Data and programs for A Macroscope of English Print Culture, 1530-1700, Applied to the Coevolution of Ideas on Religion, Science, and Institutions

The results reported in this paper were generated using multiple software packages, including Python (for pre-processing), EViews, Stata, and R. In particular, EViews results were obtained through an interactive approach (i.e., using default menus and selecting relevant options) rather than scripted execution. This approach is less amenable to rigorous documentation but is straightforward to replicate in practice. For any questions regarding the steps involved, please contact the authors.

The raw time series data are in the file timelinesWeightedWordsPopLog.csv.

These data were imported into Eviews 12 and saved as workfile timelinesweightedwordspoplogrr.wf1.

The core of the time-series analysis in the paper uses the following variables: religion science institutions, all found under the main tab in the workfile.

The results and estimates reported in the paper were then generated using standard Eviews 12 routines based on specifications as described in the paper. The key VAR, for example, may be obtained using the following script:

var myvar

myvar.ls 1 3 religion science institutions @ c t

After generating the VAR object (myvar), the impulse-responses in Figure 4 are generated using View->Impulse Response… and specifying the options as described in the paper.

The annual structural shocks presented in Figure 5 are obtained analogously, using View->Structural Residuals. The corresponding moving averages are stored under the tab rsisr. The pertinent graphical object is figure5.

The historical decompositions shown in Figures 6 and 7 are obtained in the same way, using the object myvar and applying View->Historical Decomposition.

The elements of Figure 8 are calculated as described in the paper. The pertinent series, religininstshocksperc and religinscishocksperc, are found under the main tab of the workfile.

All of the series shown in Figures 2 and 3 are in the workfile under the main tab.

Figure 1 was generated using Stata, doFileHistogramEEBOTCP.do, drawing on inputCSVtcpWithoutDocuments.dta.

Table 1 was produced using the file Mtheta.csv, containing the document-topic prevalence matrix implied by our estimated STM. In Mtheta.csv, the numbers in the column names (e.g., 9 in V9) correspond to topic numbers as listed in Appendix B. Themes can be constructed by summing up, for each document (row), the values in pertinent columns, as defined in Table 1 or Appendix C. The numbers reported in Table 1 then refer to means of the correspondingly defined variables/columns (topics and themes).

The series underlying the timelines in Appendices C and D are in timelinesWeightedWordsPopLog.csv. The reported figures were created using R (script not provided, but easily replicated using any plotting software).

The estimates in Appendix G, Figure G2, were generated using Eviews 12, with the add-in localirfs, using the specification as described in that appendix. In the workfile, the relevant graphical object containing the estimates is lpirf.