Concording PC8 products over time: Readme file

June 25, 2012

1 Introduction

This document provides practical guidelines on how to use the PC8 classification and concordance files for European 8-digit Prodcom (PC8) codes over time, as explained in Van Beveren, Bernard and Vandenbussche (2012). The concordance files can be used to concord PC8 products into a common product classification (PC8+) for all European Union countries and for any time period between 1993 and 2010. The concordance procedure relies on the algorithms developed by Pierce and Schott (2012) and Pierce and Schott (forthcoming). The original classification and correspondence files are available on the Eurostat Ramon server.[1]

If you use the concordance files, please cite:

Van Beveren, Ilke, Andrew B. Bernard, and Hylke Vandenbussche (2012). Concording EU Trade and Production Data Over Time. Tuck School of Business, mimeo.

---

[1]http://ec.europa.eu/eurostat/ramon/.

2 Concordance procedure

In order to translate the PC8 products into a common classification over time, the full list of existing PC8 codes in each year as well as the changes over time in the PC8 codes between each pair of years have been downloaded from the Eurostat Ramon server. These files are translated into usable stata files.[2] For the list of PC8 products in each year, this implies retaining only mandatory 8-digit PC8 codes (the original files additionally contain optional codes) and renaming and formatting the variables consistently for use in the concordance procedure. The original yearly concordance files (changes in PC8 classification between t-1 and t) provided by Eurostat typically include (some) optional (B- and N-list) and aggregated (Z-, T-, Q- or E-list) codes (cfr. Section 3.3 in Van Beveren et al. 2012).

The input files need to be adapted such that they only include changes that apply to mandatory PC8 codes, disregarding changes in optional and aggregated codes unless these changes have implications for the underlying mandatory codes. To this end, the input file **input_file_PC_over_time_edited.csv** contains two additional variables compared to the original input file: (i) *pcfrom_recode*, containing the mandatory or disaggregated counterpart of the obsolete optional or aggregated PC8 code; and (ii) *pcto_recode*, containing the mandatory or disaggregated counterpart of the new optional or aggregated PC8 code.[3]

In addition, whenever an optional or aggregate code is listed as an "exit" code (no new code is listed), while the underlying mandatory or disaggregated codes continue to exist, the mandatory code is entered as the "new" code (variable *pcto*).[4] Whenever a change has been made in the original input files from Eurostat, the source for the change is mentioned (Prodcom structure files in consecutive years have been used to verify the manual changes). Original Eurostat files, as well as the Prodcom structure files for different years are available in the folder **Originals Ramon**. Specifically, input files required to run the concordance procedure are:

- List of mandatory PC8 codes in each year (**PC_yyyy.dta**), where **yyyy** refers to the different years included in the concordance period.

---

[2] Files can be run in Stata 10 or higher.

[3] It can occur that an optional code later becomes mandatory, i.e. a more detailed breakdown that was optional in year t-1, becomes mandatory in year t. In these cases, the more aggregated mandatory code of t-1 will be retained throughout the sample period (essentially grouping the more detailed mandatory codes in t, to retain comparability with t-1). While these cases are rare, it is important to keep track of these changes when concording international trade and domestic production data over time. When translating PC8 and CN8 products into a consistent classification over time, these updated mandatory codes also need to be recoded in the concordance between CN8 and PC8 in the final year of the concordance.

[4] Intuitively, the code is not considered as an exit unless the mandatory code exits.

- Year-specific list of optional (B-list) PC8 codes with their mandatory counterparts (**PC_yyyy_Blist.dta**), which allows for identification and recoding of optional codes in the production data prior to concording the data. These files are only required when the time period starts prior to 2005.

- List of optional (N-list) PC8 codes (**Nlist_codes_1993_2005.dta**) with their mandatory counterparts, which allows for identification and recoding of optional codes in the production data prior to concording the data. These files are only required when the time period starts prior to 2005.

