A Concordance Between Ten-Digit U.S. Harmonized System Codes and SIC/NAICS Product Classes and Industries[*]

Justin R. Pierce[†]

Center for Economic Studies, U.S. Bureau of the Census

Peter K. Schott[‡]

Yale School of Management & NBER

November 2009

Abstract

This paper provides and describes concordances between the ten-digit Harmonized System (HS) categories used to classify products in U.S. international trade and the four-digit SIC and six-digit NAICS industries that cover the years 1989 to 2006. We also provide concordances between ten-digit HS codes and the five-digit SIC and seven-digit NAICS product classes used to classify U.S. manufacturing production. Finally, we briefly describe how these concordances might be applied in current empirical international trade research.

Keywords: International trade; industry classification

JEL classification: F1

1. Introduction

This note describes two sets of concordances between the ten-digit Harmonized System (HS) codes used by the United States to track international trade and the Standard Industrial Classification (SIC) and North American Industry Classification System (NAICS) categories used to characterize domestic production.

The first set of concordances links each ten-digit import and export Harmonized System (HS) code to a single four-digit SIC (SIC4) and six-digit NAICS (NAICS6) *industry*. These concordances are assembled from published Census data, which provide a mapping of HS to SIC and NAICS industries from 1989 to 2001 and 2000 to 2006, respectively. Our contribution here is to extend these mappings to match HS codes with SIC industries after 2001, and to match HS codes with NAICS industries before 2000. This permits construction of a long (i.e., 1989 to 2006) time series of U.S. industry-level import and export data by trading partner in either SIC or NAICS format, subject to the caveats discussed below.

The second set of concordances links ten-digit import and export HS categories to one or more five-digit SIC (SIC5) or seven-digit NAICS (NAICS7) *product classes*. These concordances are constructed using "basecodes" created by the U.S. Census Bureau ("Census"). In each year of an economic census, Census constructs two mappings: the first links HS codes to basecodes while the second links basecodes to SIC5 or NAICS7 product classes. We combine the two mappings to link HS codes to product classes. We note that due to differences between how the HS and SIC/NAICS systems classify goods, a given HS code may match to more than one SIC5/NAICS7 product class.

Finally, this paper briefly discusses how these concordances might be applied in current empirical international trade research. In particular, we provide background information useful for linking the firm-product-class domestic production data in the U.S. Census of Manufactures (CM) to the firm-product import and export data in the Longitudinal Firm Trade Transaction Database (LFTTD).[1]

Section 2 provides a brief description of HS, SIC and NAICS codes. Section 3 describes the HS to SIC4/NAICS6 industry concordance, while Section 4 describes the HS to SIC5/NAICS7 product-class concordance. Section 5 discusses how the latter can be used to link Census production and trade data. Detailed discussion of Census' "HS-Baseroot" and "Principal Differences" files, the key inputs to our HS to product class concordances, are relegated to appendices. The appendices also include a brief discussion of how our concordances relate to others that can be found on the web.

2. A Brief Description of HS, SIC and NAICS

U.S. import and export codes are based on the Harmonized System established by the World Customs Organization (WCO). The WCO assigns six-digit codes for general categories, and countries adopting the HS are allowed to characterize goods at more detailed levels by adding digits, usually two to four, for a total of eight to ten. The United States uses up to ten digits. U.S. export codes are administrated by Census and are technically

---

[1]For more detail on the former, see Bernard, Redding and Schott (2009). For more detail on the latter see Bernard, Jensen and Schott (2009).

known as Schedule B codes. U.S. import codes are administered by the U.S. International Trade Commission (USITC) and are technically known as Harmonized Tariff System (HTS) codes. In this paper, we refer to both generically as HS codes. Changes to the U.S. import or export codes occur via three routes: changes by the World Customs Organization (WCO) to the official list of international six-digit prefixes; U.S. legislation that affects U.S. eight-digit codes (imports only); or changes by the Committee for Statistical Annotation of Tariff Schedules (known as the "484(f) Committee") to statistical ten-digit codes.[2] For a concordance of HS codes over time, see Pierce and Schott (2009).

SIC codes are used to classify domestic production in the United States until the 1997 economic census, with major revisions of the SIC occurring in 1972 and 1987. Starting with the 1997 census, U.S. output is classified according to the NAICS. Census refers to the first four digits of an SIC code, and the first six digits of a NAICS code, as an "industry". It reserves the terms "product class" and "product" for the first five and seven digits of an SIC code, and the first seven and ten digits of a NAICS code, respectively. While the set of official U.S. industries is defined outside the Census Bureau, Census generally has discretion in defining product classes and products within these industries.[3]

The HS codes used to classify trade data and the SIC or NAICS codes used to classify domestic production data cannot be matched directly due to differences in how the two systems classify products. In particular, while HS codes are based solely on product characteristics, SIC and NAICS codes may also take into account the method of production. This difference has two consequences with respect to the concordances discussed below. First, it may not be possible to match a given HS category to a single SIC or NAICS category. Second, U.S. trade cannot be broken out by the full set of SIC or NAICS categories. For example, two SIC industries that cover goods with the same characteristics may have to be combined into a single, aggregate category that does not provide information as to how the goods are made.[4]

3.   Concording HS to SIC4/NAICS6 Industries

Census includes HS to SIC4/NAICS6 industry concordances with its publicly available trade statistics.[5] For reasons discussed in the next section, these are referred to here as HS to "baseroot" mappings. From 1989 to 2001, these mappings match ten-digit HS codes to four-digit SIC baseroots. From 2000 to the present, they match ten-digit HS codes to six-digit NAICS baseroots. In almost all cases, these baseroots are proper industries listed in either the SIC or NAICS.[6] We assemble a "master list" of these concordances by appending

---

[2]See http://www.census.gov/foreign-trade/aip/comb_seminar_pres.ppt, and www.census.gov/foreign-trade/faq/sb/sb0008.html for more detail.

