Correspondences of EU Product Classifications *

Glenn Magerman
ECARES, ULB and CEPR

Version: March 2022

Abstract

I provide clean correspondences of 8-digit products in the EU Combined Nomenclature (CN) and Prodcom (PC) classifications. These mappings trace (i) products within classifications over time, and (ii) products across CN and PC. Classifications tend to vary from year to year, for several non-economic reasons. Incorrectly accounting for changes in these classifications leads to spurious entry and exit of products, price biases, and incorrect price and quantity indices. I characterize all singular and non-singular mappings, and all quantity units of measurement. Combined with product-level datasets, these mappings allow to calculate plausible unit values in CN and PC, changes in unit values, indices and e.g. domestic prices for internationally traded goods. I share all codes and data for others to use through Github.

*Email: glenn.magerman@ulb.be. This research was funded by the 2018 Colloquium of the National Bank of Belgium on "Understanding inflation dynamics: the role of costs, mark-ups and expectations", and the Action de Recherche Concertées (ARC) grant number 20204850. I would like to thank Danny Delcambre at Eurostat and Ilke Van Beveren for their help with the product classification files and provision of additional information that was not readily available on the RAMON website.

1. Introduction

Products that are produced in and/or traded with the European Union are categorized into the Prodcom (PC) classification for industrial production and the Combined Nomenclature (CN) system for internationally traded goods. Eurostat uses these classifications to provide detailed and high frequency official statistics on production and trade that are comparable across EU member states, and they serve as policy inputs (e.g. to identify the customs tariffs that apply on particular products with third countries). In economic analysis, they are often used to identify multi-product firms and their product mix, the dispersion and evolution of product margins at the firm level, prices and price indices, tariff analysis, and quantity-based TFP estimation.

Unfortunately for the economist, these classifications require quite some pre-processing before they can be used in product-level datasets. First, there is a variety of international, European and national classifications that are hierarchically linked. Figure 1 illustrates how these classifications relate to each other. Some classifications are unambiguously linked by

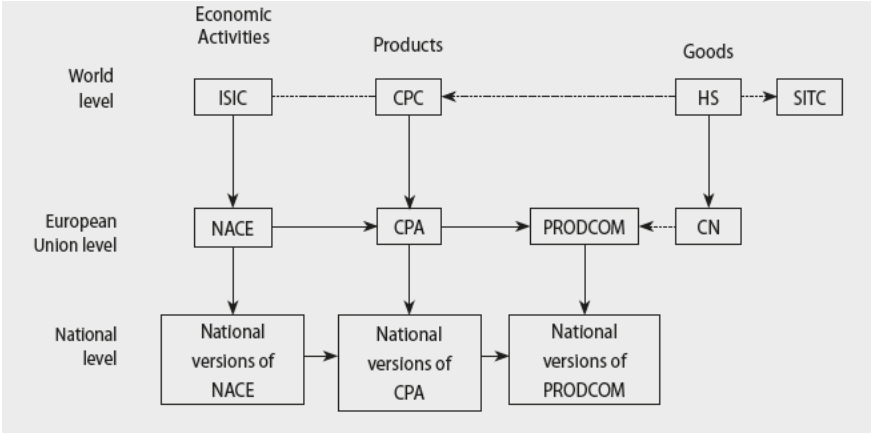

Figure 1: Relationship between international, European and national classification systems.

structure (solid arrows), while others are linked through conversion tables (dotted arrows). This implies that identical products can be coded differently in different classification systems (e.g. CN versus PC), and that similar products might be coded identical in different classifications (e.g. due to varying levels of granularity in each system). Second, product codes tend to change over time, and for reasons that are not economic in nature, such as changes in parent classifications or statistical reporting requirements.[1] These changes are prevalent,

[1]For example, the 8-digit Combined Nomenclature is the EU disaggregation of the global 6-digit Harmonized

and occur yearly in each classification system. To complicate matters, changes can be one-to-one, one-to-many, many-to-one, or many-to-many, which I generically label as $m : n$ matches. Also across classifications, correspondences can be $m : n$, as both the granularity of classifications and product scope are generally not the same.[2] Third, the PC classification contains various types of letter codes, which are either aggregate placeholders for statistical reporting purposes, or optional disaggregate codes that are only active in some member states. In many product-level datasets however, only 8-digit numeric codes are reported. These letter codes can also change over time, and again in a complex $m : n$ fashion.

Correctly dealing with these issues is imperative in empirical work. Identifying statistical changes over time avoids spurious entry/exit of products in the firm's product mix (Van Beveren et al. (2012)) and allows researchers to expand their time series dimension of analysis across structural breaks (e.g. from NACE Rev 1.1 to NACE Rev 2 in 2008, when effectively *all* PC codes changed). In the cross-section, linking classifications provides information on production vis-a-vis trade patterns of manufactured goods, e.g. carry-along trade (Bernard et al. (2019)). In the context of our own work (Duprez & Magerman (2022)), when calculating price changes within firm-products over time, I need to ensure that (i) we identify changes in product codes for the same underlying products from $t-1$ to t, and (ii) products in both years are reported in the same quantity units, even if product codes change in a $m : n$ way. Similarly, when calculating (changes in) domestic prices, I identify identical products in both CN and PC classifications to account for pricing to market in foreign destinations, and ensure all identified products are in the same unit of measurement.

