wiiw

JUNE 2024

Working Paper 248 – Appendix

The Granular Trade and Production Activities (GRANTPA) Database

Sebastien Bradley, Javier Flórez Mendoza, Mario Larch and Yoto V. Yotov

The Vienna Institute for International Economic Studies
Wiener Institut für Internationale Wirtschaftsvergleiche

The Granular Trade and Production Activities (GRANTPA) Database - Technical Appendix

SEBASTIEN BRADLEY
MARIO LARCH
JAVIER FLÓREZ MENDOZA
YOTO V. YOTOV

Sebastien Bradley is Associate Professor at the School of Economics of the LeBow College of Business, Drexel University, Philadelphia. Javier Flórez Mendoza is Economist at The Vienna Institute for International Economic Studies (wiiw). Mario Larch is Professor at the Department of Law, Business & Economics, University of Bayreuth. Yoto V. Yotov is Professor at the School of Economics of the LeBow College of Business, Drexel University, Philadelphia.

We are very grateful to the Eurostat team for research support and clarifications concerning Prodcom codes, combined nomenclature classifications, and complementary concordance files provided by Erlend Firth and Laia Guinovart.

We are grateful to Andrew Bernard, Ilke Van Beveren, and Hylke Vandenbussche for providing an earlier version of this concordance and for sharing the corresponding codes. We also thank David Zenz for his guidance in retrieving the production of manufactured goods database.

This paper expresses the authors' views solely, and Eurostat or any institution is not responsible for any errors.

To access the GRANTPA database, please e-mail us at grantpadatabase@gmail.com.

Abstract

This paper introduces the Granular Trade and Production Activities (GRANTPA) database, which covers international trade flows for 3,124 products and 247 countries over the period 1995-2019 as well as domestic trade flows and production data for the same number of products and years for a subset of 35 European economies. The original data sources that we employ are Eurostat's Comext and Prodcom databases. A gravity application delivers a large set of product-level 'home bias' estimates, which cannot be obtained without domestic trade flows. The average estimates on the standard gravity variables in our model (e.g., distance) are comparable to those from the related literature. However, our disaggregated estimates are very heterogeneous across products, thus highlighting the importance of our new database.

Keywords: Gravity Data, Structural Gravity, Domestic Trade Flows, Disaggregated Gravity Estimates, Home Bias Estimates

JEL classification: C81, F13, F14

CONTENTS

TABLES AND FIGURES

1 Concording CN8 Products Over Time: Extension

This section provides technical instructions on the approach of concording 8-digit combined nomenclature (CN8) classification files over time. We follow closely the methodology from Van Beveren, Bernard and Vandenbussche (2012) (henceforth, VBBV). While their analysis focuses on the implications of changing product classifications using Belgian firms, we aim to expand the scope to all member states in the European Union. This approach also implements the algorithms provided by Pierce and Schott (2012*a*,*b*). We extend the VBBV concordance beyond 2010 using files extracted from the Eurostat Ramon server to concord the CN8 codes into a common classification (CN8+) from 1988 to 2019.[1]

The concordance path starts with the folder "**Concordance Trade Production**". Followed by "**cn8_over_time**", which includes the CN8 classifications from 1988 to 2022. Similarly, the subfolder "**Originals Ramon Eurostat**" contains the classifications retrieved from Eurostat named "*Combined Nomenclature, yyyy (CN yyyy)*", where "*yyyy*" denotes the year of the concordance period.[2] These raw files are formatted differently depending on the group of years.[3] Based on VBBV's methodology, the input files required to concord CN8 codes into a common classification over time are the following:

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

To address the issue of differences in the format present in the original files, we created a do-file that computes the combined nomenclatures at the 8-digit level (CN8).

◇ **1_CN_1988_2022.do**

All do-files are located inside the folder named "**Stata do files**." The first do-file "**1_CN_1988_2022.do**," loops over groups of years to obtain consistent CN8 codes using the original files extracted from Eurostat.

Subsequently, the changes of CN8 codes between each pair of years are required to enable concordances of combined nomenclatures over time. The main input file is the following:

- **CN_concordances_1988_2022.csv**.

This input file can be found inside the folder "**cn8_over_time**." Moreover, the changes in CN8 codes between each pair of years can be generated in Stata by executing this do-file:

◇ **2_CN_updated.do**

The folder called "**CN_update_since_1988**" contains two supplementary files related to an updated version of the changes of CN8 concordances between pairs of years from 1988 to 2022. These two complementary files are the following:

[1] To achieve consistency in the terminology, all the names will be the same as those mentioned in Van Beveren, Bernard and Vandenbussche (2012).

[2] Original files can be downloaded from `https://circabc.europa.eu`.

[3] The file extension of the first group from 1988 to 1994 is a text document "*.txt*". The second group from 1995 to 2006 is comma-separated value "*.csv*". Lastly, the third group from 2007 to 2022 is again "*.txt*".

➔ "*CN_2022_update_of_codes.xls*" [4]

➔ "*CN_2022_update_of_codes.pdf*"

The third necessary input file for the CN8 concordances over time is the Comext international trade data at the 8-digit level, named "**goods_pt_1988_2022.dta**." This file contains bilateral trade flows at the product level from 1988 to 2022. To obtain this input file, we executed the following do-file which enables consistency in terms of the format.

◇ **3_Trade_Data_1988_2022.do**

International trade data is extracted from Eurostat using the Bulk download facility. The data is reported as "*Comext*" data. It is worth mentioning that international trade data is spread into two periods. Original data for the period between 1988 and 2001 is located at *Comext historical data*, and it can be downloaded using the following path:

➔ *Comext files* > *COMEXT_HISTORICAL_DATA* > PRODUCTS_1988_2001 > *full198852.7z*

The second period ranges from 2002 to 2022 and the access path is

➔ *Comext files* > *COMEXT_DATA* > PRODUCTS > *full200252.7z* [5]

The do-file "**3_Trade_Data_1988_2022.do**" executes a series of loops in Stata, which allows us to obtain international trade data expressed in a format consistent with VBBV. The first loop generates the original international trade data extracted from Eurostat "**full_yyyy.dta**," in Stata format from 1988 to 2022. The data with the proper format are located inside the folder "**Trade Data**." The second loop modifies the trade data in a cross-sectional setting. We renamed the variable "*product_nc*" to "*cn8*" to keep consistency with VBBV. Additionally, four new variables are created and these become important later in the final CN8 concordances over time.[6] Lastly, the final loop combines the trade data for all years. The data are saved as "**goods_pt_1988_2022.dta**."

- "*valueII*": stands for the import values inside the European Union.
- "*valueIE*": refers to the import values outside of the EU.
- "*valueXI*": stands for the export values inside the European Union.
- "*valueXE*": refers to export values outside of the EU.[7]

[4] The file "*CN_2022_update_of_codes.xls*" is extracted from Eurostat Ramon and contains three columns (*period*, *Origin code* and *Destination code*). In contrast, the file "*CN_concordances_1988_2010.csv*" provided by Van Beveren, Bernard and Vandenbussche (2012) contains four columns (*from*, *to*, *obsolete_code*, and *new_code*). The second do-file delivers consistency for the updated CN8 concordances.

[5] The original data is retrieved on the date 26 December 2022, its format is "*.dat*" and it is saved into the sub-folder "**raw**". Thus, "**3_Trade_Data_1988_2022.do**" allows us to extract the data directly from the Bulk download and import it into Stata format. Alternatively, we provide a R file named "*EUROSTAT.R*" which transforms the raw files into a readable format "*.csv*".

[6] For further details on the construction of the new variables see "**3_Trade_Data_1988_2022.do**."

[7] In addition, the file "**3_Trade_Data_1988_2022.do**" creates similar variables for quantities and supplementary quantities (*quantityII, quantityIE, quantityXI, quantityXE, sup_quantityII, sup_quantityIE, sup_quantityXI, sup_quantityXE*), see Section 2.2 for further details.

The EU Statistics register the value and quantity of goods traded *(i)* between member states of the EU (*intra-EU trade*) and *(ii)* by the EU member states with non-EU countries (*extra-EU trade*). The classification used is the 8-digit combined nomenclature or CN8.[8] Extra-EU trade data is collected from customs data, where all transactions with values higher than €1,000 or weights greater than 1,000 kg are recorded. Due to enlargements, the group of countries in the intra-EU and extra-EU declaration has changed over time. In 1995, Austria, Finland, and Sweden joined the EU. In May 2004 Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, and Slovenia joined. In 2007, Bulgaria and Romania joined. Lastly, Croatia joined the EU in 2013. In 2019, Great Britain left the EU.[9]

The do-file "**4_CN8_over_time_ext.do**" computes the CN8 concordances over time. At the beginning of the do-file, we set the starting year (*1988*) and final year (*2019*). Similarly, the do-file can automatically adapt to compute CN8 concordances over time for any chosen time within the period of investigation. The final output is called "**cn8_cn8plus_bbbb_eeee.dta**", where "**bbbb**" refers to the first year of the concordance and "**eeee**" stands for the final year. Moreover, the concordance output delivers the common classification of CN8+ codes, which can be found inside the folder "**Output**". This output file is relevant to concord international trade data and delivers a consistent classification.

The terminology used for concording combined nomenclature (CN8) over time is identical to Van Beveren, Bernard and Vandenbussche (2012) and follows Pierce and Schott (2012*a*,*b*). To keep consistency with these authors, we also refer to "***effyr***" as the year in which a particular change in classification over time becomes effective. "***obsolete***" represents the codes that are no longer used starting in the effective year (*effyr*), and "***new***" refers to the codes that will be used starting in the effective year. Furthermore, we need a unique identifier for all mappings present in a concordance file. Mappings can be "***simple***", in which one obsolete code is replaced by one new code, or one source code translates into one destination code, or "***complex***", in which one or more obsolete codes are replaced by one or more new codes, or one or more source codes translate into one or more destination codes.

There are three types of concordance procedures: (i) Developing a consistent concordance between two years (e.g., between CN8 in 1995 and 1996) or between two classifications in a single year (e.g., between CN8 and PC8 in 2003). (ii) Developing a consistent concordance over time, (e.g., CN8 classification for 1988 through 2019), this relies on the procedure in (i) plus an added procedure for the chains over time. (iii) Developing a consistent concordance between two classifications over time, (e.g., CN8 and PC8 for 1995 through 2022). This relies on the procedures in *(i)* and *(ii)*. Moreover, additional issues must be considered when concording between two classifications. Furthermore, we need a unique identifier for all mappings present in a concordance file for consistent concordances between years or between two classifications. Specifically, we create the synthetic code "***setyr***"[10] that indicates cases where two obsolete codes in $t-1$ map into one new code in t. Likewise, "*setyr*" indicates cases where one source code maps into two destination codes in a particular year. This unique identifier can be assigned to complex mappings by consecutively sorting the data on

[8]The first six digits of the CN8 codes correspond to the Harmonized System (HS6) nomenclature.

[9]When a country joins the EU, its exports and imports are declared in the intra-EU trade statistics instead of the extra-EU trade statistics. The opposite happens if a country leaves the EU.

[10]The terminology follows Pierce and Schott (2012*a*,*b*).

source (*obsolete*) and destination (*new*) codes. By consecutively sorting the data, it is possible to identify the additional source (*obsolete*) and destination (*new*) codes that need to be grouped in the synthetic codes.

