

A Concordance among Harmonized System 1996, 2002 and 2007 Classifications¹

Tolga Cebeci²

November 2012

Abstract

This note creates a concordance among Harmonized System (HS) 1996, 2002 and 2007 classifications. The main innovation it brings is that it consolidates all correlated codes instead of matching based on product expertise a single code in a classification to a single code in another classification. Hence, it does not suffer from inconsistency problems, at the expense of a tolerable loss in product detail.

I. Structure of Revisions in the Classification System

The Harmonized System (HS) is the standard classification system for countries to record the flows of goods traded across countries. The HS classification used by different countries is similar at the 6-digit level of disaggregation, enabling the user to make comparisons across countries in a given year at 6 or less digits. Since the introduction of the HS in 1988, the World Customs Organization (WCO) went through four major revisions which resulted in the initiation of new classifications in 1996, 2002, 2007 and 2012. The revisions introduced take mainly two forms:

- i. different codes for similar goods with low trade volume were merged into a smaller number of new and/or already existing codes,

¹ This note was supported by the governments of Norway, Sweden and the United Kingdom through the Multi-Donor Trust Fund for Trade and Development. I also acknowledge the generous financial support from the World Bank research support budget and the Knowledge for Change Program (KCP), a trust funded partnership in support of research and data collection on poverty reduction and sustainable development housed in the office of the Chief Economist of the World Bank (www.worldbank.org/kcp). The findings expressed in this paper are those of the author and do not necessarily represent the views of the World Bank. I thank Ana Margarida Fernandes for helpful comments on an earlier version of this note. All remaining errors are mine.

² Trade and International Integration Unit, Development Economics Research Group, World Bank (tcebeci@worldbank.org).

- ii. a code representing a good that had gained importance in world trade was split into various new and/or already existing codes, each of which representing a finer good within the original good.

As an example of a simple revision, code 030269 in the HS2002 classification, which included both swordfish and tooth fish was split into the new code 030267 (swordfish), the new code 030268 (tooth fish), and the old code 030269 (for other fish) in the HS2007 classification. For the purpose of this study, the codes in the above example are considered to form a “1 to 3” network. Tables correlating the goods across old and new classifications are published by the WCO. The total number of networks is 216 in the 1996-2002 correlation table and 340 in the 2002-2007 correlation table. Table 1 summarizes the structure of the networks included in each correlation table. Each row in the table shows how many codes are included in each type of network, the total number of networks with the same structure and the total number of codes included in each network.

Table 1: Structure of Networks

1996-2002 CORRELATION TABLE					2002-2007 CORRELATION TABLE				
Network Type		Number of Networks	Total	Total	Network Type		Number of Networks	Total	Total
Number of codes in HS1996	Number of codes in HS2002		number of codes in HS1996	number of codes in HS2002	Number of codes in HS2002	Number of codes in HS2007		number of codes in HS2002	number of codes in HS2007
1	1	43	43	43	1	1	36	36	36
1	2	70	70	140	1	2	112	112	224
1	3	12	12	36	1	3	20	20	60
1	4	16	16	64	1	4	9	9	36
1	5	6	6	30	1	5	1	1	5
1	6	2	2	12	1	7	1	1	7
1	7	2	2	14	1	14	1	1	14
1	8	4	4	32	2	1	72	144	72
1	11	1	1	11	2	2	9	18	18
1	13	1	1	13	2	3	7	14	21
2	1	25	50	25	2	4	2	4	8
2	2	7	14	14	3	1	18	54	18
2	3	2	4	6	3	2	4	12	8
2	4	1	2	4	3	3	2	6	6
2	5	1	2	5	3	5	1	3	5
2	6	1	2	6	3	6	1	3	6
3	1	3	9	3	4	1	5	20	5
3	2	2	6	4	4	2	1	4	2
3	3	1	3	3	4	3	3	12	9
3	4	1	3	4	4	5	1	4	5
3	29	1	3	29	4	10	1	4	10
4	1	3	12	3	5	1	3	15	3
4	2	1	4	2	5	2	1	5	2
4	3	1	4	3	5	3	1	5	3
4	8	2	8	16	5	4	5	25	20
5	2	1	5	2	5	5	1	5	5
5	10	1	5	10	5	6	1	5	6
6	3	1	6	3	6	1	4	24	4
6	7	1	6	7	6	3	1	6	3
12	24	1	12	24	6	4	2	12	8
14	29	1	14	29	6	5	1	6	5
29	31	1	29	31	6	8	1	6	8
					6	9	1	6	9
					7	1	2	14	2
					8	1	1	8	1
					8	2	2	16	4
					8	7	1	8	7
					8	8	1	8	8
					9	8	1	9	8
					14	1	1	14	1
					14	11	1	14	11
					25	25	1	25	25
TOTAL		216	360	628	TOTAL		340	718	718

Source: Author's calculations using WCO 2004 and WCO 2012.

