Creating Judge-Case Observations from the Songer Court of Appeals Database (As Well As Notes on the Database)

Doug M. Johnson* and David A. Gelman†

September 23, 2016

Introduction

This research note details two different but related items relating to the Songer Court of Appeals Database.[1] The first part of the note details how to create judge-case observations using the Songer Database in R.[2] The second part presents issues of case assignment and judge id codes in the database.

Generating Judge-Case Observations

The Court of Appeals Databases (CoAD) contains 18,195 cases.[3] For each case, between three and 15 (for an *en banc* review) judges sit on that case. However, in order to learn more about judicial behavior, case level observations are not sufficient. Rather, judge-case observations are needed. To do this, the CoAD needs to be transformed.

Following the Code

Here, we walk through our code step by step.[4]

First, the formalities:

*Department of Political Science, University of Rochester, Harkness Hall 333, Rochester, NY 14627 (djohns46@ur.Rochester.edu).

†Department of Political Science, University of Rochester, Harkness Hall 333, Rochester, NY 14627 (David.Gelman@Rochester.edu).

[1]The Court of Appeals Database can be accessed at `http://artsandsciences.sc.edu/poli/juri/appct.htm` and is discussed in detail in Songer (1998) and Hurwitz and Kuersten (2012).

[2]This procedure has previously been detailed for Stata in Collins (2008).

[3]The database begins in 1925 and contains 15 cases per year per circuit from 1925 to 1960 and 30 cases per year per circuit from 1961-1996.

[4]If you simply want to grab the R code and load it into R, see "CoAJudgeCaseCreation.R" located on the same page as this document.

```
rm(list=ls())
library(foreign) #to read .dta files
library(reshape) #to rename columns in dataframes
setwd('~/CoA Data Write-Up') #set your working directory
```

Next, you need to load in the data file that contains the CoAD. Go to `www.astsandsciences.sc.ede/poli/juri/appct.htm` and Download the Stata file from the "The Original U.S. Court of Appeals Courts Database 1925-1995" section and save it to the same location as you've set your working directory.

```
frame1=read.dta('cta96_stata.dta')
```

In order to create a judge-vote direction for each judge-case observation, we need to determine in what ideological direction each case was decided. For 15,060 of the 18,195, the primary issue type allowed a clean coding of a liberal or conservative direction.

```
#Identify the Conservative Votes
direct.1=subset(frame1,frame1$direct1==1)


#Identify the Liberal Votes on
direct.3=subset(frame1,frame1$direct1==3)


#Bind the 2 types back together. We now have the 15060 case on which we can
# make judgecase observations
direct1frame=rbind(direct.1,direct.3)


#Cases in which the directionality on the primary dimension is not discernible.
#1167 cases coded as a 2 or "mixed"
direct.2=subset(frame1,frame1$direct1==2)
#1966 cases we not on any conventional dimensionality
direct.0=subset(frame1,frame1$direct1==0)
```

Our next step is to subset the associated variables in the CoAD. In all there are 229 variables. We subset these down to 35 variables of interest. This entire next step can be omitted if you want to keep all 229 variables.

```
frame2=subset(frame1,select=c(casenum,year,month,day,
                              circuit,state,district,
                              method,initiate,numappel,
                              appnatpr,appbus,appnonp,
                              appfed,appsubst,appstate,
                              ap_stid,numresp,r_natpr,
                              r_bus,r_nonp,r_fed,r_subst,
                              r_state,r_stid,casetyp1,geniss,
                              direct1,majvotes,dissent,concur,
                              #The following variables are the judges for each cases vote.
                              codej1,codej2,j2vote1,j2maj1,
                              codej3,j3vote1,j3maj1,
                              codej4,j4vote1,j4maj1,
                              codej5,j5vote1,j5maj1,
                              codej6,j6vote1,j6maj1,
                              codej7,j7vote1,j7maj1,
                              codej8,j8vote1,j8maj1,
                              codej9,j9vote1,j9maj1,
                              codej10,j10vote1,j10maj1,
                              codej11,j11vote1,j11maj1,
                              codej12,j12vote1,j12maj1,
                              codej13,j13vote1,j13maj1,
                              codej14,j14vote1,j14maj1,
                              codej15,j15vote1,j15maj1))
```

Codej1 identifies the judge that wrote the majority opinion for the case. Note it does not have whether the judge is in the majority because being the opinion writer requires this.

