Gov 1000 Midterm 2: Incumbency Advantage in U.S. Senate Elections (1912–1992)

The Li'l Dictator

November 21, 2002

Abstract

This paper modifies Gelman and King's (1990) study of incumbency advantage in the U.S. House of Representatives to study incumbency advantage in the U.S. Senate. After controlling for state-specific factors, incumbency advantage remains positive and significant, averaging 3.0% from 1912 to 1992, with a 95% confidence interval of (2.1%, 4.0%).

1 Introduction

While other scholars have studied incumbency advantage in elections to the U.S. House of Representatives, incumbency advantage in the U.S. Senate remains unexplored. This paper modifies Gelman and King's (1990) study of incumbency advantage in the House to study incumbency advantage in the Senate. After controlling for state-specific factors (such as the Democratic candidate's proportion of the vote in the previous election, six years prior, and the party affiliation of the other senator), incumbency advantage in senate elections averages 3.0%, with a 95% confidence interval from 2.1% to 4.0%.

2 Problems Applying Gelman and King's Model to Senate Elections

Gelman and King propose the following model for estimating incumbency advantage for House elections in a given congressional district:

$$E(\nu_2) = \beta_0 + \beta_1 \nu_1 + \beta_2 P_1 + \psi I_2. \tag{1}$$

They define ν_t as the proportion that the Democratic candidate receives in election $t = (1,\ 2)$. P_t is the party of the winner in election 1.[1] A dummy variable I_2 is -1 for a Republican incumbent, 0 for an open seat, and 1 for a Democratic incumbent. In a manner consistent with the other literature on incumbency advantage, Gelman and King do not consider uncontested elections.

Applying this model to Senate elections is problematic because the Senate is a qualitatively different institution than the House. The framers of the Constitution intended that the Senate act as a stabilizing force against the populist House. For example, until ratification of the 17th Amendment in 1913, senators were selected by state legislatures rather than by popular vote. Senators serve six year terms and are divided into three classes, with only one class standing for election in any given congressional election year. Unlike congressional districts, states (the geographic unit of Senate representation) are not subject to either redistricting or reapportionment. Taken together, these factors create and sustain a chamber with a relatively stable composition over time, which suggests that incumbency should have at least some explanatory power for Senate election outcomes.

More specifically, a literal application of Gelman and King's specific model to Senate elections is problematic for three reasons.

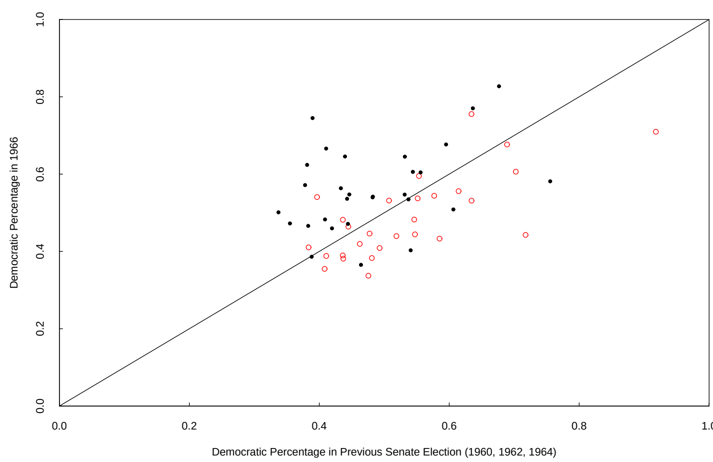
First, terms in the Senate are three times a long as terms in the House. If voters have short time horizons, then it is reasonable to hypothesize that ν_1 may have less of an influence on Senate elections than on House elections.

[1] Gelman and King actually propose $E(\nu_2) = \beta_0 + \beta_1 \nu_1 + \beta_2 P_2 + \psi I_2$, but they define P_2 as the party of the winner in the election at time $t = 1$. For clarity and consistency, I use P_1 instead.

Second, staggered Senate elections restrict the sample size for any given electoral year to a maximum of 34 states, compared to 435 districts. Regression analysis in a pair-wise comparison may not be appropriate for such limited data.

Finally, while one district elects one representative every two years, each state elects two senators on a staggered schedule such that when one senator stands for election, the other seat is not contested. Gelman and King's model for House elections does not capture this feature of Senate elections. Indeed, because the previous election (for the seat being contested in the current election) was last put to the vote six years ago, the party affiliation of the winner in the more recent, alternate election (either two or four years ago) is a better predictor for partisan swing.

Figure 1: Comparison of the party affiliation of the winner of the previous election (open circles) and the party affiliation of the winner of the alternate senate election (solid circles).



As Figure 1 illustrates, using the party affiliation of the winner six years ago produces a counter-intuitive measure of partisan swing. If incumbency were a neutral factor, the Democratic proportion of the vote in 1966 should be uncorrelated to the Democratic proportion of the vote in the 1960, 1962, or 1964 elections. The points should be randomly distributed about the 45-degree line. Using the 1960 data to predict the Democratic proportion of the vote in 1966 shows that a higher proportion of the vote in 1960 is correlated with a lower proportion in 1966; that other things being equal, incumbency is a *dis*advantage. In contrast, using the 1964 and 1962 proportions in relation to the 1966 data show a positive incumbency effect as more of the solid circles are clustered above the 45 degree line. Thus, the model proposed below includes an indicator for the party affiliation of the seat not being contested, and omits an indicator for the party affiliation of the winner of the election six years ago.

