Concording trade and production data in a single year: Readme file

July 18, 2012

1 Introduction

This document provides practical guidelines on how to apply the concordance procedures explained in Van Beveren, Bernard and Vandenbussche (2012) to concord European domestic production data (PC8 products) and international trade data (CN8 products) into a common classification (PC8+).[1] The concordance files can be used to generate a concordance for the years 2003 and 2005. 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.[2]

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] The PC8+ classification will coincide with the PC8 classification in all cases where the mapping between the CN8 and PC8 classification is many-one or one-one. For many-many and one-many mappings between CN8 and PC8 products, PC8 products are grouped into PC8+ products.

[2] `http://ec.europa.eu/eurostat/ramon/`.

2 Concordance procedure

In order to translate the PC8 and CN8 classification into PC8+ codes for a single year, the full list of existing PC8 codes and existing CN8 codes in each year as well as the concordance between the PC8 and CN8 classification have to be downloaded from the Eurostat Ramon server. Since the list of PC8 and CN8 codes is year-specific, concordance files between the two classifications are also year-specific. These files are translated into usable stata files.[3] 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, at least prior to 2005) and renaming and formatting the variables consistently for use in the concordance procedure. Similarly, the original list of CN8 codes is adapted to include only 8-digit CN products.

When concording trade (CN8) and domestic production (PC8) data, there are differences in coverage between the two classifications that need to be taken into account. The concordance procedure identifies CN8 products not covered by the PC8 classification (e.g. Fuel) in a particular year by merging the list of CN8 codes with the list of CN8 codes present in the concordance file between CN8 and PC8. Similarly, PC8 products not covered by the CN8 classification (industrial services) are identified by merging the list of PC8 codes for the particular year chosen with the concordance file between CN8 and PC8 for that year.

Additionally, the concordance procedure takes into account that certain PC8 codes feature as a more aggregated product (T-, Q-, Z-, E-list) in the concordance files between PC8 and CN8 compared to the yearly PC8 classification files. These aggregated product codes, as well as their disaggregated counterparts, are identified in the concordance procedure. By using the yearly prodcom structure files, it is possible to (manually) match aggregated and disaggregated codes and to identify industrial services. Input files for 2003 and 2005 are provided in the corresponding yearly folders (**PC8_yyyy_special_codes.csv**). They can be used to recode disaggregated PC8 codes into their more aggregated counterparts in the domestic production data and to drop the industrial services from the data prior to concording.

The do-file **CN8_PC8_cross_section.do** runs the concordance procedure. At the end of the do-file, the necessary steps that need to be taken to concord the domestic production and trade data in a common classification (PC8+) are also implemented.

Specifically, the concordance process can be summarized in 4 steps. The first step refers to the concordance of product classifications (relying on type (i) concordance procedures,

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

between two classifications in a single year, cfr. Section 3.1 in the paper). The last three steps discuss actual implementation of the concordances in the international trade and production data.

- *Step 1: Concordance from CN8 and PC8 to PC8+:* All PC8 codes that are covered by the CN8 classification are concorded into PC8+ products. Mappings between the CN8 and PC8 classification can be simple (one CN8 code maps into a single PC8 code), many-one, one-many and many-many (cfr. Table 1). A unique identifier (setyr) is assigned to each mapping. For many-many and one-many mappings between CN8 and PC8, a feedback loop derived from Pierce and Schott (2012) is used to ensure that the correct grouping procedure is applied. By merging this concordance file with the list of all CN8 codes in the corresponding year, it is possible to identify all CN8 codes that are not covered by the PC8 classification in that particular year. The final concordance file (**concordance_cn8_pc8plus_yyyy**, in dta or csv format) contains a list of CN8 codes, their corresponding (mandatory) PC8 codes and the assigned PC8+ product code, as well as a dummy variable *notpc* identifying CN8 products not covered by the PC8 classification.

- *Step 2: Concording production data:* To concord European (firm-)product production data to PC8+ products, a number of steps need to be taken prior to merging the data with the actual concordance files. First, when concording classifications prior to 2005, optional codes featuring in the production data need to be recoded into their mandatory counterparts (using input files **PC_yyyy__Blist.dta** and **Nlist_codes_1993_2005.dta**). Second, all PC8 products not covered by the CN8 classification (mostly industrial services) need to be dropped from the production data. In addition, the concordance file between CN8 and PC8 aggregates some PC8 codes into "Z-codes", i.e. groups of PC8 codes that map into one or more CN8 codes. If the more disaggregated PC8 codes (i.e. the codes mapping into the Z-aggregates) feature in the production data, they need to be recoded into their corresponding Z-code before concording the data (since the concordance file only features the Z-aggregates, not the underlying PC8 codes). Industrial services and Z-codes (and corresponding disaggregated codes) are listed in the file **PC8_yyyy_special_codes**. Once services have been dropped and Z-codes entered, the domestic production data is ready to be merged (at PC8 level) with the concordance file. To this end, the concordance file **concordance_cn8_pc8plus_yyyy.dta** needs to be adapted such that each PC8

code features only once in the concordance file, with its corresponding PC8+ product code.[4] By construction, all PC8 codes that are present in the data (after recoding optional codes and Z-aggregates and dropping services) should feature in the concordance. Since PC8+ codes are in some cases more aggregated than the PC8 codes, the production data need to be aggregated from the PC8 to the PC8+ product level in a final step.