Based on the Stata code provided by Van Beveren, Bernard and Vandenbussche (2012), we extended the concordance procedure for combined nomenclature CN8 codes until 2022 using the do-file "**4_CN8_over_time_ext.do**." Similarly, the procedure relies on the algorithms developed by Pierce and Schott (2012*a*,*b*) to sort CN8 codes consecutively using a loop in Stata. In the final concordance file, all source (*obsolete*), destination (*new*) codes, and synthetic codes (*setyr*) are collected. Consecutive changes in codes—like when new codes in some years become obsolete in later years—are chained together in "*family trees*." Family trees depend on the beginning and end year of the concordance.[11]

The concordance procedure 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 proposed by Van Beveren, Bernard and Vandenbussche (2012, Section 3.1).[12] Table 1 illustrates the replication and extension of the 8-digit combined nomenclature (CN8) classifications until 2022 (see Table 1 in VBBV). Table 2 shows that the largest number of revisions tend to occur in years when the HS6 codes are revised.[13]

- ***Step 1: Concording CN8 codes between t and t-1***: Changes in CN8 codes over time are first classified into different types of mappings. Mappings between two consecutive years can be "**simple**" (one obsolete CN8 code in *t-1* translates into one new CN8 code in *t*), or ***complex***, which can be "**one-many**" (one obsolete CN8 code maps into more than one new CN8 code); "**many-one**" (more than one obsolete CN8 code maps into one new CN8 code); and "**many-many**" (multiple obsolete codes translate into multiple new CN8 codes) (see Table 3 in VBBV).[14] A unique identifier (*setyr*) is assigned to each mapping. For many-many and one-many mapping between two years, a feedback loop derived from Pierce and Schott (2012*b*) is used to ensure that the correct grouping procedure is applied. In addition, Table 2 illustrates the years of extension, showing that these codes are identical to those presented in VBBV.

- ***Step 2: Developing a consistent classification over time***: To develop a

[11]Therefore, the concordance procedure leads to a different number of synthetic codes when the time period is changed. Increasing the time period considered will reduce the number of synthetic codes, as synthetic codes group original products that were recorded either more (shrinking family) or less (growing family) detailed in previous years. For further details on the sensitivity analysis for different time periods, see footnote 14 in VBBV. Families of codes are identified by searching for updates of new codes in later years. Only codes that have undergone multiple changes over the time period considered are retained. Hence, to end up with a unique identifier the family trees are combined with the original mappings. In other words, for each new code in a particular year, the algorithm searches for matching (identical) obsolete codes in later years. Two families of which the code is part are chained together if a new code has become obsolete in later years. The resulting family trees are then merged back into the file with all obsolete and new mappings.

[12]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. The third step discusses the actual implementation of the concordances in the international trade data.

[13]HS6 codes have been revised several times (1992, 1996, 2002, 2007, 2012, 2017 and 2022).

[14]If the mapping is one-many, the family of codes is growing, see Figure 1. If the mapping is many-one, the family of codes is shrinking, see Figure 2.

consistent concordance over time, we have to perform several additional steps. First, we use the "news loop" developed by Pierce and Schott (2012*b*) to chain subsequent code changes over time and to assign a unique identifier to these families. This ensures that codes that have changed in more than one year are chained together. We then merge these chains over time to the year-to-year concordance files developed in Step 1. We then use a similar loop to assign a unique "*setyr*" to families over time. Finally, we merge these families to the full list of existing CN8 codes in each year and translate the CN8 codes into the CN8+ classification. The concordance file "**cn8_cn8plus_bbbb_eeee.dta**" collects all CN8 codes subject to changes over time as well as their unique CN8+ code. The concordance file is dependent on the time period, i.e., the CN8+ classification will be different depending on the sample period chosen at the beginning of the do-file. Longer time periods lead to more CN8 codes that change over time, increasing the number of grouped CN8 codes. Therefore, the CN8 and CN8+ codes are year-specific, i.e., the CN8+ codes need to be merged into the data at the year-CN8 level.

- ***Step 3: Concording the international trade data***: We also need to concord European product international trade data, which is recorded in the CN8 classification, into the CN8+ classification. The name of the variable referring to the CN8 codes is "**cn8**", it is numeric and the data is sorted on the *year* and *cn*8 variables. We merge this data file with the concordance file (**cn8_cn8plus_bbbb_eeee.dta**). All CN8 codes that are in the data should also be in the concordance file.[15] In a final step, we aggregate the international trade data from the CN8 to the CN8+ product level.[16] Please see "*readme_cn8_over_time.pdf*" in Van Beveren, Bernard and Vandenbussche (2012).

Once we have the data in a consistent format, we save it as "**goods_pt_1988_2022.dta**" inside the folder "**cn8_over_time**." The final step is to execute the main do-file for CN8 concordances over time "**4_cn8_over_time_ext.do**," which is almost identical to VBBV and implements the sorting algorithms from Pierce and Schott (2012*a*,*b*) delivering CN8+ concordance from 1988 to 2022. The final concordance output file is named "**cn8_cn8plus_bbbb_eeee.dta**" (or csv format).

- We use this final file to concord (yearly) international trade data from the CN8 classification to CN8+ with the following merge variables: "*cn8*", "*year*".
- The concordance file is dependent on the time period. The time period can be set at the beginning of the do-file "**4_CN8_over_time_ext.do**".
- Variables:
  - "*year*": Numeric variable, refers to the year.
  - "*cn8*": Combined Nomenclature (8-digit) code (year-specific) is a numeric variable. CN8-year combinations are unique in the concordance file (each CN8 code features only once in each year).

[15] There are some residual categories (e.g., codes starting with or ending on 9999) and coding errors. We drop these products from the data to avoid spurious entry and exit dynamics.

[16] Van Beveren, Bernard and Vandenbussche (2012) highlight the relevance of concording the data to avoid spurious entry and exit dynamics at the product level (see Section 4.1 in VBBV).

– "*synthetic*": dummy variable equal to one if the CN8+ classification groups more than one CN8 code. It can be used to distinguish between original (ungrouped) CN8 products and sets of CN8 products grouped for consistency over time. In general, the longer the time period chosen, the higher the number of synthetic codes (numeric).

– "*cn8plus*": CN8+ code corresponding to the CN8 code in a specific year (numeric).

2 Concording PC8 Products Over Time: Extension

This section provides guidelines for the concordance procedure of Prodcom 8-digit (PC8) codes over time. Similar to Section 1, Prodcom codes can be translated into a common classification PC8+ from 1995 to 2021.[17] Likewise, the concordance approach relies on the algorithms developed by Pierce and Schott (2012*a*,*b*). The data path starts again with the folder "**Concordance Trade Production**". Prodcom classifications are obtained through Eurostat and are located inside the folder named "**pc8_over_time**". The production data for all members of the European Union is extracted using the Bulk download facility.[18]

Original files are available inside the subfolder "**Originals Ramon Eurostat**". The first input file "**input_file_PC_over_time_edited_ext.csv**" is a modified version of the original file provided by Van Beveren, Bernard and Vandenbussche (2012) named "**input_file_PC_over_time_edited.csv**". This input file contains the changes of 8-digit Prodcom codes from 1993 until 2021. To achieve the extended version of this file, we implemented the following intermediate steps.

Step 1: Checking consistency:

The file **PRODCOM_YEARLY_UPDATE_OF_CODES_SINCE_1994.xlsx** is retrieved from Eurostat Ramon and is located inside "**Originals Ramon Eurostat**". It includes the PC8 code changes between a pair of years from 1994 to 2009. In addition, all Prodcom codes are contrasted with the file "**input_file_PC_over_time_edited.csv**" provided by Van Beveren, Bernard and Vandenbussche (2012). To achieve consistency between the two sources; we generate the changes of PC8 codes between each pair of years "*Prodcom_yyyy_Prodcom_xxxx*", where "*yyyy*" refers to the current year, and "*xxxx*" is the previous year.[19] Subsequently, we manually compared each pair of codes. Lastly, we created six additional variables to

[17] In Van Beveren, Bernard and Vandenbussche (2012), the starting year of the PC8 concordance procedure over time is 1993 (see Table 4 in VBBV). However, the first year of available production data Europroms is 1995. Therefore, Prodcom concordances range from 1995 to 2021.

[18] We thank Erlend Firth and Laia Guinovart from the Eurostat team for providing us with clarifications related to the Prodcom classifications and supplementary files.

[19] The file extension of the Prodcom concordances is different depending on the pair of years. For the first year (1993), the concordance is available only in text format "*.pdf*". Then, for the period between 1994 to 2011, the file extension is comma-separated value "*.csv*". From 2012 until 2015, the format is a text document "*.txt*". For 2016 is "*.csv*". Finally, from 2017 to 2021 is "*.txt*". Moreover, the concordance file "*Prodcom_2018_Prodcom_2017*" does not exist due to no changes in classification.

highlight the discrepancies in the codes between the two files.[20]

- "*correct*": is a dummy variable, where one indicates that the code matches correctly with the Eurostat Ramon original files and zero otherwise.
- "*diff_ex_pcfrom*": is a dummy variable, where one indicates that the variable "*ex*" of the current year (*pcfrom*) is different from the Eurostat files and zero otherwise.
- "*diff_ex_pcto*": is a dummy variable, where one indicates that the variable "*ex*" of the subsequent year (*pcto*) is different from the Eurostat files and zero otherwise.
- "*diff_ab_pcfrom*": is a dummy variable, where one indicates that the variable "*ab*" of the current year (*pcfrom*) differs from the Eurostat files and zero otherwise.
- "*diff_ab_pcto*": is a dummy variable, where one indicates that the variable "*ab*" of the subsequent year (*pcto*) differs from the Eurostat files and zero otherwise.
- "*comments*": is a string variable that shows relevant comments for certain PC8 codes.

To address the discrepancies between the original files from Eurostat Ramon and Van Beveren, Bernard and Vandenbussche (2012), we generate the following do-file that allows us to highlight the main differences in PC8 concordances for each pair of years and also creates the new variables from 1993 to 2021.

◇ **5_Prodcom concordances_1993_2021.do**

The first concordance file "*prodcom 1994 - prodcom 1993*" is not compatible with Stata due to its format ("*.pdf*"). Thus, we manually introduced the corresponding values for the variable "*correct*". This concordance file contains 137 observations and 14 unmatched codes ("*correct*" is equal to zero). A modified version of VBBV file (including zeroes in 1993) named "**input_file_PC_over_time_edited.csv**" (henceforth, "input") is located inside "**Originals Ramon Eurostat**". The file "*prodcom 1995 - prodcom 1994*" shows 132 observations and 44 unmatched codes. Similarly, the following concordance file "*prodcom 1996 - prodcom 1995*", exhibits 167 observations and 11 *pcto* missing codes in "input". Moreover, "**PRODCOM_YEARLY_UPDATE_OF_CODES_SINCE_1994.csv**" (henceforth, "E-updates") reports 15 *pcto* missing codes and 4 unmatched codes. The next concordance file, "*prodcom 1997 - prodcom 1996*" does not exist due to no changes in Prodcom classification. Followed by "*prodcom 1998 - prodcom 1997*", which displays 12 observations, 4 unmatched codes, and 2 missing codes in "E-updates".

