CSSS/POLS 512:
Time Series and Panel Data for the Social Sciences

Problem Set 2

Professor: Chris Adolph, Political Science and CSSS

Spring Quarter 2026

Due via Canvas by the start of class on Monday 11 May 2026

General instructions for homeworks: Homework can be handwritten or typed. For any exercises done with R or other statistical packages, you should attach your code at the end of the document. All other materials should be collated in order by problem. The most readable and elegant format for homework answers incorporates student comments and graphics into a seamless narrative, as one would see in an article or textbook. Working in groups on R code is allowed, but (1) each member of the group must provide his or her own writeup and (2) you must list all members of your group on the first page of your assignment.

Problem 1: Analyzing US House seat shares using ARMA

[70 points total.] Since 1963, the US House of Representatives has consisted of 435 elected voting members serving two-year terms. Every seat in the House is up for election in November of even-numbered years to seat the Congress that will serve in the following two years. Thus, the 2016 election determined the 435 members of the House for the 115th Congress, serving from 2017–2018.

We will study the evolution of the time series of the number of seats won by Democrats (or by independents who caucus with the Democratic Party) in each elec-

Variable	Description
`Congress`	session of Congress (effectively a time index)
`StartYear`	the first year of each two-year session
`DemSenateSeats`	the number of Democrats (and independents caucusing with the Democratic Party) elected to the Senate in this session of Congress
`DemSenateMaj`	the size of the Democratic Senate Majority, or `DemSenateSeats` minus 50
`DemHouseSeats`	the number of Democrats (and independents caucusing with the Democratic Party) elected to the House in this session of Congress
`DemHouseMaj`	the size of the Democratic Senate Majority, or `DemHouseSeats` minus 217
`Midterm`	whether this session was elected in a midterm election (1) or a presidential election (0)
`DemPresident`	whether the president on the last election day was a Democrat (1) or a Republican (0)
`Unemployment`	the monthly unemployment rate at the time of the election of this session of Congress
`UnemDeviation`	the difference between pre-election unemployment and mean unemployment, 1963–2016 (which was 6.075%)
`Coattails`	1 if the presidency shifted to the Democrats on election day, -1 if the presidency shifted to the Republicans, and 0 if the party of the president was unchanged
`PartisanMidterm`	1 for midterms in which the Democrats hold the presidency, -1 for midterms in which there is a Republican president, and 0 in presidential elections
`PartisanUnem`	equal to `UnemDeviation` when a Democrat is president, and to $-1\times$ `UnemDeviation` when a Republican is president
`Pre1994`	1 if this Congress was elected before 1994, 0 otherwise

Table 1. Codebook for Congressional Seats data. Data are in `congress.csv`, and are taken from the Bureau of Labor Statistics (unemployment) and Wikipedia (all other raw variables), or constructed from these data by your instructor.

tion held from 1963 to 2016 (a total of 28 observations). As substantive interest focuses on the party in control and the size of their majority, we will focus our analysis on these outcomes, where positive values indicate the size of a Democratic majority and negative values the size of a Republican majority.

We will also consider three possible influences on the size of the Democratic majority. First, when one party sweeps the other party out of the presidency, they tend to bring in a wave of co-partisans to Congress who are "clinging to the president's coattails." Second, the party of the president tends to do badly in midterm elections (those that do not involve a presidential election; e.g., 2010, 2014, and 2018). The usual explanation is that voters frustrated with the president cannot replace him, but can only offset his power with opposition in Congress.[1] Third, in all elections, voters tend to attribute economic performance to the party of the president. For example, keeping unemployment below its long term average should help the president's party at the expense of the opposition, and *vice versa* when unemployment is higher than usual.

These three explanations leave a lot out: for example, changes in the use of redistricting for partisan advantage, in the partisan composition of the electorate, and especially the transition of the Southern Democrats to the Republican Party. Because incumbency provides a strong advantage to sitting members of Congress, arguably many of these changes did not act gradually, but with a "bang" when a sudden shock caused many incumbents to lose or retire. The shock in question is the 1994 midterm election; to account for the possibility it reflects a "structural break" in the level of the time series, we will also consider a control for whether our observations come before or after this watershed.

