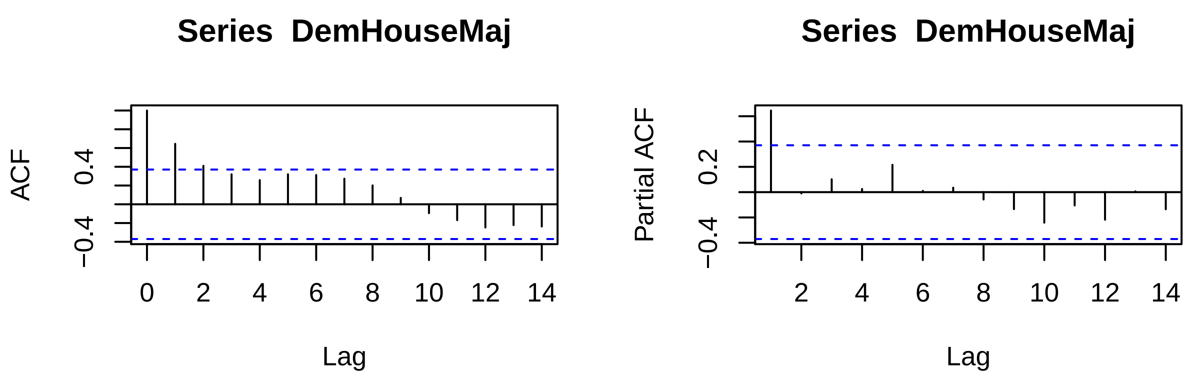
May 3, 2018

Question 1

Part A

I begin by plotting the time series for the Democratic House Majority as well as its ACF and PACF. I also run the augmented Dickey-Fuller and Phillips-Perron tests for unit roots. The correlograms below indicate the possibility of a first order autoregressive process. The unit root tests disagree in their findings; the ADF cannot reject the hypothesis of nonstationarity, while the Phillips-Perron can.





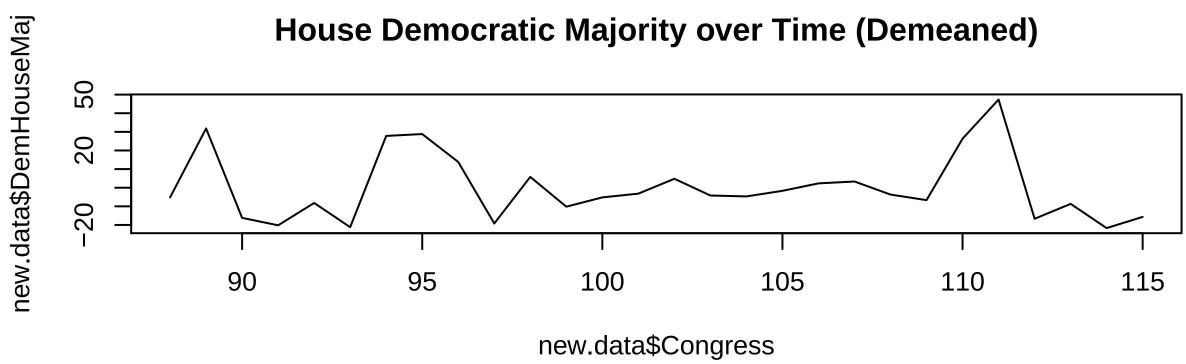

```
##
##  Augmented Dickey-Fuller Test
##
## data:  DemHouseMaj
## Dickey-Fuller = -3, Lag order = 3, p-value = 0.3
## alternative hypothesis: stationary
```

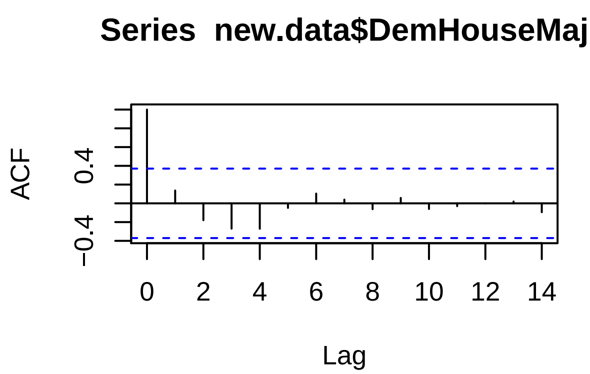
```
##
##  Phillips-Perron Unit Root Test
##
```