Furthermore, "*prodcom 1999 - prodcom 1998*" shows inconsistencies in certain codes (i.e., in "input" "232000X1" and "401010X1"). According to the Prodcom Complete Structure files (PCS), these codes should instead be reported as "232000E1" and "401010E1". The concordance file shows 157 observations and 11 *pcto* codes that are different in "E-updates". Likewise, "*prodcom 2000 - prodcom 1999*" reports 351 observations and 331 *pcto* unmatched codes in "E-updates". Then, "*prodcom 2001 -*

[20]We identified typos, duplicates, missing codes, and inconsistencies between the two sources. The main differences between Prodcom codes are shown in the file "**5_Prodcom concordances_1993_2021.do**". Additionally, the file "*input_file_PC_over_time_edited_ext.xlsx*" shows these discrepancies.

prodcom 2000" displays duplicate codes in "input"; after removing the duplicates there are 156 observations and zero unmatched codes.

The concordance "*prodcom 2002 - prodcom 2001*" shows duplicate values in "E-updates" (i.e., "21123053" and "21123059"). Based on the Prodcom Complete Structure (PCS), the code "25249023" should be reported as "29249023". The total number of observations is 147 and 26 unmatched codes plus 2 codes not present in "input". Discrepancies in the concordance "*prodcom 2003 - prodcom 2002*" are notorious; "input" reports 643 observations while "E-updates" shows 651 observations (this is relevant for the PC8 concordances over time since we included additional codes from E-updates). This concordance file exhibits significant discrepancies for the variable "*ex*". In addition, we identified typos in "*pcfrom*", i.e., the code "401010E1" in "input" is reported as "401010X1".

Then, "*prodcom 2004 - prodcom 2003*" displays duplicates for the codes *pcfrom* "21231300" and *pcto* "21231350", reports 50 observations, and shows a missing code "29242341" in "E-updates". The next concordance "*prodcom 2005 - prodcom 2004*" shows differences (i.e., "input" reports 976 observations and duplicates for the *pcto* code "24664895"). The number of unmatched codes is 670, plus 2 typos in the codes "401110X1" and "232000X1". Additionally, 20 missing codes in "input", which leads to 996 observations. All PC8 codes matched correctly; "*prodcom 2006 - prodcom 2005*" (4 observations). Similarly, all codes matched in "*prodcom 2007 - prodcom 2006*" (287 observations). The next concordance "*prodcom 2008 - prodcom 2007*", shows 4697 observations and duplicates for the code "269900Z0"; some codes should be reported in "input" as 8-digit codes (i.e., *pcto* "7101000").

Furthermore, "input" contains duplicates for the following *pcfrom* codes "28413330", "28413340", "28413370" and "28413380" in "*prodcom 2009 - prodcom 2008*"; once these duplicates are removed, there are 29 observations. The last year that allows comparisons is 2010, "*prodcom 2010 - prodcom 2009*" includes 52 observations, and all the codes matched correctly. For the extension, we extracted prodcom concordances from Eurostat Ramon. The number of observations for "*prodcom 2011 - prodcom 2010*" is 63 observations; 90 observations in "*prodcom 2012 - prodcom 2011*". "*prodcom 2013 - prodcom 2012*" shows 20 observations, "*prodcom 2014 - prodcom 2013*" reports 6, "*prodcom 2015 - prodcom 2014*" is 15, "*prodcom 2016 - prodcom 2015*" is 195, and "*prodcom 2017 - prodcom 2016*" is 161. Moreover, "*prodcom 2018 - prodcom 2017*" and "*prodcom 2020 - prodcom 2019*" are identical to the previous year (hence, the concordance file does not exist). Finally, 296 codes in "*prodcom 2019 - prodcom 2018*" and 19 codes in "*prodcom 2021 - prodcom 2020*".[21]

Step 2: Obtaining mandatory, optional and aggregated codes:

The input files required to compute the PC8 concordances over time are the following:

➔ List of mandatory PC8 codes in each year (**PC_yyyy_ext.dta**), where "**yyyy**" refers to the different years included in the concordance period.[22]

➔ Year-specific list of optional (B-list) PC8 codes with their mandatory counterparts (**PC_yyyy_Blist_ext.dta**). These input files allow for the identification and recording of optional codes used in the production data before concording the data.

[21] Total number of observations from 1993 to 2010 in VBBV "**input_file_PC_over_time_edited.csv**" is 8005. In contrast, the extended file "**input_file_PC_over_time_edited_ext.csv**" contains 8160 observations.

[22] The suffix "**_ext**" represents the extension; we compared our generated codes *w.r.t* VBBV codes.

Furthermore, these files are required if the time period in the PC8 concordance starts before 2005.

➔ List of optional (N-list) PC8 codes (**Nlist_codes_1993_2005_ext.dta**) with their mandatory counterparts. This input file enables for correct identification and recoding of optional codes used in the production data before concording the data. This file is required when the time period in the PC8 concordance starts before 2005.

➔ **input_file_PC_over_time_edited_ext.csv**: Extended version of the list of changes in mandatory PC8 codes between pairs of years.[23]

➔ List of optional PC8 codes for the period between 1993 to 2005 and their corresponding mandatory PC8 counterparts, it is important to replace optional codes before 2005 (**optional_codes_1993_2005_ext.dta**).[24]

➔ List of aggregate (T-, Z-, Q-, V-, or E-list) codes between 1993 and 2021, to be dropped from the input file (**TZQlist_codes_1993_2021_ext.dta**).[25]

The folder "**Prodcom Complete Structure**" includes the structure files from 1993 to 2021 and is located inside "**Original Ramon Eurostat**". Each file is named as follows: "*PRCSTR_yyyy_complete*", where "*yyyy*" refers to the respective year in the Prodcom classification. These files are relevant because we will use them to generate the mandatory, aggregated, and optional codes listed above by running the following do-file:

◇ **6_PC_1993_2021.do**

This do-file delivers mandatory, optional, and aggregated Prodcom codes from 1993 to 2021. It performs a loop for three different groups. The first group from 1993 to 2003 (due to the file format in 2004, it generates the required codes separately). The second group ranges from 2005 to 2010. Lastly, the third group ranges from 2011 to 2021 (*extension group*). To generate the PC8 mandatory codes ("*PC_yyyy_ext*", where "*yyyy*" refers to the respective year in the Prodcom classification) and optional B-codes (*PC_yyyy_Blist_ext*), we relied on the Prodcom Complete Structure (PCS) files. At the beginning of the do-file we set the starting (1993) and end (2003) years. This allows us to generate all the mandatory codes, optional (A-, B-, N-) codes as well as aggregated (T-, Z-, Q-, V-, E-) codes.

Additionally, we ensure that all codes are consistent with Van Beveren, Bernard and Vandenbussche (2012) using the "`cfout`" Stata command, which allows comparisons between two different data sets. After repeating this procedure for the second and third groups, we found inconsistencies in certain codes.[26] Finally, after achieving consistency in all PC8 codes, we included the years of extension as a panel (from 1993 to 2021) into

[23]This input file is generated by executing the do-file "**5_Prodcom concordances_1993_2021.do**".

[24]Optional codes are obtained from the prodcom classification extracted from the Eurostat Ramon server and contrasted with VBBV files using the do-file "**6_PC_1993_2021.do**".

[25]We drop aggregated codes here when they affect their disaggregated counterparts, as this is recorded in the concordance files using the disaggregated codes.

[26]For further details concerning discrepancies in codes with VBBV please see "**6_PC_1993_2021.do**".

the main files.[27]

Step 3: Obtaining production data:

Production data is extracted from Eurostat using the Bulk download facility and is reported as "Europroms".[28] The data for domestic production has a particular feature relative to PC8 codes; Eurostat reported the most recent data considering the last revision of Prodcom classifications in 2008 ("*NACE revision 2.0*"). This means that all PC8 codes are subject to this revision, creating additional challenges for the concordance procedure.

To address the issue of Prodcom codes with *NACE revision 2.0* for years prior to 2008, we combined previous production data which contains original PC8 codes, with an updated version of the production sold using "Easy Comext". Eurostat reports production data with original (year-specific) codes for the time period between 1995 to 2012 ("*epannsold*"), and data with PC8 codes subject to the 2008 revision from 1995 to 2014 ("*epannsold-r2*").[29] The access route for downloading follows:

➔ *Comext files* > *COMEXT_ OTHER_ DATA* > EUROPROMS > *epannsold.zip*

➔ *Comext files* > *COMEXT_ OTHER_ DATA* > EUROPROMS > *epannsold-r2.zip*

The production data "*epannsold-r2*" contains the following variables:[30]

- "*decl*": The list of reporting countries includes the EU member states, EFTA countries, and acceding and other candidate countries. Data for all the potential candidate countries are available as well. The codes 1110, 1111, 1112, and 2028 can be selected to obtain EU 15, EU 25, EU 27, and EU 28 totals respectively.
- "*period*": The annual data offers the periods 199552 on-wards, where 52 is the Comext convention for annual data.
- "*prccode*": refers to the Prodcom 8-digit (PC8) codes.[31] The list includes all codes that are valid in at least one year. If a code is selected that is not valid for the year selected, no data will be shown for that code.
- "*prodval*": This field gives the value of production in euros.
- "*prodqnt*": This field gives the volume of production in the unit indicated in "*unit*".
- "*expval*": This field gives the value of exports in euros, derived from the External Trade statistics
- "*impval*": This field gives the value of imports in euros, derived from the External Trade statistics

[27] The file "*TZQlist_ codes_ 1993_ 2021_ ext.dta*" contains the updated version of T-, Z-, Q-, V-, E-codes.

[28] Similarly to the international trade data, the file extension of the production data is "*.dat*". "**7_Production_Data_1995_2021.do**" allows us to extract the data directly from the Bulk download and import it into Stata. Alternatively, we provide an R file named "*Prodcom_ sold.R*" which transforms the raw files into a readable format "*.csv*".

[29] The last update for "*epannsold-r2*" was 24/11/2017.

[30] Europroms guides available at `https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines`

[31] All the codes in the data are subject to the *NACE Rev-2.0* which took place in 2008. Meaning that the codes prior to 2007 are subject to the PC8 2008 classification.

- "*expqnt*": This field gives the volume of exports derived from the External Trade statistics.
- "*impqnt*": This field gives the volume of imports derived from the External Trade statistics.
- "*pvalbase*": For EU totals, this gives the rounding base used if "*prod_value*" is rounded or contains a rounded element. "*prodval*" should be interpreted as lying between "*prodval*" - "*pvalbase*" and "*prodval*" + "*pvalbase*". When no rounding is applied, "*pvalbase*" is set to zero.
- "*pqntbase*": for EU totals, this gives the rounding base used if "*prodqnt*" is rounded or contains a rounded element. "*prodqnt*" should be interpreted as lying between "*prodqnt*" - "*pqntbase*" and "*prodqnt*" + "*pqntbase*". When no rounding is applied, "*pqntbase*" is set to zero.
- "*pvalflag*": This field indicates the availability of the volume data. Possible values are blank (data is available), ":" data is not available, ":C" data is confidential, ":R" the data has been rounded using the rounding base given in "*pvalbase*", "-" not applicable. Additional flags are used to indicate that a total has been constructed: e.g. EU27-EU02(R) indicates that the EU25 total has been constructed from the EU 27 minus the rounded sum of Bulgaria and Romania. The information in the flag is also given as footnotes.
- "*pqntflag*": This field indicates the availability of the volume data. Possible values are blank (data is available), ":" data is not available ":C" data is confidential, ":R" the data has been rounded using the rounding base given in "*pqntbase*", "-" not applicable. Additional flags are used to indicate that a total has been constructed: e.g. EU27-EU02(R) indicates that the EU25 total has been constructed from the EU 27 minus the rounded sum of Bulgaria and Romania. The information in the flag is also given as footnotes.

