NBER WORKING PAPER SERIES

U.S. EXPORTS, 1972-1994, WITH STATE EXPORTS AND OTHER U.S. DATA

Robert C. Feenstra

Working Paper 5990

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 1997

* The files on the accompanying CD-ROM were processed at the Social Science Data Service, Institute of Governmental Affairs, University of California Davis. I gratefully acknowledge the staff at SSDS, including Jean Stratford, Shelagh Matthews Mackay, Eric Anderson and Erick Eschker for their invaluable support, and Shunli Yao for his expert research assistance. The compilation of this data was supported by a grant from the National Science Foundation, held jointly with Harry Bowen and Robert Lipsey, who have provided excellent advice on the data and concordances. Thanks are also due to also due to those researchers who contributed datasets, including Gordon Hanson, James Levinsohn, Carsten Kowalczyk, Deborah Swenson, Matthew Slaughter, Robert Staiger and Frank Wolak.

Table of Contents

1. Introduction..... 1
2. Methodology..... 2
3. Comparison with Other Data Sources..... 3
4. Calculation of Domestic-based SIC Exports..... 12
5. Technical Information.....15

    5A. Schedule B Export Data, 1972-1977..... 17
    5B. Schedule B Export Data, 1978-1988..... 19
    5C. HS Export Data, 1989-1994..... 21
    5D. SIC Export Data, 1972-1994..... 23
    5E. Revision to SIC Import Data, 1989-1994..... 26
    5F. SITC (Rev. 1) Export Data, 1972-1977..... 27
    5G. SITC (Rev. 2 and 3) Export Data, 1978-1994..... 28
    5H. SIC and SITC Export Prices……..... 30
    5I. Schedule B Concordance, 1972-1977..... 32
    5J. Schedule B Concordance, 1978-1988..... 34
    5K. HS Concordance, 1989-1994..... 36
    5L. Schedule B to HS Concordance..... 37
    5M. SIC and SITC Names and Concordances..... 38
    5N. NAFTA Tariffs and Phase-outs..... 39
    5O. U.S. Foreign Trade Subzones Program..... 41
    5P. Materials Used by U.S. Industries, 1972-1992..... 43
    5Q. U.S. Automobile Data, 1971-1990..... 45
    5R. U.S. Antidumping Cases, 1981-1986..... 46
    5S. State-level Exports, 1991-1994.....…48
    5T. Foreign Affiliates Employment, 1981-1994..... 49
    5U. Outward Foreign Investment, 1983-1994..... 50

6. Updating the Export and Import Data.....51
References..... 53
Appendix A: Country Names..... 54
Appendix B: Units for Quantity..... 57

List of Tables

Table 1: Comparison of U.S. Merchandise Exports....……..... 4
Table 2: Comparison of U.S. Merchandise Exports to Canada…... 7
Table 3: Comparison of U.S. Manufactured Exports.....……..... 9
Table 4: Comparison of XSIC and SIC Exports, 1978.....…….. 11
Table 5: Comparison of U.S. Imports.....…… 14
Table 6: Files on the CD-ROM..... 16

1. Introduction

With the spread of economic activity around the globe, accurate data on trade between countries is of increased interest to researchers in the private and public sectors. This was recognized in the United States by the Omnibus Trade and Competitiveness Act of 1988, which mandated that increased attention be devoted to the collection and dissemination of international statistics. Resulting from the legislation was the National Trade Data Bank (NTDB) program, which distributes the data in a CD-ROM format. While the NTDB includes detailed information on U.S. imports and exports, the earliest year provided is 1989. The only source for earlier years of import and export data are magnetic tapes and printed media, which are often difficult to work with for research purposes.

The Omnibus Trade and Competitiveness Act of 1988 also mandated that the United States adopt the Harmonized System (HS) for the collection of import and export data by the Bureau of the Census. The HS is an international classification scheme for commodities. In earlier years, U.S. exports were collected under the "Schedule B" classification system, while U.S. imports were classified according to the Tariff Schedule of the United States Annotated (TSUSA). The fact that no information from either of these classification schemes is included in the NTDB means that researchers cannot easily make the link between earlier and later years. Furthermore, the Schedule B and TSUSA classifications each changed frequently, so that it is difficult to track the same commodity over successive years. These problems have meant that the Schedule B exports and TSUSA imports, which are a potentially very rich and detailed data source, have not been utilized in academic research to the extent that is possible.

The objective of this database is to provide complete information on the Schedule B and HS *export*s as collected by the Bureau of the Census for 1972-1994, as well as aggregations of the exports according to the Standard Industrial Classification (SIC) number and Standard International Trade Classification (SITC). Disk 1 of the NBER Trade Database contained complete data on U.S. *imports,* and included on this CD-ROM is a *revision* to the SIC imports for the years 1989-1994. In addition, the CD-ROM includes state-level exports and a number of other U.S. datasets contributed by various researchers, such as tariff reductions under NAFTA, antidumping cases, domestic and imported automobile data, materials consumption by industry, foreign trade zones, foreign investment, and programs used to construct and update the data.

2. Methodology for U.S. Export Data

The disaggregate export data by Schedule B or HS classification was initially obtained on magnetic tape for 1972-1989, or CD-ROM for 1990-1994 (sources are listed in section 5). These sources contained logical record files, where each record had the Schedule B or HS number, and quantitative information on: the value and quantity of exports; the destination country (designated by a Census number); whether the good was an export of domestic merchandise, or a re-export of foreign merchandise; the port of exit from the United States; and other information such as the method of shipment. Starting with these files, three changes were made:

(i) The value and quantity data for each Schedule B or HS commodity was summed over all ports of exit (and also over the methods of shipment, if these were included). This means that for each exported item and destination country, the total value and quantity of that item leaving the United States was measured (re-exports are not included on the CD-ROM);

(ii) The quantitative information on the value and quantity of exports was merged with descriptive information such as the Schedule B or HS description, and the export-based SIC or SITC classification code. This descriptive information was stored on each record, so that exported items can be identified by their alphabetic description, and export-based SIC and SITC classifications, rather than only the Schedule B or HS number;

(iii) Instead of the classification scheme for source countries used by Census, we instead introduced the United Nations (UN) system for country names and numbers.[1] In some cases, several source countries as identified by the Census classification were merged into a single country as defined by the UN.

The motivation for the first of these adjustments to the original data is that the value of exports by their port of exit is not that interesting a variable for economic analysis: the ultimate port of exit may have little relation to the location where this commodity was produced. By eliminating the disaggregation of the data according to port of exit, the size of the data files was reduced quite substantially.

---

[1] This is the six digit Standard Classification of Customs Areas and Territories, and is the same as that used by Statistics Canada in their World Trade Database. See Appendix A.

The second adjustment to the data is necessary because the Schedule B and HS numbers change frequently over time, so these codes alone cannot be used to identify a commodity over time. By including the alphabetic description of commodities, along with their export-based SIC and SITC classifications, it is much easier to track commodities over time. This descriptive information on each commodity was obtained from cumulative concordances, which were initially available for 1978-1988 and 1989-1994. A concordance for 1972-1977 was developed from the information in U.S. Bureau of the Census [4], for 1974. This information was supplemented by Schedule B numbers and alphabetic descriptions contained in U.S. Bureau of the Census [1], for 1972-1977. This time-consuming task was the only way that the quantitative information for 1972-1977 could be linked to the actual commodity names, and to the export-based SIC numbers.

The third adjustment made to the data was needed because the classification numbers for source countries used by the Census are unfamiliar, and also change over time. Typically, this classification scheme identifies countries in even more detail than the UN classification. For example, in 1988 the Census used three different numbers for France (distinguishing Andorra, Monaco, and the rest of France), three numbers for Italy (distinguishing San Marino, Vatican City, and the rest of Italy), four numbers for New Zealand (distinguishing various surrounding islands), etc. For each year, the Census numbers used for any given UN country were changed to that UN number and name, and the value and quantity of exports were summed within each UN country (and country sub-code). The complete list of UN numbers, along with country names and the corresponding Census numbers (for 1994) are given in Appendix A.

3. Comparison with Other Data Sources

To check the integrity of the data after making these adjustments, the total value of exports over all destination country and all commodities can be compared with the totals from other sources. Included in the database is the FAS (free alongside ship) export value, which reflects the value at the U.S. port of exportation, as shown in column 1 of Table 1. This is followed by the value of exports from the U.S. Bureau of the Census [1],[2],[3] which are the sources that are in principal identical to the original tapes/CD-ROM from which our data were obtained. We therefore expect the first two columns to be very close in magnitude, and they are identical or nearly so for most years.

Table 1: Comparison of U.S. Merchandise Exports ($ million, FAS value)

	Merchandise Exports		*Merchandise Exports plus Re-exports*		
	(1)	(2)	(3)	(4)	(5)=(4)-(3)
Year	Database	Bureau of the Census [1],[2],[3]	Database	Econ. Report of the President[b]	Difference
1972	48,876.2	48,876.3	49,675.6	49,900	224.4
1973	70,227.0	70,223.0	71,318.0	71,900	582.0
1974	97,143.4	97,143.5	98,506.3	99,400	893.7
1975	106,101.0	106,156.7	107,590.2	108,900	1,309.8
1976	113,321.7	113,323.1	114,995.6	116,800	1,804.4
1977	117,920.4	117,962.7	120,126.5	123,200	3,073.5
1978	141,121.2	141,154.2	143,658.3	145,800	2,141.7
1979	178,590.6	178,578.0	181,815.3	186,400	4,584.7
1980	216,592.2	216,592.2	220,704.9	225,600	4,895.1
1981	228,960.8	228,960.8	233,739.1	238,700	4,960.9
1982	207,157.6	207,157.6	212,274.5	216,400	4,125.5
1983	195,969.4	195,969.4	200,537.7	205,600	5,062.3
1984	212,057.1	212,057.1	217,888.1	224,000	6,111.9
1985	206,925.3	206,925.3	213,146.1	218,800[a]	5,653.9
1986	206,376.2	206,376.2	217,304.2	227,200[a]	9,895.8
1987	241,399.0	243,858.9	250,405.9	254,100	3,694.1
1988	308,013.5	310,346.3[b]	320,385.4	322,400	2,014.6
1989	339,432.9	339,432.9[b]	353,765.5	363,800	10,034.5
1990	374,536.6	374,536.6	392,975.8	393,600	624.2
1991	400,842.4	400,842.4	421,853.6	421,700	-153.6
1992	424,970.7	424,970.7	447,471.0	448,200	729.0
1993	439,295.3	439,295.3	464,858.4	465,900	1,041.6
1994	481,887.2	481,887.2	512,415.6	na	na

Notes na Not available
a Revised
b From *The National Trade Data Bank*, U.S. Dept. of Commerce, Economics and Statistics Administration, December 1991.

