

Constructing a Concordance Table between Harmonized System (HS) and International Classification for Standards (ICS)

Xinfen Han*
Bank of Canada

Julia Schmidt†
Banque de France

Walter Steingress‡
Bank of Canada

October 2019

Abstract

This paper describes a keyword matching procedure to create a concordance table between the 4-digit Harmonized System (HS) and the 5-digit International Classification for Standards (ICS) system. We compare our results to the concordance table between HS and ICS from the World Trade Organization (WTO). The WTO concordance table is based on member notifications of non-tariff measures (NTM).

JEL-Classification: F10, L15

Keywords: International trade, product classification, standard classification

The views expressed in this paper are those of the authors and do not reflect those of the Banque de France and the Bank of Canada.

*xhan@bankofcanada.ca

†julia.schmidt@banque-france.fr

‡wsteingress@bank-banque-canada.ca

1 Introduction

This note describes the methodological steps to map the classification of standards based on the International Classification for Standards (ICS) and product-level trade data based on the Harmonized System (HS). The aim of creating such a concordance is to allow researchers to empirically evaluate economic questions pertaining to product standards and international trade.

2 An overview of ICS and HS

The International Classification for Standards (ICS) is an international classification system for technical and product standards. The classification system was developed by the International Organization for Standardization (ISO) and designed to cover any economic sector where standards are released. The classification system consists of three levels. The first level, defined by the first two digits, denominates the field, i.e. an economic sector, technology or activity. The next three digits divide the field in different groups (5-digit), while the third level (two additional digits) divides groups into subjects that pertain to a particular aspect of the topic covered by a given group. Consider the following example code: 35.220.20. This 7-digit category describes “Magnetic Storage Devices” in the group “Data storage devices” (5-digit) and the field “Information Technology and Office Machines” (2-digit).

On the other hand, the Harmonized System (HS) was established by the World Customs Organization (WCO) and created to classify internationally traded products for the purpose of applying tariffs. The first four digits of the HS are unified across countries, while countries can characterize goods at an even more detailed level by adding more digits to the first four digits (for example, the US import code consists of 10-digit codes). Note that HS codes are used to classify trade data for tariff purposes and not to classify the domestic production process.

3 Concording 4-digit HS to 5-digit ICS

3.1 Data

We use textual analysis to link ICS and HS classes. In particular, for the ICS, we make use of keywords provided by the German Institute for Standardization (DIN) for each standard document in the database Perinorm. Perinorm is a bibliographical database that contains information on standard documents. The keywords are part of the bibliographical information as they allow users of the database to search for a particular standard. They are provided on the document-level. Each standard document in Perinorm is assigned one or several ICS classes. We aggregate all keywords on the 5-digit ICS level. We keep duplicates as we use this information in the matching procedure described below. With regards to HS, we make use of the HS description on the 6-digit level for which we extract keywords and aggregate them on the HS 4-digit level.

3.2 Generating keyword weights

We first clean the keywords via a stemming algorithm. We consider only the present tense of a verb and the singular of a noun. After having unified each word, we extract all the keywords from the respective classification and attach an importance weight to the keyword. The importance weight measures how distinctive the word is in the overall classification scheme. In particular, the importance of word w in classification scheme s is defined as the inverse document frequency (IDF):

$$IDF_s(w) = 1 + \log \left(\frac{1 + C^s}{1 + N^s(w)} \right)$$

where C^s is the total number of classes in the classification system and $N^s(w)$ counts the number of classes in the classification scheme that contain the word w . For example, there are 392 5-digit ICS classes, thus $C^s = 392$. If the word “semiconductor” appears in 50 lists that respectively describe a 5-digit ICS class, then $N^s(w) = 50$. The IDF is decreasing in the number of appearances in the classification scheme, i.e. frequent words will have little weight, while words that appear only once are considered to have high information character and hence will have a high weight.

The keywords for a particular ICS class are generated from a large number of standard documents that make up a certain ICS class. We also consider how important the keyword is for a particular class i of the ICS system and count the number of times each keyword appears in each class. We therefore define:

$$f_i(w) = N_i(w)$$

where $N_i(w)$ counts the number of occurrences of keyword w in ICS class i . For example, if the word “semiconductor” appears twice as a keyword in a 5-digit ICS code (because two standards of this ICS class contain “semiconductor” as a keyword), then $f_i(w)$ will be 2. This captures the idea that if a keyword is often repeated in a given class, it is very informative for it.

