

## Notes on Card/Della Vigna “Top 5” Data and Replication Files

1) the main data set is Top5JELPosting.dta, in stata format

2) the do file Top5JELPosting.do performs transformations and checks and creates a “working” data set “Top5JELPostingClean.dta” which has all the variables used in the paper

### *To reproduce the tables/figures of the paper:*

A. The do file “top5\_part1.do” uses the same steps as “Top5JELPosting.do” but creates less output and runs fewer checks. It also creates listings that are the data for the following figures/tables:

- a) the number of articles per journal in each year (for Figure 1 and Appendix Table 2)
- b) mean number of authors per year and mean/median normalized page lengths (for Figure 4)

It then creates a “old” format data set for transfer to sas: top5clean.dta → top5clean.sas7bdat

B. Using the sas data set:

a) **jel1.sas** prints out the mean and median numbers of citations in Figure 5 and Appendix Table 4. It also creates a data set of “relative citations” for the publications in each year, which is output to a sas data set called relcites.sas7bat. This is the “relcites” sheet of the “jelTablesFiguresPosting.xls” file, which is the basis for Figure 6.

b) **cdf2.sas** creates the cumulative distribution functions of citations by journal. These data are output to a sas data set called cdf2.sas7bdat. This is the cdf2 sheet of the “jelTablesFiguresPosting.xls” file, which is the basis for Figure 7.

c) **fields.sas** gets the mean fractions of publications in each year in each field. These are used to create Figure 8.

d) **table1.sas** runs the regressions in Table 1.

NOTE: the numbers of submissions to each journal were obtained directly from the journals themselves (and other sources) and are reported in Appendix Table 2 and Figure 1. These data are combined with the data on the number of articles published per year to create Appendix Table 3.