5/30/18

**Guide to Merged and .Fin Data Files**

**Data files.** Two different types of data files are provided as a part of the CNEP and appear on the CNEP website (u.osu.edu/cnep).

**.Fin files** are made available on the website for each country survey after it has been standardized to CNEP format. The older surveys that are available for public use appear in one section of the website. More recent surveys appear in a different section in which access is restricted to CNEP principals under the terms of the Data Sharing Agreement and in consideration of the country team’s need for their exclusive use. As individual country surveys are made available for public use, password restrictions on access to them will be removed. The .Fin files contain three different types of variables:

1. **CNEP core variables.** They are denoted by prefixes A.A. through N and typically are the same variables that are carried over to the Merged files described below.
2. **.Z variables**. Variables denoted by .Z suffixes may be similar to CNEP core variables or of interest for CNEP research even though they are not found in many other country surveys.
3. **Original survey variables**. All of the original variables from the country surveys as they appeared in the country surveys.

The **Merged file** combines into one large data file the common core variables from all of the CNEP surveys (except Germany 1990 and Japan 1992, as explained below). These variables are denoted by prefixes A.A through N. Access to the most recent version of the merged file on the CNEP website is password restricted to CNEP principals. (The CNEP co-coordinators should be contacted for the password if you do not already have it.) An earlier version of the merged file, Merge26FINAL, based on country surveys through 2008 is available for public use without password protection.

**Weights.** Weights must be used with the data.The Merged and .Fin files contain 3 different weight variables:

1. **L.WtAcross** is the country N of cases divided by 1200. It is designed to equally weight all CNEP surveys so that the surveys with larger N’s do not dominate in the merged files.
2. **L.WtWithin** is the weight assigned by the country survey team to the data for that survey. It equals 1 in the countries where the surveys were not weighted (i.e., were self weighted) and varies where the surveys were weighted typically to bring the respondent demographics in line with some standard, usually the most recent census. This weight should be used for single-country analysis.
3. **L.Combined** is the product of L.WtAcross and L.WtWithin. It is the weight that should be used in cross-national analyses of the merged files. The N of cases using this weight is 1200 X the number of surveys in the merged file at the time.

**A.A. to N. prefixes.** Each variable in the merged file and all common core variables in the .Fin files have been standardized to CNEP format. Each of them has a prefix, ranging from A.A. through N., designed to group the variables for easy reference, into the different types of variables contained in the CNEP surveys and the resulting merged file of Common Core variables. A link to the most recent version of the Common Core Questionnaire is provided on the CNEP website.

**.F Variables**. Some variable names in the merged file have a suffix of .F. These are variables that were in the Common Core Questionnaire at the time of the survey but were dropped from the Core in subsequent surveys, sometimes because the question itself was deleted and sometimes because its wording or response options were revised to make it a slightly different (and, hopefully, improved) question. Where the question was revised, two variable names appear for adjacent variables, one with a .F suffix to denote the earlier version and the other without any suffix to denote the more recent version or a “least common denominator” coding that preserves equivalence. Occasionally, a deleted variable is later returned to the core (e.g., J.Compete), resulting in a break between surveys using the .F version and those that came later. Both .F and non-.F versions of similar variables have been retained to maximize the availability of data on individual questions. Analysts will want to combine both variables if they are satisfied that the versions are not substantively different, though sometimes they will need to recode the response options to make them as equivalent as possible

**Variable name endings of 1 vs. 2**. Usually, though not always, the second is a recoded version of the first variable in the pair. Sometimes it is a number referring to a particular Government or Opposition party. Consult the variable label for the appropriate information.

**GE90 and JP92 missing**. Two of the earliest CNEP surveys, for Germany in 1990 and Japan in 1992, are not contained in the merged file. Both of them contained an insufficient number of what later became core CNEP variables to be amenable to comparative analysis. Analysts interested in using them should consult the CNEP website (u.osu.edu/cnep) for their questionnaires and should contact the CNEP co-coordinators for those original data sets.

**Party meetings and contacts**. Even where the information about party contacts and attendance at party meetings is collected through a single question, it proved advantageous to code the responses party by party.

**Multiple response variables**. Some of the questions allowed for multiple responses (e.g., candidates supported by discussants in the first and second rounds of the French election). Even though it resulted in well over 300 discrete response codes, those codes are provided as the values for the particular variables.

**Party vs. candidate.** Where the election surveyed was parliamentary, respondents were asked about the party. Where the election surveyed was presidential, respondents were asked about the candidates, but the data in the merged file are coded by party -- as Gov1 for the largest government party at the time of the election, Opp1 for the largest opposition party at the time of the election, OppNew1 for the largest new party at the time of the election, etc.

**Party and Candidate Codes.** To facilitate comparison across country surveys, parties and candidates have been assigned a common code based on whether they were in the preceding government or opposition or were new in that election. The parties/candidates are numbered within each grouping based on the votes they have received: for example, Gov1, Gov2, Opp1, Opp2, OppNew1 … The codes for each survey are posted in a document on the cnep website.

**Question not asked in country.** This is denoted by a response value of -999 in surveys where the particular question was not asked. Sometimes it was omitted from the survey even though it was in the Common Core Questionnaire at the time. More often it reflected a question that was added in subsequent versions of the Core so did not appear in earlier surveys or had been dropped from the Core for subsequent surveys.

**Missing data**. Defined missing data in the merged file almost always is restricted to 998 (missing responses) or -999 (question not asked in this country) and occasionally 995 (for inappropriate). Analysts may want to eliminate other responses as missing data as is appropriate for their analysis.

**Missing variables**. The surveys covered in the merged files contain many more variables than the Common Core variables retained in the merged file. For all of the original variables in each survey, analysts should consult the CNEP website and the .Fin data files contained therein. These data contain the core variables (prefixes of A.A. through N.) plus variables that are similar to core variables or otherwise of interest to CNEP researchers (with a .Z suffix) plus all of the original variables in that particular survey data set as submitted to CNEP.

**Technical reports.** Technical information about each CNEP survey (including sponsor, survey firm, coverage, interviewing dates, etc.) can be found in a spreadsheet file posted on the CNEP website.