

Concording U.S. Harmonized System Categories Over Time*

Justin R. Pierce[†]

U.S. Census Bureau, Center for Economic Studies

Peter K. Schott[‡]

Yale School of Management & NBER

August 2011

Abstract

Monitoring changes to product classification systems is an important component of a wide range of empirical research. In this paper we develop an algorithm for concording periodic revisions to the ten-digit Harmonized System (HS) codes used by U.S. statistical agencies to categorize international trade since 1989. We use this algorithm to construct the first comprehensive concordance of HS codes over time, and show how this concordance can be extended to incorporate future revisions. We then characterize the extent of HS-code changes since 1989 and discuss how controlling for these revisions is critical for understanding the growth of U.S. trade. Lastly, we highlight the general applicability of the algorithm to other national and international product classification systems.

JEL classification: F1

*Schott thanks the National Science Foundation (grants SES-0241474 and SES-0550190) and the Yale School of Management for research support. We thank Julie Linden of the Yale University Social Sciences Library for generous help in securing the U.S. trade data. We thank Mayumi Hairston Escalante for advice and for generous provision of unpublished Census “obsolete-new” files. The research in this paper was undertaken while Pierce was at Georgetown University. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau or the National Science Foundation.

[†]4600 Silver Hill Road, Washington, DC 20233, email: justin.r.pierce@census.gov; telephone: 301-763-6469; fax: 301-763-5935.

[‡]135 Prospect Street, New Haven, CT 06520, tel: (203) 436-4260, fax: (203) 432-6974, email: peter.schott@yale.edu.

1. Introduction

Empirical researchers including Bernard, Jensen, Redding and Schott (2009), Goldberg, Khandelwal, Pavcnik and Topalova (2010) and Pierce (2011), increasingly use product-level data to study trends in exports, imports and domestic production. These data have been particularly useful for examining the extent to which firms' growth in output or trade is due to "intensive" versus "extensive" margins, i.e., the degree to which growth takes place within surviving products or via product adding and dropping. At the same time, national statistical agencies frequently update product classification systems to incorporate new goods, drop obsolete categories and harmonize their systems with other countries. Absent a proper concordance, it can be difficult for researchers to distinguish true product-switching from spurious changes to product mix associated with product re-classifications.

In this paper we present an algorithm for constructing a concordance among revisions of the Harmonized System (HS) product codes used to track U.S. exports and imports over time. HS codes have been used by the U.S. Census Bureau since 1989 and are updated frequently. Our algorithm matches revised codes to synthetic, time-invariant identifiers that follow "families" of related products. We use our algorithm to construct the first comprehensive concordance of U.S. HS codes over time, covering the period 1989 to 2009. In an electronic appendix, we provide the Stata code used to build the concordance, thereby allowing other researchers the means to customize it or to extend it to incorporate future revisions of HS categories.

Our concordance reveals that changes in HS codes are frequent and widespread, and that they affect product categories representing a substantial portion of trade value. Indeed, of the 16,836 (8,859) import (export) codes active in 2004, 7,503 (2,929) underwent revision between 1989 and 2004—the years examined in Bernard, Jensen, Redding and Schott (2009). Furthermore, these revised codes represent 59 and 43 percent of import and export value in 2004, respectively.

The prevalence and importance of product code changes in U.S. trade underscore the need for HS code concordances in the analysis of trade flows. Using our concordance to control for changes to product categories over time, for example, Bernard, Jensen, Redding and Schott (2009) show that most of the year-to-year change in U.S. trade – as well as adjustments to "shocks" such as the 1997 Asian financial crisis – occur along the intensive margin.

The algorithm is general enough to be used to create concordances of virtually any national or international product classification system over time. This includes other international trade product classification systems such as the European Union's Combined Nomenclature or the Tariff Schedule of Japan. Moreover, the algorithm can be employed to construct concordances over time for a variety of national or international production-based product classification systems such as the North American Industry Classification System (NAICS), International Standard Industrial Classification (ISIC) or the statistical classification of economic activities in the European Union (NACE).