It should be noted that the Census collects data on two types of exports: the first *are exports of domestically produced goods,* and the second are *exports of foreign produced goods.* We shall refer to the first category as "exports," and the second category as "re-exports." A good is deemed to be a re-export if it passes through the United States without undergoing any significant processing. In this case, there are no duties charged on the original import of the good (and if these were paid, they can be refunded in what is called a "drawback"). These data on re-exports were included in the initial Census sources from which our database was constructed, but are *excluded* from the CD-ROM.[2] The reason for excluding re-exports is that they do not correspond to any production activity in the United States.[3]

Some official publications prefer to include re-exports when reporting total U.S. exports. In order to compare our data with such publications, in the third column of Table 1 we report the value of U.S. exports plus re-exports obtained from the original sources for our database; these values are nearly identical to those in U.S. Bureau of the Census [1],[2],[3]. In the fourth column we report the value of merchandise exports taken from *Economic Report of the President* (1995, Table B-108), which also includes re-exports, and in the final column we report the difference between the *Economic Report* values and those from our database. It can be seen that these figures are quite close for the later years, 1990-1993, but are significantly different in the earlier years. This difference is explained by U.S. exports to Canada. For years before 1989, exports to Canada are under-reported in the U.S. Census data, from which our database is drawn, whereas the *Economic Report* makes an adjustment to reflect this omission. Thus, the difference between columns three and four of Table 1 over 1972-1989 can be taken as an estimate of the exports to Canada that are omitted in our database and Census data.[4] For years after 1989, this problem has been corrected in the U.S. Census data, which is why the difference is quite small.

---

[2] The re-export data for 1972-1994 can be obtained from the author on request.

[3] The re-exports can be viewed as part of the balance of payments accounts, to offset these imports as they enter the United States. In the CD-ROM dealing with U.S. imports (Feenstra, 1996), we reported U.S. "imports for consumption," which *also excludes* goods that enter the United States duty-free, and are subsequently re-exported. Thus, excluding re-exports from the export data is consistent with our earlier treatment or imports. In contrast, the category "general imports" includes all goods as they enter the U.S., and whenever this is used it should be compared to U.S. exports plus re-exports.

[4] This difference is unusually small in 1987 and 1988. This is explained by a single entry into the exports to Canada in each of those years, with Schedule B number of 8189100 (and no description for this commodity). This entry takes on the extremely high values of $4.0 billion in 1987 and $8.3 billion in 1988. We can guess that this entry is an *ad hoc* attempt within the Census data to account for omitted exports to Canada in those years, and it

To verify that the difference in the final column of Table 1 is indeed due to under-reported exports to Canada, in Table 2 we report values of U.S. exports to that country. The first three columns are drawn from our database, and list the value of exports, re-exports, and their sum. These all represent the FAS (free alongside ship) values, which excludes transportation charges. For comparison, in the fourth column we report the value of U.S. exports to Canada taken from the Statistics Canada *World Trade Database*, which has been distributed on Disk 2 of the NBER Trade Database (see Feenstra, Lipsey, and Bowen, 1997). The Statistics Canada data corrects for the under-reported U.S. exports by instead using the Canadian import values from the U.S., but in so doing, it includes some of the transportation charges incurred in trade. Thus, the Statistics Canada value of U.S. exports to Canada reported column four of Table 2 should be taken as an *overstatement* of the FAS value, and actually lies between the FAS and the CIF (cost including freight) values. The difference between the Statistics Canada and our database values, reported in column five, should therefore be interpreted as an overestimate of the exports to Canada that are omitted in our database and Census data. This can be confirmed by comparing the last columns of Table 1 and Table 2, where the latter is somewhat higher reflecting the transportation charges included in the Statistics Canada data.[5]

From this comparison of the database with the *Economic Report* and Statistics Canada data, we conclude that the difference between our database value and these sources are indeed due to the under-reported U.S. exports to Canada in years before 1989. We have not attempted to correct this problem within the database, but simply alert the reader to this omission. The

---

explains why the difference in the last column of Table 2 is relatively small. For 1989, there is a single entry under the HS number 9890000000 ("adjustment for undocumented exports") of $5.6 billion. However, in later publications of this same data, this entry was revised upwards to $15.6 billion. Apparently, this revision explains the difference of $10.0 billion for 1989 reported in the last column of Table 2, and is a further *ad hoc* adjustment.

[5] The exception are 1987 and 1988, where the difference reported in column five of Table 2 is much higher than that in Table 1. As already explained in note 4, our database includes an very large adjustment for underreported U.S. exports to Canada in those two years. Surprisingly, however, the Statistics Canada data contains an unusual adjustment in exactly the same years. Thus, among the SITC codes by which the Statistics Canada data is organized, there is the single code SITC 9310, described as "special transactions and commodities, not classified by kind." For U.S. exports to Canada, this entry takes on values around $3 billion in the years 1985-1992, except for 1987 and 1988 when it has the huge values of $12.3 billion and $15.0 billion, respectively. These remarkable values explain why the U.S. exports to Canada in column four of Table 2, and the difference in column five, are so high. We conclude that both the U.S. Census and Statistics Canada data include *ad hoc* adjustments for underreported U.S. exports to Canada in 1987 and 1988, but that these adjustment are mutually inconsistent since they lead to completely different values in the last columns of Tables 1 and 2.

differences reported in the last column of Table 1 provide an estimate of the under-reported exports, and those in the last column of Table 2 are an overestimate.

Table 2: Comparison of U.S. Merchandise Exports to Canada ($ million)

Year	(1) Database Exports	(2) Database Re-Exports	(3)=(1)+(2) Exports + Re-exports	(4) Statistics Canada	(5)=(4)-(3) Difference
1972	12,058.7	333.0	12,391.7	13,513.9	1,112.2
1973	14,790.7	266.7	15,057.4	16,406.3	1,348.9
1974	19,543.1	368.0	19,911.0	21,899.9	1,988.9
1975	21,296.0	396.5	21,692.5	23,406.7	1,714.2
1976	23,555.6	510.2	24,065.9	25,700.0	1,634.1
1977	24,995.0	725.0	25,719.6	28,097.3	2,377.7
1978	27,586.9	763.8	28,350.8	31,484.8	3,134.0
1979	32,135.9	933.4	33,069.3	38,641.0	5,571.7
1980	33,968.2	1,291.0	35,259.2	40,988.6	5,729.4
1981	38,133.5	1,266.1	39,399.6	44,777.1	5,377.5
1982	32,415.2	1,216.8	33,632.0	38,673.3	5,041.3
1983	36,544.9	1,325.0	37,870.0	42,870.0	5,000.0
1984	44,515.1	1,775.5	46,290.6	53,079.5	6,788.9
1985	45,028.9	1,786.6	46,815.6	54,837.4	8,021.8
1986	42,986.1	1,857.3	44,843.4	55,781.9	10,938.5
1987	54,541.1	2,329.7	56,708.6	69,890.1	13,181.5
1988	65,910.3	2,836.2	68,746.5	79,858.6	11,112.1
1989	64,977.5	3,288.7	68,266.2	78,016.7	9,750.5
1990	78,218.0	4,748.6	82,966.5	85,261.6	2,295.1
1991	78,711.8	6,434.5	85,146.3	86,468.3	1,322.0
1992	83,217.5	6,938.5	90,156.0	92,574.0	2,418.0
1993	91,865.9	8,324.5	100,190.4	na	na
1994	103,642.8	10,611.9	114,254.7	na	na

Notes na Not available

As a further comparison, in the first column of Table 3 we report the export value from our database over just manufactured goods, identified as those commodities with a Standard Industrial Classification (SIC) number beginning with 2 or 3. In the fourth column we report the value of manufactured exports from the widely-used database of Abowd (1991). Surprisingly, the values from Abowd are much higher than those from our database, so we need to check whether this difference can be explained by factors we have already identified, or whether it reflects some other upward bias in the Abowd data.

To attempt to explain the difference between the database manufacturing exports and those from Abowd, we first note that there is one category of manufactured goods that are *not* included in the database, but might have been included by Abowd: these are the so-called "low valued exports" (less than $10,000), which are assigned the SIC number 3XXX in the Census data, and therefore cannot be allocated to a unique four-digit SIC code.[6] The effect of adding this category to the total from our database is shown in the second column of Table 3, and the effects of further adding re-exports in shown in the third column. This yields an export value that is still considerably smaller than reported by Abowd (1991), and the difference is reported in the last column. It is readily seen that this difference is considerably greater than that in the last columns either Tables 1 or 2. In other words, the large difference between the database exports and those from Abowd (1991) *cannot* be explained by either low-valued exports, re-exports, or under-reported exports to Canada: even after taking all these factors into account, the Abowd exports still exceed those from our database. Thus, there must be some other factor that accounts for the high value of exports in Abowd (1991). In fact, the differences between our database and Abowd can traced to a rather small number of SIC industries at the 4-digit level, in which it is especially difficult to measure or impute the value of exports, for the following reasons.