To download the updated version of the domestic production we require the following steps:

- Open Internet Explorer (IE) or use the IE mode through Microsoft Edge.
- Access to `https://ec.europa.eu/eurostat/comext/newxtweb/`
- Create an account using EU login.
- Click on the option "Analytical client" (at the top right of the screen).
- Install Java environment.
- Install Analytical Client v18.0.4.
- Create a query using EU Server.
- Select "Statistics on the production of manufactured goods and international trade (Europroms)" > Annual detailed data since 1995 by PRODCOM list (NACE/rev.2) > Sold production, exports and imports

- Write any Name: (i.e., "*soldprod_2021*").
- Select "AddAll" and then "Close" for *DECL*, *PERIOD*, *PRCCODE*, *INDICATORS*.
- Select Out Data > name it > make sure is Type CSV
- Select Extract and wait for COMEXT-Jobs to download your query.
- Select Tools (i.e.,"*soldprod_2021*") > File Download
- Save the file inside the folder "Prod Data" > "Prodcom" > "**Full**". The name of the domestic production is **"soldprod_2021.csv"** (the file is reshaped and formatted using "**7_Production_Data_1995_2021.do**", see lines 248-297).

To obtain production data that are consistent with the concordance procedure implemented by Van Beveren, Bernard and Vandenbussche (2012), we assemble production data suitable for the concordances both over time and in a single year using the following do-file:

◇ **7_Production_Data_1995_2021.do**

Moreover, production data from 2015 to 2021 is retrieved from Eurostat - Full access to detailed statistics on international trade in goods (*Comext*) (Easy Comext) using the option "*Analytical Client*".[32] The input file for PC8 concordances overtime is named "**production_pt_1995_2021.dta**", which combines production data from 1995 to 2007 with original PC8 codes "*epannsold*" and the most updated production data from 2008 to 2021 *"soldprod_2021.dta"* (Data retrieved from *Eurostat Easy Comext - Analytical client*). Revisions and discrepancies with respect to "*epannsold.r2"* are highlighted within the do-file.

2.1 Domestic Production Activities: Challenges

Following Van Beveren, Bernard and Vandenbussche (2012, Section 2.2), there are certain limitations and challenges when using Prodcom data. An important limitation is that PC8 codes are also subject to annual changes, i.e., the same product covered by the Prodcom survey in neighboring years might be reported under different PC8 codes. VBBV highlight three main additional challenges:

1. Coverage of the CN8 classification is constant across years, while coverage of the Prodcom list changes over time.

   – Hence, a good may be covered by a Prodcom code in one year but not covered by any Prodcom code in another year (*inconsistency*).

   – Production in these codes can not be tracked over time. Therefore, these production data are dropped when concording over time.

[32]To obtain the production data it is necessary to have installed Java with Analytical Client v18.0.4 and the browser must be Internet Explorer. "Annual detailed data since 1995 by Prodcom list (NACE/rev.2)-Sold production, exports, and imports" is downloaded through a Eurostat server. Recently, this server has been restricted to Commission internal staff. Furthermore, to achieve consistency; the data is formatted and reshaped using the do-file **7_Production_Data_1995_2021.do**.

2. The Prodcom classification system uses B-list and N-list optional codes. While B-list codes are introduced at the request of member states, N-list codes are implemented by Eurostat to allow for a finer level of disaggregation of production than the PC8 codes.

   – Only some countries used the optional codes, rendering calculation of EU totals for these optional products impossible.

   – (Van Beveren, Bernard and Vandenbussche, 2012, (Table 2)) show that optional products were gradually phased out and eliminated by 2005.

3. Existence of more aggregated versions of mandatory PC8 codes.

   – There are five types of more aggregated Prodcom codes: Q-, V-, Z-, T-, and E-list. Q-list codes refer to aggregated versions of certain textiles listings that were supposed to be reported on a quarterly basis. They were dropped in 2005.

Furthermore, note that the first six digits of the eight-digit Prodcom codes correspond to CPA6 products, or, in other words, 6-digit products classified according to the Classification of Products by Activity. PC8 codes are updated annually (with the exceptions of 1997, 2018, and 2020 when no changes were implemented in the PC8 classification), while CPA6 codes were updated in 1996, 2002, and 2008 (NACE rev.2). As not all CPA6 codes are covered by the PC list, the total number of CPA6 products does not correspond to the number of PC8 products.

The NACE 4 classification used by the European Union is a classification of economic activities that create products. It is often used to classify European firms into sectors based on their (main) economic activities.

The Prodcom list, on the other hand, is closely related to the Combined Nomenclature classification, which is used to record foreign trade statistics. The Prodcom list covers production activities in Mining, Quarrying, and Manufacturing: sections C, D, and E of the NACE 4. Prodcom does not include products that are not considered manufactured products (waste, some agricultural products where the processing is not considered as manufacturing etc.), although they belong to sections C, D, or E of NACE. The Prodcom list also does not cover fuel products.

In order to compute the PC8 codes into a common classification over time, we used the full list of existing PC8 codes in each year and the changes over time in the Prodcom codes between each pair of years. For the list of PC8 codes in each year, this implies retaining only mandatory 8-digit Prodcom codes (the original files additionally contain optional codes) and renaming and formatting the variables consistently for use in the concordance procedure. Each concordance file from Eurostat (changes in PC8 classification between *t-1* and *t*) includes optional (B- and N-list) and aggregated (Z-, T-, Q-, V- or E-list) codes. Thus, the input files need to be modified to only include changes that apply to mandatory PC8 codes, and to ignore changes in optional and aggregated codes except when these changes affect the underlying mandatory codes.

Similarly, "**input_file_over_time_edited_ext.dta**" contains two additional variables: (i) *pcfrom_recode*, includes the mandatory or disaggregated counterpart of the obsolete optional or aggregated code; and (ii) *pcto_recode*, includes the mandatory

or disaggregated counterpart of the new optional or aggregated code.[33] Moreover, if an optional or aggregate code is reported as an "*exit*" code (no new code is listed), while the mandatory or disaggregated codes still exist, the mandatory code is entered as the "new" code (variable *pcto*). Finally, the do-file "**8_pc8_over_time_ext.do**" computes the PC8 concordance procedure from 1995 to 2021, generating the concordance file "**pc8_pc8plus_bbbb_eeee.dta**", where "**bbbb**" refers to the first year of the concordance and "**eeee**" refers to the final year. This file can be found inside the folder "**Output**".

Likewise, at the beginning of the do-file we set the beginning (*1995*) and end (*2021*) year. Moreover, the concordance file provides a year-specific list of all existing mandatory PC8 products, their corresponding PC8+ code, and a dummy variable indicating which PC8 products need to be dropped due to changes in coverage over time. According to Van Beveren, Bernard and Vandenbussche (2012), all optional codes (B- and N-list) that (potentially) feature in the data need to be recoded into their mandatory counterparts before implementing the concordance procedure.

Analogously to Section 1, to concord Prodcom classifications over time, we relied on the procedure by Van Beveren, Bernard and Vandenbussche (2012), which 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) mentioned above. The last step discusses the actual implementation of the concordances in the production data expressed using the PC8 classification.

- ***Step 1: Concording PC8 codes between t and t-1***: First, changes in PC8 codes over time are classified into different types of mappings. Mappings can be *simple*, where one obsolete PC8 code in $t-1$ translates into one new PC8 code in t, *many-one*, *one-many*, and *many-many* (cfr. in VBBV, Table 4). A unique identifier (setyr) is assigned to each mapping. To ensure that the correct grouping procedure is used for many-many and one-many mappings between two years, the feedback loop derived from Pierce and Schott (2012*b*) is applied.

- ***Step 2: Developing a consistent classification over time***: To create a consistent concordance over time, the following additional steps need to be taken. First, using the "news loop" developed by Pierce and Schott (2012*a*,*b*) subsequent code changes over time are chained to assign a unique identifier to these families. Hence, codes that have changed in more than one year are chained together. These chains are then merged back into the year-to-year concordance files resulting from Step 1. As in Step 1, a unique identifier *setyr* is assigned to families over time. The families are then merged back into the full list of existing PC8 codes for each year to translate the PC8 codes into the PC8+ classification. Also, all PC8 codes that newly appear are merged into the concordance, which allows identifying all families that need to be dropped to maintain consistent coverage over time. The concordance file "**pc8_pc8plus_bbbb_eeee.dta**" collects all PC8 codes subject to changes over time, their unique PC8+ code, and a dummy indicating whether the product needs to be disregarded in the data due to changes in coverage over

[33] It can happen 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*). (cfr. footnote 3 in "*readme_pc8_over_time*" from VBBV)

time. The concordance files are specific to the time period chosen, i.e., the PC8+ classification changes depending on the sample period chosen. A longer time period chosen leads to more PC8 code changes, increasing the number of grouped PC8 codes and the number of PC8 codes that need to be dropped for consistency over time. It also implies that the PC8 and PC8+ codes in the final concordance file are year-specific, implying that the PC8+ codes need to be merged into the data at the year-PC8 level.

- ***Step 3: Concording production data***: To be able to concord European domestic production data, which uses the PC8 classification, into the PC8+ classification, all optional codes that feature in the data need to be re-coded into their mandatory counterparts.[34] The name of the variable referring to the PC8 is "*pc8*". It is a string variable before re-coding the optional codes and numeric afterward (cfr. Stata code provided at end of the do-file). Production data is sorted by *year* and *pc8*. This data file is then merged with the concordance file "**pc8_pc8plus_bbbb_eeee.dta**". All PC8 codes in the data should also feature in the concordance file. In the final step, the production data are aggregated from the PC8 to the PC8+ product level. The concordance procedure delivers the following files:

- **Nlist_codes_1993_2005_ext.dta**

  – This file re-codes optional N-list codes if they appear in the data and if the sample period starts before 2005 (the Stata code provided takes this into account automatically). The merge variable is *pc8*. Before 2004, this file is not required (and ignored) in the concordance procedure.

  – Variables:

  ■ *pc8*: Optional Prodcom N-list (10-digit) code, recorded as a string variable.

  ■ *pc_mand*: Mandatory PC8 code corresponding to N-list code (simply the 10 digits code minus the last two digits), also recorded as a string variable.

- **optional_codes_bbbb_eeee_ext.dta**

  – This file re-codes optional B-list codes if they appear in the data and if the sample period starts before 2005 (the Stata code provided takes this into account automatically). The merge variable is *pc8*. After 2004, this file is not generated by the do-file and is not required in the concordance procedure.

  – While N-list optional codes do not depend on the sample period, 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 a string variable.

  ■ *pc_mand*: Mandatory PC8 code corresponding to B-list code, also recorded as a string variable.

[34] The use of optional codes is country-specific, the Stata code allows for the existence and absence of optional codes in the data (cfr. footnote 8 in **readme_pc8_over_time** from VBBV).

- **pc8_pc8plus_bbbb_eeee (dta or csv format)**

  – This file concords (yearly) production data from the PC8 classification (mandatory codes) to PC8+. The merge variables are *pc8* and *year*.