The remainder of the paper is organized as follows. Section 2 provides a brief description of U.S. HS codes. Section 3 describes the data used to construct our concordance and Section 4 outlines the concordance algorithm. Section 5 describes the properties of a 1989 to 2004 HS-over-time concordance created using the algorithm from section 4. Section 6 shows the effect of using the HS-over-time concordance on the measurement of product-adding and dropping using year-over-year decompositions of U.S. exports as in Bernard, Jensen, Redding and Schott (2009). Section 7 describes the general applicability of the algorithm to other product classification systems. An electronic appendix on our personal websites provides concordance files in .csv format, as well as the Stata code used to generate the concordances.

2. Brief Description of HS Codes

U.S. HS codes are based on the Harmonized System established by the World Customs Organization (WCO). The WCO assigns 6-digit codes for general categories, and countries adopting the system then define their own codes to capture commodities at more detailed levels. In the United States, the most detailed level of disaggregation is ten digits. In this paper, we refer to ten-digit codes as “product” or “goods” categories. U.S. export codes—technically referred to as Schedule B codes—are administered by the United States Census Bureau (Census). U.S. import codes—technically referred to as Harmonized Tariff System (HTS) codes—are administered by the U.S. International Trade Commission (USITC). We refer to HTS and Schedule B codes together as “HS Codes” throughout this paper.

Changes to U.S. export or import product codes can occur via three routes: changes by the WCO to the official list of international six-digit prefixes; U.S. legislation that affects U.S. eight-digit codes (imports only); and changes by the Committee for Statistical Annotation of Tariff Schedules (known as the “484(f) Committee”) to statistical ten-digit codes.

HS codes are updated for several reasons. The WCO, for example, makes adjustment to the HS to reflect developments in technology and changes in trade patterns. In addition, the 484(f) Committee may split a single HS code into several new codes in order to report import or export data at a more detailed level. Similarly, producers may petition one of the official bodies noted above for code changes to obtain a higher profile for the goods they export or import.

A large number of changes in 10-digit U.S. HS codes can be attributed to the WCO’s revisions of 6-digit HS categories. The WCO has made three major revisions to the HS in 1996, 2002, 2007, with another revision planned for 2012. Each of these revisions resulted in hundreds of 6-digit HS categories being deleted, while hundreds of other 6-digit HS categories were added. The effect of the WCO’s revisions on the number of U.S. HS changes is apparent in Table 1, where a large number of HS changes are concentrated in WCO revision years.

3. Data

Each year, Census publishes documents outlining the HS codes that have become “obsolete” and the “new” codes that will take their place. We refer to these documents as Census’ “obsolete-new” files. For exports, HS code changes take effect annually in January; for imports, they can occur within as well as across years. Obsolete-new files for years before 1997 are available only in hard copy and were transcribed into electronic form as part of the construction of our concordance. These files as well as electronic versions of subsequent files were obtained from Mayumi Hairston Escalante at Census. The most recent obsolete-new files are currently posted on the Census website.

We use the terms “simple” and “complex” to describe the two basic changes to HS codes that can occur in a obsolete-new file. Simple changes make no adjustments to the actual items covered by a particular code, they just swap one ten-digit code for another. There are several possible reasons for a one-to-one re-numbering, including:

1. To align the Schedule B and HTS codes where Census finds their descriptions are the same;
2. To differentiate the Schedule B and HTS codes where Census has found them to be different;
3. To correct errors by reclassifying a commodity under a different subheading;
4. To maintain the level of statistical detail after a revision of the 6- or 8-digit codes; and

5. To accommodate a new numbering pattern, usually the result of another code being broken out.

In contrast to simple changes, complex changes alter the mix of items captured by a particular code. For these changes, the items formerly encompassed by one or more “obsolete” codes are distributed to one or more “new” codes. In 2002, for example, various types of waste oil, which previously were grouped with the fresh oils to which they were most similar, were given their own HS codes. As a result, the (now obsolete) former fresh oil product categories were linked to the new waste oil categories from which they emerged. Some new-obsolete files contain “blanket” mappings, our term for mappings that include codes ending in a series of X’s, e.g., 8486XXXXXX. These observations are dropped from our concordance, as we are unable to determine the specific HS codes to which they refer.

For each set of obsolete-new mappings in a particular obsolete-new file, we construct a synthetic HS code which we refer to as a “setyear” (`setyr` in our Stata code). This synthetic code records both the count of the change since the first change in 1989 and an identifier for when it takes place. Formally, for exports, it is defined as the count of the particular mapping plus the four-digit year in which the change occurs divided by 10,000. For imports, it is the count of the particular mapping plus six-digit year-month in which the change occurs divided by 1,000,000. The very first setyears for exports and imports, for example, are equal to 1.1989 and 1.198906.