3.3 Keyword matching

As a next step, we create a list of all links between classification systems that have at least one keyword in common. A link contains the 4-digit HS code h and 5-digit ICS code i that have a common keyword. In order to characterize the strength of the link, we multiply the respective IDFs and the frequency of occurrence in the ICS scheme and sum over all common keywords:

$$\omega_{h,i} = \sum_w IDF_h(w) \times f_h(w) \times IDF_i(w) \times f_i(w)$$

Note that the multiplicative weight gives a lot of weight to rare occasions and skews the distribution of the weights to the left.

We normalize the weight $\omega_{h,i}$ by the number of the keywords in each respective class, M^h and M^i :

$$\bar{\omega}_{h,i} = \frac{\omega_{h,i}}{M^h + M^i}$$

Finally, we also normalize the final weight $\bar{\omega}_{h,i}$ to be between 0 and 1.

3.4 Reducing the number of links

Our matching procedure produces a large number of links. There are 452129 links as every keyword, be it important or not, leads to a link between a certain ICS class and a HS class. To reduce the number of links, we use as a benchmark the concordance table established by the WTO .

The WTO concordance table is based on the Technical Barriers to Trade Information Management System (TBT IMS) database of the WTO. The TBT IMS is a publicly available database of transparency information provided by WTO members in relation to technical regulations, conformity assessment procedures and standards.¹ A typical notification of a member country consists of an explanation of why it imposes a technical barrier to trade, which partner country is affected, the ICS classification of the TBT and, in some instances, it also includes the 4-digit HS code (in some instances the 2-digit or the 6-digit code) for the products on which the measure is applied.

All the notified relationships between HS and ICS classes for the period 2000 to 2016 amount to 3775 notifications, of which several mention one or more HS and ICS classes. There are a total of 2388 links between HS and ICS, and these make up 0.5% of all possible links. Of the identified relationships. Many ICS classes are linked to several HS codes; this is partly driven by the fact that several notifications mention a 5-digit ICS class, but only a 2-digit HS class (we apply these links to all underlying 4-digit HS classes).

Our objective is to reduce the number of total links while keeping as many links as possible that are in the WTO table. We therefore define an objective function which penalizes the total number of links (weighted by the inverse of their weight) while rewarding links that are in the WTO table and choose a threshold value ω^* that maximizes this objective function:

$$\max_{\omega^*} \frac{(d_1)^5}{d_2} \quad \text{where} \quad d_1 = \sum_{\substack{\bar{\omega}_{h,i} \geq \omega^* \\ \bar{\omega}_{h,i} \in \text{WTO}}} \bar{\omega}_{h,i} \quad \text{and} \quad d_2 = \sum_{\substack{\bar{\omega}_{h,i} \geq \omega^* \\ \bar{\omega}_{h,i} \in \text{KWM}}} (1/\bar{\omega}_{h,i})$$

4 Outcome

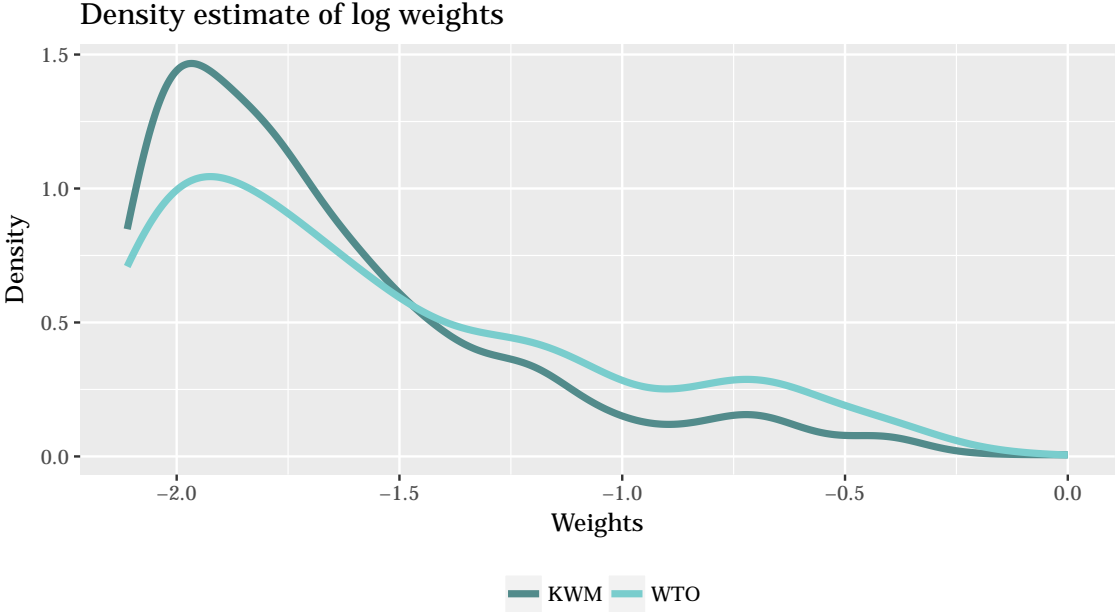
We obtain a concordance table with 994 links and can compare our procedure with the concordance table of the WTO. We can match 264 out of 392 ICS classes and 1063 out of

¹The table is available at <https://i-tip.wto.org/goods/Forms/Methodology.aspx>.