In this document, I describe these correspondences. I provide (i) clean datasets on detailed 8-digit products in both trade (CN) and production (PC) for the years 2001-2014, (ii) ready to use correspondence tables for year-on-year changes within classifications, and yearly correspondences from CN to PC, and (iii) descriptive statistics for these classifications and their concordances. I characterize all $m : n$ correspondences and retain information on the quantity units of measurement for further use. With these datasets and code, I hope to lower the

System (HS6) classification for internationally traded goods. Any change in the HS6 system, administrated by the World Customs Organization and updated each 5-6 years, has to materialize in the European CN classification, even if the underlying product scope did not change. Similarly, the 8-digit Prodcom classification is an extension of particular Sections in the 4-digit EU NACE classification, broadly covering mining and quarrying, manufactured goods and some industrial services. Large overhauls of the NACE classification also disrupt the PC classification, such as the structural break from NACE Rev 1.1. to NACE Rev 2, when all product codes changed from 2007 to 2008. Finally, product classes become more or less disaggregated, mirroring their prominence and statistical reporting relevance, and products also enter and exit, reflecting technological innovation and obsolescence or changes in statistical coverage of the classification system.

[2]The CN classification is generally more disaggregated than the PC classification, and consists of roughly 9,500 products in a given year, versus around 4,000 products in the PC system. Moreover, while the CN classification covers only traded goods, including antiques, the PC classification covers only manufactured production and some industrial services, excluding antiques.

barrier for other researchers using product-level data on production and/or trade at the EU level or its EU memberstates.

Our method differs from similar procedures (Pierce & Schott (2012a,b); Van Beveren et al. (2012); Amiti et al. (2019)) in two ways. First, I do not harmonize products over the full sample period. Some procedures collapse $m : n$ matches into more aggregated synthetic groupings ("family trees") that are constant over time. The advantage is that this delivers a panel dataset of products that can be tracked over the whole sample period. The identity of these family trees however, depends on the exact time span used in the data. The level of necessary aggregation also increases in the length of the panel, as each $m : n$ match has to be collapsed over the full panel. Our procedure tracks instead all $m : n$ matches from year to year, which makes a panel analysis elusive, and I instead resort to a year-on-year analysis in Duprez & Magerman (2022). Nevertheless, researchers can still construct family trees by just collapsing all our non-singular correspondences over time. Second, I retain the quantity units of measurement to avoid potential price bias from differences in reported quantities. All CN products are reported in weight (kilograms), and 28% of products are also reported in supplementary units (e.g. liters, square meters, per piece,...). Conversely, PC products are reported in one of 43 different units of measurement. Ignoring units of measurement creates wrong unit values or changes therein, as well as incorrect aggregation to price indices.[3]

[3]For example, using information on values and quantities, in Duprez & Magerman (2022), we calculate price changes for product i as $d \ln p_{it} = \ln\left(\frac{v_{it}}{q_{it}}\right) - \ln\left(\frac{\sum_j v_{jt-1}}{\sum_j q_{jt-1}}\right)$ where v and q are values and quantities, and j refers to the set of m products in $t-1$ that map to each i of n products in t. All quantities in both $t-1$ and t, and the sets of m and n products, have to be reported in the same units to obtain sensible price changes and to avoid price bias from quantities. Similar reasoning holds for $m : n$ matches across classifications within years.

2. Data sources

I draw from Eurostat's RAMON database, which documents the various EU statistical classifications and nomenclatures.[4] I extract information on the yearly CN and PC product lists, the yearly CN to PC convergence tables, and the year-on-year correspondences of both CN and PC for the years 2001 to 2014. I complete these tables with additional information directly obtained from Eurostat.

3. Combined Nomenclature (CN)

Information on international trade in goods in the EU is collected through *Intrastat* (for flows from one member state to another) and *Extrastat* (between member states and third countries).[5] In both cases, products are reported in the Combined Nomenclature (CN) classification, which is the 8-digit EU extension of the international 6-digit Harmonized System (HS), developed by the World Customs Organization (WCO).[6] It is used to classify all goods (services are excluded) when declared to customs in the EU, it determines the rate of common customs tariffs that apply, and how the goods are treated for statistical purposes or other policies.

Trade flows in Intrastat and Extrastat contain product-level information on values and quantities of enterprise-level shipments. Values are reported in euros, and quantities in net mass (weight in kilograms). Some products also are reported in a supplementary unit of measurement. Reporting is subject to some thresholds.[7] While the enterprise-level data is confidential, Eurostat provides detailed data on monthly and yearly trade flows at 8-digit CN codes, values and quantities for all EU members and their worldwide partners through their bulk download facility.[8]

[4] https://ec.europa.eu/eurostat/ramon.