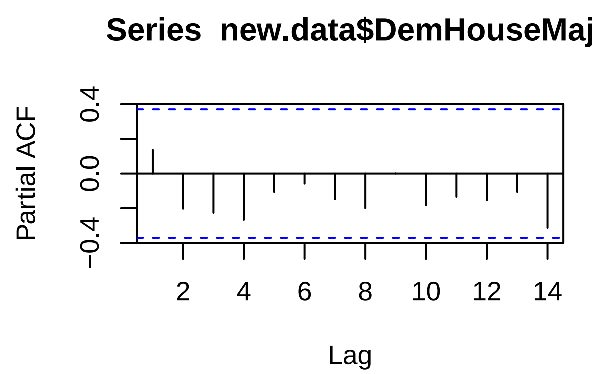
```
## data:  DemHouseMaj
## Dickey-Fuller = -4, Truncation lag parameter = 2, p-value = 0.04
```

I continue by examining the possibility of a structural break in the model at the year 1994. I demeaned the data by period, both before and after 1994, constructing a new dataset. The modified time series and ACF/PACF plots are below. After accounting for the structural break, there does not seem to be evidence of a AR or MA process. Any analysis of this time series should include a control for this break.

```r
pre1994 <- data[StartYear<1994,]
post1994 <- data[StartYear>1994,]
pre1994$DemHouseMaj <- pre1994$DemHouseMaj - mean(pre1994$DemHouseMaj)
post1994$DemHouseMaj <- post1994$DemHouseMaj - mean(post1994$DemHouseMaj)
new.data <- rbind(pre1994, post1994)
```







Part B

I fit an AR(0) regression by defining the relevant covariates and estimating the model with arima. Notably, OLS would give the same coefficient estimates for this model (but different standard errors). Relevant model information is displayed in Table 1.

```r
xcovariates <- cbind(PartisanMidterm, PartisanUnem, Coattails, Pre1994)

ar0 <- arima(DemHouseMaj, order = c(0,0,0),
             xreg = xcovariates, include.mean = TRUE)
```

Table 1: Model Evaluation

Model Components	AIC	$\hat{\sigma}^2$	N	$\hat{\beta}_{PM}$	$\hat{\beta}_{PU}$	$\hat{\beta}_{CT}$	$\hat{\beta}_{1994}$
AR(0)	239.243	195.964	28	-7.27	-2.053	18.396	47.994
				(3.823)	(1.733)	(5.331)	(5.703)

The AR(0) model suggests that the House Democratic majority was much higher in the sessions before our structural break in 1994, as expected. There is also a substantial coattail effect, with Democrats increasing their majority by about 18 seats when the Presidency switches from a Republican to a Democratic.

Part C

```r
ar1 <- arima(DemHouseMaj, order = c(1,0,0),
            xreg = xcovariates, include.mean = TRUE)

ar2 <- arima(DemHouseMaj, order = c(2,0,0),
            xreg = xcovariates, include.mean = TRUE)

ma1 <- arima(DemHouseMaj, order = c(0,0,1),
            xreg = xcovariates, include.mean = TRUE)

arma11 <- arima(DemHouseMaj, order = c(1,0,1),
            xreg = xcovariates, include.mean = TRUE)
```

Table 2: Model Comparison

Model Components	AIC	$\hat{\sigma}^2$	N	$\hat{\beta}_{PM}$	$\hat{\beta}_{PU}$	$\hat{\beta}_{CT}$	$\hat{\beta}_{1994}$
AR(0)	239.243	195.964	28	-7.27	-2.053	18.396	47.994
				(3.823)	(1.733)	(5.331)	(5.703)
AR(1)	240.221	188.565	28	-8.848	-2.42	15.364	46.656
				(3.86)	(1.75)	(5.828)	(7.023)
AR(2)	239.243	169.879	28	-10.726	-2.858	10.279	44.729
				(3.158)	(1.744)	(5.862)	(6.038)
MA(1)	239.495	183.033	28	-9.866	-2.761	13.121	45.386
				(3.743)	(1.841)	(6.315)	(7.58)
ARMA(1,1)	238.096	148.58	28	-11.673	-2.758	14.522	42.887
				(3.204)	(1.484)	(6.058)	(6.623)