3 Estimating Incumbency Advantage

For any given state in election t, let O_t indicate the party affiliation of the senator not up for reelection at time t such that if O_t is -1, the sitting senator is a Republican and 1 if the sitting senator is a Democrat. The other variables are defined as in the Gelman and King specification. For a Senate election at time t:

$$E(\nu_t) = \beta_0 + \beta_1 \nu_{t-1} + \beta_3 O_t + \psi I_t \tag{2}$$

This model is quite robust and parsimonious. Using the data set described in Appendix A, the linear regression fit to this model for the Senate elections occurring in 1966 is $E(\nu_{1966}) = 0.335 + 0.274\nu_{1960} + 0.003O_{1966} + 0.081\psi$. A β_1 of 0.274 indicates that a *ceteris paribus* one percentage point increase in the Democratic proportion of the 1960 vote will increase the Democratic proportion of the 1966 vote by 0.274 percentage points. All other factors held equal, if the senator not standing for election in 1966 is a Republican, the Democratic proportion of the the 1966 vote total will decrease by 0.3 percentage points; conversely if the other senator is a Democrat, the Democratic proportion of the 1966 vote total will increase by 0.3 percentage points. The marginal effect of a Democratic incumbent on the Democratic proportion of the 1966 vote is a positive 8.1 percentage points; an open seat has no effect; and a Republican incumbent decreases the Democratic proportion of the vote by 8.1 percentage points. The variation in the Democratic vote in the 1966 senate elections (R^2) accounted for in this model is 69.2%.

Table 1 summarizes the results of six other specifications, including a literal application of Gelman and King's model. I use the mean of the vector of R^2 statistics as a measure for the overall fittedness of the model to the data. I use the mean of ψ to evaluate whether the model over- or under-estimates ψ relative to the best model (specified in equation 2). I use the 95% confidence interval calculated from the vector of ψs as a measure of the variability in this key causal variable.

An immediate observation is that excluding ν_{t-1} in specification 4 results in a markedly poorer fit than the other specifications. Consistent with Gelman and King's analysis of House elections, my model includes ν_{t-1} to eliminate a large source of potential bias.

Specifications which include Gelman and King's P_{t-1} variable increase the variability observed in ψ_t, without an appreciable improvement in the fittedness of the regression line. For example, comparing specifications 1 and 3 or 2 and 4 show a wider 95% confidence interval for the specification including P_{t-1}. Thus, the model described by Equation 2 omits Gelman and King's variable for the party of the winner of the previous election.

Comparison of specifications 5 and 7 show that excluding O_t increases the variability in ψ by about 70% and overestimates ψ. Hence, the model includes O_t to refine and provide an accurate estimator for ψ.

Table 1. Summary statistics for seven specifications.
(n = 38 for each specification)

	Variables Included	Mean R^2	Mean ψ	95% CI for ψ
1	$\nu_{t-1}, P_{t-1}, O_t, I_t$	0.60	0.029	(0.018, 0.040)
2○	ν_{t-1}, P_{t-1}, I_t	0.57	0.027	(0.017, 0.038)
3*	ν_{t-1}, O_t, I_t	0.58	0.030	(0.020, 0.039)
4	P_{t-1}, O_t, I_t	0.46	0.031	(0.019, 0.044)
5	$\nu_{t-1}, P_{t-1}, O_t, I_{t-1}, I_t$	0.63	0.030	(0.018, 0.043)
6	$\nu_{t-1}, O_t, I_{t-1}, I_t$	0.60	0.030	(0.021, 0.040)
7	$\nu_{t-1}, P_{t-1}, I_{t-1}, I_t$	0.59	0.037	(0.015, 0.058)

* Specification given in this paper.

○ Gelman and King's specification applied to Senate elections.

Comparison of specifications 3 and 6 shows that including I_{t-1} does not have an appreciable effect on either the estimate of ψ or the 95% confidence interval. In the interests of parsimony, I_{t-1} is omitted from the specification.

4 Conclusion

Even after controlling for Democratic proportion in the previous Senate election six years prior and the party affiliation of the other senator, the coefficient for ψ remains significant and positive. Although no time-series trends were observed in a pair-wise comparison of elections, incumbency advantage from 1912 to 1992 was estimated to be 3.0% on average, with a 95% confidence interval of (2.1%, 4.0%). Further research may require the construction of a pooled data set that contains additional variables to control for systemic factors, such as the party affiliation of the president and the party in control of Congress.

Appendix A: Data Documentation

This analysis utilizes a data set which covers elections to the U.S. Senate for the period from 1912 to 1992. For each year-state combination, this data set initially contained the Democratic proportion of the presidential vote, the Democratic proportion of the Senate vote, incumbency status, the state's electoral votes, and the number of votes for the Democratic and Republican Senate candidates. I do not consider the variable for the Democratic proportion of the presidential vote in election t because it is not causally prior to the incumbency of the senator standing for election and is only available every other congressional election.

I reclassify the incumbency variable to be -1 for a Republican incumbent, 0 for an open seat, and 1 for a Democratic incumbent to be consistent with Gelman and King's definitions. I generate a dummy variable for the party affiliation of the other senator, coded 1 for a Democrat and -1 for a Republican.

(See Appendix B for details.)

Since the proportion of the vote for the Democratic Senate candidate is my dependent variable, I remove all elections missing this variable. Furthermore, because it was inserted from another source in the original dataset, I replace it with a proportion generated from the data on votes for the Democratic and Republican candidates.

Because incumbency is the primary causal effect examined, and all the incumbency variables are missing for 1916, I track down the missing incumbency variables from `http:\\www.senate.gov\`.

Prior to 1958, there was a theoretical maximum of 32 senators up for election at any one time. However, because the 17th Amendment (requiring that Senators be elected by popular vote) was not ratified until 1913, the data for 1912 is largely incomplete, with data only for seven states. For elections from 1914 to the 1950s, there are usually 25 to 30 Senate seats in the data set. After the 1950s, this increases to 30 to 34 senate seats. The size of the sample expands over time because I remove elections not contested by one of the two major parties, removing all the elections for which the Democratic proportion of the Senate vote is 0 (for an election without a Democratic candidate) or 1 (for an election without a Republican candidate). This has the practical effect that although Louisiana, South Carolina, and Mississippi had senators prior to 1950, 1956, and 1960, respectively, these were the first elections in which the Republicans fielded senatorial candidates in those states. Alaska and Hawaii did not become states until 1959, and they were not added to the data set until 1960 and 1958, respectively.