[5] For additional information, see the compilers guide on European statistics on international trade in goods, and a description of the EU trade statistics.

[6] Countries that adopt the HS system can further disaggregate the classification for their own statistical purposes. For example, the USA uses HS10, an extension of the HS6 system to 10 digits. The EU uses 10-digit Integrated Tariff (TARIC), a two-digit extension of the CN classification, which provides information on all trade policy and tariff measures applicable to EU imports from third countries (e.g. temporary suspension of duties, antidumping duties, etc), see here for more info on TARIC.

[7] The Intrastat system replaced customs declarations on Jan 1, 1993 as the EU internal market opened up. All VAT-liable entities above certain thresholds submit their Intrastat declarations monthly to their respective national offices. Thresholds are set by individual member states so that reported trade covers at least 97% of total dispatch value (intra-EU exports) and 93% of total arrival value (intra-EU imports). These thresholds can vary across member states, across arrivals and dispatches and over time. See here for the latest list of reporting thresholds by member state. Extrastat information is reported daily through customs declarations when crossing a state border. Thresholds are at 1,000 euro in value or 1,000 kg in mass, and have been constant over time. Member states then provide monthly reports to Eurostat to be used for further statistics.

[8] Available here.

I extract the yearly lists of CN8 product codes and their year-on-year correspondences, together with a verbal description of the products and optional supplementary units. Detailed information on the resulting datasets is provided in Appendix A.

To provide a sense of the level of detail of the classification, consider an example in Table 1: HS 2-digit **Chapter 18** (*Cocoa and Cocoa Preparations*) contains the 4-digit **Heading 1806** (*Chocolate and other food preparations containing cocoa*). Within that heading, the **HS6 code 1806.10** (*Cocoa powder, containing added sugar or sweetening matter*) can be further disaggregated into the **CN8 product 1806.10.15** (*Containing no sucrose or containing less than 5 % by weight of sucrose (including invert sugar expressed as sucrose) or isoglucose expressed as sucrose*). This product is close to, but different from **CN8 product 1806.10.20** (*Containing 5 % or more but less than 65 % by weight of sucrose (including invert sugar expressed as sucrose) or isoglucose expressed as sucrose*).

Table 1: CN8 hierarchy example.

Level	Code	Description
HS Chapter	18	Cocoa and Cocoa Preparations
HS Heading	1806	Chocolate and other food preparations containing cocoa
HS Subheading	1806.10	Cocoa powder, containing added sugar or sweetening matter
CN8 Code	1806.10.15	Containing no sucrose or containing less than 5 % by weight of sucrose (including invert sugar expressed as sucrose) or isoglucose expressed as sucrose
	1806.10.20	Containing 5 % or more but less than 65 % by weight of sucrose (including invert sugar expressed as sucrose) or isoglucose expressed as sucrose

Table 2 describes the CN 8-digit products by year, and their year-on-year changes. Columns 1-2 report the number of products by year. In any given year, there are around 9,500 products. Column 3 reports the number of unchanged product codes, which represent the vast majority of codes: across all years, 94% of all codes remains unchanged from one year to another. The remaining columns then describe the occurrence and nature of $m : n$ matches over time. Most changes occur in years when there are changes in the parent HS system (in our sample in 2002, 2007 and 2012) that have to be adapted in the CN classification. The remaining changes stem from smaller changes in the CN classification for statistical, technological or policy reasons.

Changes are decomposed into singular (one-to-one) and non-singular (one-to-many, many-to-one and many-to-many) matches. One-to-one mappings are reclassifications in the taxon-

Table 2: CN8 products and their correspondences.

year	products	unchanged	singular	non-singular	one-to-many		many-to-one		many-to-many	
					$t-1$	t	$t-1$	t	$t-1$	t
2001	10,274	–	–	–	–	–	–	–	–	–
2002	10,400	9,427	311	662	100	240	106	49	330	373
2003	10,404	10,384	0	20	8	16	8	4	0	0
2004	10,174	9,901	7	266	6	14	378	168	112	84
2005	10,096	9,988	5	103	13	26	168	77	0	0
2006	9,842	9,353	11	478	9	21	341	170	382	287
2007	9,720	8,640	387	693	62	154	292	125	461	414
2008	9,699	9,624	2	73	11	24	71	35	12	14
2009	9,569	9,442	0	127	6	15	237	100	14	12
2010	9,443	9,188	1	254	17	40	279	121	84	93
2011	9,294	9,161	0	133	2	4	262	116	18	13
2012	9,383	8,335	357	691	111	279	276	122	215	290
2013	9,376	9,340	1	35	5	14	32	16	5	5
2014	9,379	9,336	2	41	9	22	17	8	12	11