Substantively, all of the models are fairly similar in terms of their coefficient estimates. In terms of in-sample model fit, the ARMA(1,1) model does the best, although none of the models seems to perform particularly poorly in terms of AIC. Adding additional terms (either AR or MA), lowers $\hat{\sigma}^2$ which is not surprising.

Part D

```r
ar0.cv <- arimaCV(DemHouseMaj, order = c(0,0,0), forward=3,
                  xreg=xcovariates, include.mean = TRUE, minper = 20)

ar1.cv <- arimaCV(DemHouseMaj, order = c(1,0,0), forward=3,
```

```
                     xreg=xcovariates, include.mean = TRUE, minper = 20)

ar2.cv <- arimaCV(DemHouseMaj, order = c(2,0,0), forward=3,
                     xreg=xcovariates, include.mean = TRUE, minper = 20)

ma1.cv <- arimaCV(DemHouseMaj, order = c(0,0,1), forward=3,
                     xreg=xcovariates, include.mean = TRUE, minper = 20)

arma11.cv <- arimaCV(DemHouseMaj, order = c(1,0,1), forward=3,
                     xreg=xcovariates, include.mean = TRUE, minper = 20)
```

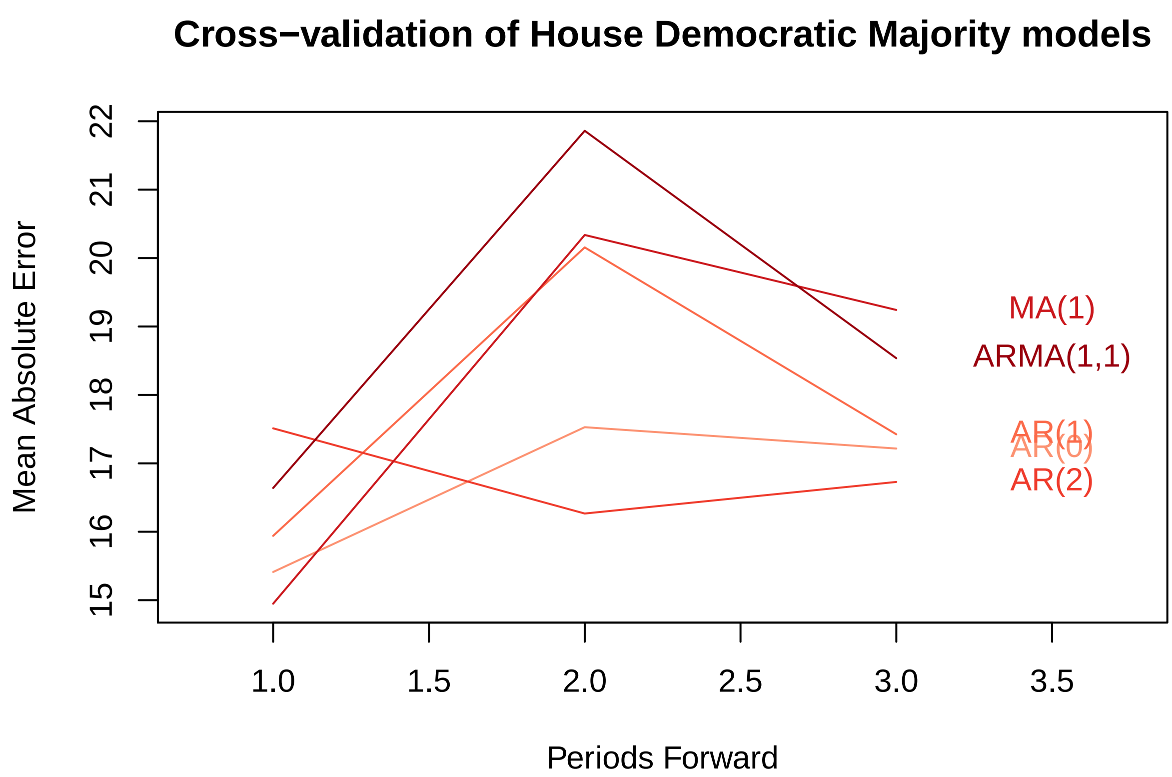


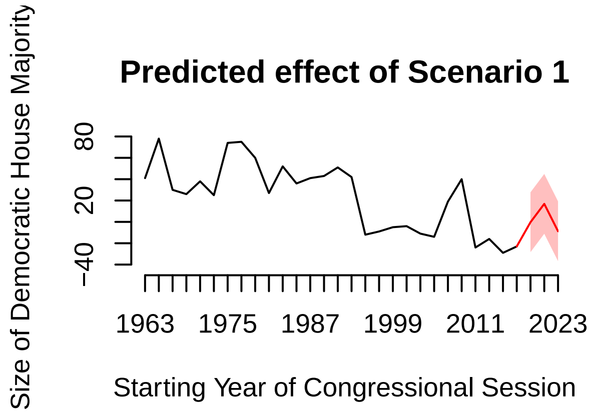
Table 3: Model Comparison

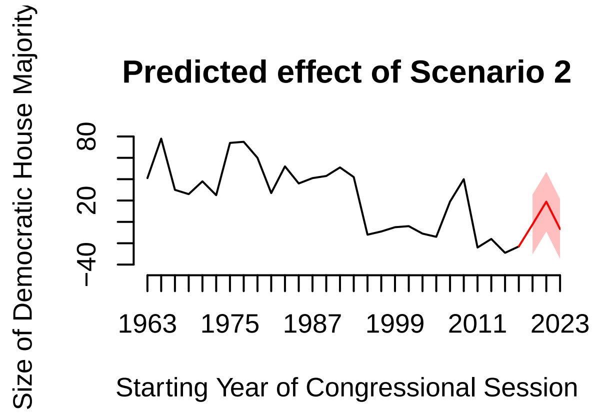
Model Components	AIC	RMSE	MAE_1	MAE_2	MAE_3	Average MAE
AR(0)	239.243	13.999	15.412	17.528	17.216	16.719
AR(1)	240.221	13.732	15.94	20.156	17.424	17.84
AR(2)	239.662	13.034	17.512	16.266	16.728	16.835
MA(1)	239.495	13.529	14.949	20.338	19.242	18.176
ARMA(1,1)	238.096	12.189	16.64	21.86	18.536	19.012

Using rolling cross-validation, we can see that the AR(0) model is actually the best model, even though it doesn't have the best in-sample fit. The AR(2) model comes close, but the AR(0) is the overall best performance and it is supported by the correlograms that we constructed that found no evidence of AR processes after taking into account the structural break.

Part E

The predicted effects of each of the three scenarios, as well as a table of counterfactual values, are below.