II. Effect of Revisions on Time-Series Analysis

The revisions made to the HS classification system create inconsistencies for analyses that use trade data spanning more than a single classification period. As an example,

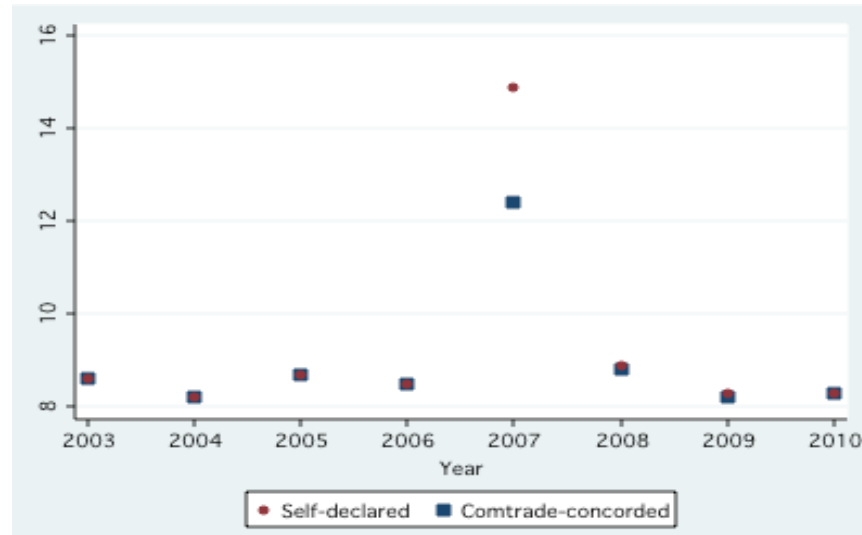
assume that country A exports swordfish in 2006 and in 2007. Swordfish appears as a new exported product for country A in 2007 relative to 2006 because it is recorded under a new code, 030269, from 2007 onwards.

In order to address the potential problems in time-series analysis created by the revisions of the HS classifications, concordance tables were prepared by numerous international organizations and researchers. As far as we know, the methodology employed to obtain all existing and available concordance tables for the HS classification dictate a 1 to 1 matching based on product expertise. However, as is obvious from Table 1, most of the networks subject to revision are composed of more than 1 code in each classification. In fact, the United Nation Statistics Division (UNSD), creator of the most widely used concordance table available, acknowledges in “Note on HS 2007 data conversion” (2009a) that *“The data conversions from HS2007 to earlier HS versions developed by UNSD assign one single code (subheading) of an earlier HS edition to each HS 2007 subheading. Yet, users should be aware that the very nature of a revision of a classification does not allow establishing a clear 1:1 correspondence for all codes (subheadings) of a new to the codes of previous versions of a classification ...”* and *“The data conversions have been developed based on the best judgment of the staff at the International Merchandise Trade Statistics Section of the UNSD but have no binding character whatsoever.”*

To exemplify the inconsistencies created by the revisions in HS classifications over time, Figure 1 presents the average product exit rates based on actual 6-digit product-level export data for the 59 countries included continuously in the UN Comtrade database during the 2002-2010 period. The product exit rate in each year is computed as the number of HS 6-digit products exported in the previous year but not in the current year divided by the number of HS 6-digit products exported in the previous year. The circles in the figure are constructed using the raw export data declared by countries. That is, the export data for the 2002-2006 period is based on the HS2002 classification whereas the export data for the 2007-2010 period is based on the HS2007 classification. The squares in the figure are constructed using the data concoded by Comtrade based on UNSD

(2008) and on UNSD (2009b). That is, the export data for the 2007-2010 period originally provided in the HS2007 classification is transformed to the HS2002 classification.

Figure 1: Effect of Revisions on Product Exit Rate (%)



Source: Author's calculations using Comtrade data.

The circles and squares suggest a smooth path with product exit rates between 8 and 9% in all years except 2007. For the 2003-2006 period, the two sets of squares and circles overlap since both make use of the same HS2002 classification. On the other hand, there is a sharp increase in the product exit rate in 2007 to about 15% according to the country self-declared data. This is an inflated rate because many goods exported in 2006 are still exported in 2007 but are no longer recorded under same codes. Using the Comtrade-concorded data, the deviation from the path in 2007 is smaller but is still considerable with a product exit rate of about 12%. Using Comtrade-concorded data solves the problem only to some extent because the concordance is based exclusively on 1 to 1 code transformation whereas revisions in the classification system usually link more than 1 code in each classification (as was shown in Table 1).