Table A. Summary statistics for U.S. Senate election data set, 1912–1992 (n = 1,178)

	Democratic Proportion	Party of Winner	Party of Other Senator	Incumbency
mean	0.520	0.085	0.097	0.075
std. deviation	0.130	0.985	0.981	0.765
median	0.508	1	1	0
minimum	0.121	-1	-1	-1
maximum	0.943	1	1	1
1st quartile	0.446	1	-1	-1
3rd quartile	0.589	1	1	1

Appendix B: Party Affiliation for the Senator Not Facing Reelection

Because each state has two senators who are elected on a staggered schedule, only one senator from a given state stands for election in any given election year. I create a dummy variable to indicate the party affiliation of the senator in the seat not up for reelection. This variable is coded 1 for a Democrat, and -1 for a Republican.

I assume that the elections alternate such that one seat is contested, then the other seat is contested, such that the seat that was contested in the previous election is not being contested in the current election. This variable may be coded incorrectly for some states due to special mid-term elections, sitting senators from third parties, and other instances where an election year is missing for the state. Before performing more detailed analysis on this variable, future researchers should check the accuracy of the coding.

The data on the party affiliation on the sitting senator for the first election is drawn from `http://www.senate.gov/pagelayout/senators/f_two_sections_with_teasers/states.htm`.