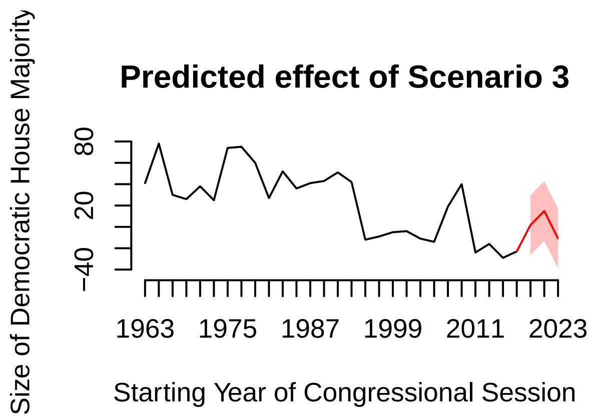


Table 4: Counterfactual Values

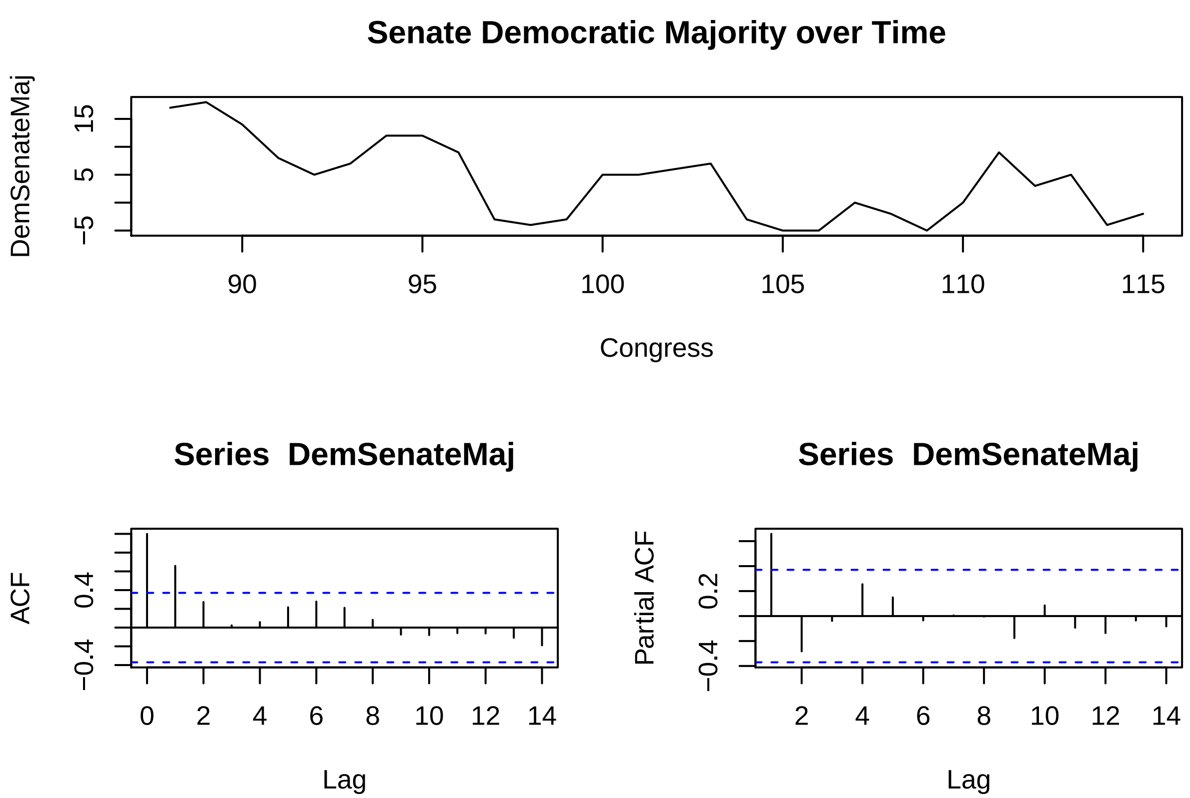
Variable	2018	2020	2022
Scenario 1			
Partisan Midterm	-1	0	1
Partisan Unemployment	1.475	-1.475	-1.475
Coattails	0	1	0
Pre-1994	0	0	0
Scenario 2			
Partisan Midterm	-1	0	1
Partisan Unemployment	2.475	-2.475	-2.475
Coattails	0	1	0
Pre-1994	0	0	0
Scenario 3			
Partisan Midterm	-1	0	1
Partisan Unemployment	0.475	-0.475	-0.475
Coattails	0	1	0
Pre-1994	0	0	0

All three secnarios predict an increase in Democratic seats in the 2018 midterm, followed by another increase in 2020 (assuming a Democratic president wins), and then a decrease in 2022 midterm. The exact predicted amount varies slightly because of the different levels of unemployment, but that difference doesn't appear to be enough to be significant. The main driver of the predictions seems to be the spike in 2020 due to the coattail effect.