[3]See Census (1992) for more detail. The SIC, for example, was the responsibility of the Technical Committee on Industrial Classification in the Office of Management and Budget.

[4]For a more detailed discussion of Census' SIC and NAICS concordance methods, see in particular the website "U.S. Commodity Exports and Imports as Related to Output (OEI)" at www.census.gov/epcd/www/intronet.html.

[5]They also include HS to Standard International Trade Classification (SITC) concordances, which we include with our other concordances in the files described below.

[6]Of the 461 NAICS baseroots in the HS-NAICS6 import concordance and 455 NAICS baseroots in the HS-NAICS6 export concordance, 10 are not real industries as defined in the NAICS. They are 11211X, 1123XX, 31131X 31181X, 31511X, 33631X, 910000, 920000, 980000, 990000. Of the 471 SIC baseroots in

the HS-baseroot mappings contained in the December trade CD-ROMs published by Census from 1989 to 2006.[7]

For certain applications, it might be useful to extend each set of mappings beyond the years for which these official concordances are available. That is, it may be useful to have an HS-SIC concordance for years after 2001 or an HS-NAICS concordance for years prior to 2000. We extend the HS-SIC and HS-NAICS mappings to incorporate the entire period from 1989 to 2006 using a three-step alogorithm based on the procedures used previously in Feenstra et al. (2002). The steps of this algorithm are described immediately below. The Stata code for steps 1 and 2, and for incorporating the results of step 3, is available in the zip archive accompanying this paper on Schott's website, with filename schott_algorithm_10.do.[8]

1. Step 1: Look at all ten-digit HS within a nine-digit category. If all assigned ten-digit HS within this category have the same SIC assignment, assign that SIC to any unassigned ten-digit HS within that nine-digit category. Repeat for eight-, seven-, etc. digit HS categories.

2. Step 2: Sort list by ten-digit HS code. Examine "gaps" between ten-digit HS codes with assignments. If a gap is preceded and succeeded by the same SIC code, use that SIC code for all unassigned ten-digit HS codes in the gap.

3. Step 3: Hand match remaining unmatched HS codes where possible. Note that any remaining unmatched ten-digit HS codes account for a very small fraction of U.S. imports or exports.

Table 1 summarizes the number of SIC and NAICS codes "filled in" using this procedure for exports and imports, respectively. The descriptions in the first column match those provided by the variable "matchtype" in the files discussed in Section 6 below.

Table 1: Extending the HS-SIC and HS-NAICS Concordances

Match Type	HS-SIC		HS-NAICS	
	Imports	Exports	Imports	Exports
From Census	23,233	10,341	19,473	9,818
From Mechanical Match 1	1,532	891	4,422	1,203
From Mechanical Match 2	207	84	526	216
From Hand Match	357	193	908	272
Total HS codes 1989-2006	25,329	11,509	25,329	11,509

the HS-SIC import concordance, 5 are not real industries as defined in the SIC. They are 314X, 9100, 9200, 9800, 9900. Of the 470 SIC baseroots in the HS-SIC export concordance, 7 are not real industries. They are 314X, 3XXX, 9000, 9100, 9200, 9800, 9900.

[7]These CDs are sold by Census's Foreign Trade Division and are often available in university libraries. The copies used here are provided generously by the Yale University Social Sciences Library. Each of the monthly CDs for imports and exports contains a dBase-formatted file (called concord.dbf) that separately relates the ten-digit import and export HS codes used in the month to four-digit SIC and/or six-digit NAICS basecodes, depending upon the year. CDs are available starting in December, 1989 for exports and January 1989 for imports. Note that the December CD for each year contains trade information for that month as well as the yearly totals.

[8]See http://www.som.yale.edu/faculty/pks4/sub_international.htm.

By aggregating the HS-baseroot mappings for all available years and extending them for the full period in which the HS was in existence, we create HS to SIC4 and HS to NAICS6 concordances for both imports and exports, for the period from 1989 to 2006. See Section 6 for a full description of the final concordance files available in the electronic appendix to this paper.