Every case will remain in this frame, as every case has an opinion writer.

```
j1frame=subset(direct1frame,select=c(casenum,year,month,day,
                        circuit,state,district,
                        method,initiate,numappel,
                        appnatpr,appbus,appnonp,
                        appfed,appsubst,appstate,
                        ap_stid,numresp,r_natpr,
                        r_bus,r_nonp,r_fed,r_subst,
                        r_state,r_stid,casetyp1,geniss,
                        direct1,majvotes,dissent,concur,
                        codej1))
```

Now we rename the codej1 variable to a simple Judge ID variable, which we will rename every judge of interest moving forward, to be binded back together with the information we care about for each set of judge votes.

```
colnames(j1frame)[32]="JID"
```

Code the direction of the judges vote. Being the majority writer, the direction of the case is the same as the direction of the judges vote.

```
j1frame$votedirect=j1frame$direct1


#Again, we don't need to check if the judge is in the majority...
j1frame$majority=c(1)
#Nor that they wrote the majority opinion...
j1frame$wroteopin=c(1)
```

Having dealt with the issue of judges who are different because of their majority opinion writing, we can loop over the rest of the steps. Prior to this, we create an object that contains the list of the names of the variables that never change that we care about. Then we create a judge-case bin that is 34 columns wide (for our variable of interest)

```
sub_vars=c("casenum","year","month","day",
           "circuit","state","district",
           "method","initiate","numappel",
           "appnatpr","appbus","appnonp",
           "appfed","appsubst","appstate",
           "ap_stid","numresp","r_natpr",
           "r_bus","r_nonp","r_fed","r_subst",
           "r_state","r_stid","casetyp1","geniss",
           "direct1","majvotes","dissent","concur")


jcbin=matrix(rep(NA),1,35)
jcbin=as.data.frame(jcbin)
names(jcbin)=names(j1frame)
```

Now we do our two loops. The first loop is for judges 2 and 3 (which every case has).
The second loop covers cases for which there were more than three judges.

```
for (i in 2:3){ #there are 15 judges, we have judge 1 done though, so start at 2.
  jvar1=paste("codej",i,sep="") #columns for the ith judge have the ith value in
   # them, so creating that value for each judge
  jvar2=paste("j",i,"vote1",sep="") #same
  jvar3=paste("j",i,"maj1",sep="") #same
  j_vars=c(jvar1,jvar2,jvar3) #making a list of the ith judge values
  sub_vars2=c(sub_vars,j_vars) #making a list on which we will subset our data
  tempframe=subset(direct1frame,select=sub_vars2) #subsetting our data
  colnames(tempframe)[32:34]=c("JID","votedirect","majority") #renaming the columns
   # to make column names consistent across all judges
  tempframe$wroteopin=c(0) #none of these judges wrote the opinion that j1 did
  jcbin=rbind(jcbin,tempframe) #bind with our bin
}
```

```
#there are 15 judges, we have judge 1 thru 3 done though, so start at 4.
for (i in 4:15){
  jvar1=paste("codej",i,sep="")
  jvar2=paste("j",i,"vote1",sep="")
  jvar3=paste("j",i,"maj1",sep="")
  j_vars=c(jvar1,jvar2,jvar3)
  sub_vars2=c(sub_vars,j_vars)
  subframe=subset(direct1frame,select=sub_vars2)
  colnames(subframe)[32:34]=c("JID","votedirect","majority")
  tempframe=subset(subframe,subframe$JID>0) #drop the values for all the cases
   #in which there was not an ith judge
  tempframe$wroteopin=c(0)
  jcbin=rbind(jcbin,tempframe)
}
jcbin=jcbin[-1,] #remove the NA row from our original bin

#Write the judges all into one frame
jcbin=rbind(j1frame,jcbin)
```