  – The concordance file is specific to the time period chosen. The time period can be set at the beginning of the do-file "**8_pc8_over_time_ext.do**".

  – Variables:

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

  ■ *pc8*: Prodcom (8-digit) code (year-specific), recorded as a 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*: dummy variable equal to one if the PC8+ classification groups more than one PC8 code. It 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 (numeric).

  ■ *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).

2.2 Concording Trade and Production in a Single Year: Extension

To concord international trade and domestic production data at the 8-digit product level, it is necessary to translate the CN8 product codes (used for international trade) and PC8 product codes (used for domestic production) into a common classification (PC8+). Moreover, due to differences in coverage between the PC8 and CN8 classifications, several things have to be taken care of when concording the trade and production data for a single year: (*i*) some PC8 codes are not covered by CN8. (*ii*) some PC8 codes are recorded as "aggregated" codes in the PC8-CN8 concordance.[35] (*iii*) not all CN8 codes feature in the Prodcom list, for example, Fuel products.

As noted by Van Beveren, Bernard and Vandenbussche (2012), (*i*) arises for certain industrial services and for activities related to installation, maintenance, repair, or processing activities.[36] Prodcom codes that feature on the Prodcom list, but not in the concordance list between CN8 and PC8, are either PC8 codes not covered by the CN8 classification (e.g., industrial services, waste products) or the disaggregated versions of codes on the Z-, T-, V-, Q- or E-list (aggregated PC8 products). After dropping industrial services and recoding optional and disaggregated PC8 products into their mandatory and aggregate counterparts, the PC8 classification can be concorded into the (PC8+) classification.

[35] These are the so-called Z-, T-, Q-, V- and E-aggregates in the Prodcom manual. The Z-, T-, Q-, V- and E-codes correspond to the grouped PC8 codes. The codes that map the underlying PC8 codes into these codes are the codes that are on the PC list (cfr. VBBV, footnote 25). Moreover, aggregate codes are grouped in the structure files under headings 99.t, 99.z, 99.q, 99.v and 99.e.

[36] Examples of industrial services include: dyeing, finishing, and printing of textiles, electronic books, coating of metals and bookbinding services.

(*iii*) implies that the set of CN8 codes that do not appear in the Prodcom list do not feature in the concordance files provided by Eurostat Ramon and need to be excluded from the international trade data. Since the Prodcom list changes every year, the list of CN8 codes that do not feature in the Prodcom list is also year-specific. All codes present in the CN8 classification but not present in the concordance from CN8 to PC8 are CN8 codes that are not covered by the Prodcom list. These CN8 codes necessarily have to be excluded from the international trade data when merging domestic production and trade data at the product level. The remaining CN8 codes are all covered by the Prodcom List in that particular year and can hence be translated into the (PC8+) classification.

When combining data on international trade and domestic production over time, it is important to take into account changes in the coverage of the Prodcom list, the difference in coverage between the PC8 and CN8 classification (industrial services, CN8 products not covered by Prodcom) and changes in both classification systems (CN8 and PC8) over time.

Following the methodology from Van Beveren, Bernard and Vandenbussche (2012), the concordance procedure can be summarized in 4 steps. The first step refers to the concordance of product classifications (relying on type (*i*) concordance procedures, between two classifications in a single year, (cfr. VBBV, Section 3.1). The last three steps discuss the actual implementation of the concordances in 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. Van Beveren, Bernard and Vandenbussche (2012, 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*b*) 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 not covered by the PC8 classification in that particular year. The final concordance file (**concordance_cn8_pc8plus_yyyy.dta**), 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 the production data***: To concord European 8-digit production data to PC8+ products, the following steps are taken. First, for the years before 2005, optional codes featuring in the production data are re-coded 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) are 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) are in the production data, they are re-coded into their corresponding Z-code 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 the PC8 level) with the concordance file. To this end, the concordance file "**concordance_cn8_pc8plus_yyyy.dta**" is adapted such that each PC8 code features only once in the concordance file, with its corresponding PC8+ product code. By construction, all PC8 codes that are present in the data (after re-coding optional codes and Z-aggregates and dropping services) should feature in the concordance. In contrast to Van Beveren, Bernard and Vandenbussche (2012), where PC8+ codes are in some cases more aggregated than the PC8 codes and thus the production data is aggregated to the PC8+, we implement a procedure that ensures the uniqueness for each product code (i.e., optional and disaggregated versions of aggregated codes are dropped to avoid double-counting in the production data).

- ***Step 3: Concording international trade data***: To concord the international trade data, the concordance file **concordance_cn8_pc8plus_yyyy.dta** is 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 is then merged (at the 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 are dropped from the data before concording (these codes have no associated PC8+ product code). To avoid double-counting in the international trade data, we implement a procedure that delivers unique codes. Contrary to Van Beveren, Bernard and Vandenbussche (2012), where PC8+ codes are (can be) more aggregated than CN8 codes, and thus the data is aggregated from the CN8 to the PC8+.

- ***Step 4: Merging domestic production and trade data***: The domestic production data and international trade data are sorted on the firm (if applicable) and product (*pc8plus*) identifier and then merged. The final data contains data on international trade and production, recorded using the PC8+ classification, allowing for product-level comparison. The concordance procedure of CN8 and PC8 into PC8+ within a single year delivers the following files:

- **Nlist_codes_1993_2005_ext.dta**

  – This file re-codes optional N-list codes if they feature in the production data and if the sample period starts before 2005 (the Stata code provided takes this into account automatically). The merge variable is "*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 a string variable.

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

- **optional_codes_bbbb_eeee.dta**

  – This file (in conjunction with the file **PC_yyyy_Blist.dta**, that identifies the year-specific B-list codes) re-codes optional B-list codes that feature in the data and if the sample period starts before 2005 (the Stata code provided takes this into account automatically). The merge variable is "*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 a string variable.

■ *pc_mand*: Mandatory PC8 code corresponding to B-list code, also recorded as a string variable.

- **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 re-code 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 a 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 with their corresponding Z-aggregate (*new_code*) and aggregated when applicable.

  ■ *new_code*: Z-aggregate for codes that are recorded at a higher level of aggregation in the concordance files between CN8 and PC8. The variable is recorded as a string (length 8).

- **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 codes (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.

  – Before 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 before concording.

Additionally, to concord the trade and production data in a single year; we created a do-file for each year named "**9_Data Construction yyyy.do**", where "**yyyy**" is a year ranging from 1995 to 2019. This do-file is year-specific and computes the following input files necessary for the concordance between classifications:

- "**CN_yyyy.dta**" List of CN8 codes the respective year (see Section 1).
- "**CN_self-explanatory text_yyyy.dta**", Descriptions in English for each CN8 code.
- "**cn_pc_yyyy.dta**", Correspondence between CN8 and PC8.
- "**PC_yyyy.dta**" List of mandatory PC8 codes in each year (see Section 2)
- "**Prodcom_Description_yyyy.dta**" Description in English for each PC8 code.
- "**PC_yyyy_Blist.dta**", List of optional (B-list) PC8 codes with their mandatory counterparts, this input file allows for identification and recoding of optional codes used in the production data prior to concording the data. Similarly, these files are required for years prior to 2005 (see Section 2).
- "**PC8_special_codes_yyyy.dta**" List of PC8 codes that have to be dropped (i.e., industrial services and codes without correspondence in the CN8 classification) and their corresponding Z-aggregate in the domestic production data.

Once these input files are saved in the sub-folder with the name of the respective year and located in the folder "**cn_pc_yearly**" we can proceed to execute the concordance procedure between classifications type (i). The do-file "**10_CN8_PC8_cross section_ext.do**" computes yearly concordances from CN8 and PC8 into PC8+ for a chosen year from 1995 to 2019. 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.

Step-by-Step Details for the Concordance CN8-PC8 into PC8+:

- ***Step 1: Concordance from CN8 to PC8+*** :
  - (A) Reading in concordance: **cn_pc_yyyy.dta**
  - (B) Identify types of codes CN8-PC8 (one-many, many-one, many-many)
  - (C) Create groupings (assign "setyr" to different mappings) identical to VBBV
  - (D) Loop to identify feedback effects for Many to Many groupings within each year. Based on the algorithm developed by Pierce and Schott (2012a). The output generated is called "**concordance_cn8_pc8plus_yyyy.dta**" and is located inside the folder "**output**".
- ***Step 2: Differences in coverage CN8-PC8***:
  - (A) List of PC8 codes not covered in concordance (auxiliary file): We need to generate a list of PC8 codes that feature in concordance, but not in the PC8 list. To achieve this, we identify the aggregated Z-aggregate codes and all PC8 codes not covered by CN8 using the Prodcom Complete

Structure (PCS) file (i.e., list of Z-aggregates - NACE 99.z). In addition, for the list of PC8 codes not covered by CN8 + PC8 codes that need to be recoded in aggregates, we contrasted the codes with the auxiliary file "**PC8_yyyy_special_codes.dta**".

(B) List of CN8 codes not covered in concordance. Taking into account which CN8 codes need to be dropped before aggregating to PC8 identifying PC8+ groups.

- ***Step 3: Concording production data***: Intermediate steps (i) year-specific.

(i) Optional codes: B-list and N-list: prepare files (see Section 2)

(A) Read in domestic production data (**production_pt_yyyy.dta**).

(B) Recode of all optional codes (B- and N- codes). In contrast to Van Beveren, Bernard and Vandenbussche (2012), to avoid double-counting in the domestic production data we dropped disaggregated B-codes and N-codes (see the do-file "**10_CN8_PC8_cross section_ext.do**" lines 462-488).

(C) Merge the production data with the file "**PC8_yyyy_special_codes.dta**" generated using the do-file "**9_Data Construction yyyy.do**" on variable "*pcyyyy*", where "*yyyy*" is the respective year.

(D) Drop products which type is "industrial services" or "no cn correspondence".

(E) Recode All products where the type is "aggregate". Again contrary to VBBV, we dropped the disaggregated Z-codes to avoid double-counting (i.e., 15711000, 15721050 in the year 1995, see lines 558-565).

(F) Concording PC8 products into PC8+ classification. In addition, we identified pc8plus codes that contain more than one PC8 code (*synthetic*).

(G) Avoid double-counting for synthetic letter (T-, Q-, V-, E-) aggregated codes. We assigned flags for each letter code (see lines 616-621). Moreover, we created additional flags for Z-aggregate codes that are synthetic (i.e., pc8 271041Z5, 271041Z6 in the year 1995) and assigned additional flags namely, (*synth_agg*, *synth_dis* and *synth_both_agg*) and modified the variable *drop_synth* accordingly to avoid double-counting in the data (see lines 624-658). The output file containing the domestic production concorded into PC8+ is the following "**production_yyyy_concorded.dta**".

- ***Step 4: Concording international trade data***:

(A) Read in international trade data (**goods_pt_yyyy.dta**).

(B) Merge
the trade data with the file "**concordance_cn8_pc8plus_yyyy.dta**" on variable "*cnyyyy*" to obtain PC8+.

(C) Drop CN8 products not covered by PC8 list. We dropped products where the dummy "*notpc*" equals one (these represent CN8 products not covered by PC8).