4. Concording HS to SIC5/NAICS7 Product Classes

4.1. Census's Procedure for Mapping HS to SIC and NAICS

Census matches trade data to domestic production data using a bridge code known as an "SIC-base" or "NAICS-base" code, which we refer to here generically as "basecodes."[9] Basecodes are eight-digit alphanumeric codes that can generally be thought of as describing product characteristics. The first four (six) digits of the SIC (NAICS) basecode represent the "root" industry of the basecode, which, loosely speaking, is Census's best match for the industry in which the traded product belongs. We refer to basecode roots here as "baseroots" and use them in constructing the industry concordances noted above. The remaining digits are internal identifiers for whether the basecode encompasses one or more product classes, and, in the latter instance, whether those product classes are from different industries. For Census year 1992, enough data are available for us to also construct a HS to SIC5/NAICS7 concordance based on basecodes. For the 1997 and 2002 HS to SIC5/NAICS7 concordances, however, we are restricted by data limitations to matching HS and SIC/NAICS product codes through baseroots only. The differences between constructing concordances using basecodes versus baseroots are discussed in detail in the next sub-section.

We match HS product codes to SIC5/NAICS7 product classes via baseroots using two complementary mappings produced by Census. The first mapping, which we refer to here as an "HS-baseroot" concordance, assigns a single baseroot to each HS code.[10] As noted above, these mappings are published in Census's monthly releases of U.S. trade data on CDs. The second mapping, known as a principle differences (PD) file, is constructed for every economic census, which take place in years ending in 2 and 7; it assigns a single baseroot to each product class in the SIC or NAICS.[11] HS product codes can then be matched to SIC5/NAICS7 product classes through their baseroots. The HS-baseroot and PD mappings are discussed in detail in Appendixes 2 and 3 below, respectively.

---

[9] As noted above, a more detailed discussion of Census' SIC and NAICS concordance methods is available at www.census.gov/epcd/www/intronet.html.

[10] Census constructs the HS-to-basecode concordance so that the Foreign Trade Division can publish trade statistics using the same industry categories it uses to publish domestic production statistics. As alluded to above and as discussed in more detail at www.censusbureau.biz/epcd/oei/view/appenda.txt, the HS to basecode mappings often make more sense for exports than for imports: "It is somewhat easier to find a reasonable statistical basis for comparing domestic output with exports than with imports. This is because there are substantial numbers of imported commodities which are not produced in the United States or are produced in very small quantities. On the other hand, the merchandise exported from the United States is ordinarily produced in this country and reflects items important to output."

[11] "Product trailer" files associated with the quinquennial CM record the set of product classes produced by U.S. manufacturing establishment in each census year. See Bernard, Redding and Schott (2009) for more detail. Similarly, the "materials trailers" record the set of products establishments use as inputs.

At this point, an example may be useful for fixing ideas. In 1992, HS code 7215200000 was used to track exports of "other bars and rods of iron or nonalloy steel, cold-formed or cold-finished, less than 0.25 percent carbon." According to the 1992 HS-baseroot concordance, this HS code – and 222 others – maps into SIC baseroot 3312. This baseroot, in turn, maps into 11 different SIC product class codes from 3 different four-digit SIC industries in the 1992 PD file: 33121, 33122, 33123, 33124, 33126, 33127, 33128, 3312C, 33167, 33168 and 33170.[12] We note that in the official Census ten-digit HS to four-digit SIC mapping discussed in Section 3, HS code 7215200000 maps uniquely to SIC industry 3312.

This example highlights the "many-to-many" nature of the HS-product-class concordances. While each HS code maps to a single baseroot, many HS codes (223 in this example) can map to a single baseroot. Similarly, while each five-digit SIC product class maps to a single baseroot, many product classes (from three different industries, in this example) may map to a single baseroot. As discussed in Section 5 below, the HS-baseroot and PD files can in theory be used to match the product classes U.S. manufacturing firms produce in each CMF year to the products they import and export in those years.

4.2. Matching on Basecodes Versus Baseroots

Matching on baseroots is appealing because HS-baseroot mappings are available in all years, allowing us to create concordances for every economic census year since 1992. As noted above, however, we have access to more disaggregated HS-basecode and SIC5-basecode mappings for 1992. The primary advantage of concordances using basecodes is a "more precise" mapping between HS and SIC5.

To illustrate what we mean by "more precise", consider once again HS code 7215200000, which we used to illustrate matching through baseroots in the previous sub-section. As mentioned above, HS code 72150200000 and 222 other HS codes matched to 11 different SIC5 product classes through baseroot 3312. When HS and SIC5 codes are matched through full basecodes, rather than baseroots, however, we find that 7215200000 is one of only 10 HS codes that map to only two SIC5s – 33128 and 33168, defined under basecode 33128B00 – described as "cold-finished steel bars/bar shapes (made in mills)" and "cold-finished steel bars/bar shapes (not made in mills)."[13] Because HS code 7215200000 is described as "other bars and rods of iron or nonalloy steel, cold-formed or cold-finished, less than 0.25 percent carbon", it appears that assigning SIC5s based on a full basecode, rather than a baseroot, has provided a better match, by dropping unrelated SIC5 products like sheet and strip, pipe and tube and rails.

HS code 7215200000 was matched to 9 additional SIC5s when we matched HS codes to

---

[12]The product descriptions for these SIC5 product-classes are as follows: 33121 - Coke over and blast furnace products; 33122 - Steel ingot and semifinished shapes; 33123 - Hot-rolled sheet and strip including tin-milled products; 33124 - Hot-rolled bars and bar shapes, plates, structural; 33126 - Steel pipe and tubes (made in steel mills); 33127 - Cold-rolled steel sheet and strip (made in mills); 33128 - Cold-finished steel bars/bar shapes (made in mills); 3312C - Other steel mill products, including steel rails; 33167 - Cold-rolled steel sheet and strip (not made in mills); 33168 - Cold-finished steel bars and bar shapes (not made in mills); 33170 - Steel pipe and tubes.