Question 2

Part A

```
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
plot(Congress, DemSenateMaj, type="l", main="Senate Democratic Majority over Time")
acf(DemSenateMaj)
pacf(DemSenateMaj)
```




```
adf.test(DemSenateMaj)
```

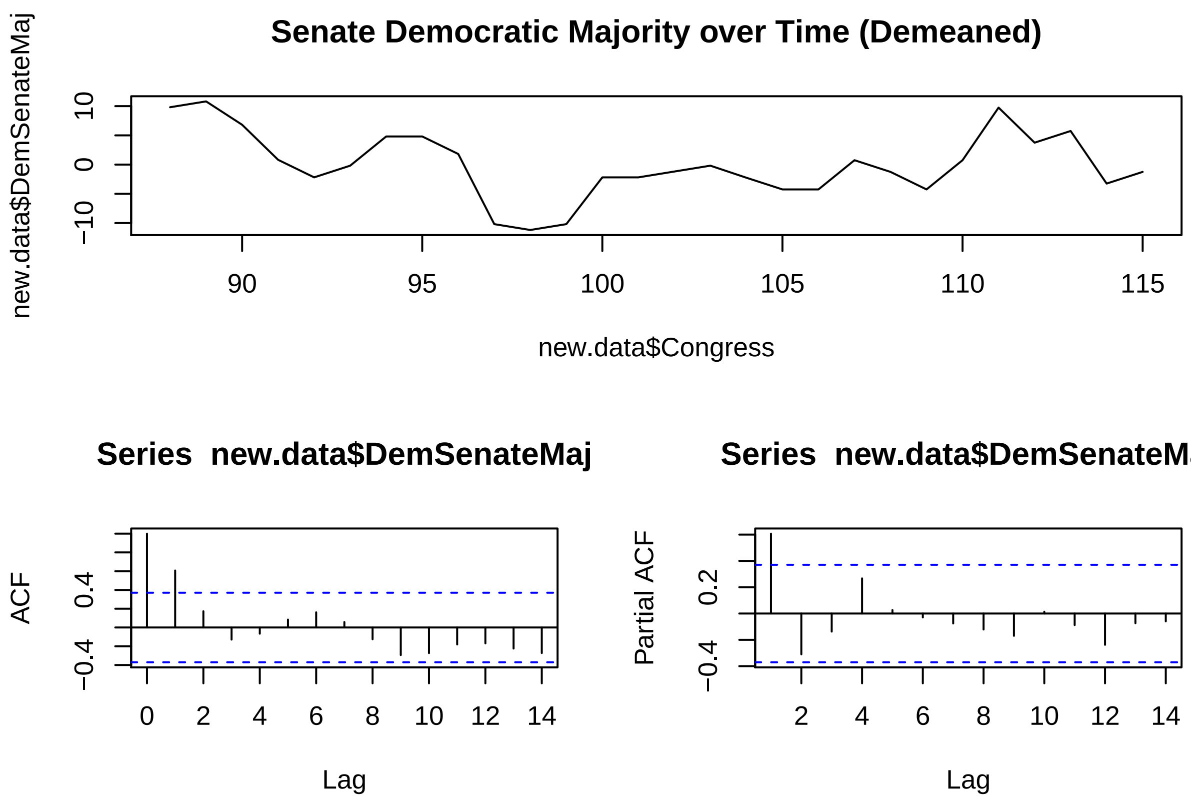
```
##
##  Augmented Dickey-Fuller Test
##
## data:  DemSenateMaj
## Dickey-Fuller = -2, Lag order = 3, p-value = 0.5
## alternative hypothesis: stationary
```

```r
PP.test(DemSenateMaj)
```

```
##
##  Phillips-Perron Unit Root Test
##
## data:  DemSenateMaj
## Dickey-Fuller = -3, Truncation lag parameter = 2, p-value = 0.3
```

```r
pre1994 <- data[StartYear<1994,]
post1994 <- data[StartYear>1994,]
pre1994$DemSenateMaj <- pre1994$DemSenateMaj - mean(pre1994$DemSenateMaj)
post1994$DemSenateMaj <- post1994$DemSenateMaj - mean(post1994$DemSenateMaj)
new.data <- rbind(pre1994, post1994)

layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
plot(new.data$Congress, new.data$DemSenateMaj, type="l", main="Senate Democratic Majority over Time (De
acf(new.data$DemSenateMaj)
pacf(new.data$DemSenateMaj)
```



There does seem to be some evidence of a structural break. After demeaning the data, the underlying process looks to be an AR(1).