The export data from Census [1],[2] is not collected or concorded to a *domestic* SIC basis. Rather, this data is concorded to a so-called *export-based* SIC. This distinction is necessary because a number of industries in the domestic SIC are identified by their method of processing, and this may be unknown to the exporter. For example, the SIC industry 2011, meat packing plants, and industry 2013, sausages and other prepared meats, produce many of the same

[6] Included within SIC 3XXX are low-valued exports (less than $10,000 in value), as well as exports of goods donated by private agencies, and nonidentified military goods. Also omitted from the database manufactured exports was another SIC code 23XX, which consists of military clothing.

Table 3: Comparison of U.S. Manufactured Exports ($ million)

Year	(1) Database Exports	(2) Plus Low-Valued Items[a]	(3)=(1)+(2) Plus Re-exports	(4) Abowd (1991)	(5)=(4)-(3) Difference
1972	38,899.8	39,993.4	40,443.2	41,636.5	1,193.3
1973	52,225.1	53,636.6	54,293.4	55,446.8	1,153.4
1974	72,978.5	74,982.7	75,938.8	77,876.7	1,937.9
1975	80,561.3	82,543.7	83,598.3	86,691.5	3,093.2
1976	88,035.1	89,657.5	90,888.3	94,549.3	3,661.0
1977	92,966.0	94,707.8	96,388.4	101,787.1	5,398.7
1978	111,869.4	113,667.1	115,551.6	120,564.9	5,013.3
1979	141,390.5	143,465.0	145,909.2	156,513.7	10,604.5
1980	169,449.4	171,849.8	174,889.5	183,135.4	8,245.9
1981	181,939.2	184,626.7	188,292.4	211,576.9	23,284.5
1982	166,679.8	168,742.5	172,743.2	189,711.5	16,968.3
1983	158,052.2	160,019.2	163,902.2	181,854.4	17,952.2
1984	168,948.7	173,660.3	178,782.1	190,111.4	11,329.3
1985	169,644.6	175,750.4	181,209.8	194,182.2	12,972.4
1986	173,252.1	179,401.3	189,056.7	na	na
1987	199,420.4	208,437.6	215,863.7	na	na
1988	251,313.3	262,016.3	272,299.5	na	na
1989	280,807.3	292,272.9	304,323.2	na	na
1990	322,001.7	332,643.7	347,584.5	na	na
1991	350,545.0	362,137.9	380,168.7	na	na
1992	373,853.3	385,610.5	405,355.1	na	na

Notes na Not available

[a] Includes low-valued items (less than $10,000), private donations, nonidentified military equipment and military clothing.

products. What makes them different is that industry 2011 slaughters while industry 2013 uses purchased carcasses. Most of the exports are grouped into the *export-based* SIC code (or XSIC) 2011, with only a very small amount of exports in XSIC 2013. The magnitude of exports in each of these categories for 1978 is shown in Table 4.

Listed in the first and second columns of Table 4 are the value of exports from our database, along with the correspondingly *export-based* SIC (or XSIC) number. These values can be directly compared to Census [4], *U.S. Commodity Exports and Imports as Related to Output*. In that publication, the value of exports is compared to U.S. industry supply, so that the *export shares* defined as (export value)/(domestic supply) can be constructed. To achieve this, the SIC data on domestic output is concorded to the XSIC data on exports. For example, in the third column of Table 2 are listed the SIC numbers that correspond to the XSIC codes given in the first column: as already noted, most of the prepared meats are grouped into XSIC 2011, which incorporates SIC 2011+2013. The exception to this is SIC 2013A, natural sausage casings, which are given the XSIC number 2103 upon exportation. It is apparent that there is a very close correspondence between the export values in columns two and four. In column five, the export shares calculated over the corresponding XSIC and SIC groups are given.

In the last columns of Table 2, we show the 1978 export values for from Abowd (1991), which are listed according to *domestic-based* SIC numbers. Because exports are not collected separately by the SIC 2011 and 2013 categories, Abowd has imputed the exports to each in such a way that the export shares in the two SIC categories are equal. We will use the same kind of imputation in our database - assigning equal export shares across SIC categories - when there is not a one-to-one correspondence between the XSIC and SIC. The total value of exports in these two industries from Abowd's data is somewhat higher than in either our dataset or Census [4], which we cannot explain in this case. There are other cases, however, where the correspondence between the XSIC and SIC categories becomes more difficult, and it is these cases that account for much of the overstatement in the total export value from Abowd.

The case with the greatest overstatement of total exports is shown next in Table 4. The SIC industry 333 deal with the *primary* refining and smelting of copper (3331), lead (3332), zinc (3333), aluminum (3334) and materials not elsewhere classified (3339), whereas the SIC industry 3341 deals with the *secondary* smelting and refining of these metals. The type of processing

Table 4: Comparison of XSIC and SIC Exports, 1978 ($ million)

Database		Census [4]			Abowd (1991)		
XSIC	Value	SIC	Value	Share	SIC	Value	Share
Meat Packing Plants, Sausages and Prepared Meats							
2011	1,400.7	2011+2013	1399.8	3%	2011	1290.5	3.4%
2013	23.4	2013A	23.4	64%	2013	323.6	3.4%
Total	1,424.1		1,423.2			1,614.1	
Primary and Secondary Smelting and Refining of Nonferrous Metals							
3331	155.2	3331+33412	155.2	4%	3331	155.4	3.7%
3332	7.1	3332+33413	7.1	1%	3332	4.5	0.5%
3333	1.3	3333+33414	1.3	<0.5%	3333	1.0	0.2%
3334	134.5	3334+33417	134.5	2%	3334	99.7	1.9%
3339	1,300.0	3339+33415,6	1302.7	64%[a]	3339	609.3	58.4%
					3341	2212.6	58.4%
Total	1,598.1		1,600.8			4483.2	

Notes [a] Estimated from data provided in Census [4].

(through primary or secondary smelting) cannot be distinguished when these products are exported, and instead, the export data is simply collected by the type of metal, as listed in the first two columns of Table 2. The SIC numbers in the third column show clearly that the exports in XSIC 3331, for example, are drawn from both primary and secondary smelting of copper (SIC 3331+33412), exports in XSIC 3332 include both primary and secondary smelting of zinc (SIC 3332+33413), and similarly for the other metals. Note that the total value of exports from the database and Census [4] are in very close agreement.

Turning to the final columns, Abowd (1991) imputed exports into each of the primary smelting SIC categories 3331,2,3,4,9, using export shares that are quite close to those from Census [4]. The problem arises, however, in imputing an export value for *secondary* smelting (SIC 3341), where is no separate export data. Faced with the range of export shares for each type of metal in the final column of Table 2, Abowd evidently chose the *maximum value* of these

shares (58.4%), and applied this to the value of domestic production, to impute exports for secondary smelted metal of $2,212.6 million. When summing this together with the other imputed exports, the total value of exported nonferrous metals from Abowd is $4.5 billion, which exceeds the value of exported nonferrous metals in our database or Census [4] by nearly $2 billion. In other words, this single industry accounts for about 40% of the overstatement in the total value of manufacturing exports for 1978, shown in the final column of Table 3.

We have identified one other industry - aircraft (SIC 3721) - that has an export value from Abowd (1991) in 1978 that exceeds the value in our database by about $1.5 billion, and two industries - industrial organic chemicals (SIC 2869) and oil field machinery (SIC 3533) - where the export value from Abowd for 1978 exceeds that in our database by about $0.5 billion. In each of these cases, the value in our database is nearly exactly the same as that in Census [4], which indicates that these values are reliable. There are another twenty industries where the value from Abowd exceeds that in our database by at least $100 million in 1978, and of course, a number of industries where the value from Abowd is less than that in our database. Taken together, these various industries account for the bulk of the overstatement in Abowd's export values. We conclude that the number of industries in which the export values are greatly overstated are fairly small in number, and even in these cases, the *export shares* appear to lie within the range that occur in other industries. This means that studies that have used Abowd's data in share form (such as some chapters in Abowd and Freeman, 1991), may not be affected to any significant extent by the overstatement. But it is potentially more serious for studies that have relied on the absolute magnitude of exports, such as those that estimate the factor-content of trade and its impact on the labor market.

4. Calculation of Domestic-based SIC Exports

Included in the database are the value of exports by source country for all manufacturing industries at the 4-digit SIC level, using the 1972 version of the domestic SIC. These values were computed by first summing the export data according to the 4-digit XSIC values. Then a concordance between the XSIC and SIC values was applied to convert the export data to the SIC basis. This concordance was itself developed by starting with the concordance of Abowd (1991), and making adjustments to it based on two sources of information: Census [4], and a remarkable Appendix in Census [5] entitled "Principal Differences Between the SIC-Based Output and

Export Product Codes." As suggested by the title, it lists those cases in which the SIC and XSIC numbers differ for the same underlying commodity. This information is actually provided at the 5-digit SIC and XSIC level, though for the purpose of the concordance we worked at the 4-digit level. Then there are three types of matches that can occur between the SIC and XSIC numbers: (1) an "exact match" between one XSIC and one SIC; (2) one XSIC category corresponds to several SIC industries; (3) several XSIC categories correspond to several SIC industries.

In the first case, the "exact match" between a XSIC and SIC means that either these two categories really were identical, or that any differences were judged as too small to be important. This occurred for 314 cases out of the 437 4-digit SIC industries with positive exports. In the second case, the XSIC exports are allocated to the various SIC industries in proportion to their domestic output, which means that the export share in each of the SIC industries would be identical. This occurred for another 45 4-digit SIC industries. In the third case, we impute exports to various SIC industries according to the following rules: (a) the total value of exports across the XSIC and SIC categories is identical; (b) exports are allocated across the SIC industries in proportion to their domestic output. An exception is the smelting and refining of nonferrous metals, detailed in Table 2. In that case, the allocation of total exports across *primary* and *secondary* smelting is done according to rules (a) and (b). But given this allocation, the division of primary smelting exports across the various types of metals (SIC 3331,2,3,4,9) can be done in proportion to the exports in the like-numbered XSIC categories.