[13]Note that this example provides a good illustration of how HS codes may match to more than one SIC5, since the SIC considers the method of production when assigning product classifications. The difference between these product classes is whether or not they are made in steel mills.

SIC5 codes with baseroots versus basecodes. This matching of HS codes to additional SIC5s when matching with baseroots is not uncommon, as illustrated in the following analysis of 1992, the only year for which we can do both types of mappings. Of the 16,022 import HS codes in use in 1992, 9,289 are matched to additional SIC5s when using baseroot matching. The mean number of additional SIC5s matched to each import HS is 2.35. Similarly, of the 8,054 export HS codes in use in 1992, 5,396 are matched to extra SIC5s under baseroot matching. The mean number of additional SIC5s matched to each export HS is 2.72. Table 2 displays the number of extra SIC5s associated with HS10 import and export codes for 1992.

Table 2: Additional SIC5s Associated With Each HS Under Baseroot Matching

Exports		Imports	
HS10	Additional SIC5	HS10	Additional SIC5
2,658	0	6,733	0
1,090	1	1,818	1
1,102	2	1,983	2
742	3	1,294	3
731	4	1,225	4
389	5	618	5
181	6	289	6
400	7	819	7
283	8	368	8
212	9	402	9
42	10	107	10
25	11	38	11
74	12	112	12
50	13	75	13
33	14	60	14
23	16	57	16
11	17	13	17
1	20	1	20
3	24	6	24
4	27	4	27

Notes: Table displays the number of "Additional" SIC5s associated with HS10 export and import codes in 1992. Additional SIC5s are SIC5 product codes that are associated with a particular HS10 when a concordance is constructed with 4-digit baseroots, rather than a full 8-digit basecode.

For some types of research, matching HS and SIC5/NAICS7 codes through full basecodes might be useful. Pierce (2009), for example, identifies U.S. manufacturing establishments that received antidumping protection by matching the HS10s used to classify products in antidumping investigations to the SIC5 product-classes that establishments reported producing in the CMF. In this case, matching on baseroots, rather than full basecodes, would likely lead to some unprotected plants being incorrectly identified as recipients of antidumping protection.

Unfortunately, Census published a full HS10-basecode mapping only for 1992. As a result, matching on basecodes can only be performed in a somewhat limited time period. In the electronic appendix, we provide HS10 to SIC5 concordances constructed with basecode matching for 1992 with filenames m_basecode_92.csv and x_basecode_92.csv for imports and exports, respectively.[14] In the import concordance 16,022 HS codes are matched to 1,564

---

[14]This concordance might be extended to prior and subsequent years using the HS-over-time concordance provided in Pierce and Schott (2009).
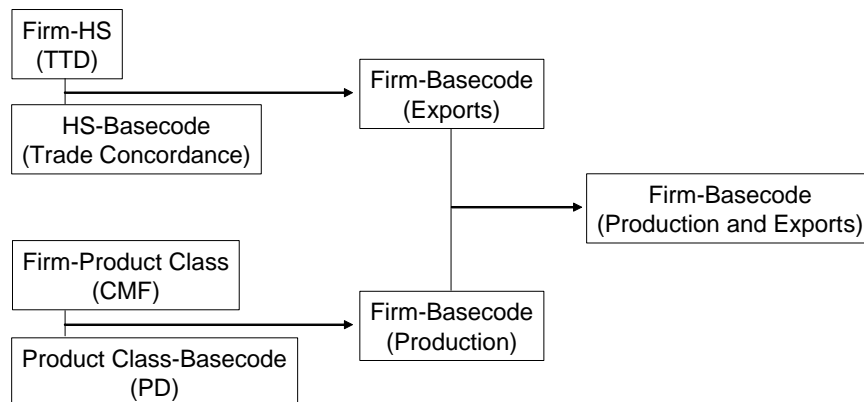
Figure 1: Linking the TTD to the CMF at the Firm-Basecode Level

SIC5 codes through 812 basecodes.[15] In the export concordance 8,053 HS codes are matched to 1,555 SIC5 codes through 806 basecodes.[16]

5. Concording the LFTTD and CM

The concordance procedures described above can be used to link firm-level domestic production data contained in the Census of Manufactures (CM) with the transaction-level trade data housed in the Longitudinal Firm Trade Transactions Database (LFTTD). The resulting merged dataset would contain a firm-baseroot-level dataset recording production, imports and exports at the same level of aggregation (i.e., according to SIC or NAICS baseroots) for a particular census year.

The merged trade and production dataset can be constructed as follows. First, the LFTTD is merged with the trade concordance for the appropriate year, which yields a firm-baseroot version of the LFTTD. Second, the product trailer files of the CM–which contain data on output by product for every U.S. manufacturing establishment–are merged with the PD file for the appropriate year, yielding a firm-baseroot version of the CM. Lastly, these two datasets are merged by baseroot. This process is illustrated in Figure 1.