(i) Include the non-EU countries that report production. We included non-EU countries' trade flows that record production as a declarant in the trade data. Availability for Norway (1995-2019), Iceland (1997-2019),

Turkey (2006-2010), Bosnia and Herzegovina (2011-2019), Serbia (2011-2019), Montenegro (2011-2019) and Macedonia (2011-2019). We switched between declarant and partner and renamed variables to achieve consistency (see lines 711-774).

(ii) Consistency between trade and production data. Modification for Belgium and Luxembourg in the Comext data before 1999. We assigned the trade values of both countries to Belgium from 1995 to 1999. Finally, we modified the variable *declarant* to achieve consistency between trade and production and also assigned labels to the main variables. The output file which contained international trade data concorded into PC8+ is named "**goods_pt_concorded_yyyy.dta**".

- ***Step 5: Merging domestic production and trade data***:

  Sort international trade (**goods_pt_concorded_yyyy.dta**) and production data (**production_yyyy_concorded.dta**) and merge the two data sets on the common product identifier (*pc8plus*). The output file is the following "**trade_prod_pc8plus_yyyy.dta**". (End of VBBV procedure)

- ***Step 6: Extension: Obtaining consistent codes over time***:

  (A) Obtain the PC8+ codes consistent over time: To obtain product codes consistent over time, we merge by (*pc8*) using the PC8 concordances over time generated in Section 2 named "**pc8_pc8plus_bbbb_eeee_ext.dta**". Moreover, PC8+ (*pc8plus_over_time*) can only be obtained for the "mandatory" codes.

  (B) Obtain the CN8+ codes consistent over time: Similarly, we obtain product codes consistent over time, merging by (*cn8*) using the CN8 concordances over time generated in Section 1 named "**cn8_cn8plus_bbbb_eeee.dta**".

  Finally, the output file is named "**trade_prod_yyyy_concorded.dta**" where "**yyyy**" denotes the respective year. This file can be found inside the folder "**cn_pc_yearly**" (include all variables description).

3 The GRANTPA Database: Coverage and Limitations

To construct the GRANTPA (Granular Trade and Production Activities) database we execute the do-file named "**11_GRANTPA.do**", which allows us to automate the data cleaning. The input file "**trade_prod_yyyy_concorded.dta**" developed in Section 2.2 allows product comparisons between Combined Nomenclature and Prodcom codes. However, these data contain several duplicates. To avoid double-counting we implemented the following steps:

- ***Step 1: Prepare final trade data***:

  (a) Imports:

  (i) Keep only imports in the data and drop disaggregated categories of letter codes (T-,Q-,V-,E-) by *drop_synth*. Done to avoid double counting.

(ii) Combine data for internal European trade and external trade and generate import data.

(iii) Rename some variables. (i.e., *decl* as *importer*).

(iv) Trade type: Reporting Intra-EU and Extra-EU (i.e., Hungary 2013, flows reported as intra-EU and extra-EU). We identify duplicate observations based on the CN8 code. Then, we generate a new variable *int_ext* that indicates whether the trade is intra-EU or extra-EU.

(v) Take care of duplicate values. Specific cases for certain years. (i.e., 2005 *product_over_time* "714.2008"; 2009 *product_over_time* "40.2007"). Done to achieve consistency between product codes over time for trade and production. Moreover, we remove duplicate observations with the same country codes and product codes to avoid double-counting.

(vi) Compare with raw data to make sure the procedure is valid. We use the Stata command "*cfout*" to contrast the imports after cleaning with the original data extracted from Eurostat. After comparing both data sets, we achieve unique observations for all years and imports are identical to the original source.

(b) Exports: repeat the same steps for Exports

(i) Drop disaggregated categories of letter codes (T-,Q-,V-,E-). Done to avoid double counting.

(ii) Combine data for internal European trade and external trade and generate export data.

(iii) Rename some variables. (i.e., *decl* as *exporter*)

(iv) Trade type: Reporting Intra-EU and Extra-EU (i.e., Hungary 2013 Intra-EU and Extra EU flows).

(v) Take care of duplicate values to avoid double-counting.

(vi) Compare with raw data to make sure the procedure is valid. Again we achieve unique observations for all years and exports are identical to the original source.

(c) Combine exports and imports. In some cases, we have no imports, but only exports reported. In other cases, we have zero imports, but non-zero exports. We use reported exports to fill in in order to gather as much information as possible. The alternative is to leave these values missing.

(d) Construct theory-consistent trade for estimation.

(e) Construct trade with exports as the base. For the construction of domestic trade.

(f) Construct total exports. Achieved by collapsing "*trade*".

(g) Report list of exporters for the specific year.

(h) Report list of importers for the specific year.

(i) Report list of products ("*pc8_plus_over_time*") for the specific year.

(j) Combine exporters for all years.

(i) Prepare a list of reporters, i.e., countries for which there should be output.

(ii) Prepare a list of all countries. (see lines 393-469).

(iii) List of countries in DGD.

(iv) Combine and inspect country all ISO3 country codes.

(v) Combine importers for all years (from 1995 to 2019).

(vi) Combine products ("*pc8_plus_over_time*") for all years. Where 83.4% are common product codes. Main discrepancies due to synthetic codes, either the codes were introduced at some point or just disappeared.

- ***Step 2: Prepare final production data***:

  (a) Similarly, we dropped disaggregated categories of letter codes (T-,Q-,V-,E-) by *drop_synth*. Done to avoid double counting.

  (b) Take into account the codes that were recorded in the trade data.

  (c) Keep only data on the declarant level, i.e., keeping the dimensions of *epannsold.dta* and *epannsold_2021.dta* consistent.

  (d) Compare with raw data to make sure the procedure is valid. Data is identical to the original files extracted from Eurostat.

  (e) Create product codes ("GRANTPA codes") and country codes for all years.

  (f) Create final country coverage. see Table 4.

  (g) Flag alternative country samples. (all, medium, small – 'flags')

- ***Step 3: Prepare total exports and construct domestic trade***:

  (a) Construct total exports from original Europroms data. Using the input file "**trade_prod_yyyy_concorded.dta**" we are able to construct domestic trade for certain products. Similarly, we avoid double-counting by dropping duplicates using the flag *drop_synth*. Then we are able to achieve the production dimension *epannsold* and *epannsold_2021* and collapse the data in a proper way accounting for missing and zeroes.

  (b) Compare with raw data to make sure the procedure is valid. The data is identical for all years.

  (c) Construct total exports from bilateral data. see lines 827-828.

  (d) Combine two total export variables

  (e) Add production data and construct domestic trade. Domestic trade is constructed as the difference between production sold (*lev3_prod*) and exports (*trade*) see lines 835-852.

  (f) Drop unnecessary data.

- ***Step 4: Combine international and domestic trade data***:

  (a) Construct single "full-large" GRANTPA database, including all countries for which data are available.

  (b) The 'flag' variable serves as a dummy variable, where one denotes the countries and corresponding years in the database with domestic trade data, specifically the 35 European economies for which domestic trade is available.

4 Tables and Figures

Table 1: Structure of the Combined Nomenclature (CN8) Classification (Extended)

Combined Nomenclature 8-digit(CN8)		*Harmonized System 6-digit (HS6)*
Year	*# of CN8 products*	
1988	9506	
1989	9579	HS6 1988
1990	9695	(# HS6 = 5019)
1991	9743	
1992	9837	
1993	9906	HS6 1992
1994	10108	(# HS6 = 5018)
1995	10448	
1996	10495	
1997	10606	
1998	10587	HS6 1996
1999	10428	(# HS6 = 5113)
2000	10314	
2001	10274	
2002	10400	
2003	10404	HS6 2002
2004	10174	(# HS6 = 5224)
2005	10096	
2006	9841	
2007	9720	
2008	9699	HS6 2007
2009	9569	(# HS6 = 5051)
2010	9443	
2011	9294	
2012	9383	
2013	9376	HS6 2012
2014	9379	(# HS6 = 5205)
2015	9386	
2016	9414	
2017	9528	
2018	9533	HS6 2017
2019	9533	(# HS6 = 5387)
2020	9483	
2021	9494	
2022	9736	HS6 2022 (# HS6 = 5612)

Note: All classification files are obtained from Eurostat Ramon server.

Table 2: Changes in the Combined Nomenclature Classification over time: Extension

Effective year	Number of obsolete codes	Number of new codes	Number of families (including simple changes)	Number of simple (one-one) changes
1989	76	149	58	1
1990	122	238	111	11
1991	85	133	64	8
1992	128	222	85	2
1993	276	345	171	14
1994	233	435	197	11
1995	531	871	383	31
1996	1257	1304	792	435
1997	170	281	130	0
1998	334	315	175	0
1999	303	144	132	3
2000	223	109	96	0
2001	90	50	42	1
2002	847	973	504	311
2003	16	20	12	0
2004	503	273	211	7
2005	186	108	95	5
2006	743	489	281	11
2007	1202	108	630	387
2008	96	75	54	2
2009	257	127	111	0
2010	381	255	151	1
2011	282	133	124	0
2012	959	1048	637	357
2013	43	36	24	1
2014	40	43	22	2
2015	18	25	11	0
2016	27	55	18	0
2017	766	876	414	133
2018	13	18	9	0
2019	9	9	4	0
2020	104	54	42	1
2021	9	20	9	0
2022	535	769	332	135

Note: This table reports the number of obsolete and new codes each year, the number of families (shrinking, growing, or simple), and the number of simple changes (one-one). The effective year is the year in which the change becomes effective. HS6 codes have been revised in 1992, 1996, 2002, 2007, 2012, 2017 and 2022. The main changes in the combined nomenclature (CN8) classification over time are obtained from Eurostat Ramon server as shown in Van Beveren, Bernard and Vandenbussche (2012).

Table 3: Changes in the Prodcom Classification over time: Extension

Effective year	Number of obsolete codes	Number of new codes	Number of families (including simple changes)	Number of simple (one-one) changes	Number of codes that are dropped (exit)	Number of codes that are new on the list (entry)
1994	32	46	29	17	4	3
1995	33	52	15	12	19	29
1996	118	80	54	12	14	15
1997	0	0	0	0	0	0
1998	2	0	1	0	2	0
1999	68	92	31	2	3	62
2000	16	12	9	1	0	0
2001	113	76	57	0	0	0
2002	82	54	29	3	1	3
2003	363	296	215	190	1	13
2004	35	24	17	1	1	2
2005	305	105	91	0	67	1
2006	4	2	2	0	0	0
2007	184	131	76	13	3	9
2008	4396	3864	3651	3258	52	19
2009	28	15	15	1	1	1
2010	45	26	23	4	0	0
2011	61	28	28	0	0	0
2012	68	53	40	11	0	5
2013	11	8	1	0	11	8
2014	4	2	1	0	4	2
2015	9	6	1	0	9	6
2016	141	135	95	72	23	28
2017	105	43	1	0	105	43
2018	0	0	0	0	0	0
2019	79	217	1	0	79	217
2020	0	0	0	0	0	0
2021	5	12	1	0	5	12

Note: This 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. Some PC8 codes are not covered throughout the whole sample period, resulting in new codes (entry) appearing on the list and old codes (exit) disappearing from the list. All changes in the PC8 classification over time are obtained from Eurostat Ramon server. Following closely Van Beveren, Bernard and Vandenbussche (2012), optional codes have been removed (or replaced by their mandatory aggregates) to ensure comparability over time and across countries.