This methodology can be used to obtain exports by source country for the 4-digit SIC industries for 1972-1988, since our export data for those years include the XSIC numbers on a 1972 basis. However, for 1989 and later years, the export data includes only the XSIC numbers on a 1987 basis, which is the year in which the SIC codes were substantially revised. In order to convert the export totals for 1989 and later from the 1987 XSIC to the 1972 XSIC basis, we proceeded as follows. For the single year 1988, each Schedule B commodity in the export data includes a listing for both the 1972-basis XSIC and 1987-basis XSIC numbers. By aggregating across these, we obtain a "weighting matrix" for each source country between the 4-digit 1972-basis XSIC numbers and 4-digit 1987-basis XSIC numbers. For later years, we first sum the data according the 1987-basis XSIC numbers, then multiply by the weighting matrix for each source country, and then apply our concordance between the 1972-basis XSIC and 1972-basis SIC

numbers. This gives us the value of exports for the year in question organized by the 1972-basis SIC industries, and distinguishing the various destination countries for each industry.

In this calculation to convert the 1987-basis XSIC exports to the 1972-basis, the "weighting matrix" must include every 1987-basis XSIC code that appears in the 1989-1992 data. To achieve this, numerous corrections to the 1987-basis XSIC code within the 1978-1988 concordance were made. In addition, since our construction of the "weighting matrix" relies on 1988 trade values, when there were exports to *new* destination countries in later years, the matrix would not contain the information needed to allocate the 1987-basis XSIC exports to the appropriate 1972-basis XSIC number(s). So for these new destination countries, we allocated the 1987-basis XSIC values to the *single* 1972-basis XSIC industry that had the highest weight over those countries that actually received the export in 1988.

For Disk 1 of the NBER Trade Database, which dealt with U.S. imports, these corrections and imputations to the 1987-basis XSIC numbers had not been made. Accordingly, the SIC import data has been *revised* for these years to reflect these changes. In Table 6, we report the valued of imports over 1989-1992 for all merchandise trade, and then the previous and revised value of manufactured imports from our database. It can be seen that the revision has raised the overall level of SIC imports by about $15 billion annually. The revised SIC import data for 1989-1992 is included on this CD-ROM. Also included are the programs used to construct the SIC export and import data, as described in section 6, and these programs can be used to update the SIC export and import data for years later than 1992.

Table 5: Comparison of U.S. Imports ($ million)

	All Merchandise (Customs Value)			*Manufactured Imports (CIF)*	
Year	Database	Bureau of the Census [1],[2]	Econ. Report of the President	Database (Previous)	Database (Revised)
1989	468,012.0	468,012.0	473,200	396,746.1	411,144.7
1990	490,553.7	490,553.7	495,300	407,020.3	421,899.4
1991	483,027.9	482,083.1	488,500	405,566.3	420,556.0
1992	525,091.4	525,260.2	532,700	441,935.7	459,499.5

A word should also be said about the thirteen SIC industries in which there are zero values for exports.[7] In these cases, there may actually be some exports, but because they are not identified as belonging to this SIC industry when they cross the border, they have actually been aggregated with some other, related 4-digit SIC industry. An example is "marking devices" (SIC 3953), for which no distinct export category exists in most years.[8] It is most likely that these exports are included within "pens, mechanical pencils and parts" (SIC 3951). Thus, in each of the SIC industries with zero exports in the database, these values are not actually zero, but are presumably small, and have been included within some other related industry.[9]

5. Technical Information

All of the data, concordances, and documentation in the export database are stored as ASCII files. The data and concordances have the extension *.ASC, and in all cases are logical record files. For each datafile (or group of files) and each concordance, there is documentation with the identical name, but the extension *.TXT. The documentation is the same as that found on the following pages. The ASCII datafiles are kept to a maximum of 10 megabytes in size, so that they can be transferred more easily and can be viewed using conventional text editors. This means that the Schedule B and HS export data for each year is broken into four separate files, with _1, _2, _3, _4 appended onto the file names. This disaggregate export data includes both values and quantities, so that unit-values(=value/quantity) can be constructed. Disk 1 of the NBER Trade Database contained complete data on U.S. imports, and included on this CD-ROM is a *revision to* the SIC imports over 1989-1994. In addition, the CD-ROM includes state-level exports and a number of other U.S. datasets contributed by various researchers. Not listed in

---

[7] In these cases exports are zero in all years, but in another two cases (SIC 3362 and 3369) zero exports occur only in the years 1989-1992, because the concordance in those years did not identify these SIC industries.

[8] For the early years, 1972-1977, there was a Schedule B number assigned to this industry. But because there was no such Schedule B or Harmonized system number in later years, we re-assigned those exports for 1972-1977 to the industry "pens, mechanical pencils and parts" (SIC 3951). The value of exports was $3-12 million in those years.

[9] A special case occurs for "fresh or frozen packaged fish and seafoods" (SIC 2092), which is difficult to distinguish from the non-manufactured product "finfish" (SIC 0912). The export-based SIC code 2092 should be divided between *both* the categories SIC 2092 and SIC 0912, but we have no basis on which to make this division. Following Abowd, we have allocated the export-based SIC category 2092 *entirely* to the SIC code 2092, whereas on the import side the SIC code was given a *zero* value. Thus, the value of exports in SIC 2092 clearly overstates the true value, whereas the value of imports in SIC 2092 understates the true value.

Table 6 are several programs included on the CD-ROM that can be used to construct and update the SIC export and import data; these programs are discussed in section 6 of this paper.

Table 6: Files on the CD-ROM

Directory and File	Description
Schedule B export data, 1972-1988:	
EXP72_1.ASC, EXP72_2.ASC, EXP72_3.ASC,	1972 Export data by Sched. B commodity
EXP72_4.ASC **....**	(same for 1973-1988 years)
EXP72_77.TXT	Documentation for 1972-1977 Sched. B data
EXP77_88.TXT	Documentation for 1972-1977 Sched. B data
Harmonized System export data, 1989-1994:	
EXP89_1.ASC, EXP89_2.ASC, EXP89_3.ASC	1989 Export data by HS commodity
EXP89_4.ASC **.....**	(same for 1990-1994 years)
EXP89_94.TXT	Documentation for 1989-1994 HS data
SIC (1972 basis, 4-digit) export data, 1972-1992:	
SIC72.ASC,....,SIC92.ASC	SIC exports by destination country and year
SICPRICE.ASC, SICPRICE.TXT	SIC export prices and documentation
SIC58_92.ASC	SIC imports, exports, & shipments, 1958-1992
SIC.TXT	Documentation for 1972-1992 SIC data
Revised SIC (1972 basis, 4-digit) import data, 1989-1992:	
IMPSIC89.ASC,...,IMPSIC92.ASC	Revised SIC imports by country, 1989-1992
SITC (Rev. 1, 2 and 3, 5-digit) export data, 1972-1994:	
SITC72.ASC,....,SITC77.ASC	SITC Rev. 1 exports by destination and year
SITC78.ASC,....,SITC94.ASC	SITC Rev. 2 & 3 exports by destination and year
SITC7277.TXT, SITC7894.TXT	Documentation for SITC export data
SITPRICE.ASC, SITPRICE.TXT	SITC export prices and documentation
Concordances:	
CON72_77, ASC, CON72_77.TXT	Concordance for 1972-1977 Sched. B data
CON78_88, ASC, CON78_88.TXT	Concordance for 1978-1988 Sched. B data
CON89_94.ASC, CON89_94.TXT	Concordance for 1989-1994 HS data
SCH_B_HS.ASC, SCH_B_HS.TXT	Concordance between Sched. B and HS
SIC_XSIC.ASC, SIC_XSIC.TXT	Concordance between SIC and XSIC
SITCR1_2.ASC, SITCR2_3.ASC, etc.	Concordances between SITC Revisions
Other:	
COUNTRY.ASC, COUNTRY.TXT	List of country names and codes
SICNAME.ASC, SITCREV*.ASC	Industry/product names for SIC and SITC
UNIT7288.TXT, UNIT8994.TXT	Units of quantity used in EXPASC
NAFTA_*.ASC, NAFTA.TXT	NAFTA Tariffs and Phase-outs
ZONES.TXT	U.S. Foreign Trade Subzones Program
MAT7292.ASC, MAT7292.TXT	Materials Used by U.S. Industries, 1972-92
AUTOAS.ASC, AUTOS.TXT	U.S. Automobile Data, 1971-1990
DUMP1.ASC, DUMP2.ASC	Antidumping Cases in the U.S., 1980-1986
STATE.ASC, STATE.TXT	State-level Exports, 1991-1994
FDIEMP1.TXT, FDIEMP2.TXT	Employment of U.S. Foreign Affiliates
FDIOUT*.GZ, FDIOUT*.TXT	Outward FDI from the U.S., 1983-1994

5A. Schedule B Exports, 1972-1977 (EXP*.ASC)

For 1972-1977, exports to the United States at a disaggregate level were measured according to the Schedule B classification. These datafiles contain the U.S. export data according to Schedule B number, distinguished by destination country, and including both quantitative information about exports and descriptive information about each commodity.

The files EXPYR_1.ASC, EXPYR_2.ASC, EXPYR_3.ASC and EXPYR_4.ASC contain U.S. export data for 1972-1977, sorted by Schedule B number, with YR = {a two digit number in the range 72-77}. The first of these files, EXPYR_1.ASC, includes commodities with a Sched. B number beginning with 0-5; the second, EXPYR_2.ASC, contains those commodities with a Sched. B number beginning with 6; the third, EXPYR_3.ASC, contains those commodities with a Sched. B number beginning with 7; and the fourth, EXPYR_4.ASC, contains those commo-dities with a Sched. B number beginning with 8-9.