6. Downloads

All files described here are available in a zip archive accompanying this paper on Schott's website.[17]

---

[15]Six-hundred twenty-five import HS codes have basecodes with no SIC5 match. Two SIC5 codes have basecodes with no import HS match.

[16]Four-hundred eight export HS codes have basecodes with no SIC5 match. Eleven SIC5 codes have basecodes with no export HS match.

[17]See http://www.som.yale.edu/faculty/pks4/sub_international.htm.

6.1. HS-SIC4/NAICS6 Concordance Files

The HS-SIC4/NAICS6 industry concordances for 1989 to 2006 are available in two files for exports and imports, named, respectively, hs_sic_naics_imports_89_106_20091016.dta and hs_sic_naics_exports_89_106_20091016.dta, where 89 represents the beginning year of 1989, 106 represents the ending year of 2006 and 20091016 represents the version date.

1. HS: ten-digit HS import or export code

2. SIC: corresponding four-digit SIC code

3. NAICS: corresponding six-digit NAICS code

4. SIC_MATCHTYPE: description of match origin (see Table 1)

5. NAICS_MATCHTYPE: description of match origin (see Table 1)

6. COMMODITY: a string version of HS, with leading zeroes, where applicable

The Stata do-files used to create these concordances are also available in the electronic appendix with filename schott_algorithm_10.do.

6.2. HS-SIC5 (1992, Using Full Basecodes)

The HS-SIC5 (basecode) concordances for 1992 are available in two files named m_basecode_92.csv for imports and x_basecode_92.csv for exports. These files contain the following variables:

1. HS: ten-digit HS import or export code

2. Basecode: eight-character basecode associated with HS

3. SIC5: The SIC5s associated with a particular HS and basecode.

Note that there may be multiple entries for a single HS code when it matches to more than one SIC5. The Stata do-file used to create these concordances is also available in the electronic appendix with hs_sic5_basecodes_02.do.

6.3. HS-SIC5 (1992, Using Baseroots)

The HS-SIC5 (baseroot) concordances for 1992 are available in two files named hs_sic5_imports_92.csv for imports and hs_sic5_exports_92.csv for exports. These files contain the following variables:

1. HS: ten-digit HS import or export code

2. SICBASEROOT: four-character SIC baseroot associated with HS

3. SIC5: The SIC5s associated with a particular HS and basecode.

Note that there may be multiple entries for a single HS code when it matches to more than one SIC5. The Stata do-file used to create these concordances is also available in the electronic appendix with filename hs_sic5_naics7_baseroots_04.do.

6.4. HS-NAICS7 (1997 and 2002, Using Baseroots)

The HS-NAICS7 (baseroot) concordances for 1997 and 2002 are available in four files named hs_naics7_imports_yy.csv for imports and hs_naics7_exports_yy.csv for exports, where yy is the last two digits of the year. These files contain the following variables:

1. HS: ten-digit HS import or export code

2. NAISBASEROOT: six-character NAICS baseroot associated with HS

3. NAICS7: The NAICS7 associated with a particular HS and basecode.

Note that there may be multiple entries for a single HS code when it matches to more than one NAICS7. The Stata do-file used to create these concordances is also available in the electronic appendix with filename hs_sic5_naics7_baseroots_04.

6.5. HS-SITC Concordance Files

Census's mapping of HS and SITC codes from its published trade data are available in two files named hs_sitc_imports.csv for imports and hs_sitc_exports.csv for exports. These files contain the following variables:

1. HS: ten-digit HS import or export code

2. Corresponding five-digit revision 3 SITC code.

References

Bernard, Andrew B., Stephen J. Redding and Peter K. Schott (2009), "Multi-Product Firms and Product Switching", *American Economic Review*, forthcoming.

Bernard, Andrew B., J. Bradford Jensen and Peter K. Schott (2009), "Importers, Exporters and Multinationals: A Portrait of Firms in the U.S. that Trade Goods" in T. Dunne, J.B. Jensen and M.J. Roberts (eds.), Producer Dynamics: New Evidence from Micro Data (University of Chicago Press).

Feenstra, Robert C. (1996), "U.S. Imports, 1972-1994: Data and Concordances," NBER Working Paper no. 5515.

March 1996, with CD-ROM.

Feenstra, Robert C., Romalis, John and Schott, Peter K. (2002), "U.S. Imports, Exports and Tariff Data, 1989 to 2001," NBER Working Paper 9387.

Pierce, Justin R. and Peter K. Schott (2009), "Concording U.S. Harmonized System and Schedule B Codes Over Time," NBER Working Paper 14837.

Pierce, Justin R. (2009), "Spare the Rod, Spoil the Firm? A Plant-Level Analysis of the Effects of Temporary Protection," available online at http://ssrn.com/abstract=1443467.

A   Appendix 1: Other Concordances

This appendix discusses the relationship between the concordances developed above and two other HS-SIC/NAICS concordances that can be found on the web.