Appendix C: R Code

```
##  This function loads the data from each "sXXX.txt" file into one data set.

> load.data <- function (end.year) {
+   result <- data.frame()
+   for (year in seq(912, end.year, by = 2)) {
+    file <- paste("s", year, ".txt", sep = "")
+    x <- read.table(file, col.names = c("year", "state", "dem.pres", "dem.sen",
"incum", "e.votes", "dem.votes", "rep.votes"), na.strings = "-9")
+    x$year <- 1000 + year
+    result <- rbind (result, x)
+   }
+  result
+ }

##  I save this data object as "data1".

> data1 <- load.data(992)
> dim(data1)
[1] 2091    8

##  Performing summary(data1) shows me that there are several problems
##  with the data.  The incumbency variable is coded on a 0 to 2 scale
##  instead of a -1 to 1 scale.  There are a lot of missing values in
##  the dem.sen, incum, dem.votes, and rep.votes columns.

> summary(data1)
      year          state          dem.pres           dem.sen
 Min.   :1912   Min.   : 1.00   Min.   :   0.1173   Min.   :  0.0000
 1st Qu.:1932   1st Qu.:23.00   1st Qu.:   0.4098   1st Qu.:  0.4494
 Median :1952   Median :43.00   Median :   0.4911   Median :  0.5186
 Mean   :1952   Mean   :40.88   Mean   :   0.5072   Mean   :  0.5525
 3rd Qu.:1972   3rd Qu.:61.00   3rd Qu.:   0.5807   3rd Qu.:  0.6247
 Max.   :1992   Max.   :82.00   Max.   :   1.0000   Max.   :  1.0000
                                NA's   :1181.0000   NA's   :828.0000
     incum           e.votes         dem.votes          rep.votes
 Min.   : 0.000   Min.   :  3.00   Min.   :      0   Min.   :      0
 1st Qu.: 0.000   1st Qu.:  4.00   1st Qu.: 111939   1st Qu.:  87101
 Median : 1.000   Median :  8.00   Median : 273779   Median : 220350
 Mean   : 1.183   Mean   : 10.94   Mean   : 520561   Mean   : 471693
 3rd Qu.: 2.000   3rd Qu.: 13.00   3rd Qu.: 634324   3rd Qu.: 577184
 Max.   : 9.000   Max.   : 54.00   Max.   :5173467   Max.   :5143409
 NA's   :843.000  NA's   :161.00   NA's   :    803   NA's   :    798
```

```
##  I begin by redefining the incumbency variable to match Gelman and
##  King's definition.

> new.incum <- data2$incum
> new.incum[data2$incum == 0] <- -1
> new.incum[data2$incum == 2] <- 0
> data2$incum <- new.incum

##  Now, I check to see if the variable representing the Democratic
##  proportion of a two-party vote (dem.sen).  Because this data was
##  entered from another source, I want to see how closely it matches
##  the proportion calculated from the data in the data set.

> check <- (clean$dem.votes/(clean$dem.votes + clean$rep.votes)) - clean$dem.sen
> sum(check)
[1] 4.766263

##  Because dem.sen seems to be off, I generate a new variable to represent
##  the Democratic proportion of a two party vote from the data and replace
##  dem.sen.  I check to make sure that the new dem.sen is consistent with the data.

> data2$dem.sen <- data2$dem.votes/(data2$dem.votes + data2$rep.votes)
> check.new <- data2$dem.sen - data2$dem.votes/(data2$dem.votes + data2$rep.votes)
> sum(check.new)
[1] 0

##  I subset out the uncontested elections.

> data2 <- data2[! data2$dem.sen %in% c(0,1),]


##  I generate Gelman and King's variable for the party affiliation of the
##  winner of the previous election and check to make sure that it is entered
##  correctly.

> dem.win <- ifelse(data2$dem.sen > 0.5, 1, -1)
> summary(dem.win)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1.0000 -1.0000  1.0000  0.0619  1.0000  1.0000
> data2 <- cbind(data2, dem.win)

##  I excerpt out the electoral votes variable.

> data3 <- data2[c("year", "state", "dem.pres", "dem.sen", "incum", "dem.votes", "rep.votes'

##  I generate a dummy variable for each senate class (to indicate when a
```

```
##  particular senate seat comes up for reelection).  This is for my reference
##  in gathering additional data.

> c1 <- as.integer(data3$year %in% seq(1916, 1988, by = 6))
> c2 <- as.integer(data3$year %in% seq(1912, 1990, by = 6))
> c3 <- as.integer(data3$year %in% seq(1914, 1992, by = 6))
> c2 <- c2*2
> c3 <- c3*3
> class <- c1 + c2 + c3
> data3 <- cbind(data3, class)

##  I fix this class variable for the two midterm elections in the data set.

> data3[data3$state == 82 & data3$year == 1990, c("class")] <- 1
> data3[data3$state == 82 & data3$year == 1990, c("class")] <- 3

##  I see that there is no incumbency data for any of the 1916 elections.
##  This interfers with my programs, so I insert this data.

> summary(data3[data3$year == 1916,c("incum")])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
                            NaN                      30
> fix.incum <- data3[data3$incum %in% c(NA),]
> data4 <- data3[! data3$incum %in% c(NA),]
> fix.incum <- read.table("incum.txt", col.names = c("year", "state",
"dem.pres", "dem.sen", "incum", "dem.votes", "rep.votes",
"dem.win", "class"), na.strings = "-9")
> data5 <- rbind(data4, fix.incum)

##  The following function generates the dummy variable for the party affilation
##  of the senator not up for election.

> other.sen.fn <- function (dataframe1) {
+   states <- c(1:6, 11:14, 21:25, 31:37, 40:49, 51:54, 56, 61:68, 71:73, 81:82)
+   result <- data.frame(year = dataframe1$year, state = dataframe1$state)
+   many.states <- data.frame()
+   for (s in states) {
+    one.state <- dataframe1[dataframe1$state == s,]
+    lag <- one.state$dem.win
+    x <- length(lag)
+    z <- x + 1
+    y <- array(NA, z)
+    y[2:z] <- lag
+    other.sen <- y[1:x]
+    one <- cbind(one.state, other.sen)
+    one.state <- one[c("year", "state", "other.sen")]
```

```
+    many.states <- rbind(many.states, one.state)
+    }
+   merge(result, many.states, by = c("year", "state"))
+   }
> other <- other.sen.fn(data5)
> keep <- other[!other$other.sen %in% c(NA),]
> replace <- read.table("firsts.txt", col.names =
c("year", "state", "other.sen"))
> other <- rbind(keep, replace)
> data6 <- merge(data5, other, by = c("year", "state"))

##  I insert additional variables for systemic comparisons.

> additional <- read.table("additional.txt", col.names =
c("year", "pres", "house", "senate", "div.gov"))
> data7 <- merge(data6, additional, by = c("year"))

##  I define the variable congress to be 1 for Democratic control
##  of both chambers, 0 if one party controls one and the other the
##  other, and -1 for Republican control of both chambers.

> congress <- as.integer(data7$house + data7$senate == 0)
> congress <- congress*-1
> congress <- congress + as.integer(data7$house + data7$senate == 2)
> data8 <- cbind(data7, congress)
> data8 <- data8[c("year", "state", "dem.sen", "dem.win", "other.sen",
"pres", "congress", "div.gov", "incum")]
> clean <- data8

##  I save this data object and begin my analysis.

> save(clean, file = "Senate.Rdata")
> dim(clean)
[1] 1178    9

##  I generate summary statistics (Table A in Appendix A) for this dataset.

> summary(clean)
      year          state         dem.sen         dem.win           other.sen           pres
 Min.   :1912   Min.   : 1.0   Min.   :0.121   Min.   :-1.0000   Min.   :-1.0000   Min.   :(
 1st Qu.:1934   1st Qu.:22.0   1st Qu.:0.446   1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:(
 Median :1956   Median :41.0   Median :0.508   Median : 1.0000   Median : 1.0000   Median :(
 Mean   :1954   Mean   :39.4   Mean   :0.520   Mean   : 0.0849   Mean   : 0.0968   Mean   :(
 3rd Qu.:1974   3rd Qu.:61.0   3rd Qu.:0.589   3rd Qu.: 1.0000   3rd Qu.: 1.0000   3rd Qu.:1
 Max.   :1992   Max.   :82.0   Max.   :0.943   Max.   : 1.0000   Max.   : 1.0000   Max.   :1
    congress         div.gov          incum
```

```
 Min.   :-1.000   Min.   :0.000   Min.   :-1.0000
 1st Qu.: 0.000   1st Qu.:0.000   1st Qu.:-1.0000
 Median : 1.000   Median :0.000   Median : 0.0000
 Mean   : 0.469   Mean   :0.309   Mean   : 0.0747
 3rd Qu.: 1.000   3rd Qu.:1.000   3rd Qu.: 1.0000
 Max.   : 1.000   Max.   :1.000   Max.   : 1.0000
> sd(clean$dem.sen)
[1] 0.130
> sd(clean$dem.win)
[1] 0.985
> sd(clean$other.sen)
[1] 0.981
> sd(clean$incum)
[1] 0.765

##  I generate a sample for a sample pair-wise comparison.

> e1966 <- clean[year == 1966,]
> e1960 <- clean[year == 1960,]
> sample2 <- merge(e1966, e1960, by = c("state"), suffixes = c("", ".last"))
> e1964 <- clean[year == 1964,]
> sample2a <- merge(e1966, e1964, by = c("state"), suffixes = c("", ".last"))
> e1962 <- clean[year == 1962,]
> sample2b <- merge(e1966, e1962, by = c("state"), suffixes = c("", ".last"))

##  I plot this data to show partisan swing.  (Figure 1)

> postscript("oneyear.ps")
> plot.default(sample2$dem.sen.last, sample2$dem.sen, col = 2,
xlab = "Democratic Percentage in Previous Senate Election
(1960, 1962, 1964)", ylab = "Democratic Percentage in 1966",
main = "", xlim = 0:1, ylim = 0:1, axes = TRUE, xaxs = "i",
yaxs = "i", tcl = 0.25)
> points(sample2a$dem.sen, sample2a$dem.sen.last, pch = 20)
> points(sample2b$dem.sen, sample2b$dem.sen.last, pch = 20)
> abline(0,1)
> dev.off()
null device
          1
##  The following regression generates the coefficients
##  for the sample pair-wise comparison.

> summary(lm(dem.sen ~ dem.sen.last + other.sen +
incum, data = sample2))

Call:
```

```
lm(formula = dem.sen ~ dem.sen.last + other.sen +
incum, data = sample2)

Residuals:
     Min       1Q   Median       3Q      Max
-0.11846 -0.03886 -0.00184  0.02922  0.16221

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.33464    0.07314    4.58  0.00012 ***
dem.sen.last  0.27384    0.13881    1.97  0.06015 .
