Research Spotlight

Transforming the United States Courts of Appeals Databases in Stata[*]

Paul M. Collins, Jr.
University of North Texas
pmcollins@unt.edu

In the Winter 2006 issue of *Law and Courts,* I presented Stata code allowing researchers to reliably transform the unit of analysis in Spaeth's (2005) Original U.S. Supreme Court Judicial Database from the case to the justice-vote (Collins 2006). The purpose of making this transformation is to enable scholars to tackle questions at the level of the individual justice (Giles and Zorn 2000). Since the publication of that code, I have developed comparable code allowing researchers to transform the unit of analysis in Songer's (2007) United States Courts of Appeals Data Base from the case to the judge-vote.[1]

Like the Original U.S. Supreme Court Database, the United States Courts of Appeals Data Base is structured such that the unit of analysis is the case, which makes some research questions more easily investigated than others. By transforming the unit of analysis from the case to the judge-vote, students of law and courts are enabled to address a multitude of new inquiries at the level of the individual judge. For example, using the judge-vote as the unit of analysis allows researchers to examine acclimation effects, factors related to the decision to dissent or concur, the influence of Chief Judges on their fellow panel members, in addition to a host of other topics.

Graphically, the code transforms the Courts of Appeals Data Base from this format:

case	j1vote1	j2vote1	j3vote1	j1maj1	j2maj1	j3maj1
1	1	3	1	1	2	1
2	1	3	1	1	2	1

To this format:

case	judgeid	jvote1	jmaj1
1	1	1	1
1	2	3	2
1	3	1	1
2	1	1	1
2	2	3	2
2	3	1	1

The transformation of the Courts of Appeals Data Base is accomplished in the following steps[2]:

```
. set mem 75m
. use "cta96_stata.dta"
```

The database that is about to be created will be substantially larger in size than the Courts of Appeals Data Base in its original form. As such, users must allocate sufficient memory. Employing a number of alternatives, I found that at least 75m must be allocated (though allocating more memory will speed up the process). Memory must be allocated prior to opening the Data Base.

```
. gen j1vote1=direct1
. gen j1vote2=direct2
. gen j1maj1=1
. gen j1maj2=1 if direct2~=.
```

This series of commands creates four new variables that contain information on judge 1's voting behavior. In its original form, the Data Base does not contain these variables as judge 1 is by definition a part of the majority and, therefore, judge 1's behavior corresponds exactly to the ideological direction of the majority's decision. Because Stata will transform the data based on characters contained in the judge-specific variables, new variables for judge 1 must be created. **j1vote1** and **j2vote2** contain information on the ideological direction of judge 1's votes. **j1maj1** and **j1maj2** indicate that judge 1 was a member of the majority. Note that **j1maj2** will only contain values if the ideological direction of the second case type is specified in the data.

```
. gen caseid = _n
```

This creates a unique identification number for each case in the database.[3]

```
. reshape long codej@ j@vote1 j@vote2 j@maj1 j@maj2, i (caseid) j (judgeid)
```

This command reshapes the data from wide format (unit of analysis = case) to long format (unit of analysis = judge-vote).[4] Using the **@** character, Stata locates the judge-numbers (1-16) and recognizes that these variables vary within cases. Note that there is no need to tell Stata which variables are constant within each case; Stata will recognize these variables on its own. **i (caseid)** tells Stata that each individual case is uniquely identified by the **caseid** variable. **j (judgeid)** tells Stata to create a new variable, **judgeid**, that corresponds to each judge's number (1-16).

```
. drop if codej==. & jvote1==. & jvote2==. & jmaj1==. & jmaj2==.
```

The final step is to clean-up the data. The previous code transforms the data such that there are 16 judge-number observations for each case, increasing the number of observations in the data from 18,195 to 291,120. This code purges the data of these missing observations, leaving the researcher with a final dataset of 55,575 observations.

The availability of publicly accessible databases has marked a dramatic change with regard to the ability of scholars to conduct empirically-oriented research on law and courts. Yet, the structure of many of these databases makes certain lines of inquiry more easily broached than others. The purpose of this note is to provide an easily implemented and reliable code for transforming the unit of analysis in the original and updated United States Courts of Appeals Data Bases. While the code is targeted toward these specific datasets, its basic logic can be extended to a variety of similarly organized databases. As such, I encourage researchers to share extensions to this code and welcome other innovations in data transformation, which can save us all some frustration from time to time (e.g., Benesh and Zorn 1998:12).

