

TRANSFORMING THE ORIGINAL U.S. SUPREME COURT JUDICIAL DATABASE: AN ALTERNATIVE
 APPROACH FOR USE WITH STATA

Paul M. Collins, Jr.*

University of Houston

pmcollins@uh.edu

Virtually all quantitative research on the U.S. Supreme Court utilizes one or more of the Spaeth Supreme Court judicial databases. To promote their intelligent use, Sara Benesh (2005) provides a superb treatment of the management of these datasets. Explicit in her discussion is the need for common standards in the selection of, for example, units of analyses to promote the comparability and reliability of Supreme Court research. The purpose of this article is to contribute to this discourse by providing an easily implemented code that allows users to reliably transform the unit of analysis in ALLCOURT, the Original U.S. Supreme Court Judicial Database, which runs from 1953 to the end of the Court's last completed term.¹

The structure of ALLCOURT is set up such that the case is the unit of analysis, which makes certain types of research questions more easily broached than others. Seven years ago, Sara Benesh and Chris Zorn (1998) performed a great service to the law and courts community by publishing code that allows users of the database to alter its unit of analysis from the case to the justice-vote. Currently, three flipped datasets, based on the Benesh and Zorn code, are available on the Ulmer Project's website, containing information on the Warren (1953-1969), Burger (1969-1985), and Rehnquist (1986-1998) Courts, respectively.² While the availability of these datasets allows researchers to tackle new questions, analyzing judicial behavior based on these Justice-centered databases is limited in two regards. First, because the justice-centered databases are separate files, researchers are required to merge these datasets in order to correspond to roughly the time frame of ALLCOURT. Second, the Justice-Centered Rehnquist database is only current to the 1998 term and, as such, does not include information from the 1999-2004 terms. To alleviate these problems, I offer new code for Stata³ that allows researchers to transform the unit of analysis in ALLCOURT from the case to the justice-vote. In addition to allowing researchers to transform all of ALLCOURT, this new code, based on updates made to Stata since 1998, substantially expedites and simplifies the data transformation process.

Graphically, the code transforms ALLCOURT from this format:

case	rehndir	stevdir	ocondir	scaldir	kendir	soutdir	thomdir	gindir	brydir
1	0	0	0	0	1	1	1	1	1
2	1	1	1	1	0	0	0	0	0

To this format:

case	justice	dir
1	rehn	0
1	stev	0
1	ocon	0
1	scal	0
1	ken	1
1	sout	1
1	thom	1
1	gin	1
1	bry	1
2	rehn	1
2	stev	1
2	ocon	1
2	scal	1
2	ken	0
2	sout	0
2	thom	0
2	gin	0
2	bry	0

The code presented below makes several improvements over the code published by Benesh and Zorn. First, in order to flip the entire ALLCOURT database, the new code requires researchers to create only a single new variable, compared to 204 in the Benesh and Zorn procedure (although the new code does require that 33 variables be renamed). Second, the new code allows researchers to flip the entire database in a single process. Benesh and Zorn's method is limited in the fact that only 10 variables can be flipped at a time, thus requiring that the data be flipped five times (as there are 44 variables in ALLCOURT that are not justice-specific) and merged back together to transform the entire database. Third, the new method does not require researchers to transform any variables in the database from string (i.e., alphabetic) to numeric. Finally, the new code can be implemented in substantially fewer commands, thus reducing the likelihood of user error. Transformation of ALLCOURT based on the new code can be completed in the following steps:

```
. set mem 65m
. use "allcourt_stata.dta", clear
```

The database that is about to be created will be substantially larger in size than ALLCOURT in its original form. Accordingly, users must allocate adequate memory. Using a variety of alternatives, I have discovered that at least 65m should be allocated (though allocating more memory will speed up the process). Memory must be allocated prior to opening ALLCOURT.

```
. rename har harb
. rename blc blcb
. rename doug doughb
...
. rename bry bryb
```

This series of commands renames the variables that describe the individual justice's voting behavior. It is necessary to rename these variables as the new code seeks out j suffixes in order to make the transformation (in the above example the j suffix is "b" for behavior).

```
. rename dir courtdirection
. rename lctdir lctdirection
. rename mult_mem multcase
. rename term termofcase
```

This series of commands renames variables that correspond to j suffixes. In particular, **dir** and **lctdir** are renamed as the code will identify the ideological direction of the individual justice's voting behavior based on the "dir" suffix attached to each justice's abbreviation. Similarly, **mult_mem** and **term** are renamed as the code will identify each justice's majority or minority voting based on the "m" suffix attached to each justice's abbreviation.

```
. gen caseid = _n
```

This creates a unique identification number for each case in the database.⁴

```
. reshape long @b @a1 @a2 @v @o @dir @m, i (caseid) j (justid) string
```

This is the command that reshapes the data from wide format (unit of analysis = case) to long format (unit of analysis = justice-vote). Using the @ character, Stata locates j suffixes and recognizes that those variables vary within cases. Note that there is no need to tell Stata which variables are constant within each case; Stata recognizes this and keeps those variables without a separate command. **i (caseid)** tells Stata that each individual case is uniquely identified by the **caseid** variable. **j (justid)** tells Stata to create a new variable, **justid**, that corresponds to each justice's abbreviation and **string** tells Stata that this variable is a string (alphabetic) variable.

```
. drop if b==" & o==. & a1==" & a2=="
```

The final step is to perform some data clean-up. The previous code transforms the data such that there are 29 justice-observations for each case, thus increasing the number of observations in the data (in its raw form) from 12,577 to 364,733.

This code purges the data of all missing observations resulting in a final data set of 111,971 observations.

In 1998, Sara Benesh and Chris Zorn graciously provided Stata code allowing researchers to transform the Spaeth data from using the case as the unit of analysis to using the justice-vote as the unit of analysis. Recent advances in technology now make this process even easier. Despite the fact that this code can be relatively easily implemented, I join in Benesh and Zorn's sentiment in encouraging others to develop and share new and creative ways to transform data, which still "offers the prospect of saving us all a few headaches now and then! (1998, 12)"

*I am grateful to Sara Benesh, Wendy Martinek, and Mitch Pickerill for helpful comments and encouragement and to Harold Spaeth for his continued work on the Supreme Court judicial databases.

Notes

¹ The Spaeth judicial databases, along with several other datasets, are available online at The S. Sidney Ulmer Project for Research in Law and Judicial Politics, which is generously maintained by Kirk Randazzo.

(<http://www.as.uky.edu/polisci/ulmerproject/databases.htm>; accessed October 21, 2005).

² In addition to containing information from ALLCOURT, the Justice-Centered Warren and Burger Court databases also include conference vote data.

³ This code works with both Stata 8.0 and Stata 9.0

⁴ Prior to this step, users may wish to select a desired unit of analysis (e.g., case citation, docket number, etc.). However, the desired unit of analysis can be selected after the transformation is complete.

REFERENCES

- Benesh, Sara C. 2005. "Becoming an Intelligent User of the Spaeth Supreme Court Databases." *Law and Courts* 15(4): 15-21.
- Benesh, Sara C. and Christopher J.W. Zorn. 1998. "Data Transformation for the Analysis of Judicial Voting: Part I, The Supreme Court Judicial Database." *Law and Courts* 8(3): 9-12.
- 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 8.0. 2003-2005. College Station, TX: Stata Corporation.

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