Table 1 summarizes the number of obsolete-new mappings in the raw data for export and import codes, respectively. Results for export codes are displayed in the left panel while those for import codes are displayed in the middle and right panels. The first column of each panel notes the year-month in which the noted changes take place. The second and third columns report the total number of retired and replacement codes encompassed by the number of sets reported in column four. Note that the number of sets in column four of each panel is smaller than the numbers of HS codes in columns two and three because multiple codes are often involved in a particular change (i.e., a particular set). The fifth column reports the number of changes that are “simple” in the sense outlined above.

As indicated in the table, HS codes are updated unevenly in the sense that some years (e.g., 2002) encompass substantially more changes than others (e.g. 2000).

4. An Algorithm for Creating an HS Concordance

Concording HS codes over time is complicated by the existence of chains of HS-code changes across months and years, which we refer to as “family trees”. There are two basic types of family tree. We refer to the first case, displayed in Figure 1, generically as a “growing family tree”. In this case, code a from period t may become obsolete and be mapped to new codes b and c in period $t + 1$. Then, in period $t + 2$, codes b and c may become obsolete and be mapped to new codes e and f , and g and h , respectively. Our concordance of the period t to period $t + 2$ HS codes assigns a common synthetic code to all HS codes in a growing family tree. Such an assignment may result in potentially many more HS codes being mapped to a given synthetic code in the final year of the concordance than in the first year. In 1997, for example, 7802000000 is mapped to 7802000030 and 7802000060. In a 1996 to 1997 concordance, we would assign a single synthetic HS code to all of these actual HS codes. For this reason, it may be useful for some analyses to restrict a concordance to a narrower set of years than the 1989 to 2009 concordance provided below.

The second type of family tree, which we refer to generically as a “shrinking family tree”, is displayed in Figure 2. In this case, codes a and b , and c and d , from period t separately become obsolete and mapped to codes e and f , respectively, in period $t + 1$. Then, in period $t + 2$, codes e and

Table 1: HS Code Changes by Year-Month

Exports					Imports										
Date	Obsolete	New	Sets	Simple	Date	Obsolete	New	Sets	Simple	Date	Obsolete	New	Sets	Simple	
1989_01	234	310	157	92	1989_06	2	12	2	0	1999_01	81	88	53	16	
1990_01	156	201	96	60	1989_07	112	196	91	27	1999_07	54	70	33	5	
1991_01	186	313	131	34	1990_01	346	724	295	15	2000_01	16	29	13	0	
1992_01	37	60	29	9	1990_05	16	20	16	12	2000_03	11	30	11	0	
1993_01	64	126	60	19	1990_07	133	256	119	25	2000_04	10	17	7	0	
1994_01	109	181	77	25	1990_08	38	49	30	17	2000_07	6	13	6	1	
1995_01	137	205	113	63	1990_10	70	121	47	6	2000_12	24	45	24	3	
1996_01	787	1,071	532	349	1991_01	69	194	45	0	2001_01	119	113	55	1	
1997_01	216	232	145	107	1991_02	15	24	15	6	2001_07	19	25	9	3	
1998_01	128	138	101	76	1991_05	11	20	11	2	2002_01	1,122	1,542	874	595	
1999_01	23	29	22	17	1991_07	247	393	190	77	2002_07	86	84	66	49	
2000_01	6	15	6	0	1992_01	85	138	50	0	2002_08	5	10	5	0	
2001_01	16	9	7	0	1992_05	28	29	28	27	2003_01	26	44	20	0	
2002_01	717	1,031	531	323	1992_07	117	194	109	42	2003_02	1	2	1	0	
2003_01	97	87	81	74	1993_01	135	218	74	7	2003_04	5	4	4	3	
2004_01	11	14	10	5	1993_02	42	51	42	33	2003_07	45	67	37	11	
2005_01	43	82	38	8	1993_06	3	5	2	0	2004_01	46	38	23	2	
2006_01	3	4	2	0	1993_07	7	8	7	6	2004_02	5	7	4	0	
2007_01	1,140	1,030	821	631	1993_08	33	53	25	0	2004_04	4	4	2	0	
2008_01	64	68	65	61	1993_11	8	10	2	0	2004_07	44	87	37	1	
2009_01	15	15	11	4	1993_12	1	2	1	0	2005_01	42	72	39	11	
					1994_01	667	1,082	468	176	2005_07	32	45	26	9	
					1994_04	13	43	13	0	2005_11	4	8	4	0	
					1994_06	66	112	47	0	2006_01	19	38	19	0	
					1995_01	1,933	2,187	1,162	555	2006_03	2	2	2	2	
					1995_07	38	73	31	0	2006_04	4	5	4	3	
					1995_09	77	168	33	12	2006_06	49	58	9	0	
					1996_01	1,164	1,485	798	523	2006_07	63	59	35	0	
					1996_06	5	8	5	4	2007_01	2,026	1,896	1543	1,220	
					1996_07	4	12	4	0	2007_07	25	35	16	3	
					1996_11	18	31	18	3	2008_01	19	39	13	0	
					1997_01	148	198	107	66	2008_04	12	8	6	0	
					1997_02	11	11	11	11	2008_07	15	34	15	0	
					1997_06	18	33	18	3	2008_10	12	26	12	0	
					1997_07	231	319	190	89	2009_01	42	61	28	1	
					1997_08	55	65	33	1	2009_07	20	39	20	3	
					1998_01	52	85	47	18						
					1998_03	4	8	2	0						
					1998_04	3	3	3	3						
					1998_07	6	8	6	4						
					1998_08	9	23	9	0						