Record Layout:

The variables included in EXP*.ASC are:

columns 1-7	-	Schedule B number
columns 9-14	-	Country code United Nations codes are used (see Appendix A)
columns 16-23	-	Country name
columns 25-29	-	Revision 1 SITC number (5 digit)
columns 31-38	-	1972 export-based SIC number (8 digit)
columns 40-42	-	Units of quantity (see Appendix B)
columns 44-54	-	Quantity
columns 56-65	-	Value of exports (dollars)
columns 67-86	-	Description of Schedule B commodity (uses first 20 characters of Description in CON72_77.ASC)
columns 88-89	-	Year

Missing Values:

Missing values for any alphabetic variable are indicated by a blank field, as occurs especially for the Units of quantity, indicating that either the units could not be measured, or were simply missing. When the units could not be measured, there will be a zero value for Quantity, but positive entries for Value.

Special Considerations:

(i) The 1972 export-based SIC numbers are not the same as the SIC numbers used to identify U.S. industries. This is because industries in the United States are sometimes defined in terms of the processing that occurs in them, whereas the method of processing may not be known to the exporter. As a result, a condensed set of SIC numbers called "export based SIC" are used, and these are included in the EXP*.ASC files.

(ii) It should be noted that the Schedule B numbers for any commodity change over time, so the only sure way to keep track of a given commodity is by its full alphabetic description.

(ii) The Schedule B numbers used in these files for 1972-1977 differ very substantially from the Schedule B numbers used after 1977, even though they are referred to by the same name. The Schedule B numbers for 1972-1977 are similar to those used for the SITC, Revision 1; whereas the Schedule B numbers for 1978-1988 are similar to those used in the Tariff Schedules of the United States.

Related Files:

(i) The file CON72_77.ASC contains a complete list of the Schedule B numbers for 1972-1977, along with various information about that commodity.

(ii) The source country for each exported commodity is identified by the name and United Nations (UN) code. The complete list of names and UN codes, along with a correspondence to the country codes used by the U.S. Census, is provided in Appendix A.

Size: Each file EXP*.ASC is between 3 and 7 megabytes

Sources:

The data for 1972-1977 were obtained from:

National Archives and Record Administration, Annual Export Databank, EA622, Record group 29, Washington, D.C. [magnetic tapes], 1972-1977.

The same data are reported in printed form in:

U.S. Foreign Trade, Exports, Commodity by Country, Schedule B, FT410, Bureau of the Census, Washington, D.C., 1972-1977.

Additional information on the variables listed above can be obtained from:

Guide to Foreign Trade Statistics, Bureau of the Census, Department of Commerce, Washington, D.C., 1983.

5B. Schedule B Exports, 1978-1988 (EXP*.ASC)

For 1978-1988, exports to the United States at a disaggregate level were measured according to the Schedule B classification. These datafiles contain the U.S. export data according to Schedule B number, distinguished by destination country, and including both quantitative information about exports and descriptive information about each commodity.

The files EXPYR_1.ASC, EXPYR_2.ASC, EXPYR_3.ASC and EXPYR_4.ASC contain U.S. export data for 1978-1988, sorted by Schedule B number, with YR = {a two digit number in the range 78-88}. The first of these files, EXPYR_1.ASC, includes commodities with a Sched. B number beginning with 0-3; the second, EXPYR_2.ASC, contains those commodities with a Sched. B number beginning with 4-5; the third, EXPYR_3.ASC, contains those commodities with a Sched. B number beginning with 60-67; and the fourth, EXPYR_3.ASC, contains those commodities with a Sched. B number beginning with 68-69 and 7-9.

Record Layout:

The variables included in EXP*.ASC are:

columns 1-7	-	Schedule B number
columns 9-14	-	Country code United Nations codes are used (see Appendix A)
columns 16-23	-	Country name
columns 25-29	-	Revision 2 SITC number (5 digit)
columns 31-33	-	Revision 3 SITC number (3 digit)
columns 35-42	-	1972 export-based SIC number (8 digit)
columns 44-46	-	Units of quantity (see Appendix B)
columns 48-58	-	Quantity
columns 60-69	-	Value of exports (dollars)
columns 71-98	-	Description of Schedule B commodity
columns 100-101	-	Year

Missing and Imputed Values:

(i) Missing values for any alphabetic variable are indicated by a blank field, as occurs especially for the Units of quantity, indicating that either the units could not be measured, or were simply missing. When the units could not be measured, there will be a zero value for Quantity, but positive entries for Value.

(ii) The Rev. 2 SITC numbers were missing for many of the commodities with Schedule B numbers that were only used only in 1978. These Rev. 2 SITC numbers have been imputed by using the first 5 digits of the corresponding Schedule E number from CON78_88.ASC, or by using the description and SITC number for closely matched commodities.

Special Considerations:

(i) The 1972 export-based SIC numbers are not the same as the SIC numbers used to identify U.S. industries. This is because industries in the United States are sometimes defined in terms of the processing that occurs in them, whereas the method of processing may not be known to the exporter. As a result, a condensed set of SIC numbers called "export based SIC" are used, and these are included in the EXP*.ASC files.

(ii) It should be noted that the Schedule B numbers for any commodity change over time, so the only sure way to keep track of a given commodity is by its full alphabetic description.

Related Files:

(i) The file CON78_88.ASC contains a complete list of the Schedule B numbers for 1978-1988, along with various information about that commodity.

(ii) The source country for each exported commodity is identified by the name and United Nations (UN) code. The complete list of names and UN codes, along with a correspondence to the country codes used by the U.S. Census, is provided in Appendix A.

(iii) For years after 1988, exported commodities are identified by the Harmonized System (HS) numbers. A concordance of these numbers is contained in CON89_94.ASC, as described in CON89_94.TXT. A cross-reference between the Schedule B and HS numbers is contained in SCH_B_HS.ASC, as described in SCH_B_HS.TXT.

Size: Each file EXP*.ASC is between 4 and 8 megabytes

Sources:

The data for 1978-1988 were obtained from:

National Archives and Record Administration, Annual Export Databank, EA622, Record group 29, Washington, D.C. [magnetic tapes], 1978-1988.

The same data are reported in printed form in:

U.S. Exports, Schedule B, Commodity by Country, FT446, Bureau of the Census, Washington, D.C., 1978-1988.

Additional information on the variables listed above can be obtained from:

Guide to Foreign Trade Statistics, Bureau of the Census, Department of Commerce, Washington, D.C., 1983.

5C. Harmonized System Exports, 1989-1994 (EXP*.ASC)

For 1989-1994, exports to the United States at a disaggregate level were measured according to the Harmonized System (HS) classification. These datafiles contain U.S. export data according to their HS number, distinguished by destination country, and including both quantitative information about exports and descriptive information about each commodity.

The files EXPYR_1.ASC, EXPYR_2.ASC, EXPYR_3.ASC and EXPYR_4.ASC contain U.S. export data for 1989-1994, sorted by HS number, with YR = {a two digit number in the range 89-94}. The first of these files, EXPYR_1.ASC, includes commodities with a HS number beginning with 0-3; the second, EXPYR_2.ASC, contains those commodities with a HS number beginning with 4-7; and the third, EXPYR_3.ASC, contains those commodities with a HS number beginning with 80-84; and the fourth, EXPYR_4.ASC contains those commodities with a HS number beginning with 85-89 and 9.

Record Layout:

The variables included in EXP*.ASC are:

columns 1-10 - Harmonized System (HS) number
columns 12-17 - Country code
United Nations codes are used (see Appendix A)
columns 19-26 - Country name
columns 28-32 - Revision 2 SITC number (5 digit)
columns 34-38 - Revision 3 SITC number (5 digit)
columns 40-43 - 1987 export-based SIC number (4 digit)
columns 45-47 - Units of quantity (see Appendix B)
columns 49-59 - Quantity
columns 61-70 - Value of exports (dollars)
columns 72-91 - Description of HS commodity
(first 20 characters of Description from CON89_94.ASC)
columns 93-94 - Year

Missing Values:

Missing values for any alphabetic variable are indicated by a blank field, as occurs especially for the Units of quantity, indicating that either the units could not be measured, or were simply missing. When the units could not be measured, there will be a zero value for Quantity, but positive entries for Value.

Special Considerations:

(i) The 1987 export-based SIC numbers are not the same as the SIC numbers used to identify U.S. industries. This is because industries in the United States are sometimes defined in terms of the processing that occurs in them, whereas the method of processing may not be known to the exporter. As a result, a condensed set of SIC numbers called "export based SIC" are used, and these are included in the EXP*.ASC files.

(ii) It should be noted that the HS numbers for any commodity change over time, so the only sure way to keep track of a given commodity is by its full alphabetic description.

Related Files:

(i) The file CON89_94.ASC contains a complete list of the HS numbers for 1989-1994, along with various information about that commodity.

(ii) The source country for each exported commodity is identified by the name and United Nations (UN) code. The complete list of names and UN codes, along with a correspondence to the country codes used by the U.S. Census, is provided in Appendix A.

(iii) For years before 1989, exported commodities are identified by the Schedule B numbers. A concordance of these numbers is contained in CON78_88.ASC, as described in CON78_88.TXT. A cross-reference between the Schedule B and HS numbers is contained in SCH_B_HS.ASC, as described in SCH_B_HS.TXT.

Size: Each file EXP*.ASC is between 6 and 9 megabytes.

Sources:

The data for 1989 was obtained from:

National Archives and Record Administration, Annual Export Databank, EA645, Record group 29, Washington, D.C. [magnetic tape], 1989.