other.sen     0.00339    0.01478    0.23  0.82043
incum         0.08147    0.02117    3.85  0.00077 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.063 on 24 degrees of freedom
Multiple R-Squared: 0.692,Adjusted R-squared: 0.654
F-statistic:   18 on 3 and 24 DF,  p-value: 2.46e-06

##  These are the specifications summarized in Table 1.  The functions return
##  data frames that allow me to easily view and manipulate the coefficients
##  in other statistical programs and functions.  I save the data frames with
##  names like "test1" to correspond to "spec1"

> spec1
function(df) {
  elec <- seq(1918, 1992, by = 2)
  results <- data.frame(year = elec, year.norm = NA, beta1 = NA, Tb1 = NA,
beta2 = NA, Tb2 = NA, beta3 = NA, Tb3 = NA, psi = NA, Tpsi = NA,
R2 = NA)
  for (y in elec) {
    this.elec <- df[df$year %in% c(y),]
    last.elec <- df[df$year %in% c(y - 6),]
    new <- merge(this.elec, last.elec, by = c("state"),
suffixes = c("", ".last"))
    lm.obj <- lm(dem.sen ~ dem.sen.last + dem.win.last + other.sen +
incum, data = new)
    sum.obj <- summary(lm.obj)
    x <- c(sum.obj$coefficients)
    results$beta1[results$year == y] <- x[2]
    results$beta2[results$year == y] <- x[3]
    results$beta3[results$year == y] <- x[4]
    results$psi[results$year == y] <- x[5]
    results$Tb1[results$year == y] <- x[12]
    results$Tb2[results$year == y] <- x[13]
```

```
    results$Tb3[results$year == y] <- x[14]
    results$Tpsi[results$year == y] <- x[15]
    results$R2[results$year == y] <- sum.obj$r.squared
    results$year.norm[results$year == y] <- y - 1917
  }
  results
 }
> test1 <- spec1(clean)
> spec2
function(df) {
  elec <- seq(1918, 1992, by = 2)
  results <- data.frame(year = elec, year.norm = NA,  beta1 = NA,
Tb1 = NA, beta2 = NA, Tb2 = NA, psi = NA, Tpsi = NA, R2 = NA)
  for (y in elec) {
    this.elec <- df[df$year %in% c(y),]
    last.elec <- df[df$year %in% c(y - 6),]
    new <- merge(this.elec, last.elec, by = c("state"),
suffixes = c("", ".last"))
    lm.obj <- lm(dem.sen ~ dem.sen.last + dem.win.last +
incum, data = new)
    sum.obj <- summary(lm.obj)
    x <- c(sum.obj$coefficients)
    results$beta1[results$year == y] <- x[2]
    results$beta2[results$year == y] <- x[3]
    results$psi[results$year == y] <- x[4]
    results$Tb1[results$year == y] <- x[10]
    results$Tb2[results$year == y] <- x[11]
    results$Tpsi[results$year == y] <- x[12]
    results$R2[results$year == y] <- sum.obj$r.squared
    results$year.norm[results$year == y] <- y - 1917
  }
  results
 }
> test2 <- spec2(clean)
> spec3
function(df) {
  elec <- seq(1918, 1992, by = 2)
  results <- data.frame(year = elec, year.norm= NA, beta1 = NA,
Tb1 = NA, beta3 = NA, Tb3 = NA, psi = NA, Tpsi= NA, R2 = NA)
  for (y in elec) {
    this.elec <- df[df$year %in% c(y),]
    last.elec <- df[df$year %in% c(y - 6),]
    new <- merge(this.elec, last.elec, by = c("state"),
suffixes = c("", ".last"))
    lm.obj <- lm(dem.sen ~ dem.sen.last + other.sen + incum,
data = new)
```

```
    sum.obj <- summary(lm.obj)
    x <- c(sum.obj$coefficients)
    results$beta1[results$year == y] <- x[2]
    results$beta3[results$year == y] <- x[3]
    results$psi[results$year == y] <- x[4]
    results$Tb1[results$year == y] <- x[10]
    results$Tb3[results$year == y] <- x[11]
    results$Tpsi[results$year == y] <- x[12]
    results$R2[results$year == y] <- sum.obj$r.squared
    results$year.norm[results$year == y] <- y - 1917
  }
  results
 }
> test3 <- spec3(clean)
> spec4
function(df) {
  elec <- seq(1918, 1992, by = 2)
  results <- data.frame(year = elec, year.norm = NA, beta2 = NA,
Tb2 = NA, beta3 = NA, Tb3 = NA, psi = NA, Tpsi = NA, R2 = NA)
  for (y in elec) {
    this.elec <- df[df$year %in% c(y),]
    last.elec <- df[df$year %in% c(y - 6),]
    new <- merge(this.elec, last.elec, by = c("state"),
suffixes = c("", ".last"))
    lm.obj <- lm(dem.sen ~ dem.win.last + other.sen + incum,
data = new)
    sum.obj <- summary(lm.obj)
    x <- c(sum.obj$coefficients)
    results$beta2[results$year == y] <- x[2]
    results$beta3[results$year == y] <- x[3]
    results$psi[results$year == y] <- x[4]
    results$Tb2[results$year == y] <- x[10]
    results$Tb3[results$year == y] <- x[11]
    results$Tpsi[results$year == y] <- x[12]
    results$R2[results$year == y] <- sum.obj$r.squared
    results$year.norm[results$year == y] <- y - 1917
  }
  results
 }
> test4 <- spec4(clean)
> spec5
function(df) {
  elec <- seq(1918, 1992, by = 2)
  results <- data.frame(year = elec, year.norm = NA, beta1 = NA,
Tb1 = NA, beta2 = NA, Tb2 = NA, beta3 = NA, Tb3 = NA,
beta4 = NA, Tb4 = NA, psi = NA, Tpsi = NA, R2 = NA)
```

```
  for (y in elec) {
   this.elec <- df[df$year %in% c(y),]
   last.elec <- df[df$year %in% c(y - 6),]
   new <- merge(this.elec, last.elec, by = c("state"),
suffixes = c("", ".last"))
   lm.obj <- lm(dem.sen ~ dem.sen.last + dem.win.last + other.sen +
incum.last + incum, data = new)
   sum.obj <- summary(lm.obj)
   x <- c(sum.obj$coefficients)
   results$beta1[results$year == y] <- x[2]
   results$beta2[results$year == y] <- x[3]
   results$beta3[results$year == y] <- x[4]
   results$beta4[results$year == y] <- x[5]
   results$psi[results$year == y] <- x[6]
   results$Tb1[results$year == y] <- x[14]
   results$Tb2[results$year == y] <- x[15]
   results$Tb3[results$year == y] <- x[16]
   results$Tb4[results$year == y] <- x[17]
   results$Tpsi[results$year == y] <- x[18]
   results$R2[results$year == y] <- sum.obj$r.squared
   results$year.norm[results$year == y] <- y - 1917
  }
  results
 }
> test5 <- spec5(clean)
> spec6
function(df) {
  elec <- seq(1918, 1992, by = 2)
  results <- data.frame(year = elec, year.norm = NA, beta1 = NA,
Tb1 = NA, beta3 = NA, Tb3 = NA, beta4 = NA, Tb4 = NA,
psi = NA, Tpsi = NA, R2 = NA)
  for (y in elec) {
    this.elec <- df[df$year %in% c(y),]
    last.elec <- df[df$year %in% c(y - 6),]
    new <- merge(this.elec, last.elec, by = c("state"),
suffixes = c("", ".last"))
    lm.obj <- lm(dem.sen ~ dem.sen.last + other.sen +
incum.last + incum, data = new)
    sum.obj <- summary(lm.obj)
    x <- c(sum.obj$coefficients)
    results$beta1[results$year == y] <- x[2]
    results$beta3[results$year == y] <- x[3]
    results$beta4[results$year == y] <- x[4]
    results$psi[results$year == y] <- x[5]
    results$Tb1[results$year == y] <- x[12]
    results$Tb3[results$year == y] <- x[13]
```

```
    results$Tb4[results$year == y] <- x[14]
    results$Tpsi[results$year == y] <- x[15]
    results$R2[results$year == y] <- sum.obj$r.squared
    results$year.norm[results$year == y] <- y - 1917
  }
  results
 }
> test5 <- spec5(clean)
> spec6
function(df) {
  elec <- seq(1918, 1992, by = 2)
  results <- data.frame(year = elec, year.norm = NA, beta1 = NA,
Tb1 = NA, beta3 = NA, Tb3 = NA, beta4 = NA, Tb4 = NA,
psi = NA, Tpsi = NA, R2 = NA)
  for (y in elec) {
    this.elec <- df[df$year %in% c(y),]
    last.elec <- df[df$year %in% c(y - 6),]
    new <- merge(this.elec, last.elec, by = c("state"),
suffixes = c("", ".last"))
    lm.obj <- lm(dem.sen ~ dem.sen.last + other.sen +
incum.last + incum, data = new)
    sum.obj <- summary(lm.obj)
    x <- c(sum.obj$coefficients)
    results$beta1[results$year == y] <- x[2]
    results$beta3[results$year == y] <- x[3]
    results$beta4[results$year == y] <- x[4]
    results$psi[results$year == y] <- x[5]
    results$Tb1[results$year == y] <- x[12]
    results$Tb3[results$year == y] <- x[13]
    results$Tb4[results$year == y] <- x[14]
    results$Tpsi[results$year == y] <- x[15]
    results$R2[results$year == y] <- sum.obj$r.squared
    results$year.norm[results$year == y] <- y - 1917
  }
  results
 }
> test6 <- spec6(clean)
> spec7
function(df) {
  elec <- seq(1918, 1992, by = 2)
  results <- data.frame(year = elec, beta1 = NA, Tb1 = NA,
beta2 = NA, Tb2 = NA, beta4 = NA, Tb4 = NA, psi = NA,
Tpsi = NA, R2 = NA)
  for (y in elec) {
    this.elec <- df[df$year %in% c(y),]
    last.elec <- df[df$year %in% c(y - 6),]
```

```
    new <- merge(this.elec, last.elec, by = c("state"),
suffixes = c("", ".last"))
    lm.obj <- lm(dem.sen ~ dem.sen.last + dem.win.last +
incum.last + incum, data = new)
    sum.obj <- summary(lm.obj)
    x <- c(sum.obj$coefficients)
    results$beta1[results$year == y] <- x[2]
    results$beta2[results$year == y] <- x[3]
    results$beta4[results$year == y] <- x[4]
    results$psi[results$year == y] <- x[5]
    results$Tb1[results$year == y] <- x[12]
    results$Tb2[results$year == y] <- x[13]
    results$Tb4[results$year == y] <- x[14]
    results$Tpsi[results$year == y] <- x[15]
    results$R2[results$year == y] <- sum.obj$r.squared
    results$year.norm[results$year == y] <- y - 1917
  }
  results
 }
> test7 <- spec7(clean)

##  The following function generates the mean and 95% confidence
##  intervals on the vector of psi values for any given specification.

> psi.ci <- function(test) {
+   coeff <- c(summary(lm(psi ~ 1, data = test))$coefficients)
+   a <- coeff[1]
+   b <- coeff[2]
+   c <- coeff[1] - 1.96*b
+   d <- coeff[1] + 1.96*b
+   cat("The mean, and lower and upper bounds on the 95% confidence
interval are:\n", a, c, d)
+  }

##  The 95% confidence intervals for psi from each specification are
##  summarized in Table 1.

> psi.ci(test1)
The mean, and lower and upper bounds on the 95% confidence
interval are:
 0.0291 0.0182 0.04

> psi.ci(test2)
The mean, and lower and upper bounds on the 95% confidence
interval are:
 0.0274 0.0171 0.0377
```