Notes: Table reports changes to export (left panel) and import (middle and right panel) HS codes in noted year-month. Obsolete is number of codes retired from prior year. New is number of codes replacing these retirements. Sets is a count of the overall number of obsolete-new matches. Simple refers to re-numberings of individual codes.

f become obsolete and are assigned to new code g . In this case, the number of HS codes mapped to the family’s common synthetic code declines over time. In 1997, for example, 8506800010 and 8506800050 are mapped to 8506800000. In a 1996 to 1997 concordance, we would assign a single synthetic HS code to all of these actual HS codes.

The algorithm we develop for concording HS codes between arbitrary beginning and ending year-months accounts for both types of family trees, as well as combinations of the two types. Though specific details about how the algorithm is implemented can be determined by examining the Stata code in the electronic Appendix, the basic steps are as follows:

1. Read in raw obsolete-new mappings;
2. Assign a single setyear to each obsolete-new mapping appearing in the raw files;
3. Choose a beginning and end year for the concordance;
4. Identify family trees extending between the beginning and end years of the concordance; and

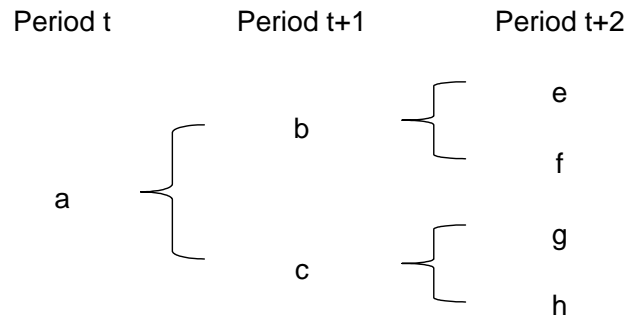


Figure 1: Growing Family Tree

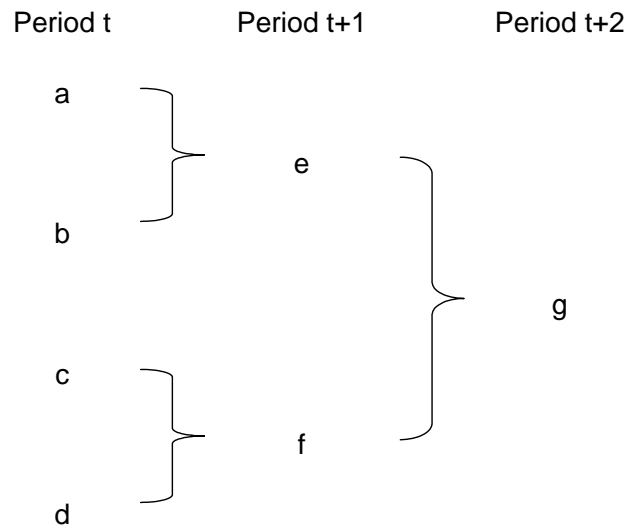


Figure 2: Shrinking Family Tree

5. Assign all members of a family tree the minimum setyear among family members within the time-frame of the concordance. Note that the part of the setyear after the decimal point identifies the year in which the family tree starts (i.e., period t in Figures 1 and 2 above). In the Stata code below, a separate variable (named `effyr`) identifies the year in which a particular obsolete-new mapping occurs. For example, in 1998 export code 8531800035 from 1997 is mapped to code 8531804000. Then, in 2002, codes 8531804000 and 8527908015 from 2001 are mapped into 8527908600. The `setyr` for the family is 1404.1998. The integer part of this `setyr` indicates that the first mapping in the family, from 8531800035 to 8531804000, is the 1404th mapping since 1989. The part after the decimal point indicates it occurs in 1998. The `effyr` for the two mappings are 1998 and 2002, respectively.

Step four is accomplished by successively merging subsequent obsolete-new mappings to all periods' obsolete-new mappings between the beginning and end years of the concordance. To bridge codes used from 1989 onwards, for example, the chained file is constructed as follows. First, merge the new codes in the 1990 file to the obsolete codes in 1991 file, dropping any codes that are unique to 1991. Second, merge the obsolete codes in the 1992 file to the new codes in the previously merged 1990-1991 file, again dropping any codes unique to 1992. And so on. Note that this successive merging has to be done starting with every year-month between the beginning and ending year-month because chains can begin in any year-month, and they would be missed otherwise given the dropping just mentioned. After these chains are created, they are appended into a single file and added to all obsolete-new mappings that are not parts of a chain.