Notes: Column 2 reports the number of CN8 product codes in a given year. Column 3 reports the number of codes that remain unchanged from the previous year to the current. Column 4 reports singular mappings (one-to-one), and column 5 reports non-singular correspondences. The remaining columns decomposes the nature of the non-singular mappings (one-to-many, many-to-one and many-to-many). Each of these is split into the number of codes in $t-1$ and the resulting number of codes in t.

omy. The verbal descriptions of the CN8 codes are either identical for old and new codes, or reflect a change in hierarchy of the taxonomy while describing the same product. For example: **CN8 code 8112.30.20** in 2006 (8112.30 *"Germanium"*; 8112.30.20 *"Unwrought; powders"*) changed to **CN code 8112.92.95** in 2007 *(*8112.92*: "Other"; *8112.92.95*: "Germanium"*).

Non-singular correspondences are also pervasive. For example, **CN8 code 5702.52.00** in 2004 *(*5702*: Carpets and other textile floor coverings, woven, not tufted or flocked, whether or not made-up, including 'Kelem', 'Schumacks', 'Karamanie' and similar hand-woven rugs;* 5702.52.00*: Of man-made textile materials)* split into **CN8 code 8112.53.90** (*"Other"*) and **CN8 code 5792.52.10** (*"Of polypropylene"*) in 2005. For one-to-many correspondences, the median change is one code that splits into two codes from $t-1$ to t, with a maximum of 16 new codes for one old code. Similar examples accrue for many-to-one links. Many-to-many correspondences are the remaining non-singular correspondences. For example, from 2003 to 2004, **CN8 code 0205.00.11** ("*Meat of horses, fresh or chilled")* and **CN8 code 0205.00.90** ("*Meat of asses, mules or hinnies, fresh, chilled or frozen"*) transformed into **code 0205.00.20** (0205.00: *"Meat of horses, asses, mules or hinnies, fresh, chilled or frozen*"; 0205.00.20: "*fresh or chilled*") and **code 0205.00.80** ("*frozen*").

I also retain the units of measurements for the CN8 codes by year. Table 3 shows the distribution of the units of measurement observed in the data. Over the period 2001-2014, 28% of CN8 products are reported in a supplementary unit, and there are 30 additional units, such as *per piece*, *square meters*, *liters*, but also *grams*, *cubic meters*, *per 1000 pieces* etc. For chemicals, supplementary units are mostly expressed in *kilograms of active substance*.

4. Prodcom (PC)

The Prodcom ("Production Communautaire") classification provides information on industrial production in the EU plus EFTA countries since 1993.[9] It broadly covers sections B and C of NACE Rev. 2: activities in mining, quarrying and manufacturing.[10] It refers to manufactured products, but also some industrial services are included, such as treatment, repairs and maintenance, and assembly work. Some activities such as waste management and some agricultural products are not covered. Products are identified by an 8-digit code, of which the first 4 digits are the same as the NACE 4-digit classification of economic activities, and the first 6 digits correspond to the CPA, the European Classification of Products by Activity. It is used to measure firm-level production and sales of industrial products, and to compare pro-

[9] For additional information, see Eurostat Prom, the Prodcom user guide, and additional explanatory notes.
[10] These are NACE Sections C, D and E in NACE Rev 1.1.

Table 3: Supplementary units reported in CN8 (2001-2014).

suppl. unit	obs.	suppl. unit	obs.	suppl. unit	obs.
none	91,586	*g*	682	*kg met.am.*	13
100 p/st	26	*gi F/S*	91	*kg/net eda*	71
1000 kWh	13	*kg 90% sdt*	215	*l*	1,621
1000 l	65	*kg C5H14C1NO*	10	*l alc. 100%*	728
1000 m³	13	*kg H2O2*	13	*m*	201
1000 p/st	139	*kg K2O*	70	*m²*	6,035
GT	45	*kg KOH*	13	*m³*	1,367
TJ	26	*kg N*	161	*p/st*	21,895
c/k	65	*kg NaOH*	13	*pa*	1,333
ce/el	90	*kg P2O5*	70		
ct/l	48	*kg U*	52		

Notes: Description of units of measurement. *p/st*: per piece, *l:* litres, m^3: cubic metres, *kWh*: kilo-watt hour, GT: gross tonnage, *TJ*: terajoule (gross calorific value), *c/k*: carat, *ce/el*: number of cells, *ct/l:* carrying capacity in tonnes, *g*: gram, *gi F/S*: gram of fissile isotopes, *kg 90% sdt*: kilogram of substance 90 % dry, *kg met.am.*: kilogram of methylamines, *kg/net eda*: kilogram drained net weight, *l alc. 100%*: litre pure (100 %) alcohol, *pa*: number of pairs.

duction with trade of goods. Derived statistics are used as economic indicators for economic policy, to monitor trends, and as inputs for national accounts of EU member states.