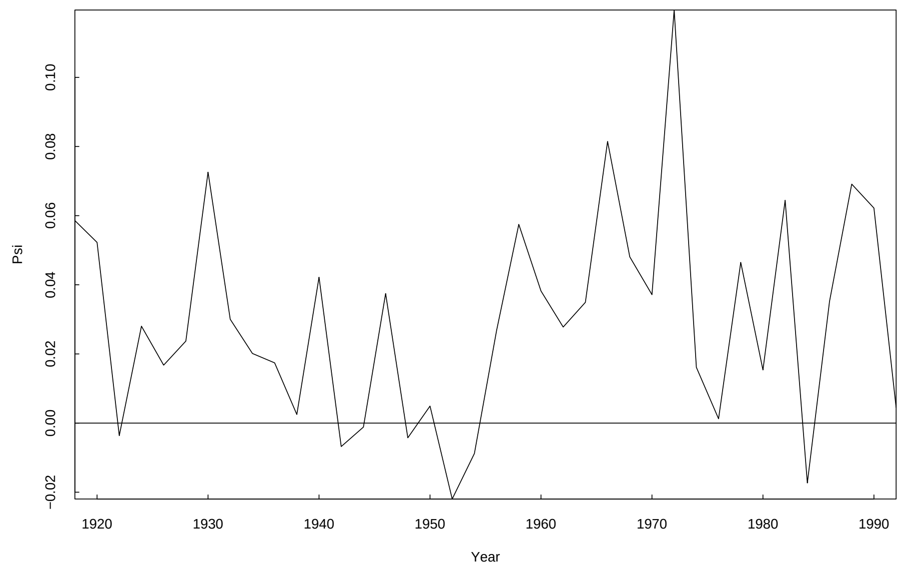
```
> psi.ci(test3)
The mean, and lower and upper bounds on the 95% confidence
interval are:
 0.0297 0.0201 0.0394

> psi.ci(test4)
The mean, and lower and upper bounds on the 95% confidence
interval are:
 0.0313 0.0192 0.0435

> psi.ci(test5)
The mean, and lower and upper bounds on the 95% confidence
interval are:
 0.0301 0.0175 0.0427

> psi.ci(test6)
The mean, and lower and upper bounds on the 95% confidence
interval are:
 0.0304 0.0205 0.0404

> psi.ci(test7)
The mean, and lower and upper bounds on the 95% confidence
interval are:
 0.0366 0.0154 0.0579

##  The mean R^2 values are also summarized in Table 1.

> mean(test1$R2)
[1] 0.604
> mean(test2$R2)
[1] 0.57
> mean(test3$R2)
[1] 0.576
> mean(test4$R2)
[1] 0.458
> mean(test5$R2)
[1] 0.625
> mean(test6$R2)
[1] 0.602
> mean(test7$R2)
[1] 0.591

##  The 70% figure on page 4, in the first full paragraph, is from:

> (0.0366-0.0154)/(0.0301-0.0175)
```

```
[1] 1.68

##  For Appendix A, this function returns the first year in the
## data set:

> first.elec <- function(df) {
+    states <- c(1:6, 11:14, 21:25, 31:37, 40:49, 51:54,
56, 61:68, 71:73, 81:82)
+    results <- data.frame(state = states, first.elec = NA)
+    for (s in states){
+    x <- df[df$state == s,]
+    y <- min(x$year)
+    results$first.elec[results$state == s] <- y
+    }
+   results
+  }
> first.elec(clean)
   state first.elec
1      1       1914
2      2       1912
3      3       1916
4      4       1914
5      5       1916
6      6       1916
7     11       1916
8     12       1916
9     13       1914
10    14       1914
11    21       1914
12    22       1914
13    23       1916
14    24       1914
15    25       1914
16    31       1914
17    32       1912
18    33       1912
19    34       1914
20    35       1916
21    36       1914
22    37       1914
23    40       1922
24    41       1914
25    42       1914
26    43       1916
27    44       1918
28    45       1950
```

```
29    46       1960
30    47       1914
31    48       1956
32    49       1916
33    51       1914
34    52       1914
35    53       1912
36    54       1916
37    56       1916
38    61       1914
39    62       1912
40    63       1914
41    64       1912
42    65       1914
43    66       1916
44    67       1914
45    68       1916
46    71       1914
47    72       1912
48    73       1914
49    81       1960
50    82       1958

##  The following function counts the number of states in each
##  election congressional election year.

> elections <- table(clean$year, clean$state)
> check.elec <- function(tb) {
+ elec <- seq(1912, 1992, by = 2)
+ results <- data.frame(year = elec, states = NA)
+ n <- nrow(tb)
+ for (i in 1:n) {
+  s <- sum(tb[i,])
+  results[i, c("states")] <- s
+  }
+ results
+ }
> check.elec(elections)
   year states
1  1912      7
2  1914     27
3  1916     30
4  1918     25
5  1920     29
6  1922     29
7  1924     28
```

```
8  1926     28
9  1928     28
10 1930     25
11 1932     29
12 1934     26
13 1936     25
14 1938     29
15 1940     27
16 1942     24
17 1944     29
18 1946     31
19 1948     27
20 1950     28
21 1952     27
22 1954     25
23 1956     28
24 1958     31
25 1960     29
26 1962     33
27 1964     32
28 1966     30
29 1968     32
30 1970     30
31 1972     33
32 1974     31
33 1976     30
34 1978     30
35 1980     33
36 1982     32
37 1984     31
38 1986     34
39 1988     33
40 1990     31
41 1992     33

##  The following graph shows that there is no trend in
##  psi over time.

> postscript("psitime.ps")
> plot.default(test3$year, test3$psi, type = "l", xlab = "Year", ylab = "Psi", axes = TRUE,
> abline(0,0)
> dev.off()
null device
          1
```