5. A 1989-to-2004 Concordance

This section describes a 1989 to 2004 concordance constructed using the algorithm described above, which was employed in Bernard, Redding and Schott (2009). The first and second columns of Table 2 summarize total U.S. exports in 1989 and 2004 and the total number of HS product categories exported in those two years, respectively. Columns three and four provide analogous detail with respect to U.S. imports. As indicated in the table, (nominal) exports more than double while (nominal) imports more than triple over the fifteen-year interval. The number of pre-concordance export and import HS codes observed in each year of data grows 13 percent and 21 percent, respectively.

Table 2: Trade in 1989 and 2004

	Exports		Imports	
	Value	Codes	Value	Codes
1989	354	7,853	468	13,941
2004	818	8,859	1,460	16,836

Notes: Export and import values in billions of U.S. dollars. Number of codes refers to number of original ten-digit HS categories in the raw trade data.

Table 3 reports two decompositions of export and import codes. The first three rows of the Table show how many of the original HS codes in each year survive versus being replaced by synthetic codes. The remaining rows in the table decompose the actual plus synthetic codes that remain after the concordance into those which are common across years and those which are idiosyncratic to a particular year.

Of the 7,853 original HS codes appearing in the 1989 U.S. export data, for example, 1,917 are replaced by synthetic codes. Since the same synthetic code is often assigned to more than one

Table 3: Distribution of HS Codes in Matched 1989 to 2004 Trade Data

	Exports				Imports			
	1989	2004	1989	2004	1989	2004	1989	2004
Original HS Codes	7,853	100	8,859	100	13,941	100	16,836	100
Not Replaced by Synthetic Codes	5,936	76	5,930	67	9,383	67	9,333	55
Replaced by Synthetic Codes	1,917	24	2,929	33	4,558	33	7,503	45
Actual + Synthetic Codes After Concordance	7,162	91	7,157	81	12,527	90	12,534	74
Actual Codes	5,936	76	5,930	67	9,383	67	9,333	55
Common to both years	5,904	75	5,904	67	9,047	65	9,047	54
Appear in only one year	32	0	26	0	336	2	286	2
Synthetic Codes	1,226	16	1,227	14	3,144	23	3,201	19
Common to both years	1,221	16	1,221	14	3,057	22	3,057	18
Appear in only one year	5	0	6	0	87	1	144	1

Notes: Table decomposes the number of original HS codes in each year into those replaced by a synthetic code versus not, and total surviving HS plus synthetic codes in each year into noted sub-groups. All replacements are with respect to a 1989 to 2004 concordance. Even columns display values as a percent of first row in preceding column.

original code, the resulting concorded dataset contains 7,162 actual plus synthetic codes. Of these, 5,936 and 1,226 are actual and synthetic, respectively. Each of these totals, in turn, can be broken down into actual codes which are common to both 1989 and 2004 (5,904), synthetic codes that are common to both 1989 and 2004 (1,221), actual codes unique to 1989 (32) and synthetic codes that are unique to 1989 (5). These breakdowns reveal that the number of actual and synthetic export and import goods actually added and dropped between 1989 and 2004 is relatively small.