Enterprises that have industrial activities submit monthly to the national *Prodcom Survey*, and report values and quantities at 8-digit product levels in the PC classification, sold or produced during the reference period. Values are reported in euros, and quantities are reported in one of several units of measurement (e.g. kilograms, square meters, pairs). Reporting is subject to some thresholds.[11] While enterprise-level data is confidential, Eurostat provides yearly statistics on 8-digit PC products, their valued and quantities for each member state.[12]

I extract the yearly lists of PC8 product codes and their year-on-year correspondences, together with a verbal description of the products and their quantity units of measurement. Detailed information on the resulting datasets is provided in Appendix A.

[11] The survey is a cut-off sample, and contains all enterprises, authorities and organizations that carry out any target industrial activity (Sections B and C of NACE Rev. 2) with at least 20 employees. At least 90% of production in each (four-digit) class of NACE Rev. 2 must be recorded.

[12] Available here.

The main complexity in the PC classification is the existence of letter codes. These letter codes are used for statistical purposes, and are optional disaggregated or aggregated codes. In most product-level datasets however, PC codes are reported as 8-digit numeric codes. The following letter codes exist over our panel period: B, N, Z, E, V, Q, and T. B-list and N-list codes are optional disaggregated codes, requested respectively by individual member states or by Eurostat, and allow for more disaggregated classifications than the mandatory PC8 codes. These phased out and are eliminated by 2005. I provide all correspondences for the mandatory codes and mappings from B- and N-list codes to their mandatory counterparts in separate datasets. Users that have optional codes in product-level datasets can then first aggregate to the mandatory codes using these mappings, and subsequently run all correspondences.

All other letter codes are aggregates of PC8 codes. Z, E and V codes are aggregates reflecting the same level of aggregation as their CN counterpart codes, which are further disaggregated in PC for statistical purposes. In order to maintain the link to CN for such codes, special aggregates are created, while the underlying disaggregate codes are also reported. These include Z and E codes, which respectively have a "Z" or an "E" in the 7th digit of the PC code. V aggregates have a "V" in the 7th position of the code, and are found in NACE 27 for the collection of iron and steel statistics, and are grouped together for statistical reasons. Q codes (with a "Q" in the 7th position) refer to aggregates of certain textile codes that have been reported quarterly. Q, E and V codes all phased out by 2005. Finally, T codes are also headers of PC8 codes, to map more closely to their CN8 counterparts, and are mostly used in the correspondences between CN and PC codes. I disaggregate all these letter codes to their 8-digit numerical codes. I then concord the PC8 codes over time using the year-to-year correspondence tables for PC at RAMON. Correspondences can be singular or non-singular.

To illustrate the granularity of the classification, consider the following example in Table 4: **NACE Rev. 2 code 24.20** (*Manufacture of tubes, pipes, hollow profiles and related fittings, of steel*) contains **CPA code 24.20.12** (*Casing, tubing and drill pipe, of a kind used in the drilling for oil or gas, seamless, of steel*). Within that code, **PC products 24.20.12.10** (*..., of stainless steel*), and **24.20.12.50** (*..., of steel other than stainless steel*) are distinct.

Changes in PC8 codes over time can come from three sources: (i) changes in the parent NACE classification, (ii) changes in the products covered by the Prodcom Survey, and (iii) mappings from (dis-)aggregations in the statistical classification of Prodcom. First, over our sample period, there have been changes in the NACE classification from NACE Rev. 1 (1990) to NACE Rev. 1.1 (2002) and to NACE Rev. 2 (2008).[13] While the change from Rev. 1 to

[13]See the Eurostat publication on NACE here.

Table 4: PC8 hierarchy example.

Level	Code	Description
NACE Rev. 2	24.20	Manufacture of tubes, pipes, hollow profiles and related fittings, of steel
CPA Code	24.20.12	Casing, tubing and drill pipe, of a kind used in the drilling for oil or gas, seamless, of steel
PC8 Code	24.20.12.10	Casing, tubing and drill pipe, of a kind used in the drilling for oil or gas, seamless, of stainless steel
	24.20.12.50	Casing, tubing and drill pipe, of a kind used in the drilling for oil or gas, seamless, of steel other than stainless steel

Rev. 1.1 was minor, there has been a major overhaul towards NACE Rev. 2. Several changes occurred at the highest levels of the classification, including reclassifications and completely new sections, creating a large structural break in the time series as all product codes changed from 2007 to 2008. Second, the choice of products covered in the PC lists generally depends on their economic importance, and these lower-tier changes can occur independent of changes in the NACE classification. Third, singular and non-singular correspondences arrive over time, as in the CN classification.

Table 5 reports the number of PC8 numeric codes per year and all correspondences. Columns 1-2 report the number of PC 8-digit numeric codes per year. There are roughly 4,000 identified PC8 codes in a given year, while this number has been declining over time. Column 3 reports the number of unchanged product codes. The vast majority of codes remains unchanged from one year to the next, except in 2008, when all codes change due to the large overhaul of the NACE classification.

