Policy Agendas Project

Codebook for the *Encyclopedia of Associations* 1970-2005 Imputed Annual Values for Associations by PAP Major Topic

Frank R. Baumgartner, UNC-Chapel Hill John D. McCarthy, Penn State University Shaun Bevan, University of Mannheim January 11, 2012

This dataset consists of estimated numbers of associations per year for each of the major topics of the Policy Agendas Project. The main EA dataset has entries for each association listed in five year intervals from 1970 through 2005. This dataset provides an estimated number of associations per year. Estimations were done as follows. Note that we have actual observations for every 5th year, so the imputed values are only for years not ending in a 0 or a 5. In addition, we know the actual number of groups in each annual volume (though not the breakdown by topic area). We use both bits of information in the following procedure. First, we use a linear imputation for each group type, based on the five-year interval data. So if we have 100 groups in yeart and 120 groups in yeart+5 then we impute that $1/4^{\text{th}}$ of the growth came in each year: 104, 108, 112, 116 are the initial imputed values. But we also know that overall growth in the EA was not necessarily equal in each year; in fact we have precise numbers for the overall growth in the EA in each successive volume. So our second step is to calculate the percentage of the annual total which is represented by each topic, for the years where we have full data. This means for example that a particular topic code might be 7.2 percent of the total number of groups in year, and then 7.3 percent in year, Finally, we multiply these percentages of the total number of groups, for each year (imputed as well as observed years), times the actual number of groups in the entire EA for that year. The total number of groups in the EA is easily calculated from annual volumes because each association is given a sequential id number.¹ This process, combining information about the total size of the EA in each year as well as shifting proportions of groups by topic, is slightly more accurate than the first step alone, which is a linear imputation within topic area.

Variable definitions are below.

edition	Edition number of the printed copyright year
year	Copyright year of the printed edition.
observed_year	Dummy variable coded 1 if the year is included in the raw data and 0 if it was not and is therefore imputed.
topic	Major topic code according to the New York Times dataset version of the Policy Agendas Project topic codebook.
associations	Rounded imputation of the number of associations in the accompanying year/edition and topic.

¹ We noted errors in the allocation of ID numbers in fewer than 10 cases out of typically 20,000 or so in each annual volume.

total_associations Total number of groups across all topics in the associated year/edition.