A1.   The Feenstra (2002) Concordance

Feenstra et al. (2002) provide background for U.S. HS10-level trade data for 1989 to 2001. Those data have subsequently been extended to 2006 and are available on Feenstra's website. Of the 26,277 ten-digit HS codes used to track U.S. imports (exports) in the Feenstra et al. (2002) 1989 to 2001 dataset, Census provided a baseroot concordance for all but 1,222. Of these 1,222 HS codes, 898 were assigned to a four-digit SIC category using a HS to 1987-revision MSIC concordance from Feenstra (1996). Though in principle MSIC codes differ from SIC codes, a number of MSIC codes map directly into regular SIC codes. The remaining 324 products were assigned to industries via an algorithm similar to that described in Section 3 above.

The set of HS codes found in the Feenstra et al. concordances differs slightly from that of the master list described in Section 3. Of the 25,329 (11,509) unique import (export) HS codes that result from merging Feenstra et al.'s concordances with our own, we find that 24,947 (11,472) are in common while 382 (37) are only in the Feenstra et al (2002) concordance. We don't have an explanation for the codes unique to the Feenstra et al (2002) dataset though we suspect they may be due to Census' periodic revisions of the trade data.

A2.   The EIIT Concordance

A five-digit SIC to ten-digit HS concordance of unknown origin is posted to the EIIT website.[18] This concordance does not distinguish between import or export HS categories and it does not note the years to which either its HS codes apply.

The EIIT concordance contains 17,436 HS codes and maps them to 805 five-digit SIC categories, 741 of which are in manufacturing. If collapsed to the four-digit SIC level, this list comprises 439 four-digit SIC codes, 386 of which are in manufacturing. This compares with the 1,440 five-digit and 459 four-digit manufacturing SIC codes contained in the 1987 revision of the SIC, and the 1,462 five-digit and 459 four-digit manufacturing SIC codes in the 1992 revision of the SIC. The 386 unique manufacturing codes in the EIIT concordance are similar to the 386 "super-sic" codes described in Feenstra et al (2002).

The EIIT concordance appears to be a close cousin of the concordance described in Section 4. Of the 8,215 (15,120) export (import) HS codes which appear in both concordances, 6,058 (10,762) have the same four-digit root.

B   Appendix 2: Census's HS-Baseroot Concordances

Census produces an HS-basecode concordance only for the years in which there is an economic census. However, it provides more aggregate, HS-baseroot concordances with its monthly published trade statistics. As discussed above, we assemble a "master list" of these

---

[18] See www.macalester.edu/research/economics/page/haveman/Trade.Resources/Concordances/FromHS/10hs5sic87.txt.

mappings by appending the HS-baseroot concordances contained in the December trade CD-roms. The Stata files containing these lists are discussed in Section 6. They are available on Schott's website.[19]

B1. HS-SIC

Census's HS-baseroot concordances virtually always map HS codes to a single four-character SIC root. As noted above, these roots are the first four characters of an eight-character SIC basecode.[20] For the most part, these baseroots are proper industries, but there are some (e.g., 3XXX) that reflect the difficulties noted in Sections 3 and 4 above. We note the following:

- As indicated in Table 3, the number of unique HS export (import) codes in the master list that have SIC basecodes associated with them in at least one year ranges from 7,908 (14,402) in 1989 to 8,629 (17,183) in 2001. 2001 is the final year in which SIC codes appear in the concordance.

- The number of unique SIC codes to which these export (import) HS codes match ranges from 429 (443) in 1989 to 449 (450) in 2001.[21]

- Some of the SIC basecodes to which HS codes are assigned are incomplete (e.g., 23XX), while others are outside manufacturing (e.g., 0273). As noted in the third column of each panel in Table 3, the number of manufacturing SIC basecodes to which these export (import) codes match ranges from 371 (386) in 1989 to 391 (392) in 2001. The fact that there are fewer than the official number of 459 manufacturing SIC codes in the concordance files is consistent with the discussion in Sections 3 and 4 above.

B2. HS-NAICS

As with the SIC, Census's concordance files virtually always map HS codes to a unique six-digit NAICS baseroot. For the most part, these baseroots are proper NAICS industries, but there are some that reflect the difficulties noted in Sections 3 and 4 above. We note the following:

- As summarized in Table 4, the number of HS export (import) codes in the master list that have NAICS basecodes associated with them in at least one year ranges from 8,628 (16,897) in 2000 to 8,972 (17,746) in 2006. 2000 is the first year that NAICS codes appear in the concordance files.

---

[19] Though we have not explored whether reliance on the December files makes a difference (e.g., whether there is within-year variance of SIC mappings to a particular HS), we note that, very rarely, the baseroots Census assigns to an HS code do change across years.

[20] Though the concordance files included with the monthly trade data do not include the full, internal-to-Census basecode, that mapping is available for 1992 at http://www.census.gov/epcd/www/intronet.html (see second paragraph).

[21] There are 459 "official" four-digit SIC manufacturing codes in 1992 and 1997 economic censuses. For a complete list, see http://www.censusbureau.biz/epcd/oei/view/sic-sht2.txt.