The values of U.S. exports and imports associated with each of the cells in Table 3 are reported in Table 4. As indicated below, synthetic codes account for the majority of import value in both 1989 and 2004.

Table 4: Distribution of Value in Matched 1989 to 2004 Trade Data

	Exports				Imports			
	1989	2004	1989	2004	1989	2004	1989	2004
Original HS Codes	353,765	100	817,936	100	468,012	100	1,460,160	100
Not Replaced by Synthetic Codes	222,293	63	467,854	57	196,051	42	600,941	41
Replaced by Synthetic Codes	131,472	37	350,082	43	271,961	58	859,219	59
Actual + Synthetic Codes After Concordance	353,765	100	817,936	100	468,012	100	1,460,160	100
Actual Codes	222,293	63	467,855	57	196,051	42	600,942	41
Common to both years	204,570	58	448,183	55	193,451	41	588,628	40
Appear in only one year	17,723	5	19,672	2	2,600	1	12,314	1
Synthetic Codes	131,472	37	350,082	43	271,962	58	859,219	59
Common to both years	131,405	37	347,416	42	270,859	58	855,029	59
Appear in only one year	67	0	2,666	0	1,103	0	4,190	0

Notes: Table decomposes U.S. export and import value according to whether HS codes are original or synthetic. All replacements are with respect to a 1989 to 2004 concordance. Values are in millions of U.S. dollars. Even columns display values as a percent of first row in preceding column.

Tables 3 and 4 also underscore the prevalence of changes in HS codes over time. As of 2004, 45 percent of import products and 33 percent of export products had been involved in an HS code change since 1989. Moreover, trade in products with code changes accounted for 59 percent of the value of U.S. imports and 43 percent of the value of U.S. exports in 2004.

We note that two features of Census' new-obsolete mappings complicate the identification of new product introductions (e.g., iPods). First, new HS codes always emerge from predecessor HS codes. Second, new HS codes' emergence may take place an unknown period of time after an underlying good has been introduced. Statistical agencies may wait to establish a new HS category

until it reaches a certain size or until manufacturers apply sufficient lobbying.

6. The Effect of the Concordance on Measurement of Product Adding and Dropping

In this section we illustrate the importance of controlling for HS code reclassifications when measuring product adding and dropping in U.S. export data. In Table 5 below, we present the value and share of U.S. exports associated with product adding and dropping, both with and without controlling for changes in HS codes over time. The top portion of the table reports results with unadjusted HS codes and the bottom portion reports results after controlling for HS code reclassifications using our concordance. We report these results for two-year periods between 1993 and 2003 as in Bernard, Jensen, Redding and Schott (2009).

The figures reported in Table 5 were generated using publicly-available product-level U.S. export data. At this level of data aggregation, product adding refers to an instance in which the U.S. does not export a product p in the beginning year of the period, but does export that product in the end year. Similarly, product dropping refers to an instance in which the U.S. did export a product in the beginning year, but did not export that product in the end year.

Table 5: Value of Exports Associated with Product Adding and Product Dropping: With and Without Concordance

		1993-1994	1994-1995	1995-1996	1996-1997	1997-1998	1998-1999	1999-2000	2000-2001	2001-2002	2002-2003
No Concordance	Added Products	11,934	63,662	108,544	15,735	25,009	4,338	1,484	4,593	92,395	4,587
	Added Products (% Beginning Year Exports)	2.6%	12.4%	18.6%	2.5%	3.6%	0.6%	0.2%	0.6%	12.6%	0.7%
	Dropped Products	11,028	52,010	102,890	16,547	24,907	4,114	1,954	4,920	101,289	5,357
	Dropped Products (% Beginning Year Exports)	2.4%	10.1%	17.6%	2.7%	3.6%	0.6%	0.3%	0.6%	13.9%	0.8%
	Dropped Products	360	53	963	713	522	220	477	208	683	420
With Concordance	Dropped Products (% Beginning Year Exports)	0.1%	0.0%	0.2%	0.1%	0.1%	0.0%	0.1%	0.0%	0.1%	0.1%
	Added Products	276	15	900	26	2,172	2,573	6	1,937	44	41
	Added Products (% Beginning Year Exports)	0.1%	0.0%	0.2%	0.0%	0.3%	0.4%	0.0%	0.2%	0.0%	0.0%