Data for 1990 and later years were obtained from:

U.S. Exports of Merchandise on CD-ROM [machine-readable data file] / prepared by the Bureau of the Census. - Washington: The Bureau [producer and distributor], 1990-1994.

The same data in printed form are reported in:

U.S. Exports, Harmonized System, Commodity by Country, FT447, Bureau of the Census, Washington, D.C., 1989-1994.

5D. SIC (1972 basis, 4-digit) Export Data, 1972-1992 (SIC*.ASC)

These files contain the U.S. export data according to a 4-digit Standard Industrial Classification (SIC) basis, for 1972-1992, and related information. Later years are not included because converting the data to an SIC basis requires data on domestic shipments by 4-digit SIC industries, which is available only up until 1992. Also, note that this conversion is done only for manufactured goods, which are those having SIC numbers beginning with 2 or 3.

(1) Conversion from Export-based SIC to Domestic SIC

Using the files EXP*.ASC for 1972-1992, the export data are summed according to the 4-digit Standard Industrial Classification (SIC) codes, using the 1972-basis export-based SIC for 1972-1988 and the 1987-basis export-based SIC for 1989-1994. As explained in the main text of the documentation, the export-based SIC codes - or XSIC - found in EXP.* differ from the domestic-based SIC because many domestic industries are classified according to the method of processing in the industry, which may be unknown for exports. Thus, the XSIC will sometimes combine several SIC codes, or overlap in more complicated ways. In order to convert the 4-digit XSIC data to a 4-digit SIC basis, a concordance between the XSIC and SIC was developed, as described in the main text. This concordance is stored in SIC_XSIC.ASC.

Record Layout for SIC_XSIC.ASC:

columns 1-5 - 4-digit SIC (1972 basis)
column 8-11 - 4-digit XSIC (1972 basis)
column 15 - Type of match between SIC and XSIC
1 = Several XSIC overlap with several SIC
2 = One XSIC corresponds to several SIC
3 = Exact match between one XSIC and one SIC
4 = No XSIC corresponding to that SIC

Size: SIC_XSIC has 550 records.

(2) 1972-basis SIC Export Data

Using this concordance, the export data summed according to 4-digit XSIC was converted to a 4-digit SIC basis for 1972-1988. For 1989 and later years, the XSIC was on a 1987-basis, and so it first had to be converted to a 1972-basis, as described below. These data are stored in the files SIC72.ASC,...,SIC92.ASC.

Record Layout for SIC72.ASC,...,SIC92.ASC:

columns 1-4 - 4-digit SIC number (1972-basis)
columns 6-13 - Destination country name
columns 15-20 - Country code (UN codes are used, as listed in Appendix A)
columns 22-33 - Export value (millions of dollars)
columns 35-36 - Year

Size: Each file SIC*.ASC is 1-2 megabytes.

Special Values:

Included in these files are the value of exports summed over all source countries, in which case the Country name is listed as WORLD, and the Country code is 100000.

(3) Conversion from 1987-basis to 1972-basis XSIC

In order to convert the export values for 1989 and later to the 1972-basis XSIC, the "weights" found in XSIC8772.ASC were used. Specifically, for each 1987-basis XSIC and each source country, the file lists the corresponding 1972-basis XSIC number(s) and the fraction of value for the 1987-basis XSIC coming from that 1972-basis XSIC. Thus, by construction the fractions shown will sum to unity over all the 1972-basis XSIC within each 1987-basis XSIC and each source country. This file has been constructed from the export values found in EXP*.ASC for 1988, since in that year the file CON78_88.ASC provides a complete listing of the 1972-basis XSIC and the 1987-basis XSIC numbers. For later years, the export data in EXP*.ASC has only the 1987-basis XSIC numbers. In order to convert these to a 1972-basis, one would first sum up the export values according to 4-digit 1987-basis XSIC; then merge the resulting file with XSIC8772.ASC; then multiply the 1987-based XSIC values by the "weights" found in XSIC8772.ASC (for each country), so as to obtain the export values on a 1972-basis XSIC.

Record Layout for XSIC8772.ASC:

columns 1-4 - 1987-basis XSIC number (4-digit)
columns 6-9 - 1972-basis XSIC number (4-digit)
columns 11-18 - Destination country name
columns 20-25 - Destination country code (UN codes are used; see Appendix A)
columns 27-35 - Weight for Export value

Size: XSIC8772.ASC has 42,216 records

(4) WORLD Import and Export Values, 1958-1992

The Value for WORLD imports and exports have been gathered together in the single file SIC58_92.ASC, which also makes use of the 1958-1974 value of exports from Abowd (1991). In order to splice these datasets together, the ratio of the WORLD export values from SIC72.ASC and from Abowd were computed for each 4-digit SIC industry, and then the export value for 1972 and earlier years from Abowd were multiplied by that ratio. In addition, SIC58_92.ASC includes: (i) WORLD imports by SIC industry, which have been *revised* for 1989-1994 as compared to those included on Disk 1 of the NBER Trade Database; (ii) the value of domestic shipments from Eric J. Bartelsman and Wayne B. Gray (NBER Productivity Database) for 1958-1991, which were updated to 1992 using data from the *Annual Survey of Manufactures* provided by Gordon H. Hanson, Department of Economics, University of Texas, Austin.

Record Layout for SIC58_92.ASC:

columns 1-4	- 4-digit SIC number (1972 basis)
columns 6-7	- Year
columns 9-20	- Import value (CIF, millions of dollars)
columns 22-33	- Export value (millions of dollars)
columns 35-43	- Industry Shipments (millions of dollars)

Size: SIC58_92.ASC has 15,750 records.

Missing Values:

For thirteen industries the value of exports are missing in SIC72.ASC,...,SIC88.ASC, and for fifteen industries the value of exports are missing in SIC89.ASC,...,SIC92.ASC, as indicated by zero values in SIC58_92.ASC. See the discussion in the main text (end of section 4) concerning missing values.

Related Files:

(i) The prices for 3-digit and selected 4-digit SIC exports can be found in SICPRICE.ASC.

(ii) The commodity names corresponding to each 4-digit manufacturing SIC category can be found in SICNAME.ASC.

(iii) The data in SI58_94.ASC is identical to the WORLD values found in the annual files SIC*.ASC for 1972-1992, and IMPSIC*.ASC for 1989-1992, while the import data for 1972-1988 is identical to that found on Disk 1 of the NBER Trade Database.

5E. Revised SIC (1972 basis, 4-digit) Import Data, 1989-1992 (IMPSIC*.ASC)

These files contain U.S. import data according to the 4-digit Standard Industrial Classification (SIC) basis, that have been *revised* for 1989-1992 as compared to the data included on Disk 1 of the NBER Trade Database. This revision was needed to take account of: (i) 1987-basis MSIC codes that did not appear in the 1972-1988 concordance of TSUSA commodities; (ii) imports of commodities from new countries, that did not supply those goods to the U.S. in 1988. For both these cases, the earlier conversion from the 1987-basis to the 1972-basis MSIC led to an understatement of total U.S. manufactured imports of about $15 billion annually over 1989-1992. Appropriate corrections for both these features have now been introduced into the programs that compute the SIC imports (see section 6), and the 1989-1992 data have been revised; no revision is needed for earlier years, when the data is initially assigned a 1972-basis MSIC number.

Included on the CD_ROM are the revised import data for 1989-1994, by source country:

Record Layout for IMPSIC89.ASC,...,IMPSIC92.ASC:

columns 1-4	- 4-digit SIC number (1972-basis)
columns 6-13	- Source country name
columns 15-20	- Country code (UN codes are used, as listed in Appendix A)
columns 22-33	- CIF import value (millions of dollars)
columns 35-46	- Customs import value (millions of dollars)
columns 48-49	- Year

Size: Each file IMPSIC*.ASC is 1-2 megabytes.

Special Values:

Included in these files are the value of imports summed over all source countries, in which case the Country name is listed as WORLD, and the Country code is 100000.

Missing Values:

For eighteen industries the value of imports are missing in SIC89.ASC,...,SIC92.ASC, as indicated by a zero value in SIC58_92.ASC. See the discussion in the main text (end of section 4) concerning missing value.

Related Files:

The import data in SI58_94.ASC is identical to the WORLD values found in the annual files IMPSIC*.ASC for 1989-1992, while the import data in SI58_94.ASC for 1972-1988 is identical to that found on Disk 1 of the NBER Trade Database.

5F. SITC (Revision 1, 5-digit) Export Data, 1972-1977 (SITC*.ASC)

Using the files EXP*.ASC for 1972-1977, the export data are summed according to the 5-digit Standard International Trade Classification (SITC), Revision 1 codes. The results are stored in SITC72.ASC,....,SITC77.ASC.

Record Layout:

The variables included in SITC*.ASC are:

columns 1-5 - SITC Rev. 1 number (5-digit)
columns 7-12 - Country code (UN codes are used, as listed in Appendix A)
columns 14-21 - Source country name
columns 23-33 - Value (dollars)
columns 35-36 - Year

Missing Observations:

There are some observations with missing values for the SITC code, which are indicated with a blank field.

Special Considerations:

Also included in these files are the variables listed above summed over all source countries, in which case the Country name is listed as WORLD, and the Country code is 100000.

Size: Each file SITC*.ASC is between 2 and 3 megabytes.

Related Files:

(i) The commodity names corresponding to each SITC code, for each revision of the SITC, are found in SITCREV1.ASC, SITCREV2.ASC, and SITCREV3.ASC.

(ii) Concordances between the various revisions of the SITC are found in SITCR2_1.ASC, SITCR2_3.ASC and SITCR3_2.ASC.

(iii) The data in SITC*.ASC is a summation of that found in EXP*.ASC. Thus, the details about each variable provided in the documentation to EXP*.ASC also applies to SITC*.ASC.