Table 4: GRANTPA: country coverage

country_name	country_iso	iso3	from_year	to_year	country_num	database
Afghanistan	AF	AFG	01/01/1976	31/12/2500	660	2
Albania	AL	ALB	01/01/1976	31/12/2500	70	2
Algeria	DZ	DZA	01/01/1976	31/12/2500	208	2
American Samoa	AS	ASM	01/01/2001	31/12/2500	830	2
Andorra	AD	AND	01/01/1976	31/12/2500	43	2
Angola	AO	AGO	01/01/1976	31/12/2500	330	2
Anguilla	AI	AIA	01/01/1987	31/12/2500	446	2
Antarctica	AQ	ATA	01/01/2001	31/12/2500	891	2
Antigua and Barbuda	AG	ATG	01/01/1983	31/12/2500	459	2
Argentina	AR	ARG	01/01/1976	31/12/2500	528	2
Armenia	AM	ARM	01/01/1992	31/12/2500	77	2
Aruba	AW	ABW	01/01/1987	31/12/2500	474	2
Australia	AU	AUS	01/01/1976	31/12/2500	800	2
Austria	AT	AUT	01/01/1976	31/12/2500	38	1
Azerbaijan	AZ	AZE	01/01/1992	31/12/2500	78	2
Bahamas, The	BS	BHS	01/01/1976	31/12/2500	453	2
Bahrain	BH	BHR	01/01/1976	31/12/2500	640	2
Bangladesh	BD	BGD	01/01/1976	31/12/2500	666	2
Barbados	BB	BRB	01/01/1976	31/12/2500	469	2
Belarus	BY	BLR	01/01/1992	31/12/2500	73	2
Belgium	BE	BEL	01/01/1976	31/12/2500	17	1
Belize	BZ	BLZ	01/01/1976	31/12/2500	421	2
Benin	BJ	BEN	01/01/1976	31/12/2500	284	2
Bermuda	BM	BMU	01/01/1976	31/12/2500	413	2
Bhutan	BT	BTN	01/01/1976	31/12/2500	675	2
Bolivia	BO	BOL	01/01/1976	31/12/2500	516	2
Bonaire, Sint Eustatius and Saba	BQ	BES	01/01/2013	31/12/2500	477	2
Bosnia and Herzegovina	BA	BIH	01/01/1992	31/12/2500	93	3
Botswana	BW	BWA	01/01/1976	31/12/2500	391	2
Bouvet Island	BV	BVT	01/01/2001	31/12/2500	892	2
Brazil	BR	BRA	01/01/1976	31/12/2500	508	2
British Indian Ocean Ter.	IO	IOT	01/01/1976	31/12/2500	357	2
British Virgin Islands	VG	VGB	01/01/1987	31/12/2500	468	2
Brunei	BN	BRN	01/01/1976	31/12/2500	703	2
Bulgaria	BG	BGR	01/01/1976	31/12/2500	68	1
Burkina Faso	BF	BFA	01/01/1976	31/12/2500	236	2
Burundi	BI	BDI	01/01/1976	31/12/2500	328	2
Cambodia	KH	KHM	01/01/1976	31/12/2500	696	2
Cameroon	CM	CMR	01/01/1976	31/12/2500	302	2
Canada	CA	CAN	01/01/1976	31/12/2500	404	2
Cape Verde	CV	CPV	01/01/1976	31/12/2500	247	2
Cayman Islands	KY	CYM	01/01/1976	31/12/2500	463	2
Central African Republic	CF	CAF	01/01/1976	31/12/2500	306	2
Chad	TD	TCD	01/01/1976	31/12/2500	244	2
Chile	CL	CHL	01/01/1976	31/12/2500	512	2
China	CN	CHN	01/01/1976	31/12/2500	720	2
Christmas Island	CX	CXR	01/01/2001	31/12/2500	834	2
Cocos (Keeling) Islands	CC	CCK	01/01/2001	31/12/2500	833	2
Colombia	CO	COL	01/01/1976	31/12/2500	480	2
Comoros	KM	COM	01/01/1976	31/12/2500	375	2
Congo, Democratic Republic of the	CD	COD	01/01/1976	31/12/2500	322	2
Congo, Republic of the	CG	COG	01/01/1976	31/12/2500	318	2
Cook Islands	CK	COK	01/01/2001	31/12/2500	837	2
Costa Rica	CR	CRI	01/01/1976	31/12/2500	436	2
Cote d'Ivoire	CI	CIV	01/01/1976	31/12/2500	272	2
Croatia	HR	HRV	01/01/1992	31/12/2500	92	1
Cuba	CU	CUB	01/01/1976	31/12/2500	448	2
Curacao	CW	CUW	01/01/2013	31/12/2500	475	2
Cyprus	CY	CYP	01/01/1976	31/12/2500	600	1
Czech Republic	CZ	CZE	01/01/1993	31/12/2500	61	1
Czechoslovakia	CS	CSK	01/01/1976	31/12/1992	94	2
Denmark	DK	DNK	01/01/1976	31/12/2500	8	1
Djibouti	DJ	DJI	01/01/1976	31/12/2500	338	2
Dominica	DM	DMA	01/01/1980	31/12/2500	460	2
Dominican Republic	DO	DOM	01/01/1976	31/12/2500	456	2
East Timor	TP	TLS	01/01/2001	31/12/2002	626	2
Ecuador	EC	ECU	01/01/1976	31/12/2500	500	2
Egypt, Arab Rep.	EG	EGY	01/01/1976	31/12/2500	220	2
El Salvador	SV	SLV	01/01/1976	31/12/2500	428	2
Equatorial Guinea	GQ	GNQ	01/01/1976	31/12/2500	310	2
Eritrea	ER	ERI	01/01/1994	31/12/2500	336	2
Estonia	EE	EST	01/01/1992	31/12/2500	53	1
Ethiopia (excludes Eritrea)	ET	ETH	01/01/1976	31/12/2500	334	2
Faeroe Islands	FO	FRO	01/01/1976	31/12/2500	41	2
Falkland Islands	FK	FLK	01/01/1976	31/12/2500	529	2
Fiji	FJ	FJI	01/01/1976	31/12/2500	815	2
Finland	FI	FIN	01/01/1976	31/12/2500	32	1
France	FR	FRA	01/01/1976	31/12/2500	1	1
French Guiana	GF	GUF	01/01/1976	31/12/1996	496	2
French Polynesia	PF	PYF	01/01/1976	31/12/2500	822	2
French Southern Territories	TF	ATF	01/01/2001	31/12/2500	894	2
Gabon	GA	GAB	01/01/1976	31/12/2500	314	2
Gambia, The	GM	GMB	01/01/1976	31/12/2500	252	2
Gaza Strip	XP	GAZ	01/01/1995	31/12/2000	625	2
Georgia	GE	GEO	01/01/1992	31/12/2500	76	2
Germany	DE	DEU	01/01/1976	31/12/2500	4	1
Ghana	GH	GHA	01/01/1976	31/12/2500	276	2
Gibraltar	GI	GIB	01/01/1976	31/12/2500	44	2
Greece	GR	GRC	01/01/1976	31/12/2500	9	1
Greenland	GL	GRL	01/01/1976	31/12/2500	406	2
Grenada	GD	GRD	01/01/1976	31/12/2500	473	2
Guadeloupe	GP	GLP	01/01/1976	31/12/1996	458	2
Guam	GU	GUM	01/01/2001	31/12/2500	831	2
Guatemala	GT	GTM	01/01/1976	31/12/2500	416	2
Guinea	GN	GIN	01/01/1976	31/12/2500	260	2

Continued on next page ...

Table 4: GRANTPA: country coverage

country_name	country_iso	iso3	from_year	to_year	country_num	database
Guinea-Bissau	GW	GNB	01/01/1976	31/12/2500	257	2
Guyana	GY	GUY	01/01/1976	31/12/2500	488	2
Haiti	HT	HTI	01/01/1976	31/12/2500	452	2
Heard Island and McDonald Islands	HM	HMD	01/01/2001	31/12/2500	835	2
Honduras	HN	HND	01/01/1976	31/12/2500	424	2
Hong Kong	HK	HKG	01/01/1976	31/12/2500	740	2
Hungary	HU	HUN	01/01/1976	31/12/2500	64	1
Iceland	IS	ISL	01/01/1976	31/12/2500	24	3
India	IN	IND	01/01/1976	31/12/2500	664	2
Indonesia	ID	IDN	01/01/1976	31/12/2500	700	2
Iran	IR	IRN	01/01/1976	31/12/2500	616	2
Iraq	IQ	IRQ	01/01/1976	31/12/2500	612	2
Ireland	IE	IRL	01/01/1976	31/12/2500	7	1
Israel	IL	ISR	01/01/1976	31/12/2500	624	2
Italy	IT	ITA	01/01/1976	31/12/2500	5	1
Jamaica	JM	JAM	01/01/1976	31/12/2500	464	2
Japan	JP	JPN	01/01/1976	31/12/2500	732	2
Jordan	JO	JOR	01/01/1976	31/12/2500	628	2
Kazakhstan	KZ	KAZ	01/01/1992	31/12/2500	79	2
Kenya	KE	KEN	01/01/1976	31/12/2500	346	2
Kiribati	KI	KIR	01/01/1976	31/12/2500	812	2
Korea, North	KP	PRK	01/01/1976	31/12/2500	724	2
Korea, South	KR	KOR	01/01/1976	31/12/2500	728	2
Kosovo	XK	KSV	01/01/2005	31/12/2500	95	2
Kuwait	KW	KWT	01/01/1976	31/12/2500	636	2
Kyrgyzstan	KG	KGZ	01/01/1992	31/12/2500	83	2
Laos	LA	LAO	01/01/1976	31/12/2500	684	2
Latvia	LV	LVA	01/01/1992	31/12/2500	54	1
Lebanon	LB	LBN	01/01/1976	31/12/2500	604	2
Lesotho	LS	LSO	01/01/1976	31/12/2500	395	2
Liberia	LR	LBR	01/01/1976	31/12/2500	268	2
Libya	LY	LBY	01/01/1976	31/12/2500	216	2
Liechtenstein	LI	LIE	01/01/1995	31/12/2500	37	2
Lithuania	LT	LTU	01/01/1992	31/12/2500	55	1
Luxembourg	LU	LUX	01/01/1999	31/12/2500	18	1
Macao	MO	MAC	01/01/1976	31/12/2500	743	2
Macedonia	MK	MKD	01/01/1993	31/12/2500	96	3
Madagascar	MG	MDG	01/01/1976	31/12/2500	370	2
Malawi	MW	MWI	01/01/1976	31/12/2500	386	2
Malaysia	MY	MYS	01/01/1976	31/12/2500	701	2
Maldives	MV	MDV	01/01/1976	31/12/2500	667	2
Mali	ML	MLI	01/01/1976	31/12/2500	232	2
Malta	MT	MLT	01/01/1976	31/12/2500	46	1
Marshall Islands	MH	MHL	01/01/1992	31/12/2500	824	2
Martinique	MQ	MTQ	01/01/1976	31/12/1996	462	2
Mauritania	MR	MRT	01/01/1976	31/12/2500	228	2
Mauritius	MU	MUS	01/01/1976	31/12/2500	373	2
Mayotte	YT	MYT	01/01/1977	31/12/2013	377	2
Mexico	MX	MEX	01/01/1976	31/12/2500	412	2
Micronesia, Federated States of	FM	FSM	01/01/1992	31/12/2500	823	2
Moldova	MD	MDA	01/01/1992	31/12/2500	74	2
Mongolia	MN	MNG	01/01/1976	31/12/2500	716	2
Montenegro	ME	MNE	01/01/2007	31/12/2500	97	3
Montserrat	MS	MSR	01/01/1995	31/12/2500	470	2
Morocco	MA	MAR	01/01/1976	31/12/2500	204	2
Mozambique	MZ	MOZ	01/01/1976	31/12/2500	366	2
Myanmar	MM	MMR	01/01/1976	31/12/2500	676	2
Namibia	NA	NAM	01/01/1990	31/12/2500	389	2
Nauru	NR	NRU	01/01/1976	31/12/2500	803	2
Nepal	NP	NPL	01/01/1976	31/12/2500	672	2
Netherlands	NL	NLD	01/01/1976	31/12/2500	3	1
Netherlands Antilles	AN	ANT	01/01/1976	31/12/2012	478	2
New Caledonia	NC	NCL	01/01/1976	31/12/2500	809	2
New Zealand	NZ	NZL	01/01/1976	31/12/2500	804	2
Nicaragua	NI	NIC	01/01/1976	31/12/2500	432	2
Niger	NE	NER	01/01/1976	31/12/2500	240	2
Nigeria	NG	NGA	01/01/1976	31/12/2500	288	2
Niue	NU	NIU	01/01/2001	31/12/2500	838	2
Norfolk Island	NF	NFK	01/01/2001	31/12/2500	836	2
Northern Marianas	MP	MNP	01/01/1994	31/12/2500	820	2
Norway	NO	NOR	01/01/1976	31/12/2500	28	3
Oman	OM	OMN	01/01/1976	31/12/2500	649	2
Pakistan	PK	PAK	01/01/1976	31/12/2500	662	2
Palau	PW	PLW	01/01/1995	31/12/2500	825	2
Palestine	PS	PSE	01/01/2001	31/12/2500	625	2
Panama	PA	PAN	01/01/1976	31/12/2500	442	2
Papua New Guinea	PG	PNG	01/01/1976	31/12/2500	801	2
Paraguay	PY	PRY	01/01/1976	31/12/2500	520	2
Peru	PE	PER	01/01/1976	31/12/2500	504	2
Philippines	PH	PHL	01/01/1976	31/12/2500	708	2
Pitcairn	PN	PCN	01/01/1981	31/12/2500	813	2
Poland	PL	POL	01/01/1976	31/12/2500	60	1
Portugal	PT	PRT	01/01/1976	31/12/2500	10	1
Qatar	QA	QAT	01/01/1976	31/12/2500	644	2
Reunion	RE	REU	01/01/1976	31/12/1996	372	2
Romania	RO	ROU	01/01/1976	31/12/2500	66	1
Russia	RU	RUS	01/01/1992	31/12/2500	75	2
Rwanda	RW	RWA	01/01/1976	31/12/2500	324	2
Saint Barthelemy	BL	BLM	01/01/2013	31/12/2500	466	2
Saint Helena, Ascension, and Tristan da Cunha	SH	SHN	01/01/1976	31/12/2500	329	2
Saint Kitts and Nevis	KN	KNA	01/01/1986	31/12/2500	449	2
Saint Lucia	LC	LCA	01/01/1980	31/12/2500	465	2
Saint Pierre and Miquelon	PM	SPM	01/01/1976	31/12/2500	408	2
Saint Vincent and the Grenadines	VC	VCT	01/01/1980	31/12/2500	467	2
Samoa	WS	WSM	01/01/1976	31/12/2500	819	2