Columns 4-5 report entry and exit of PC8 codes, where entry is defined as non-existing in $t-1$ but existing in t, and exit is existing in $t-1$ but no longer in t. For example, in 2012, **PC8 code 1020.42.50** (*"Fish heads, tails and maws, other edible fish offal: dried, salted or in brine, smoked"*) enters the classification. Conversely, from 2004 to 2005, **PC code 1710.42.31** (*"Yarn of carded wool or fine animal hair, n.p.r.s., for carpets and floor coverings"*) exits the classification, together with all 8-digit codes within **CPA code 1710.42** (*"Wool yarn, not put up for retail sale"*) as they are no longer covered by the classification.

Next, changes in existing PC codes are split into singular (one-to-one) and non-singular mappings in columns 6-7. Singular mappings are either reclassifications of the taxonomy, as with the CN mappings above, or local aggregations/disaggregations of PC products. For example, from 2011 to 2012, **PC8 code 2120.21.20** (*"Antisera and other blood fractions"*) changed to **PC code 2120.21.25** (*"Antisera, other immunological products which are directly involved in the regulation of immunological processes and other blood fractions"*).

Table 5: PC8 products and their mappings.

year	products	unchanged	entry	exit	singular	non-singular	one-to-many		many-to-one		many-to-many	
			in t	in $t-1$			$t-1$	t	$t-1$	t	$t-1$	t
2001	4,793	–	–	–	–	–	–	–	–	–	–	–
2002	4,764	4,711	2	0	4	47	2	4	36	17	40	26
2003	4,693	4,398	13	1	196	86	4	12	22	5	143	69
2004	4,683	4,659	2	0	1	21	1	2	22	11	10	8
2005	4,489	4,415	1	164	0	73	6	14	125	50	8	9
2006	4,487	4,485	0	0	0	2	0	0	4	2	0	0
2007	4,418	4,287	7	3	13	111	6	17	103	46	59	48
2008	3,864	0	1	52	3,299	564	44	111	891	314	145	139
2009	3,851	3,836	1	1	1	13	0	0	26	13	0	0
2010	3,832	3,806	0	0	4	22	2	4	36	16	3	2
2011	3,799	3,770	0	0	1	28	0	0	61	28	0	0
2012	3,779	3,726	5	0	10	38	3	6	46	22	10	10
2013	3,781	3,773	8	9	0	0	0	0	0	0	0	0
2014	3,779	3,777	2	4	0	0	0	0	0	0	0	0

Notes: Column 2 reports the number of PC8 product codes in a given year. Column 3 reports the number of codes that remain unchanged from the previous year to the current. Column 4 reports new product codes that enter in t, while column 5 reports codes that existed in $t-1$ but no longer in t. Column 6 reports singular mappings (one-to-one), and column 7 reports non-singular correspondences. The remaining columns decomposes the nature of the non-singular mappings (one-to-many, many-to-one and many-to-many). Each of these is split into the number of codes in $t-1$ and the resulting number of codes in t.

Non-singular mappings are also prevalent, reflecting changes in disaggregation and grouping of products over time. The remaining columns in the table split these non-singular mappings into one-to-many, many-to-one, and many-to-many, and for each, the number of products in $t-1$ and in t. As an example of one-to-many, **PC8 code 1551.11.30** in 2004 (*"Skimmed milk (milk and cream with a fat content by weight <= 1%; not concentrated or sweetened) (excluding curdled; fermented or acidified)"*) split into two codes in 2005: **code 1551.11.33** (*"Milk and cream of a fat content by weight of $\leq$ 1%, not concentrated nor containing added sugar or other sweetening matter, in immediate packings of a net content $\leq$ 2 l"*) and **code 1551.11.37** (*"[...], in immediate packings of a net content > 2 l"*). Similar examples can be given for the remaining mappings.

I also retain the units of measurement for further use in the construction of indices and/or changes. Most products are reported in kilograms, but there are 43 quantity units in total.[14] Table 6 describes the prevalence of each over our data period. Some product types do not report quantities, only values, hence no quantity unit is available.

5. Combined Nomenclature (CN) to Prodcom (PC)

The Prodcom classification is constructed precisely to compare production and trade of goods, hence almost all PC8 codes relate to one or more CN8 codes. Some types goods appear only in one classification. Not all existing CN products are linked to PC codes (e.g. fuel, antiques), and conversely not all PC codes are linked to CN codes (e.g. industrial services related to installation, repair, maintenance or processing activities, and waste activities). In some circumstances, PC products are broken down to a greater level of detail than the CN codes. This occurs when a greater level of detail than the one provided in CN is required for analytical purposes. Some of these detailed codes are also the legacy of a time when some CN codes could be linked to several CPA headings. Therefore they continue to be collected to maintain long time series. In order to maintain the link to CN for such headings, special aggregates are created, see the discussion on letter codes in the PC classification in Section 4.