Table 3: HS and Four-Digit SIC Codes in the "Master List"

	Exports			Imports		
	HS10	SIC4	Man SIC4	HS10	SIC4	Man SIC4
1989	7,908	429	371	14,402	443	386
1990	7,971	447	387	15,214	446	387
1991	8,110	448	387	15,414	446	387
1992	8,107	448	387	15,430	448	388
1993	8,167	449	391	15,502	447	389
1994	8,239	449	391	15,980	447	389
1995	8,308	449	391	16,630	447	389
1996	8,593	449	391	16,882	447	389
1997	8,609	449	391	17,345	447	389
1998	8,620	449	391	17,099	447	389
1999	8,626	449	391	17,179	450	392
2000	8,635	449	391	17,215	450	392
2001	8,629	449	391	17,183	450	392

Notes: Table displays number of ten-digit HS codes, four digit SIC codes, and four-digit manufacturing SIC codes appearing in the concordance files accompanying the U.S. monthly trade statistics sold by the U.S. Census Bureau.

- The number of NAICS basecodes to which these export (import) codes match ranges from 454 in 2000 to 453 in 2006 for both imports and exports.[22]

- Some of the NAICS basecodes to which HS codes are assigned are incomplete, while others are outside manufacturing. As noted in the third column of each panel of Table 4, the number of manufacturing NAICS industry codes to which these export (import) codes match ranges from 387 in 2000 to 386 in 2006. As with the SIC, these numbers of manufacturing codes are lower than the 473 official manufacturing industries in the NAICS.

Table 4: HS and Six-Digit NAICS Codes in the "Master List"

	Exports			Imports		
	HS10	NAICS	Man NAICS	HS10	NAICS	Man NAICS
2000	8,628	454	387	16,897	454	387
2001	8,622	453	386	16,910	453	386
2002	8,940	453	386	17,351	453	386
2003	8,930	454	386	17,390	454	386
2004	8,933	454	386	17,382	454	386
2005	8,971	453	386	17,717	453	386
2006	8,972	453	386	17,746	453	386

Notes: Table displays number of ten-digit HS codes, six-digit NAICS codes and six-digit manufacturing NAICS codes appearing in the concordance files accompanying the U.S. monthly trade statistics sold by the U.S. Census Bureau.

C   Appendix 3: Census's Principle Differences (Product Class-Basecode) Concordances

This section summarizes Census' PD files for 1992, 1997, 2002 and 2007.

---

[22]There are 473 "official" six-digit NAICS manufacturing codes in the 2002 economic census. For a complete list of the six-digit codes, see http://www.census.gov/epcd/naics02/naico602.txt.

C1. 1992 Economic Census

The 1992 PD file maps five-digit SIC product classes to eight-digit (SIC-based) basecodes and is available in the electronic appendix with filename pd92.csv. We note the following:

- 814 unique basecodes match to a product class (PC) in the 1992 PD file, 768 of which are in manufacturing. Table 5 summarizes the distribution of these basecodes according to the number of five-digit SIC product classes into which they map. As a group, the eight-digit basecodes contain 418 unique four-character basecode roots, 391 of which are in manufacturing. Note that there are 459 unique four-digit SIC manufacturing industries in 1992.[23]

Table 5: Number of Product Classes per Basecode and Basecode Root (1992)

Product Classes	Overall		Manufacturing	
	Basecodes	Basecode Roots	Basecodes	Basecode Roots
1	549	117	520	111
2	109	60	103	54
3	59	76	53	69
4	36	50	32	46
5	20	39	19	37
6	17	25	17	24
7	9	13	9	13
8	2	7	2	7
9	2	9	2	8
10	2	5	2	5
11	1	2	1	2
12	2	4	2	4
14		3		3
15		1		1
16	1		1	
18	1	3	1	3
20	1		1	
21		1		1
22	1		1	
23	1	1	1	1
25	1	1	1	1
28		1		1
Total	814	418	768	391

Notes: Table displays distribution of basecodes and basecode roots according to the number of product classes into which they map, overall and for manufacturing.

- 1,566 unique five-digit SIC product classes are matched to an eight-digit basecode in the 1992 PD file. The official list of SIC categories for the 1992 CMF encompasses 1,462 five-digit product classes for manufacturing.[24]

---

[23]The set of four-digit SIC manufacturing industries in 1992 is identical to the set used in 1987. See www.census.gov/prod/2/manmin/mc92-r-1.pdf.

[24]See Census (1992) at http://www.census.gov/prod/2/manmin/mc92-r-1.pdf.

– A merge of the unique five-digit SIC codes from the PD concordance into the official list from Census (1992) reveals that 1400 codes match exactly and that they are all in manufacturing. The largest portion (24) of the 62 in the official list but not in the PD concordance end in "9", and their descriptions indicate they are generally receipts for contract work on the good categorized by the first four digits. Code 22579 in the PD file, for example, is "contract and commission receipts for knitting only or knitting and finishing weft (circular) knit fabrics". Code 22573, which appears in both the PD and the official list, by comparison, is "finished weft (circular) knit fabrics, excluding hosiery".

– There are 166 five-digit SIC codes that are matched to HS codes in the PD concordance but do not appear in the official SIC list. Of the 166, 102 end in "0" and 95 are in manufacturing. We suspect that the 102 codes ending in "0" are used to facilitate the matching of SIC and basecodes by capturing a range of goods spread across five-digit codes with the same four-digit root. For example, 20220 is in the PD file but not on the official list, and is described as "cheese, natural and processed, not specified as to kind", versus 20223 and 20224, both of which are in both the PD and the official list but which break cheese down into natural and processed cheese, respectively. All three of these codes map into the same basecode, 20223B00, which maps to HS codes beginning with 0406, i.e., "cheese and curd".