Optional Figure. ψ over time for specification three.




```
##  The following function shows that there is no trend in
##  psi over time.

> psi
function(test) {
 x1 <- summary(lm(psi ~ 1, data = test))
 x2 <- summary(lm(psi ~ year.norm, data = test))
 x3 <- summary(lm(psi ~ 1, data = test[1:17,]))
 x4 <- summary(lm(psi ~ year.norm, data = test[1:17,]))
 x5 <- summary(lm(psi ~ 1, data = test[18:38,]))
 x6 <- summary(lm(psi ~ year.norm, data = test[18:38,]))
 return(x1$call, x1$coefficients, x2$call, x2$coefficients,
x3$call, x3$coefficients, x4$call, x4$coefficients,
x5$call, x5$coefficients, x6$call, x6$coefficients)
 }

##  While the intercept for psi is significant, the rate
##  of variation for psi against year is not.  Thus, even
##  when I subset the data into before 1950 and 1950 and after,
##  I find no time-trends on psi.
```

```
> psi(test3)
lm(formula = psi ~ 1, data = test)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.0297    0.00492    6.03 5.69e-07

lm(formula = psi ~ year.norm, data = test)
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.021830   0.009872    2.21   0.0334
year.norm   0.000207   0.000225    0.92   0.3633

lm(formula = psi ~ 1, data = test[1:17, ])
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.023    0.00575       4  0.00103

lm(formula = psi ~ year.norm, data = test[1:17, ])
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.04206   0.010437    4.03  0.00109
year.norm   -0.00112   0.000532   -2.11  0.05207

lm(formula = psi ~ 1, data = test[18:38, ])
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.0351    0.00752    4.68 0.000145

lm(formula = psi ~ year.norm, data = test[18:38, ])
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.012333    0.03545   0.348    0.732
year.norm   0.000415    0.00063   0.659    0.518

##  The following function generates a pooled data set.

> pooled <- function(df) {
+  elec <- seq(1918, 1992, by = 2)
+  results <- data.frame()
+  for (y in elec) {
+    this.elec <- df[df$year %in% c(y),]
+    last.elec <- df[df$year %in% c(y - 6),]
+    new <- merge(this.elec, last.elec, by = c("state"),
suffixes = c("", ".last"))
+    results <- rbind(new, results)
+    }
+  results
+  }
> data9 <- pooled(clean)
> all.years <- data9[c("year", "state", "dem.sen", "dem.sen.last",
"dem.win.last", "pres", "div.gov", "other.sen", "
```

```
congress", "incum")]
> summary(all.years)
      year          state          dem.sen        dem.sen.last    dem.win.last          pres
 Min.   :1918   Min.   : 1.0   Min.   :0.127   Min.   :0.121   Min.   :-1.0000   Min.   :0.0
 1st Qu.:1938   1st Qu.:21.0   1st Qu.:0.444   1st Qu.:0.446   1st Qu.:-1.0000   1st Qu.:0.0
 Median :1958   Median :37.0   Median :0.504   Median :0.503   Median : 1.0000   Median :0.0
 Mean   :1957   Mean   :39.1   Mean   :0.514   Mean   :0.513   Mean   : 0.0609   Mean   :0.4
 3rd Qu.:1976   3rd Qu.:62.0   3rd Qu.:0.582   3rd Qu.:0.579   3rd Qu.: 1.0000   3rd Qu.:1.0
 Max.   :1992   Max.   :82.0   Max.   :0.929   Max.   :0.943   Max.   : 1.0000   Max.   :1.0
    div.gov       other.sen          congress          incum
 Min.   :0.00   Min.   :-1.0000   Min.   :-1.000   Min.   :-1.0000
 1st Qu.:0.00   1st Qu.:-1.0000   1st Qu.: 0.000   1st Qu.:-1.0000
 Median :0.00   Median : 1.0000   Median : 1.000   Median : 0.0000
 Mean   :0.33   Mean   : 0.0455   Mean   : 0.451   Mean   : 0.0658
 3rd Qu.:1.00   3rd Qu.: 1.0000   3rd Qu.: 1.000   3rd Qu.: 1.0000
 Max.   :1.00   Max.   : 1.0000   Max.   : 1.000   Max.   : 1.0000
> row.names(all.years) <- seq(nrow(all.years))
```

