Concording HS6 products over time: Readme file

July 18, 2012

1 Introduction

This document provides practical guidelines on how to concord data, recorded using the European Combined Nomenclature (CN8) classification, into a common product classification (HS6+) for all European Union countries and for any time period between 1988 and 2010. The concordance procedure is based on the algorithms developed by Pierce and Schott (2012) and Pierce and Schott (forthcoming) and applied to European product classifications by Van Beveren, Bernard and Vandenbussche (2012). 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 CN8 products into a common classification over time, the full list of existing CN8 codes in each year as well as the changes over time in the CN8 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 CN8 products in each year, this implies retaining only the 8-digit CN8 codes (the original files contain all levels of aggregation in most years) and renaming and formatting the variables consistently for use in the concordance procedure. Original Eurostat files are available in the folder **Originals Ramon**.

Input files required to run the concordance procedure are:

- List of CN8 codes in each year (**CN_yyyy.dta**), where **yyyy** refers to the different years included in the concordance period.

- Changes in CN8 codes between pairs of years (**CN_concordances_1988_2010.csv**)

Since the first six digits of the CN8 classification coincide with the HS6 classification, the CN8 concordance files can be used to generate lists of HS6 products in each year as well as a list of changes in the HS6 classification over time. Table 1 provides an overview of the changes in the HS6 classification over time. While the official HS6 classification has only been revised in 1992, 1996, 2002 and 2007 during the period 1988-2010, some of the changes in the CN8 classification over time have implications for the HS6 classification in other years. However, at most 4 HS6 codes are subject to changes in years that are not official revision years, as can be seen in the table. The concordance procedure takes all changes in the HS6 classification over time into account, including changes that take place in years in which the HS6 classification was not subject to an official revision.

The do-file **HS6_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 **hs6_hs6plus_bbbb_eeee_.dta** in the output folder, where **bbbb** refers to the first year of the concordance and **eeee** refers to the final year.[3] This concordance file can be used to concord European HS6 (CN8) product codes into a consistent HS6+ classification over

---

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

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

time. At the end of the do-file, the stata code necessary to implement the concordance in the data is also given.

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 Beveren, Bernard and Vandenbussche (2012).[4] The last step discusses actual implementation of the concordances in the international trade data.

- *Step 1: Concording HS6 codes between t and t-1:* Changes in HS6 codes over time are first classified into different types of mappings. Mappings between two consecutive years can be simple (one obsolete HS6 code in t-1 translates into one new HS6 code in t), 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 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. Finally, these families are merged into the full list of existing HS6 codes in each year to translate the HS6 codes into the HS6+ classification. All HS6 codes subject to changes over time as well as their unique HS6+ code are recorded in the concordance file **hs6_hs6plus_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 HS6+ 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 HS6 codes will be subject to changes over time, hence the number of grouped HS6 codes will increase as the sample period lengthens. This also implies that the HS6 and HS6+ codes in the final concordance file are year-specific, i.e. the HS6+ codes need to be merged into the data at the year-HS6 level.

---

[4]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.

- *Step 3: Concording international trade data:* To concord European (firm-)product international trade data recorded in the CN8 classification into the HS6+ classification, the data need to be translated into HS6 products (first six digits of CN8 product = HS6 product). The name of the variable referring to the HS6 codes should be *hs6*, it should be numeric and the data should be sorted on the *year* and *hs6* variable. This data file can then be merged with the concordance file (**hs6_hs6plus_bbbb_eeee.dta**). In principle, all HS6 codes that feature in the data should also feature in the concordance file.[5] HS6+ codes are (can be) more aggregated than the HS6 codes, hence the international trade data need to be aggregated from the HS6+ product level in a final step.

3   Final concordance files

3.1   hs6_hs6plus_bbbb_eeee (dta or csv format)

- This file can be used to concord (yearly) international trade data from the HS6 classification to HS6+, merge variables: *hs6 year*.

- The concordance file is specific to the time period chosen. The time period can be set in the beginning of the do-file **HS6_over_time.do**.

- Variables:

  - *year*: Numeric variable, refers to the year.

  - *hs6* : Harmonized System (HS6) code (year-specific), recorded as numeric variable. HS6-year combinations are unique in the concordance file (each HS6 code features only once in each year).

  - *synthetic* (numeric): dummy variable equal to one if the HS6+ classification groups more than one HS6 code, can be used to distinguish between original (ungrouped) HS6 products and sets of HS6 products grouped for consistency over time. In general, the longer the time period chosen, the higher the number of synthetic codes.

  - *hs6plus*: HS6+ code corresponding to the HS6 code in a specific year (numeric).

---

[5]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.

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: Changes in HS6 classification over time

Effective year	Number of obsolete codes	Number of new codes	Number of families (including simple changes)	Number of simple (one-one) changes
1989	2	2	1	0
1990	3	2	2	1
1992	**30**	**32**	**11**	**2**
1993	2	4	2	0
1994	1	2	1	0
1995	1	2	1	0
1996	**418**	**509**	**213**	**74**
1999	2	2	1	0
2001	1	1	1	1
2002	**381**	**500**	**181**	**51**
2003	1	2	1	0
2004	3	2	2	1
2007	**747**	**586**	**253**	**75**
2008	1	1	1	1
2010	1	1	1	1

The table shows the number of obsolete and new codes in each year, as well as the number of families (shrinking, growing, simple, entry or exit) and the number of simple changes (one-one). The effective year refers to the year in which the change became effective. All changes in the HS6 classification over time are derived from changes in the CN8 classification over time. Rows highlighted in bold indicate years in which the HS6 classification has officially changed. Changes in the other years are implied changes derived from changes in the CN8 classification over time. Original files are obtained from the Eurostat Ramon server. The number of HS6 codes varies between 5018 and 5224 between 1988 and 2010.