- **input_file_PC_over_time_edited.csv**: Edited list of changes in mandatory PC8 codes between pairs of years

- List of optional PC8 codes between 1993 and 2005 and their corresponding mandatory PC8 counterparts, to replace optional codes in the input file with changes over time (**optional_codes_1993 _2005.dta**)

- List of aggregate (T-, Z-, Q- or E-list) codes between 1993 and 2010, to be dropped from the input file[5] (**TZQ_codes_1993_2010.dta**)

The do-file **PC8_over_time.do** runs the concordance procedure. The beginning and end year are set at the beginning of the file, the rest of the program automatically adapts to the chosen time period. The concordance procedure will generate the concordance file **pc8_pc8plus_bbbb_eeee_.dta** in the output folder, where **bbbb** refers to the first year of the concordance and **eeee** refers to the final year.[6] The concordance file provides a year-specific list of all existing mandatory PC8 products and their PC8+ code, as well as a dummy indicating which PC8 products need to be dropped due to changes in coverage over time. Prior to implementing the concordance in the production data at year-PC8 product level, all optional codes (B- and N-list) that (potentially) feature in the data need to be recoded into their mandatory counterparts.

Specifically, the concordance process can be summarized in three steps. The first two steps refer to the concordance of the product classification, they coincide with concordance types (i) and (ii) of the generic concordance procedure discussed in Section 3.1 of Van Bev-

---

[5]If changes in aggregated codes affect their disaggregated counterparts, this is recorded in in the concordance files using the disaggregated codes. Hence aggregated codes can be dropped here.

[6]The output folder additionally contains comma-separated (.csv) files for several different time periods for non-Stata users.

eren, Bernard and Vandenbussche (2012).[7] The last step discusses actual implementation of the concordances in international production data expressed using the PC8 classification.

- *Step 1: Concording PC8 codes between t and t-1:* Changes in PC8 codes over time are first classified into different types of mappings. Mappings between two consecutive years can be simple (one obsolete PC8 code in t-1 translates into one new PC8 code in t), many-one, one-many and many-many (cfr. Table 4 in Van Beveren, Bernard and Vandenbussche, 2012). A unique identifier (setyr) is assigned to each mapping. For many-many and one-many mappings between two years, a feedback loop derived from Pierce and Schott (2012) is used to ensure that the correct grouping procedure is applied.

- *Step 2: Developing a consistent classification over time:* To create a consistent concordance over time, a number of additional steps need to be taken. First, the "news loop" developed by Pierce and Schott (2012) is used to chain subsequent code changes over time and to assign a unique identifier to these families. This procedure ensures that codes that have undergone changes in more than one year are chained together. These chains over time are then merged back into the year-to-year concordance files developed in step 1. A similar loop as the one used in Step 1 is then used to assign a unique setyr to families over time. Next, these families are merged into the full list of existing PC8 codes in each year to translate the PC8 codes into the PC8+ classification. Finally, all PC8 codes that are subject to entry or entry during the chosen time period are merged into the concordance, allowing for identification of all families that need to be dropped to maintain consistent coverage over time. All PC8 codes subject to changes over time, as well as their unique PC8+ code and a dummy indicating whether the product needs to be disregarded in the data due to changes in coverage over time, are recorded in the concordance file **pc8_pc8plus_bbbb_eeee.dta** (csv-format is also generated automatically). It should be noted that the concordance files are specific to the time period chosen, i.e. the PC8+ classification will be different depending on the sample period chosen in the beginning of the do-file. The longer the time period chosen, the more PC8 codes will be subject to changes over time, hence the number of grouped PC8 codes will increase as the sample period lengthens (the number of PC8 codes that need to be dropped for consistency over time will also increase). This also implies that the PC8 and PC8+ codes in the final concordance

---

[7]Type (i) refers to the development of a consistent concordance between two years, while type (ii) refers to the creation of a common classification system over time.

file are year-specific, i.e. the PC8+ codes need to be merged into the data at the year-PC8 level.