III. Methodology

The methodology that I develop is based on the condition that all correlated codes in the HS1996, 2002 and 2007 classifications to be represented by a single code. Therefore, the

concordance in this note is more appropriately termed as a “consolidation”. A similar methodology was followed by Pierce and Schott (2012) for concurring US 10-digit codes over the 1989-2004 period and by Wagner and Zahler (2011) for concurring HS1992, 1996 and 2002 classifications as an intermediate step for their analysis to identify new exports. Beyond belonging to a different classification, the main difference between my concordance and Pierce and Schott’s methodology is that their concordance allows the user to choose the start and end data of the concordance within the 1989-2004 period. This would not be a useful approach for my study since most revisions in HS 6-digit codes happened in two specific dates: 2002 and 2007. Moreover, unlike in the US data, for developing countries it is quite frequent to observe HS codes from a previous classification system in the current year data, very likely because those countries were not able to implement the HS classification revisions in the recording system at their customs timely.³

One should note that a sequential transformation process based on the aforementioned correlation tables is not possible since many codes subject to transformation exist more than once in tables. To be more specific, although there is a total of 628 codes from the HS2002 classification that are included in the HS1996-HS2002 correlation table, the number of unique codes is 508. The total and unique number of codes from the HS2007 classification that are included in the HS2002-HS2007 correlation table are 718 and 596, respectively. Moreover, 176 codes are included in both correlation tables. In light of this information, I implement the following procedure on the correlation tables to generate synthetic codes:

1. combine the HS1996-HS2002 and HS2002-HS2007 correlation tables,
2. identify the networks and assign an id to each network,
3. reshape the data by HS code and list all the networks that a code belongs to,
4. reassign the same id to all the networks that include the specific HS code under consideration,

³ This issue is frequent in the raw exporter-level customs data of several African countries included in the Exporter Dynamics Database of the World Bank (see Cebeci, Fernandes, Freund, and Pierola, 2012).

5. repeat the operations in 3 and 4 above until no further change in network ids remains,
6. assign a single synthetic code to all the HS codes that belong to the same network.⁴

IV. Outputs and the Effect of Concordance

The output file for my concordance is presented in Appendix I (available online) and includes HS 6-digit codes and the corresponding synthetic codes. At the end of the process described above, 1674 unique codes existing in the HS1996-HS2002 and/or the HS2002-HS2007 correlation tables are consolidated into 369 synthetic codes.

Table 2: Synthetic Codes by Number of HS 6-digit Codes They Include

Number of HS codes included in the synthetic code	2	3	4	5	6	7	8	9	10	11	13	14	15	16	17	18	21	25	35	91	108	171
Number of synthetic codes	164	115	31	20	6	4	5	8	1	1	1	1	2	1	1	1	2	1	1	1	1	1

The breakdown of synthetic codes is presented in Table 2. Three quarters of the synthetic codes include either 2 or 3 HS 6-digit codes. At the other end, one synthetic code includes 91 HS 6-digit codes, another 108 HS 6-digit codes and yet another 171 HS 6-digit codes.

Table 3: Effect of Concordance by HS Classification

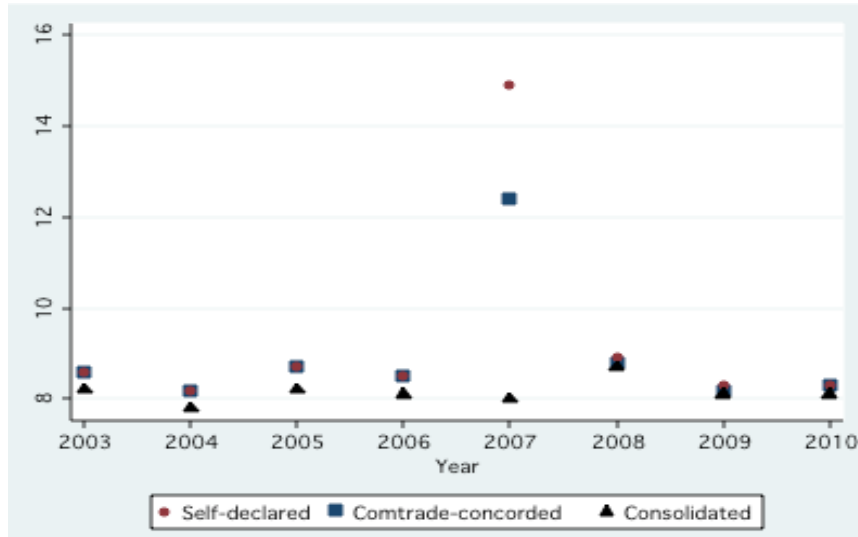
	HS1996	HS2002	HS2007
Original number of HS-6 digit codes ⁵	5132	5224	5052
Number of HS codes replaced by 369 synthetic codes	1097	1189	1017
Number of codes (HS 6 digit+ synthetic) after consolidation	4404	4404	4404

⁴ The same methodology was followed while building the concordance of HS classifications used for the Exporter Dynamics Database with a few exceptions: (i) HS Chapter 27, hydrocarbons, were excluded from the concordance in that Database; (ii) transformations across HS 2-digits were not allowed in that Database; and (iii) instead of a synthetic code, one of the 6-digit codes in the network was used as the final code in that Database. Appendix II exemplifies the differences between the consolidation methodologies in this note and that Database.