Appendix C: LaTeX Code

```
\begin{table}[h]
\begin{center}
Table A. Summary statistics for U.S. Senate election dataset, 1912 - 1992 \\(n = 1,178)}
\begin{tabular}{lcccc}
\\
& Democratic  & Party of  & Party of  &     \\
& Proportion & Winner & Other Senator & Incumbency  \\
\hline
mean & 0.520 & 0.085 & 0.097 & 0.075 \\
std. deviation & 0.130 & 0.985 & 0.981 & 0.765 \\
median   & 0.508 & 1 & 1 & 0 \\
minimum  & 0.121 & -1 & -1 & -1 \\
maximum  & 0.943 & 1 & 1 & 1 \\
1st quartile  & 0.446 & 1 & -1 & -1 \\
3rd quartile  & 0.589 & 1 & 1 & 1 \\
\hline
\end{tabular}
\end{center}
\end{table}
```

References

Gelman, A. and G. King. 1990. "Estimating Incumbency Advantge without Bias." *American Journal of Political Science* 34:1142–64.

Table B. Party affiliation of the senator not up for election, for the first election.

State	ICPSR	Year of First Election	Party Affiliation of Non-Contested Seat
Connecticut	1	1914	0
Maine	2	1912	-1
Massachusetts	3	1916	1
New Hampshire	4	1914	1
Rhode Island	5	1916	1
Vermont	6	1916	1
Delaware	11	1916	1
New Jersey	12	1916	1
New York	13	1914	1
Pennsylvania	14	1914	-1
Illinois	21	1914	0
Indiana	22	1914	1
Michigan	23	1916	1
Ohio	24	1914	1
Wisconsin	25	1914	0
Iowa	31	1914	0
Kansas	32	1912	0
Minnesota	33	1912	0
Missouri	34	1914	1
Nebraska	35	1916	1
N. Dakota	36	1914	0
S. Dakota	37	1914	0
Virginia	40	1922	1
Alabama	41	1914	1
Arkansas	42	1914	1
Florida	43	1916	1
Georgia	44	1918	1
Louisiana	45	1950	1
Mississippi	46	1960	1
N. Carolina	47	1914	1
S. Carolina	48	1956	1
Texas	49	1916	-1
Kentucky	51	1914	1
Maryland	52	1914	1
Oklahoma	53	1912	1
Tennessee	54	1916	1
W. Virginia	56	1916	1
Arizona	61	1914	1
Colorado	62	1912	1
Idaho	63	1914	0
Montana	64	1912	1
Nevada	65	1914	1
New Mexico	66	1918	-1
Utah	67	1914	0
Wyoming	68	1916	-1
California	71	1914	0
Oregon	72	1912	1
Washington	73	1914	0
Alaska	81	1960	1
Hawaii	82	1958	1