To map CN8 codes to PC8 numeric codes, I drop the optional B- and N-list codes, and disaggregate the remaining letter codes to their numeric counterparts. I also record the supplementary unit of CN and the quantity unit of measurement of PC. While the vast majority of units are the same in CN and PC, it is possible some units are different in both classifica-

[14] For example, clothing is measured per unit or per pair. Wood is in cubic meters or kilograms. Production of chemicals is mostly reported in kilograms of main chemical ingredient or in kilograms. Pesticides and related are reported in kilogram of active substance. Wicks are in kilometers. Strong liquor is in liters of alcoholic content at 100%, while Champagne, wines and beers are reported in liters. Cigarettes are per piece, while tobacco is in kilograms. Finally, some PC codes have to be reported only in values, but not quantities, mostly representing industrial manipulations or services instead of products (e.g. bleaching of leather, dying of textile).

Table 6: Units reported in PC8 (2001-2014).

unit	obs.	unit	obs.	unit	obs.	unit	obs.
none	7,011	*kg 90% sdt*	119	*kg K2O*	48	*kg TiO2*	42
1000 kWh	32	*kg Al2O3*	21	*kg KOH*	22	*act. subst.*	360
CGT	126	*kg B2O3*	20	*kg N*	243	*km*	14
GT	18	*kg BaCO3*	8	*kg Na2CO3*	14	*l*	512
TJ	40	*kg Cl*	21	*kg Na2S2O5*	14	*l alc. 100%*	112
c/k	42	*kg F*	14	*kg NaOH*	28	*m*	21
ce/el	52	*kg H2O2*	13	*kg P2O5*	104	m^2	1,361
ct/l	12	*kg H2SO4*	2	*kg PbO*	14	m^3	431
g	99	*kg HCl*	14	*kg S*	7	*p/st*	18,929
kW	35	*kg HF*	14	*kg SO2*	38	*pa*	522
kg	28,428	*kg K2CO3*	7	*kg SiO2*	28		

Notes: Description of units of measurement. *p/st*: per piece, *l:* litres, m^3: cubic metres, *kWh*: kilowatt hour, GT: gross tonnage, *TJ*: terajoule (gross calorific value), *c/k*: carat, *ce/el*: number of cells, *ct/l:* carrying capacity in tonnes, *g*: gram, *gi F/S*: gram of fissile isotopes, *kg 90% sdt*: kilogram of substance 90 % dry, *kg met.am.*: kilogram of methylamines, *kg/net eda*: kilogram drained net weight, *l alc. 100%*: litre pure (100 %) alcohol, *pa*: number of pairs.

tions.[15]

Table 7 reports the correspondences from CN8 to PC8. Generally, the PC8 classification is less disaggregate than the CN8 classification. Hence, in most cases, many CN8 products map to one PC8 product. One-to-one mappings are less common, and one-to-many mappings are rare. Here as well, many-to-many mappings exist.

[15]These are identified with "@" in the Prodcom manuals.

Table 7: CN8 to PC8 mappings.

year	mapped CN codes	singular	non-singular	one-to-many	many-to-one	many-to-many
2001	9,264	2,464	6,800	17	5,466	1,317
2002	9,360	2,422	6,938	16	5,567	1,355
2003	9,532	2,539	6,993	13	6,107	873
2004	9,319	2,568	6,751	13	5,937	801
2005	9,157	2,595	6,562	2	6,261	299
2006	8,977	2,703	6,274	2	5,970	302
2007	8,863	2,652	6,211	2	5,979	230
2008	8,758	2,000	6,758	3	6,402	353
2009	8,645	2,010	6,635	3	6,273	359
2010	8,521	2,028	6,493	4	6,135	354
2011	8,374	2,024	6,350	5	5,993	352
2012	8,364	2,010	6,354	5	5,883	466
2013	8,377	2,009	6,368	6	5,896	466
2014	8,380	2,009	6,371	6	5,898	467

Notes: Column 2 reports the number of CN8 product codes that correspond to PC8 codes in a given year. Column 3 reports singular mappings (one-to-one) from CN to PC, and the remaining columns report non-singular correspondences and their nature (one-to-many, many-to-one and many-to-many).

References

Amiti, M., Itskhoki, O., & Konings, J. (2019). International shocks, variable markups, and domestic prices. *The Review of Economic Studies*, 86(6), 2356–2402. 4

Bernard, A. B., Blanchard, E. J., Van Beveren, I., & Vandenbussche, H. (2019). Carry-along trade. *The Review of Economic Studies*, 86(2), 526–563. 3

Duprez, C. & Magerman, G. (2022). Price updating with production networks. 3, 4

Pierce, J. & Schott, P. (2012a). A concordance between u.s. harmonized system codes and sic/naics product classes and industries. *Journal of Economic and Social Measurement*, 37(1-2), 61–96. 4

Pierce, J. & Schott, P. (2012b). Concording u.s. harmonized system codes over time. *Journal of Official Statistics*, 28(1), 53–68. 4

Van Beveren, I., Bernard;, A. B., & Vandenbussche, H. (2012). Concording eu trade and production data over time. *NBER working paper 18604*. 3, 4

A. Description of the datasets

I briefly describe the final datasets. All datasets are provided in both Stata 17 (.dta) format, and non-proprietary tab-separated values (.tsv).