In the file `congress.csv`, you will find the variables described in Table 1. Examine the data file, and note well the behavior of these variables over time. Then work through the following exercises:

a. [10 points.] Plot the time series `DemHouseMaj` and plot its ACF and PACF. Perform augmented Dickey-Fuller and Phillips-Peron tests for unit roots. Describe your findings, being sure to describe what kind(s) of time series process may be at work. Now "demean" the data by period, removing the pre-1994 mean from cases before 1994, and the post-1994 mean from cases after 1994. Make new time series, ACF, and PACF plots. If 1994 represents a "structural break" in the level

1 More subtle theories exist; see Alberto Alesina and Howard Rosenthal, 1995, *Partisan Politics, Divided Government, and the Economy*, Cambridge University Press.

of the Democratic majority, what effect does that have on your diagnosis of the behavior of the time series?

b. [10 points.] Fit an AR(0) regression to the time series `DemHouseMaj` controlling for the covariates `PartisanMidterm`, `PartisanUnem`, and `Coattails`, which test the three theories mentioned above. Also control for `Pre1994` to allow for a structural break. Present the results in a table, being sure to note the coefficients, their standard errors, the AIC for the entire model, the standard error of the regression, and the number of observations. Format the table nicely, as if for a paper, and describe what you have found substantively as well as you can.

c. [15 points.] Now fit the following additional models and add them to the table you made in part **b**: (*i.*) an AR(1) model; (*ii.*) an AR(2) model; (*iii.*) an MA(1) model; (*iv.*) an ARMA(1,1) model. Make sure to include the same four controls as in part **b.** Discuss the substantive and statistical similarities and differences across all five fitted models.

d. [10 points.] Perform a rolling-windows cross-validation of all five models using a window of 20 periods and forecasting forward 3 periods. Place in a table the following goodness of fit statistics for all five models: AIC, in-sample root mean squared error, and the cross-validation mean absolute error (MAE) up to 1, 2, and 3 periods ahead, respectively, as well as the average of these three cross-validation MAEs. Based on these statistics, select a final "best" model.

e. [25 points.] Using the model you selected in part **d.**, forecast the size of the Democratic majority in the US House in the 2018, 2020, and 2022 elections for three scenarios. For all three scenarios, assume the factually correct scenario that the Democrats recapture the presidency in 2020, compute appropriate values of `PartisanMidterm` and `Coattails`, and set `Pre1994` to 0. For unemployment, assume the following:

Scenario	Counterfactual
1	unemployment stays at 4.6% for all three elections
2	unemployment falls to 3.6% for all three elections
3	unemployment rises to 5.6% for all three elections

For each scenario, report or graph the predicted Democratic majority and its 95% confidence (or predictive) interval for the 2018, 2020, and 2022 elections.

Describe the substantive impact of your forecast results in as much detail as you feel comfortable, as well as how much confidence we should have in the forecasts (you are welcome, but not required, to compare your forecasts to actual history).

NB: As a check on your work, for each scenario and year also report the table of counterfactual covariate values you used to make your forecasts. Be very careful when constructing these values to capture to logic of the covariates; each one is tricky in its own way. To carry out the forecasts, you may use either `predict()` or the `simcf` library's `ldvsimev()`.

Problem 2: Analyzing US Senate seat shares using ARMA

[30 points total.] Since 1963, the US Senate has consisted of 100 elected voting members serving staggered six-year terms. Roughly one-third of the seats in the Senate are up for election in each even-numbered year. As a result, the Senate has three "classes" of seats. For example, the class of 2012 was up for re-election in 2018. Because 2012 was a good year for Democrats (due to Barack Obama's coattails, among other factors), that meant the Democrats had many seats to defend in 2018, putting them at a disadvantage relative to 2016.

a. [10 points.] Plot the time series `DemSenateMaj` and plot its ACF and PACF. Perform augmented Dickey-Fuller and Phillips-Peron tests for unit roots. Now "demean" the data by period, removing the pre-1994 mean from cases before 1994, and the post-1994 mean from cases after 1994. Make new time series, ACF, and PACF plots and compare your results. Diagnosis the time series, accounting for the possibility of a structural break.

b. [10 points.] Repurpose your code from Problem 1 to model the time series `DemSenateMaj`. In particular, control for `PartisanMidterm`, `PartisanUnem`, `Coattails`, and `Pre1994`, and consider five models: an AR(0) model, an AR(1) model, an AR(2) model, an MA(1) model, and an ARMA(1,1) model. Recreate the two tables you made in Problem 1 (the table of coefficients and the table of goodness of fit statistics) for the Senate data. How do the substantive results compare to the House models?

c. [10 points.] Now estimate a sixth model: an AR(1)AR(1)$_3$. Add this model to the two tables you made in part **b.** How well does the new model do? What

model is best overall? Can you provide a substantive rationale for using either an MA(1) or an AR(1)AR(1)$_3$ model to model the US Senate, but not the House?