Continued on next page ...

Table 4: GRANTPA: country coverage

country_name	country_iso	iso3	from_year	to_year	country_num	database
San Marino	SM	SMR	01/01/1994	31/12/2500	47	2
Sao Tome and Principe	ST	STP	01/01/1976	31/12/2500	311	2
Saudi Arabia	SA	SAU	01/01/1976	31/12/2500	632	2
Senegal	SN	SEN	01/01/1976	31/12/2500	248	2
Serbia	XS	SRB	01/01/2005	31/12/2500	98	3
Serbia and Montenegro	CS	SCG	01/01/2004	31/12/2005	94	2
Seychelles	SC	SYC	01/01/1976	31/12/2500	355	2
Sierra Leone	SL	SLE	01/01/1976	31/12/2500	264	2
Singapore	SG	SGP	01/01/1976	31/12/2500	706	2
Sint Maarten	SX	SXM	01/01/2013	31/12/2500	479	2
Slovakia	SK	SVK	01/01/1993	31/12/2500	63	1
Slovenia	SI	SVN	01/01/1992	31/12/2500	91	1
Solomon Islands	SB	SLB	01/01/1980	31/12/2500	806	2
Somalia	SO	SOM	01/01/1976	31/12/2500	342	2
South Africa	ZA	ZAF	01/01/1976	31/12/2500	388	2
South Georgia and South Sandwich Islands	GS	SGS	01/01/2001	31/12/2500	893	2
South Sudan	SS	SSD	01/01/2013	31/12/2500	225	2
Spain	ES	ESP	01/01/1976	31/12/2500	11	1
Sri Lanka	LK	LKA	01/01/1976	31/12/2500	669	2
Sudan	SD	SDN	01/01/1976	31/12/2500	224	2
Suriname	SR	SUR	01/01/1976	31/12/2500	492	2
Svalbard and Jan Mayen Islands	SJ	SJM	01/01/1995	31/12/1996	27	2
Sweden	SE	SWE	01/01/1976	31/12/2500	30	1
Switzerland	CH	CHE	01/01/1976	31/12/2500	39	2
Syria	SY	SYR	01/01/1976	31/12/2500	608	2
Taiwan	TW	TWN	01/01/1976	31/12/2500	736	2
Tajikistan	TJ	TJK	01/01/1992	31/12/2500	82	2
Tanzania	TZ	TZA	01/01/1976	31/12/2500	352	2
Thailand	TH	THA	01/01/1976	31/12/2500	680	2
Togo	TG	TGO	01/01/1976	31/12/2500	280	2
Tokelau	TK	TKL	01/01/2001	31/12/2500	839	2
Tonga	TO	TON	01/01/1976	31/12/2500	817	2
Trinidad and Tobago	TT	TTO	01/01/1976	31/12/2500	472	2
Tunisia	TN	TUN	01/01/1976	31/12/2500	212	2
Turkey	TR	TUR	01/01/1976	31/12/2500	52	3
Turkmenistan	TM	TKM	01/01/1992	31/12/2500	80	2
Turks and Caicos Islands	TC	TCA	01/01/1976	31/12/2500	454	2
Tuvalu	TV	TUV	01/01/1980	31/12/2500	807	2
U.S. Minor Outlying Islands	UM	UMI	01/01/2001	31/12/2500	832	2
U.S. Virgin Islands	VI	VIR	01/01/1976	31/12/2500	457	2
Uganda	UG	UGA	01/01/1976	31/12/2500	350	2
Ukraine	UA	UKR	01/01/1992	31/12/2500	72	2
United Arab Emirates	AE	ARE	01/01/1976	31/12/2500	647	2
United Kingdom	GB	GBR	01/01/1976	31/12/2500	6	1
United States	US	USA	01/01/1976	31/12/2500	400	2
Uruguay	UY	URY	01/01/1976	31/12/2500	524	2
Uzbekistan	UZ	UZB	01/01/1992	31/12/2500	81	2
Vanuatu	VU	VUT	01/01/1976	31/12/2500	816	2
Vatican City	VA	VAT	01/01/1976	31/12/2500	45	2
Venezuela	VE	VEN	01/01/1976	31/12/2500	484	2
Vietnam	VN	VNM	01/01/1976	31/12/2500	690	2
Wallis and Futuna Islands	WF	WLF	01/01/1976	31/12/2500	811	2
Western Sahara	EH	ESH	01/01/2013	31/12/2500	229	2
Yemen	YE	YEM	01/01/1976	31/12/2500	653	2
Yugoslavia	YU	YUG	01/01/1976	31/12/2003	94	2
Zambia	ZM	ZMB	01/01/1976	31/12/2500	378	2
Zimbabwe	ZW	ZWE	01/01/1976	31/12/2500	382	2

Figure 1: Growing Family Tree

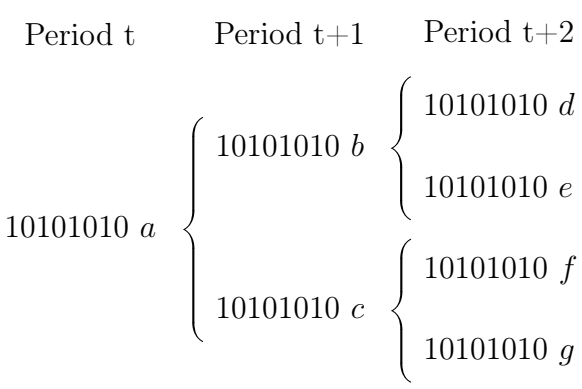
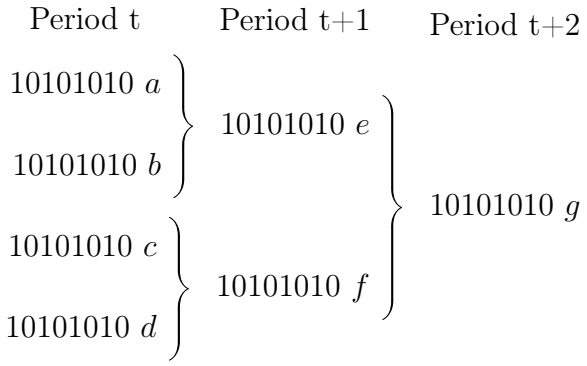


Figure 2: Shrinking Family Tree



References

Pierce, Justin R., and Peter K. Schott. 2012*a*. "A Concordance Between Ten-digit US Harmonized System Codes and SIC/NAICS Product Classes and Industries." *Journal of Economic and Social Measurement*, 37(1-2): 61–96.

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

Van Beveren, Ilke, Andrew B. Bernard, and Hylke Vandenbussche. 2012. "Concording EU Trade and Production Data over Time." *National Bureau of Economic Research*, Working Paper (18604).

IMPRESSUM

Herausgeber, Verleger, Eigentümer und Hersteller:
Verein „Wiener Institut für Internationale Wirtschaftsvergleiche“ (wiiw),
Wien 6, Rahlgasse 3

ZVR-Zahl: 329995655

Postanschrift: A 1060 Wien, Rahlgasse 3, Tel: [+431] 533 66 10, Telefax: [+431] 533 66 10 50
Internet Homepage: www.wiiw.ac.at

Offenlegung nach § 25 Mediengesetz: Medieninhaber (Verleger): Verein "Wiener Institut für Internationale Wirtschaftsvergleiche", A 1060 Wien, Rahlgasse 3. Vereinszweck: Analyse der wirtschaftlichen Entwicklung der zentral- und osteuropäischen Länder sowie anderer Transformationswirtschaften sowohl mittels empirischer als auch theoretischer Studien und ihre Veröffentlichung; Erbringung von Beratungsleistungen für Regierungs- und Verwaltungsstellen, Firmen und Institutionen.

wiiw

wiiw.ac.at

https://wiiw.ac.at/p-6912.html