A.1 cn8_byyear

This dataset contains all CN 8-digit codes for each year from 2001 to 2014. For each product code, I also provide a brief verbal description, and any optional supplementary units. The variables are: *year, cn8*, *desc*, *suppl_unit*.

A.2 cn8_concord

This dataset provides the correspondences for CN 8-digit codes from one year to another. For each product code j in $t-1$, there is a corresponding code i in t. There is no entry or exit of products over time, only a change in product codes. The dataset spells out whether codes are unchanged, change in a singular fashion (one-to-one), or non-singular (one-to-many, many-to-one, many-to-many). One point of care is that a code that is the same in $t-1$ and t is not necessarily unchanged, but could be part of a non-singular mapping. That is correctly accounted for in the preparation of the data. For each product code in t, I also provide a brief verbal description, and any optional supplementary units. The variables are: *year, cn8, lcn8, lyear, lcount, count, unchanged, oto, otm, mto, mtm, desc suppl_unit.*

A.3 pc8_byyear

This dataset contains all PC *numerical* 8-digit codes for each year from 2001 to 2014. For each product code, I also provide a verbal description, the unit of measurement in words and in the official code. Some products can officially be reported in a second unit, for which I also provide the unit of measurement in words and code. Finally, each product is also labeled by a production type, either sold (S), industrial services (I), production intended for sale (C) (phased out after 2004), or total production (T) (for products that are likely used for further processing). The variables are: *year, pc8, unit_code, unit_label, desc, unit_code2, unit_label2, prodtype.*

A.4 pc8_by_year_incl_letters

This auxiliary dataset contains the raw PC codes and their descriptions, including product codes that are numeric only, and those with letter codes. The variable prcaggr contains information to further split letter codes into their numeric counterparts. The variables are: *year, pc8, desc, unit_code, unit_label, unit_code2, unit_label2, prodtype, blist, prcaggr.*

A.5 pc8_concord

This dataset provides the correspondences for PC 8-digit *numeric* codes from one year to another. For each product code j in $t-1$, there is a corresponding code i in t. Contrary to the CN classification, there is entry and exit of products, so some codes are not mapped to other codes in either $t-1$ or t. The dataset spells out whether codes are unchanged, change in a singular fashion (one-to-one), or non-singular (one-to-many, many-to-one, many-to-many). Again, a code that is the same in $t-1$ and t is not necessarily unchanged, but could be part of a non-singular mapping. That is correctly accounted for in the preparation of the data. The variables are: *year, lyear, pc8, lpc8, entry, exit, lcount, count, unchanged, oto, otm, mto, mtm.*

A.6 B-, N- and T-lists

These three datasets contain information on the mapping between PC numerical 8-digit codes and letter codes that exist in the PC classification system.

B-list and N-list codes are optional disaggregated codes, requested respectively by individual member states or Eurostat, to allow for more disaggregated classifications than the mandatory PC8 codes. These phased out and are completely eliminated by 2005.

The B-list dataset contains a mapping between the standard PC 8-digit numerical codes, and the B-list disaggregates. These more disaggregated codes are also numeric. For example, code "1411.11.50" in 2001 has three optional B-list codes: "1411.11.53", "1411.11.55" and "1411.11.57". The variables are: *year, pc8, optional.*

The N-list codes are also optional codes with a futher disaggregation of PC 8-digit codes. In this case, these 10-digit optional codes contain a letter N in the 9th position, and a numeric disaggregation in the last position. For example, code "1710.30.30" in 2001 has two optional N-list codes: "17103030N1" and "17103030N2". The variables are: *year, pc8, optional.*

Finally, T-list codes are aggregators for PC 8-digit codes. They do not have a numerical 8-digit PC counterpart, but are used in mapping PC products to CN codes. The variables are: *year, pc8*. There is also a separate dataset that maps T-list codes from one year to the next. The

variables are: *pc8, lpc8, year, lyear.*

A.7 cn8_pc8_concord

This final dataset contains the mappings in year t from CN to PC codes. All codes and mappings are numeric. I provide information on the year, CN8 and PC8 codes, the units of measurement in PC, optional supplementary units in CN, whether the units are the same in both classifications, and the singular/ non-singular nature of the mappings. The variables are: *year, cn8, pc8, unit_label, unit_label2, suppl_unit, same_unit, same_2nd_unit, cn_count, pc_count, oto, otm, mto, mtm.*