Part B

Table 5: Model Comparison

Model Components	AIC	$\hat{\sigma}^2$	N	$\hat{\beta}_{PM}$	$\hat{\beta}_{PU}$	$\hat{\beta}_{CT}$	$\hat{\beta}_{1994}$
AR(0)	183.044	26.333	28	1.792	0.358	3.403	8.503
				(1.402)	(0.635)	(1.954)	(2.091)
AR(1)	172.3	16.315	28	-0.341	-0.268	1.892	8.151
				(0.945)	(0.459)	(1.392)	(3.507)
AR(2)	183.044	12.698	28	-0.806	-0.884	0.103	6.611
				(0.635)	(0.367)	(1.091)	(2.581)
MA(1)	170.826	14.053	28	-0.464	-1.224	1.784	7.377
				(0.767)	(0.339)	(1.467)	(2.204)
ARMA(1,1)	168.501	11.651	28	-0.72	-1.309	0.807	9.131
				(0.576)	(0.24)	(1.134)	(2.171)

The best model from in-sample tests appears to be the ARMA(1,1), but again we run the chance that the model may have overfit the data. Cross-validation techniques will help us figure out which model is actually most reliable. Substantively, only the indicator of the structural break in 1994 is significant across the models. This suggests that while there was a meaningful change in the Senate in 1994 as there was in the House, for the most part Senators are insulated from the other factors that have stronger effects on House elections. This makes sense as only a third of the Senate is up for reelection at a time, so the previous majority share is likely a good predictor of the next majority share.