Notes: Table displays the value of U.S. exports associated with added and dropped products over two-year time periods where products are defined both without and with the HS-over-time concordance. Rows for "Added Products" and "Dropped Products" are measured in Millions of U.S. Dollars. Additional rows report the value associated with added and dropped products as a share of the total value of exports in the beginning year of each two-year period.

As can be seen in the table, the value of exports associated with product adding and dropping is greatly overstated in the “no concordance” case with unadjusted HS codes. The reason for this overstatement is intuitive—some of the products that appeared and disappeared during each two-year period were due to changes in HS codes, rather than the U.S. starting or stopping exporting those products. This phenomenon is particularly pronounced in time periods with many HS code changes such as 1995-1996 and 2001-2002. In the period from 1995-1996, for example, export data with unadjusted HS codes indicate that product adding (dropping) equaled 19 percent (18 percent) of the value of 1995 exports. After using the concordance, the shares of 1995 exports associated with product adding and dropping were 0.2 percent each.

This example illustrates the importance of properly controlling for changes in HS codes in research examining product-adding and dropping. Indeed, accounting for these changes in HS codes contributed to Bernard, Jensen, Redding and Schott’s (2009) finding that most of the year-

to-year changes in U.S. trade values occurred along the intensive margin associated with surviving products, rather than the extensive margin associated with product-adding and dropping.

7. Applicability of the Algorithm to Other National and International Product Classification Systems

The algorithm described in this paper can be used to create a concordance for any product classification system over time so long as the associated statistical agency periodically makes available mappings of obsolete and new codes. Given this information, the process of assigning product codes to families will be identical to that described above, and it should be fairly simple to adapt our Stata code to cover any idiosyncrasies.

For example, the algorithm could be applied to other international trade product classification systems such as the European Union’s Combined Nomenclature (CN) codes. Changes to the CN are published annually in the L-series of the Official Journal of the European Communities. Application of our method would permit evaluation of the EU’s product-level exports and imports on a consistent basis over time. Moreover, it is possible to apply the algorithm to more aggregated levels of international trade product classification systems, such as the 6-digit HS codes defined by the WCO.

Our algorithm can also be applied to track changes in production-based industry classification systems such as NAICS (North America) or NACE (EU). The U.S. Census Bureau, for example, publishes correspondence tables for the various revisions to NAICS, and these can be used to identify “families” of industry codes over time. The analogous information for NACE is published by Eurostat with each NACE revision.

8. Conclusion

Controlling for changes in product codes over time is critical in the growing body of research examining firms’ product-mix choices. In this paper, we present a concordance algorithm that can be used to track changes in product codes and generate time-consistent “synthetic” codes. We use this algorithm to generate the first complete concordance of changes in U.S. HS codes over time. We also describe the prevalence of changes in HS codes over time, underscoring the importance of controlling for these changes in empirical research. Lastly, we provide an electronic appendix containing the final concordance files, as well as Stata code that can be used to customize this and other product code concordances.

References

- Bernard, Andrew B., Stephen J. Redding and Peter K. Schott, (2010a). “Multi-Product Firms and Product-Switching.” *American Economic Review*. 100, 70-97
- Bernard, Andrew B., Stephen J. Redding and Peter K. Schott, (2010b). “Multi-Product Firms and Trade Liberalization.” *Quarterly Journal of Economics*, forthcoming.
- Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding and Peter K. Schott, (2009). “The Margins of U.S. Trade (Long Version).” NBER Working Paper 14662.
- Goldberg, Pinelopi K., Amit Khandelwal, Nina Pavcnik and Petia Topalova, (2010). “Imported Intermediate Inputs and Domestic Product Growth: Evidence from India.” *Quarterly Journal of Economics*, forthcoming.
- Pierce, Justin R., (2011). “Plant-Level Responses to Antidumping Duties: Evidence from U.S. Manufacturers.” *Journal of International Economics*, forthcoming.