- *Step 3: Concording international trade data:* To concord the international trade data, the concordance file **concordance_cn8_pc8plus_yyyy.dta** needs to be adapted such that each CN8 code features only once in the concordance file, with its corresponding PC8+ product code. The trade data file for 2005 then needs to be merged (at CN8 level) with the unique CN8 codes in the concordance file to translate the CN8 products into PC8+ products. By construction, all CN8 codes that feature in the data should also feature in the concordance file. All CN8 products for which the dummy "notpc" equals one need to be dropped from the data prior to concording (these codes have no associated PC8+ product code). Since PC8+ codes are (can be) more aggregated than CN8 codes, the data need to be aggregated from the CN8 to the PC8+ product level in a final step.

- *Step 4: Merging domestic production and trade data:* Sort the international trade and domestic production data on the firm (if applicable) and product (pc8plus) identifier and merge the two data sets. The final data contain data on international trade and production, recorded using the PC8+ classification, allowing for product-level comparison between the two.

3 Final concordance files

3.1 Nlist_codes_1993_2005.dta

- This file can be used to recode optional N-list codes in the production 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:

[4] The do-file provides the stata code that can be used to implement the concordance in the data.

- *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 (in conjunction with the file **PC_yyyy_Blist.dta**, that identifies the year-specific B-list codes), 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 generated by the do-file and is not required in the concordance procedure.
- 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_yyyy_special_codes (dta or csv format)

- This file can be used to identify the PC8 codes that have to be dropped (industrial services and codes without correspondence in the CN8 classification) and to recode the underlying PC8 codes into their corresponding Z-aggregate in the *domestic production data, prior to concording the data.*
- Variables:
  - *pcyyyy* : prodcom code for the year 2003 or 2005 (yyyy), recorded as string (length 8) variable.
  - *type*: three different types: "industrial services", "aggregate" and "no cn correspondence". PC8 codes with type "industrial services" or "no cn correspondence" need to be dropped from the domestic production data prior to concording the data. PC8 codes with type "aggregate" need to be replaced their corresponding Z-aggregate (*new_code*) and aggregated when applicable.
  - *new_code*: Z-aggregate for codes that are recorded at higher level of aggregation in the concordance files between CN8 and PC8. Variable is recorded as string (length 8).

3.4 concordance_cn8_pc8plus_yyyy (dta or csv format)

- This file can be used to concord domestic production data (PC8) and international trade data (CN8) into the PC8+ classification. Merge variables: *pcyyyy* (PC8 code for the chosen year) for domestic production and *cnyyyy* (CN8 code for the chosen year) for international trade data.
- Variables:
  - *pcyyyy* : prodcom code for 2003 or 2005 (yyyy), recorded as string (length 8) variable. PC8 codes are not unique in the concordance file.
  - *cnyyyy* : CN8 code for 2003 or 2005 (yyyy), recorded as string (length 8) variable. CN8 codes are not unique in the concordance file.
  - *pc8plus*: PC8+ code corresponding to the CN8 and PC8 codes (string with length 8).
  - *synthetic*: dummy variable identifying groups of PC8 products (PC8+).
  - *notpc*: dummy variable identifying CN8 codes not covered by the PC8 classification.
- Prior to concording the production data, industrial services need to be identified and dropped in the production data and certain PC8 codes need to be recoded into their corresponding Z-codes. The file **PC8_yyyy_special_codes** identifies these codes. CN8 products not covered by the PC8 classification similarly need to be dropped from the trade data prior to concording.

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.

Table 1: Concordance CN8 to PC8+ for 2005

Type of match CN-PC	*Number of CN8 products*	*Number of PC8 products*	*Number of PC8+ products*
Simple (one-one) CN8-PC8	2588	2588	2588
Many-one CN8-PC8	6383	1582	1582
One-many CN8-PC8	3	6	15
Many-many CN8-PC8	183	44	
Total	*9157*	*4220*	*4185*

Overall, there are 10096 CN8 products in 2005. However, only 9157 CN8 products are covered by the Prodcom List in 2005. Similarly, there are 4489 (mandatory) PC8 products on the Prodcom List for 2005, 4242 of these are covered by the CN8 concordance. Some PC8 products are aggregated into a Z- or T-aggregate for the purpose of concording them to the CN8 classification, resulting in 4220 PC8 categories in the correspondence file between PC8 and CN8.