⁵ Number of standard codes, i.e., codes in HS Chapters 01-97.

Table 3 summarizes the effect of the concordance on the number of product codes. Roughly 20% codes are transformed into 369 synthetic codes in each classification, resulting in a total of 4404 codes for all classifications.

Figure 2: Effect of Revisions on Product Exit Rate (%) after Consolidation



Source: Author’s calculations using Comtrade data.

I repeat the example in Figure 1 after I concord HS codes using the concordance developed in this note. The triangles in Figure 2 show the product exit rates for the consolidated codes. Two important changes emerge. First, the problem of the deviation from the path in 2007 is fixed. Second, the product exit rate is slightly lower using the concordance relative to the other for non-problematic years, i.e., for years except 2007. This slight decrease after the consolidation is due to the fact that we now have broader product categories with a roughly 20% lower number of codes. This loss in detail is the cost of addressing the problem of a very large spike in the product exit rate in the year of the HS classification revision.

REFERENCES

Schott, P. and J. Pierce (2012). “Concording U.S. Harmonized System Categories over Time,” *Journal of Official Measures* 28: 53-68.

United Nations Statistics Division (2008). Correspondence between HS 1996 and HS 2002, United Nations Statistics Division, New York, USA. Available online at <http://unstats.un.org/unsd/cr/registry/regdnld.asp?Lg=1>

United Nations Statistics Division (2009a). Note on HS 2007 data conversion in UN Comtrade. United Nations Statistics Division, New York, USA. Available online at <http://unstats.un.org/unsd/trade/>

United Nations Statistics Division (2009b). Correspondence between HS 2002 and HS 2007, United Nations Statistics Division, New York, USA. Available online at <http://unstats.un.org/unsd/cr/registry/regdnld.asp?Lg=1>

United Nations Statistics Division (2012). *UN COMTRADE*. International Merchandise Trade Statistics, United Nations Statistics Division, New York, USA. Available online at <http://comtrade.un.org/>

Wagner, R. and A. Zahler (2011). “New Exports from Emerging Markets: Do Followers Benefit from Pioneers?,” MPRA Paper No. 30312, University Library of Munich, Germany.

World Customs Organization (2004). 1996-2002 Correlation Table, World Customs Organization, Brussels, Belgium. Available online at http://www.wcoomd.org/home_hsoverviewboxes_tools_and_instruments_hsnomenclature.htm

World Customs Organization (2012). 2002-2007 Correlation Table, World Customs Organization, Brussels, Belgium. Available online at http://www.wcoomd.org/home_hsoverviewboxes_tools_and_instruments_hsnomenclature.htm

APPENDIX I

Consolidation file in MS Excel (available online).

APPENDIX II

Below I provide an example of the differences between the consolidation methodology used in this Note and the consolidation used in the Exporter Dynamics Database. Table A1 presents a sample network which includes codes that belong to different HS-2 digits, e.g., “13” and “29”. In addition to being connected to 130211, 292911 is connected to 293910 and 293919 as well. Therefore, we have a network of 4 HS-6 digit codes.

Table A1: A Network with Transformation Across different HS 2 digit Codes

HS1996	HS2002
130211	130211
	293911
293910	293911
	293919

Source: WCO 2004.

Using the methodology in this Note I transform the 4 HS-6 digit codes in Table A1 into a single synthetic code (*039). Note that out of a total of 369 synthetic codes, 9 of them include HS-6 digit codes from different HS-2 digits.

Table A2: Comparison of Outputs in the Note and the Database

Original Code	Final Code in the Database	Final Code in the Note
130211	130211	*039
293910	293910	
293911		
293919		

On the other hand, for the HS 6-digit consolidated classification used in the Exporter Dynamics Database, I identify one of the codes in the network as the final code (293910) instead of using a synthetic code as the final code. In addition, I do not allow HS-6 digit codes under different HS-2 digit codes to transform to each other. As a result, unlike the case in this Note, 130211 is not subjected to any transformation and is therefore untouched in the HS 6-digit classification used in the Exporter Dynamics Database.

The question of which of these methodologies is more suitable depends on the type of the analysis being conducted. For a regression analysis that includes all sectors, the methodology in this Note is preferable since it is slightly better in terms of consistency over time. For a sector-specific policy analysis that spans only one HS classification period it is preferable to use the codes in the HS 6-digit classification used in the Exporter Dynamics Database in order not to mix across different sectors and have a more detailed picture.