A Appendix

This appendix describes the files contained in the electronic appendix available online at: http://www.som.yale.edu/faculty/pks4/sub_international.htm. All files are contained in a zip folder with filename `hs_concordance_20101020.zip`.

A1. *Stata Programs That Create the HS-Over-Time Concordances*

The files `hts.do` and `schedule_b.do` contain our algorithm for creating import and export HS concordances, respectively, for arbitrary beginning and ending year-months between 1989 and 2009. Those comfortable with Stata programming should find these files relatively easy to manipulate. Those unfamiliar with Stata programming can instead use one of the output files described below.

A2. *A Stata Program To Match the HS-Over-Time Concordances to U.S. Trade Data*

The file `trade_merge.do` is a Stata program that matches our HS-over-time concordances to publicly available U.S. trade data. Researchers may find this example useful when employing the concordances in their own research. In addition, this Stata program produces some of the output files described below.

A3. *A File Tracking “Raw” HS Code Changes*

Each Stata program requires as an input a data file containing the raw obsolete-new mappings discussed in the main text. These input files are named `sch_b_concordances_20100522_02.dta` and `hts_concordances_20100522_02.dta`, respectively, where 20100522 is the user-defined version date. The basic structure of these input files resembles the raw obsolete-new files; i.e., each set of obsolete HS codes is followed by the new set of HS codes into which they map. In this sense, researchers who wish to examine a simple record of changes to HS codes, as reported in the official obsolete-new releases may find these files useful. The files contain the following variables:

- `obsolete`: old HS codes that become obsolete as of effective date;
- `new`: new HS codes replacing the obsolete codes;
- `setyr`: synthetic code to which new and obsolete codes belong, as defined in main text; and
- `effyr`: date the mapping is effective.

A4. *Concordances for Changes in Import and Export HS Codes from 1989 to 2009*

The Stata programs described above produce the output files that can be used to concord HS codes in U.S. import and export data. Specifically, the code produces output files:

- `sch_b_concordances_20100522_BEG_END.dta`, and
- `hts_concordances_20100522_BEG_END.dta`

where `BEG` and `END` reflect beginning and end years (exports: 1989_2009) or year-months (imports: 198906_200907), respectively. These concordances include the same variables as the input files, but with `setyr` and `effyr` standardized across family trees, as described in Section 4 above. Variables in the concordance output files include:

- obsolete: old HS codes that become obsolete as of effective date;
- new: new HS codes replacing the obsolete codes;
- setyr: synthetic code to which new and obsolete codes belong, as defined in main text; and
- effyr: year (export) or year-month (import) in which the particular obsolete-new mapping first appears in the raw data.

A5. Simple Versions of the Concordances for Changes in Import and Export HS Codes from 1989 to 2009

The files `simple_hts_198906_200907.dta` and `simple_schedule_b_1989_2009.dta` provide the setyear for all HS codes that have experienced changes between 1989 and 2009 for imports and exports, respectively. The files have a simple two-column format where the first column reports the HS code that has experienced a change between 1989 and 2009 and the second column provides the setyear for that HS code. Researchers can merge this file by HS code with product-level trade data and easily assign a setyear to any HS codes that have been changed. HS codes not appearing in these output files are consistent across all years of the data.

In almost every case, this simple concordance is one-to-one, in the sense that each HS code maps to a single setyear. However, six (two) HTS (Schedule B) codes were listed as obsolete in one year and then "reappeared" as new codes in a later year with a different setyear. Each of these HS codes, therefore, has two setyears. The dates given in the setyear indicate the years in which they became active. These duplicate HS codes are: HTS - 2905492000, 5112196010, 5112196020, 5112196040, 5112196050, 7304390040; Schedule B - 481190900, 9027501000.

A6. A Record of HS Codes Associated with Each "setyear" Synthetic HS Code

The files `setyr_x_1989_2009.dta` and `setyr_m_1989_2009.dta`, provide a record of every HS code associated with every setyear that appears in the 1989-2009 concorded data. The first column of each file lists the setyears, sorted from low to high. Each additional column lists the actual HS codes appearing in a particular year of the trade data that should be replaced by the setyear. These actual HS codes also are sorted from low to high in each year. To concord U.S. trade data from 1989 to 2009, one would just replace all codes listed in the table with the synthetic setyear, and then collapse the data according to these setyears. HS codes not appearing in these output files are consistent across all years of the data.