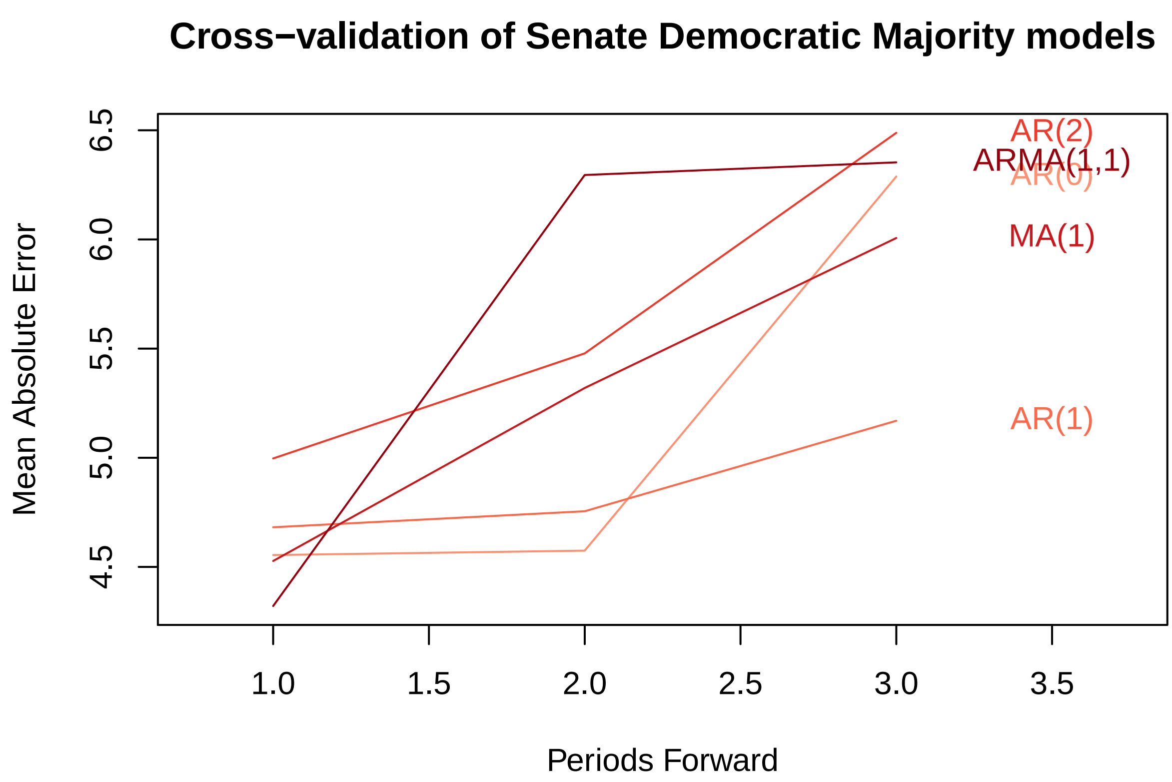


Table 6: Model Comparison

Model Components	AIC	RMSE	MAE_1	MAE_2	MAE_3	Average MAE
AR(0)	183.044	5.132	4.555	4.575	6.288	5.139
AR(1)	172.3	4.039	4.682	4.755	5.169	4.869
AR(2)	168.047	3.563	4.997	5.478	6.488	5.654
MA(1)	170.826	3.749	4.527	5.32	6.006	5.284
ARMA(1,1)	168.501	3.413	4.321	6.295	6.353	5.656

Using cross-validation techniques, the AR(1) model stands out as the best fitting model. As explained above, this makes theoretical sense as 2/3 of the Senate stays the same for each new session.

Part C

```
ar1ar1.3 <- arima(DemSenateMaj, order = c(1,0,0),
                  seasonal = list(order = c(1,0,0), period = 3),
                  xreg = xcovariates, include.mean = TRUE)

ar1ar1.3.cv <- arimaCV(DemSenateMaj, order = c(1,0,0), forward=3,
                       seasonal = list(order= c(1,0,0), period=3),
                  xreg=xcovariates, include.mean = TRUE, minper = 20)
```

Table 7: Model Comparison

Model Components	AIC	$\hat{\sigma}^2$	N	$\hat{\beta}_{PM}$	$\hat{\beta}_{PU}$	$\hat{\beta}_{CT}$	$\hat{\beta}_{1994}$
AR(0)	183.044	26.333	28	1.792	0.358	3.403	8.503
				(1.402)	(0.635)	(1.954)	(2.091)
AR(1)	172.3	16.315	28	-0.341	-0.268	1.892	8.151
				(0.945)	(0.459)	(1.392)	(3.507)
AR(2)	183.044	12.698	28	-0.806	-0.884	0.103	6.611
				(0.635)	(0.367)	(1.091)	(2.581)
MA(1)	170.826	14.053	28	-0.464	-1.224	1.784	7.377
				(0.767)	(0.339)	(1.467)	(2.204)
ARMA(1,1)	168.501	11.651	28	-0.72	-1.309	0.807	9.131
				(0.576)	(0.24)	(1.134)	(2.171)
$AR(1)AR(1)_3$	169.047	13.117	28	-0.494	-0.301	0.728	6.887
				(0.641)	(0.388)	(1.029)	(3.045)

Table 8: Model Comparison

Model Components	AIC	RMSE	MAE_1	MAE_2	MAE_3	Average MAE
AR(0)	183.044	26.333	4.555	4.575	6.288	5.139
AR(1)	172.3	16.315	4.682	4.755	5.169	4.869
AR(2)	168.047	12.698	4.997	5.478	6.488	5.654
MA(1)	170.826	14.053	4.527	5.32	6.006	5.284
ARMA(1,1)	168.501	11.651	4.321	6.295	6.353	5.656
$AR(1)AR(1)_3$	169.047	13.117	4.802	5.482	5.659	5.314

The new model does not outperform the AR(1) model. Overall, it doesn't even rank in the top half of the models we fit based on MAE. There's certainly a theoretical reason why the AR(1)AR$(1)_3$ model might fit in the Senate. As explained previously, only a third of the Senate is up for reelection every year, so a 3-period seasonal cycle would capture any recurring "cohort" patterns. This is a pattern that wouldn't make sense in the House. However, ultimately we don't see strong evidence of such a pattern and the AR(1) remains the best model.