1244 HS classes. The WTO table has a fewer matches in terms of classes covered: 211 ICS classes and 916 HS classes. In the WTO table, the mean number of HS links per ICS is 11.3. In comparison, in the keyword matching table, this number is 11.4. Similarly, in the WTO table, the mean number of ICS links per HS is 2.6. In comparison, in the keyword matching table, this number is 2.8.

We plot the density of the weights $\bar{\omega}_{h,i}$ for the links we can match with the WTO against the weights of our concordance table (“KWM” which stands for keyword matching). Figure 1 shows that the keyword matching approach contains a higher mass of links with less informative weights.

Figure 1: Relation between keyword matching table and WTO concordance



In tables 1 and 2, we show the top ten and bottom ten matches (ranked by their weight $\bar{\omega}_{h,i}$). One can see that the mapping performs reasonably well.

Table 1: Top ten matches

HS code	ICS code	HS description	ICS description	Weight
3407	11.060	Modelling pastes, including those put up for children's amusement; preparations known as dental wax or as dental impression compounds, put up in sets, in packings for retail sale or in plates, horseshoe shapes, sticks or similar forms; other preparati	Dentistry	1.0000
8517	33.070	Electrical apparatus for line telephony or line telegraphy, including line telephone sets with cordless handsets and telecommunication apparatus for carrier-current line systems or for digital line systems; videophones.	Mobile services	0.8564
3306	11.060	Preparations for oral or dental hygiene, including denture fixative pastes and powders; yarn used to clean between the teeth (dental floss), in individual retail packages.	Dentistry	0.7640
1102	67.060	Cereal flours other than of wheat or meslin.	Cereals, pulses and derived products	0.7242
3816	81.080	Refractory cements, mortars, concretes and similar compositions, other than products of heading 38.01.	Refractories	0.7052
2608	73.060	Zinc ores and concentrates.	Metalliferous minerals and their concentrates	0.6842
2610	73.060	Chromium ores and concentrates.	Metalliferous minerals and their concentrates	0.6717
2603	73.060	Copper ores and concentrates.	Metalliferous minerals and their concentrates	0.6667
2614	73.060	Titanium ores and concentrates.	Metalliferous minerals and their concentrates	0.6660
2607	73.060	Lead ores and concentrates.	Metalliferous minerals and their concentrates	0.6647

Table 2: Bottom ten matches

HS code	ICS code	HS description	ICS description	Weight
8474	73.060	Machinery for sorting, screening, separating, washing, crushing, grinding, mixing or kneading earth, stone, ores or other mineral substances, in solid (including powder or paste) form; machinery for agglomerating, shaping or moulding solid mineral fuels,	Metalliferous minerals and their concentrates	0.1215
8702	43.060	Motor vehicles for the transport of ten or more persons, including the driver.	Internal combustion engines for road vehicles	0.1215
8516	97.100	Electric instantaneous or storage water heaters and immersion heaters; electric space heating apparatus and soil heating apparatus; electro-thermic hair-dressing apparatus (for example, hair dryers, hair curlers, curling tong heaters) and hand dryers; ele	Domestic, commercial and industrial heating appliances	0.1215
2301	67.060	Flours, meals and pellets, of meat or meat offal, of fish or of crustaceans, molluscs or other aquatic invertebrates, unfit for human consumption; greaves.	Cereals, pulses and derived products	0.1213
7507	77.150	Nickel tubes, pipes and tube or pipe fittings (for example, couplings, elbows, sleeves).	Products of non-ferrous metals	0.1212
7216	77.120	Angles, shapes and sections of iron or non-alloy steel.	Non-ferrous metals	0.1212
6206	61.020	Women's or girls' blouses, shirts and shirt-blouses.	Clothes	0.1212
7615	97.040	Table, kitchen or other household articles and parts thereof, of aluminium; pot scourers and scouring or polishing pads, gloves and the like, of aluminium; sanitary ware and parts thereof, of aluminium.	Kitchen equipment	0.1212
0708	67.080	Leguminous vegetables, shelled or unshelled, fresh or chilled.	Fruits. Vegetables	0.1212
2106	67.250	Food preparations not elsewhere specified or included.	Materials and articles in contact with foodstuffs	0.1211