- *Step 3: Concording production data:* To concord European (firm-)product domestic production data recorded in the PC8 classification into the PC8+ classification, all optional codes that feature in the data[8] need to be recoded into their mandatory counterparts. The name of the variable referring to the PC8 codes should be *pc8*, it should be a string variable prior to recoding the optional codes and numeric afterwards (cfr. Stata code provided at end of the do-file). Data need to be sorted on the *year* and *pc8* variable. This data file can then be merged with the concordance file (**pc8_pc8plus_bbbb_eeee.dta**). In principle, all PC8 codes that feature in the data should also feature in the concordance file.[9] PC8+ codes are (can be) more aggregated than the PC8 codes, hence the international trade data need to be aggregated from the PC8 to the PC8+ product level in a final step.

3 Final concordance files

3.1 Nlist_codes_1993_2005.dta

- This file can be used to recode optional N-list codes in the data, to the extent that they actually feature in the data and if the sample period starts prior to 2005 (the Stata code provided takes this into account automatically), merge variable: *pc8*. If the sample period starts after 2004, this file is not required (and ignored) in the concordance procedure.

- Variables:

    - *pc8* : Optional Prodcom N-list (10-digit) code, recorded as string variable.
    - *pc_mand*: Mandatory PC8 code corresponding to N-list code (simply the 10d code minus the last two digits), also recorded as string variable.

3.2 optional_codes_bbbb_eeee.dta

- This file can be used to recode optional B-list codes in the data, to the extent that they actually feature in the data and if the sample period starts prior to 2005 (the

---

[8]The use of optional codes is country-specific, the stata code allows for the existence and absence of optional codes in the data.

[9]Possible exceptions are some residual categories (e.g. codes starting with or ending on 9999) and coding errors. These products should be dropped from the data to avoid spurious entry and exit dynamics.

Stata code provided takes this into account automatically), merge variable: *pc8*. If the sample period starts after 2004, this file is not generated by the do-file and is not required in the concordance procedure.

- Contrary to the N-list optional codes, optional B-list codes are specific to the sample period chosen, to take into account that in some (rare, cfr. supra) cases optional codes can become mandatory in later years, in which case the more aggregated code is applied in all years.

- Variables:

    - *pc8* : Optional Prodcom B-list (8-digit) code, recorded as string variable.
    - *pc_mand*: Mandatory PC8 code corresponding to B-list code, also recorded as string variable.

3.3   pc8_pc8plus_bbbb_eeee (dta or csv format)

- This file can be used to concord (yearly) production data from the PC8 classification (mandatory codes) to PC8+, merge variables: *pc8 year*.

- The concordance file is specific to the time period chosen. The time period can be set in the beginning of the do-file **PC8_over_time.do**.

- Variables:

    - *year*: Numeric variable, refers to the year.
    - *pc8* : Prodcom (8-digit) code (year-specific), recorded as numeric variable. PC8-year combinations are unique in the concordance file (each PC8 code features only once in each year). Only mandatory PC8 codes are included.
    - *synthetic* (numeric): dummy variable equal to one if the PC8+ classification groups more than one PC8 code, can be used to distinguish between original (ungrouped) PC8 products and sets of PC8 products grouped for consistency over time. In general, the longer the time period chosen, the higher the number of synthetic codes.
    - *pc8plus*: PC8+ code corresponding to the PC8 code in a specific year (numeric).
    - *exit*: dummy equal to one if the PC8 code needs to be dropped from the data to maintain consistency over time (i.e. dummy marking changes in coverage during the time period considered).

References

Pierce, Justin R. and Peter K. Schott, "Concording US Harmonized System Categories over Time," *Journal of Official Statistics*, 2012, *28* (1), 53–68.

_ **and** _ , "A concordance between ten-digit US Harmonized System codes and SIC/NAICS product classes and industries," *Journal of Economic and Social Measurment*, forthcoming.

Van Beveren, Ilke, Andrew B. Bernard, and Hylke Vandenbussche, "Concording EU Trade and Production Data over Time," *Tuck School of Business, mimeo*, 2012.