5G. SITC (Revision 2 and 3, 5-digit) Export Data, 1978-1994 (SITC*.ASC)

Using the files EXP*.ASC for 1978-1994, the export data are summed according to the 5-digit Standard International Trade Classification (SITC), Revision 2 and 3 codes. The results are stored in SITC78.ASC,....,SITC94.ASC.

Record Layout:

The variables included in SITC*.ASC are:

columns 1-5 - SITC Rev. 2 number (5-digit)
columns 7-11 - SITC Rev. 3 number
A 3-digit Revision 3 code is used in the 1978-1988 files, and
a 5-digit Revision 3 code for 1989-1994
columns 13-18 - Country code (UN codes are used, as listed in Appendix A)
columns 20-27 - Source country name
columns 29-39 - Value (dollars)
columns 41-42 - Year

Special Considerations:

(i) Also included in these files are the variables listed above summed over all source countries, in which case the Country name is listed as WORLD, and the Country code is 100000.

(ii) The Rev. 2 SITC numbers were missing for many of the commodities with Schedule B numbers that were only used only in 1978. These Rev. 2 SITC numbers have been imputed by using the first 5 digits of the corresponding Schedule E number from CON78_88.ASC, or by using the description and SITC number for closely matched commodities.

(iii) For 1978-1988, the data were initially organized on a 5-digit SITC Rev. 2 basis. Then the SITC Rev. 3 codes were imputed using the concordance contained in SITCR3_2.ASC. Because this imputation is difficult to perform when going from Rev. 2 to 3 (because Rev. 3 is more disaggregate at the 5-digit level), only the 3-digit Rev. 3 codes are used. This imputation may cause some irregular movement in the export values across 3-digit SITC Rev. 3 categories over 1988-1989. In cases where a 3-digit SITC Rev. 3 code could not be imputed, a 2-digit code is used instead, with zero appended as the third digit. Any given 3-digit SITC Rev. 3 number will generally appear multiple times in a year, corresponding to different 5-digit SITC Rev. 2 numbers. Thus, to compute the value of exports by 3-digit Rev. 3 code, it is necessary to sum over all records for each such code.

(iv) For 1989-1994, the data are initially organized on a 5-digit SITC Rev. 3 basis. In this case the concordance SITCR3_2.ASC gives a unique 5-digit SITC Rev. 2 code for each of the Rev. 3 codes. So these Rev. 2 codes were added into the file for each year, and then the records were re-sorted according to the Rev. 2 codes. As is apparent by inspection, each 5-digit Rev. 2 code can appear more than once. Thus, to compute the value of exports by 5-digit Rev. 2 code, it would be necessary to sum over all records for each such code.

Size: Each file SITC*.ASC is between 3 and 7 megabytes.

Related Files:

(i) The prices for 2-digit and selected 3-digit Rev. 3 SITC exports can be found in SITPRICE.ASC, as described in SITPRICE.TXT.

(ii) The commodity names corresponding to each SITC code, for each revision of the SITC, are found in SITCREV1.ASC, SITCREV2.ASC, and SITCREV3.ASC.

(iii) Concordances between the various revisions of the SITC are found in SITCR2_1.ASC, SITCR2_3.ASC and SITCR3_2.ASC.

(iv) The data in SITC*.ASC is a summation of that found in EXP*.ASC. Thus, the details about each variable provided in the documentation to EXP*.ASC also applies to SITC*.ASC.

5H. SIC and SITC Export Prices (SICPRICE.ASC, SITPRICE.ASC)

Most of the data included on this CD-ROM does not include export prices. In some files, there is an export value and an export quantity, which can be used to construct a unit-value (=value/quantity). But this unit-value will generally combine products that are somewhat heterogeneous, so that it cannot be treated as the price of a single product. The only source for actual prices of U.S. imports and exports is the Division of International Prices, U.S. Bureau of Labor Statistics. That office collects price by repeat interviewing of U.S. firms that import or export goods. These prices are then aggregated into SIC or SITC categories, and these indexes are reported in the files SICPRICE.ASC and SITPRICE.ASC, respectively.

The file SICPRICE.ASC contain export price indexes for the U.S. at the 3-digit 1972-basis SIC level (annual and quarterly) over 1980-1992, though data for some industries is over a shorter period. Selected 4-digit industries are also included. More recent data at the SIC level is not available. The record layout for the file is as follows:

columns 1-4 - SIC code (2, 3, or 4 digits)
columns 6-7 - Year
columns 9-11 - Quarter, preceded by Q (Q01, Q02, etc.)
columns 13-17 - Price index

Special Values:

The SIC code 2A3M is the aggregate for SIC 2 and 3, which is then the total for all manufacturing. It is reported on lines 2440-2477 of the file.

Size: SICPRICE.ASC has 5,944 records.

The file SITPRICE.ASC contains the export price indexes at the 2-digit SITC Revision 3 level (annual, quarterly and monthly for recent years) for various starting years up until 1995. Selected 3-digit industries are also included. The record layout for the file is as follows:

columns 1-3 - SITC code (1, 2, or 3 digits)
columns 5-6 - Year
columns 8-10 - Month, preceded by M (M01, M02, etc.)
columns 12-16 - Price index

Missing and Special Values:

(i) For many of the series, only quarterly data is available, and is included in months 03, 06, 09, and 12. The price index is missing for the other months, as indicated by a blank.

(ii) Some SITC codes are a residual category, made up of all other products for which price information is available but is not used elsewhere in the price indexes. These SITC codes are appended with R, such as 05R.

Size: SITPRICE.ASC has 22,477 records.

Related Files:

Since the 1972-basis SIC is used for the export price indexes, they can be used with the export values in SIC58_92.ASC, which are also on a 1972-basis SIC, but at the 4-digit level. Thus, the export values would have to be aggregated to a 3-digit level to be fully comparable to the export prices. Since the SITC price indexes are on a Rev. 3 basis, they can be used with the export values reported in SITC89.ASC,...,SITC94.ASC, which is also on a Rev. 3 basis but at the 5-digit level. For years before 1989, the SITC export values are reported on a Rev. 3 basis, 3-digit level, in SITC78.ASC,...,SITC88.ASC.

Source:

These data and the interview process are described in:

William Alterman, "Price Trends in U.S. Trade: New Data, New Insights," in Peter Hooper and J. David Richardson, eds. *International Economic Transactions*, Chicago: Univ. of Chicago Press and NBER, 1991, pp. 109-139.

The export price indexes are published by the Division of International Prices, Bureau of Labor Statistics. The data in this file was downloaded by anonymous FTP from the site stats.bls.gov.

5I. Schedule B Concordance, 1972-1977 (CON72_77.ASC)

For 1972-1977, exports to the United States at a disaggregate level were measured according to the Schedule B classification. The Schedule B numbers differ very substantially from those used after 1977, though they are still referred to by the same name. The Schedule B numbers for 1972-1977 are similar to those used for the SITC, Revision 1; the Schedule B numbers for 1978-1988 are similar to those used in the Tariff Schedules of the United States; and the Schedule E numbers for 1978-1988 are similar to those used for the SITC, Revision 2. The file CON72_77.ASC contains a complete list of the Schedule B numbers for 1972-1977, along with various information about that commodity.

Record Layout:

Each record in CON72_77.ASC contains the variables:

columns 1-7	-	Schedule B number
columns 9-16	-	1972 export-based SIC number
columns 18-22	-	Revision 1 SITC number
columns 24-27	-	End Use Classification
columns 29-30	-	First year Schedule B number is used
columns 32-33	-	Last year Schedule B number is used
columns 35-37	-	Units of quantity (see Appendix B)
columns 39-128	-	Description of Schedule B commodity
column 130	-	End of record indicator (the number 1)

Missing and Imputed Values:

(i) Some of the units and descriptions are missing, and are indicated by blank fields.

(ii) For those commodities with missing descriptions, or descriptions in lower-case, the SITC Rev. 1 numbers, 1972 export-based SIC numbers, and end-use classifications were imputed. This was generally done by using the same codes as appeared for the lagged commodity within the concordance.

Special Considerations:

It should be noted that the Schedule B numbers for any commodity change over time, so the only way to keep track of a given commodity is by its alphabetic description. One way to identify all the Schedule B numbers corresponding to a given commodity is to *sort* the data by the

description. However, the descriptions themselves also change slightly over time due to variations in spelling, etc. This means that the set of Schedule B numbers for a given description will not equal all the Schedule B numbers for that item, since there may be slightly different alphabetic descriptions for essentially the same item. Some attempt has been made to resolve this issue by using consistent alphabetic description for commodities over various years, but cases of slight changes in spelling still remain.

Related Files:

The first 20 characters of the Description has been used to identify the Schedule B commodities in the files EXPYR_1.ASC, EXPYR_2.ASC, EXPYR_3.ASC, and EXPYR_4.ASC for the years YR=72,73,...,77. The SIC and SITC numbers, and units of quantity in this concordance are identical to those in EXP*.ASC for 1972-77.

Size: CON72_77.ASC contains 4,175 records.

Sources:

For the years 1972-1977, the Schedule B numbers and commodity descriptions for these years were typed into the concordance from the annual source:

U.S. Foreign Trade, Exports, Commodity by Country, FT410, Bureau of the Census, Washington, D.C., 1972-1977.

The SIC and Rev. 1 SITC numbers corresponding to the 1972-1977 Schedule B numbers were obtained from:

U.S. Foreign Trade Statistics. Classifications and Cross-Classifications, 1974. Washington, D.C.: U.S. Dept. of Commerce, Bureau of the Census, December 1975.

5J. Schedule B Concordance, 1978-1988 (CON78_88.ASC)

For 1978-1988, exports to the United States at a disaggregate level were measured according to the Schedule B classification. The Schedule B numbers differ very substantially from those used before 1978, though they are referred to by the same name. The Schedule B numbers for 1972-1977 are similar to those used for the SITC, Revision 1; the Schedule B numbers for 1978-1988 are similar to those used in the Tariff Schedules of the United States; and the Schedule E numbers for 1978-1988 are similar to those used for the SITC, Revision 2. The file CON72_77.ASC contains a complete list of the Schedule B numbers for 1972-1977, along with various information about that commodity. The file CON78_88.ASC contains a complete list of the Schedule B numbers for 1978-1988, along with various information about that commodity.