Appendix

The following code can be implemented to transform the updated Courts of Appeals Data Base (Kuersten and Haire 2007). Since the intuition behind the code is identical to that discussed above, I will forgo a discussion of each step in the process.

```
. set mem 75m
. use "KH_update_stata.dta"

.gen j1_vote1=direct_1
.gen j1_vote2=direct_2
.gen j1_maj1=1
.gen j1_maj2=1 if direct_2~=.

.rename j2vote2 j2_vote2
.rename j2maj1 j2_maj1
.rename j3_maj_1 j3_maj1
.rename j4_vote_1 j4_vote1
.rename j4_vote_2 j4_vote2
.rename j4maj1 j4_maj1
.rename j5_vote_1 j5_vote1
.rename j5_vote_2 j5_vote2
.rename j5_maj_2 j5_maj2
.rename j6_vote_1 j6_vote1
.rename j7_vote_1 j7_vote1
.rename j7_vote_2 j7_vote2
.rename j8_vote_1 j8_vote1
.rename j8_vote_2 j8_vote2
.rename j9_vote_1 j9_vote1
.rename j10_vote_1 j10_vote1
.rename j10_vote_2 j10_vote2
.rename j13_vote_1 j13_vote1
.rename j13_vote_2 j13_vote2

.gen id = _n

.reshape long j@ j@_vote1 j@_vote2 j@_maj1 j@_maj2 , i (id) j (judgeid)

.drop if j==. & j_vote1==. & j_vote2==. & j_maj1==. & j_maj2==.
```

(Notes)

1. The United States Courts of Appeals Data Base (1925-1996) is available on-line at The S. Sidney Ulmer Project for Research in Law and Judicial Politics, which is graciously maintained by Kirk Randazzo. Note also that the Ulmer Project contains an updated version of the Courts of Appeals Data Base (1996-2002), based on the efforts of Ashlyn Kuersten and Susan Haire. The code contained here is designed for use with the "original" Courts of Appeals Data Base. The Appendix reports the code for transforming the updated Data Base.

2. The code appearing below performs equally well in Stata 9.0 and Stata 10.0.

3. Prior to this step, users may wish to weight the cases. However, weights can be assigned after the transformation is complete.

4. After running this line of code, Stata will return a message indicating that codej16, j16vote2, and j16maj2 were not found. This is not an error message, but rather reflects the fact that entries for those variables, corresponding to judge 16, are missing from the Data Base.

(References)

Benesh, Sara C., and Christopher J. W. Zorn. 1998. "Data Transformation for the Analysisof Judicial Voting: Part I, The Supreme Court Judicial Database." *Law and Courts* 8(3): 9-12.

Collins, Paul M., Jr. 2006. "Transforming the Original U.S. Supreme Court Judicial Database: An Alternative Approach for Use with Stata." *Law and Courts* 16(1): 22-24.

Giles, Micheal W., and Christopher Zorn. 2000. "Gibson versus Case-Based Approaches: Concurring in Part, Dissenting in Part." *Law and Courts* 10(2): 10-16.

Kuersten, Ashlyn K., and Susan B. Haire. 2007. *Update to the Appeals Court Data Base (1997-2002).* Department of Political Science, Western Michigan University. Available at: http://www.as.uky.edu/polisci/ulmerproject/appctdata.htm.

Songer, Donald R. 2007. *The United States Courts of Appeals Data Base.* Columbia, SC: Department of Political Science, University of South Carolina. Available at: http://www.as.uky.edu/polisci/ulmerproject/ appctdata.htm.

Spaeth, Harold J. 2005. *The Original United States Supreme Court Database, 1953-2004 Terms.* East Lansing, MI: Department of Political Science, Michigan State University. Available at: http://www.as.uky.edu/polisci/ ulmerproject/sctdata.htm.

Stata release 9.0. 2005. College Station, TX: Stata Corporation.

Stata release 10.0. 2005. College Station, TX: Stata Corporation.