We've now reshaped all of the first dimension cases and can turn to the second dimension cases. This yields a further 3,133 judge-case observations.

```
direct2frame=rbind(direct.0,direct.2)

#Identify the Conservative Votes (74 obs)
direct.1=subset(direct2frame,direct2frame$direct2==1)

#Identify the Liberal Votes on (66)
direct.3=subset(direct2frame,direct2frame$direct2==3)

#Bind the 2 types back together. We now have the 15060 cases on which we
```

```
# can make judge-case observations
direct2frame=rbind(direct.1,direct.3)
```

Having defined the directionality of each second dimension case, we just repeat the looping steps we did on the first dimension cases earlier.

```
sub2_vars=c("casenum","year","month","day",
                      "circuit","state","district",
                      "method","initiate","numappel",
                      "appnatpr","appbus","appnonp",
                      "appfed","appsubst","appstate",
                      "ap_stid","numresp","r_natpr",
                      "r_bus","r_nonp","r_fed","r_subst",
                      "r_state","r_stid","casetyp1","geniss",
                      "direct2","majvotes","dissent","concur")
j1frame2=subset(direct2frame,select=c(sub2_vars,"codej1"))
colnames(j1frame)[32]="JID"
j1frame2=rename(j1frame2,c(codej1="JID"))

#The majority writing judge...
j1frame2$votedirect=j1frame2$direct2
j1frame2$majority=c(1)
j1frame2$wroteopin=c(1)

#make our bin
jcbin2=matrix(rep(NA),1,35)
jcbin2=as.data.frame(jcbin2)
names(jcbin2)=names(j1frame2)

#judges 2 and 3 there are 15 judges, we have judge 1 done though, so start at 2
for (i in 2:3){
```

```
  jvar1=paste("codej",i,sep="")
  jvar2=paste("j",i,"vote1",sep="")
  jvar3=paste("j",i,"maj1",sep="")
  j_vars=c(jvar1,jvar2,jvar3)
  sub_vars2=c(sub2_vars,j_vars)
  tempframe=subset(direct2frame,select=sub_vars2)
  colnames(tempframe)[32:34]=c("JID","votedirect","majority")
  tempframe$wroteopin=c(0)
  jcbin2=rbind(jcbin2,tempframe)
}


#and judges 4-5, because there are no cases on the 2nd dimension in which more than 5
# judges took part in the case. we have judge 1 through 3 done though, so start at 4
for (i in 4:5){
  jvar1=paste("codej",i,sep="")
  jvar2=paste("j",i,"vote1",sep="")
  jvar3=paste("j",i,"maj1",sep="")
  j_vars=c(jvar1,jvar2,jvar3)
  sub_vars2=c(sub2_vars,j_vars)
  subframe=subset(direct2frame,select=sub_vars2)
  colnames(subframe)[32:34]=c("JID","votedirect","majority")
  tempframe=subset(subframe,subframe$JID>0)
  tempframe$wroteopin=c(0)
  jcbin2=rbind(jcbin2,tempframe)
}
#remove the NA row
jcbin2=jcbin2[-1,]
#Write the judges all into one frame.
```

```
jcbin2=rbind(j1frame2,jcbin2)
```

With all the looping done for the second dimension cases, it's time to merge all the judge-case observations together. The first and second dimension cases have different column names for the case direction so we need to rename both in order to merge the data properly. Next, bind the two dataframes together. Finally, write the final dataframe as a .csv file to your working directory. This file is called "songer_judgecases.csv". The final file has 46,448 judge-case observations. Please note that the judge-case observations are uncleaned and contain several errors as a result of mis-attributed judges on cases in the original CoAD.

```
jcbin=rename(jcbin,c(direct1="issdirect"))
jcbin2=rename(jcbin2,c(direct2="issdirect"))


jcfinal=rbind(jcbin,jcbin2)


write.csv(jcfinal,"songer_judgecases.csv")
```

Data Issues in the CoAD

More on this soon...

References

Collins, Paul M. 2008. "Transforming the United States Courts of Appeals Databases in Stata." *Law and Courts* 18(1):19–22.

Hurwitz, Mark S and Ashlyn Kuersten. 2012. "Changes in the circuits: Exploring the Courts of Appeals databases and the federal appellate courts." *Judicature* 96:23–34.

Songer, Donald R. 1998. "The Multi-User Database on the United States Courts of Appeals, 1925-1996.".
URL: *http://artsandsciences.sc.edu/poli/juri/appct.htm*