C2. 1997 Economic Census

The 1997 PD file maps seven-digit NAICS product classes to eight-digit (NAICS-based) basecodes and is available in the electronic appendix with filename pd97.csv.[25] We note the following:

• 841 unique basecodes are matched to a product class (PC) in the 1997 PD file, 763 of which are in manufacturing (i.e., begin with a "3"). Table 6 summarizes the distribution of these basecodes according to the number of seven-digit NAICS product classes into which they map. As a group, the eight-digit basecodes contain 451 unique six-character basecode roots, 388 of which are in manufacturing.

• 1559 unique seven-digit NAICS product classes are matched to an eight-digit basecode in the 1997 PD file, of which 1418 are in manufacturing. The official list of NAICS categories for the 1997 CMF encompasses 1469 seven-digit product classes in manufacturing.

C3. 2002 Economic Census

The 2002 PD file maps seven-digit NAICS product classes to eight-digit (NAICS-based) basecodes and is available in the electronic appendix with filename pd02.csv. We note the following:

---

[25] We thank Alvin Venning of the U.S. Census Bureau for providing us with a copy of the 1997, 2002 and 2007 PD files.

Table 6: Number of Product Classes per Basecode and Basecode Root (1997)

Product Classes	Overall		Manufacturing	
	Basecodes	Basecode Roots	Basecodes	Basecode Roots
1	576	143	518	105
2	130	91	120	80
3	55	71	50	65
4	24	44	21	40
5	25	38	23	35
6	7	16	7	15
7	6	12	6	12
8	5	7	5	7
9	4	6	4	6
10	3	4	3	4
11		4		4
12		4		4
13	1	1	1	1
14		2		2
16		1		1
17		1		1
18	2	2	2	2
19		1		1
24	1	1	1	1
30	1		1	
36		1		1
44	1	1	1	1
Total	841	451	763	388

Notes: Table displays distribution of basecodes and basecode roots according to the number of product classes into which they map, overall and for manufacturing.

- 832 unique basecodes are matched to a product class (PC) in the 2002 PD file, 754 of which are in manufacturing (i.e., begin with a "3"). Table 7 summarizes the distribution of these basecodes according to the number of seven-digit NAICS product classes into which they map. As a group, the eight-digit basecodes contain 450 unique six-character basecode roots, 388 of which are in manufacturing.

- 1,547 unique seven-digit NAICS product classes are matched to an eight-digit basecode in the 1997 PD file, of which 1,406 are in manufacturing. The official list of NAICS categories for the 2002 CMF encompasses 1,450 seven-digit product classes in manufacturing.

C4. 2007 Economic Census

The 2007 PD file maps seven-digit NAICS product classes to eight-digit (NAICS-based) basecodes and is available in the electronic appendix with filename pd07.csv. We note the following:

- 799 unique basecodes are matched to a PC in the 2007 PD file, 724 of which are in manufacturing (i.e., begin with a "3"). Table 8 summarizes the distribution of these basecodes according to the number of seven-digit NAICS product classes into which they map. As a group, the eight-digit basecodes contain 454 unique six-character basecode roots, 390 of which are in manufacturing.

Table 7: Number of Product Classes per Basecode Root

Product Classes	Overall		Manufacturing	
	Basecodes	Basecode Roots	Basecodes	Basecode Roots
1	567	143	509	105
2	132	90	122	80
3	53	73	48	67
4	25	42	22	39
5	25	39	23	36
6	7	19	7	17
7	6	11	6	11
8	4	4	4	4
9	4	6	4	6
10	2	5	2	5
11		2		2
12		4		4
13	2	2	2	2
14		1		1
15		2		2
17		2		2
18	1		1	
19	1	2	1	2
23	1	1	1	1
31	1		1	
37		1		1
43	1	1	1	1
Total	832	450	754	388

Notes: Table displays distribution of basecodes and basecode roots according to the number of product classes into which they map, overall and for manufacturing.

- 1,496 unique seven-digit NAICS product classes are matched to an eight-digit basecode in the 2007 PD file, of which 1,383 are in manufacturing. The official list of NAICS categories for the 2007 CMF encompasses 1,435 seven-digit product classes in manufacturing.

Table 8: Number of Product Classes per Basecode Root (2007)

Product Classes	Overall		Manufacturing	
	Basecodes	Basecode Roots	Basecodes	Basecode Roots
1	558	170	498	123
2	115	82	107	75
3	48	67	46	65
4	23	42	19	37
5	25	34	24	32
6	8	19	8	18
7	7	9	7	9
8	5	5	5	5
9	3	5	3	5
10	1	4	1	4
11		3		3
12	1	5	1	5
13	1	1	1	1
14		2		2
15		1		1
18		1		1
20	1	1	1	1
22	1	1	1	1
29	1	1	1	1
30	1		1	
36		1		1
Total	799	454	724	390

Notes: Table displays distribution of basecodes and basecode roots according to the number of product classes into which they map, overall and for manufacturing.