Record Layout: The variables included in CON78_88 are:

columns 1-7	-	Schedule B number
columns 9-15	-	Schedule E number
columns 17-24	-	1972 export-based SIC number
columns 26-33	-	1987 export-based SIC number
columns 35-39	-	Revision 2 SITC number
columns 41-43	-	Revision 3 SITC number
columns 45-49	-	End-use classification
columns 51-52	-	First year Schedule B number is used
columns 54-55	-	Last year Schedule B number is used
columns 57-59	-	Units of quantity (see Appendix B)
columns 61-88	-	Description of Schedule B commodity
column 90	-	End of record indicator (the number 1)

Missing and Imputed Values:

(i) Some of the units of quantity are missing, and are indicated by blanks.

(ii) The first 8 commodities, with Schedule B numbers 0000001,...,0000008, are for U.S. goods returned, and do not have any SIC or SITC codes, as indicated by blanks. This is also true for the ninth commodity, number 0000009, which is for foreign merchandise returned to the U.S.

(iii) The Rev. 2 SITC numbers were missing for many of the commodities with Schedule B numbers that were only used only in 1978. These Rev. 2 SITC numbers have been imputed by using the first 5 digits of the corresponding Schedule E number, or by using the description and SITC number for closely matched commodities.

(iii) The 1987 export-based SIC numbers were revised for a number of commodities. The revised codes all end with the digits "XXXX".

(iv) The 1987 export-based SIC numbers contain a number of cases where "Z" is used to indicate that the classification is not known. For example, "23ZZZZZZ" indicates that the commodity belongs to the 2-digit SIC category 23, but more precise information is not available.

Special Considerations:

It should be noted that the Schedule B numbers for any commodity change over time, so the only way to keep track of a given commodity is by its alphabetic description. One way to identify all the Schedule B numbers corresponding to a given commodity is to *sort* the data by the description. However, the descriptions themselves also change slightly over time due to variations in spelling, etc. This means that the set of Schedule B numbers for a given description will not equal all the Schedule B numbers for that item, since there may be slightly different alphabetic descriptions for essentially the same item. Some attempt has been made to resolve this issue by using consistent alphabetic description for commodities over various years, but cases of slight changes in spelling still remain.

Related Files:

(i) The Description has been used to identify the Schedule B commodities in the files EXPYR_1.ASC, EXPYR_2.ASC, EXPYR_3.ASC, and EXPYR_4.ASC for the years YR=78,79,...,88. The SIC and SITC numbers, and units of quantity in this concordance are identical to those in EXP*.ASC for 1978-88.

(ii) For years after 1988, exported commodities are identified by the Harmonized System (HS) numbers. A concordance of these numbers is contained in CON89_94.ASC, as described in CON89_94.TXT. A cross-reference between the Schedule B and HS numbers is contained in SCH_B_HS.ASC, as described in SCH_B_HS.TXT.

Size: CON78_88.ASC contains 5,494 records.

Source: The concordance for 1978-1988 was obtained from:

U.S. International Trade Administration, COMPRO database [machine-readable file].

5K. Harmonized System Concordance, 1989-1994 (CON89_94.ASC)

In 1989 and later years, the Harmonized System (HS) of commodity classification has been used to measure disaggregate U.S. exports and imports. It replaces the Schedule B system used in earlier years for exports. The file CON89_94.ASC is a concordance that contains a complete list of the HTS numbers used identify U.S. exports over 1989-1994, along with various information about each of these commodities.

Record Layout:

columns 1-10 - Harmonized System (HS) number
columns 12-15 - 1987 export-based SIC 4-digit code
columns 17-21 - Revision 2 SITC code (5-digits)
columns 23-27 - Revision 3 SITC code (5-digits)
columns 29-33 - End-Use Classification
columns 35-36 - First year that this HS number was used
columns 38-39 - Last year that this HS number was used
columns 41-43 - Units of quantity (see Appendix B)
columns 45-94 - HS description
column 96 - End of record indicator (the number 1)

Missing and Imputed Values:

(i) Some units of quantity are missing, which are indicated by a blank.

(ii) The first 8 commodities, with HS numbers 0000000001,...,0000000008, are for U.S. goods returned, and do not have any SIC or SITC codes, as indicated by blanks. This is also true for the ninth commodity, HS 0000000009, which is for foreign merchandise returned to the U.S.

(iii) A number of export-based SIC codes within the concordance were revised to conform more closely to the commodity names. These revisions would need to be taken into account when merging the concordance with any future years of data. The revisions were as follows:

```
if hs='3100000000' then sic_87='2874'; if hs='3101000000' then sic_87='2874';
if hs='3102290000' then sic_87='2874'; if hs='3102240000' then sic_87='2874';
if hs='3102600000' then sic_87='2874'; if hs='3102800000' then sic_87='2874';
if hs='4203300000' then sic_87='2387'; if hs='8412100010' then sic_87='3764';
if hs='8536690060' then sic_87='3678'; if hs='8546900000' then sic_87='3089';
if hs='8703240020' then sic_87='3716'; if hs='8703240030' then sic_87='3716';
if hs='8703330020' then sic_87='3716'; if hs='8703330030' then sic_87='3716';
if hs='8710000030' then sic_87='3795'; if hs='8710000060' then sic_87='3795';
if hs='8713300090' then sic_87='3795'; if hs='9306900020' then sic_87='3761';
if hs='9306900060' then sic_87='3769'; if hs='9306900040' then sic_87='3483';
if hs='9306900080' then sic_87='3483';
```

Related Files:

(i) The first 20 characters of the HS description, along with the SIC and SITC numbers, and units of quantity from this concordance are identical to those used in the files EXPYR_1.ASC, EXPYR_2.ASC, EXPYR_3.ASC, and EXPYR_4.ASC for the years YR=89,90,...,94.

(ii) For years before 1989, exported commodities are identified by the Schedule B classification. A concordance of these numbers is contained in CON78_88.ASC, as described in CON78_88.TXT. A cross-reference between the Schedule B and HS numbers is contained in SCH_B_HS.ASC, as described in SCH_B_HS.TXT.

Size: CON89_94.ASC contains 8,978 records.

Source: U.S. International Trade Administration, COMPRO database [machine-readable file].

5L. Concordance between the Schedule B and Harmonized System (SCH_B_HS.ASC)

In 1989, the United States changed its system for collecting export (and import) data. Prior to that year, the export data were collected according to the Schedule B classification, whereas beginning in 1989 the data were collected according to the Harmonized System (HS). The file SCH_B_HS.ASC contains a cross-reference between the Schedule B codes for 1987, and the HS codes proposed at that time. The actual conversion from the Schedule B codes and the HS codes occurred in 1989, so this file can be used to obtain the linkage between a 1987 or 1988 Schedule B code, and the corresponding HS code in 1989. (HS codes that were introduced in years later than 1989 would not be reflected in this cross-reference).

Record Layout: Each record of SCH_B_HS.ASC contains:

columns 1-7	-	Schedule B number
columns 9-18	-	Harmonized System number
column 20	-	Asterisk in some records, indicating that the HS classification is taken taken from two or more Schedule B classifications.

Size: SCH_B_HS.ASC contains 14,058 records.

Source:

U.S. Dept. of Commerce, Bureau of the Census, Schedule B: Statistical Classification of Domestic and Foreign Commodities Exported from the United States, 1988 edition, Appendix, Table 1.

5M. SIC and SITC Names and Concordances (SICNAME.ASC, SITCREV*.ASC, SITCR1_2.ASC, SITCR2_1.ASC, SITCR2_3.ASC, SITCR3_2.ASC)

The file SICNAME.ASC contains the commodity codes and names for the Standard Industrial Classification (SIC), 1972 basis, for manufactured goods only (SIC codes beginning with 2 and 3). These SIC codes are used in the files SIC72.ASC,...,SIC92.ASC, SIC58_92.ASC. The organization of the records is self-explanatory. This file was scanned from the document: *Standard Industrial Classification Manual*, 1972, Executive Office of the President, Office of Management and Budget, Government Printing Office.

The file SITCREV1.ASC contains the commodity codes and names for the Standard International Trade Classification (SITC), Rev. 1, and was obtained from Harry P. Bowen, National Bureau of Economic Research. The file SITCREV2.ASC contains the codes and names for the SITC Rev. 2, and was obtained from Bruce A. Blonigen, Dept. of Economics, University of Oregon. The file SITCREV3.ASC contains the codes and names for the SITC Rev. 3, and was copied from: *U.S. Exports of Merchandise on CD-ROM* [machine-readable data file]/prepared by the Bureau of the Census. - Washington: The Bureau [producer and distributor], 1994.

Concordances between SITC Rev. 2 and 1, and between Rev. 3 and 2, are contained in SITC2_1.ASC and SITC3_2.ASC, respectively. Both of these files were obtained from Robert E. Lipsey, National Bureau of Economic Research. In general, it is easier to go from a later revision of the SITC to an earlier edition, rather than vice-versa, because the number of categories tends to expand over time. However, in some cases it was also necessary to go from Rev. 1 to Rev. 2, or Rev. 2 to Rev. 3. To achieve this, the concordances SITCR1_2.ASC and SITCR2_3.ASC was developed from SITCR2_1.ASC and SITCR3_2.ASC. In the first of these, SITCR1_2.ASC, each 5-digit Rev. 1 number can correspond to several different Rev. 2 numbers at the 5-digit or 4-digit level. In cases where a 5-digit Rev. 1 number does not correspond to a unique 5-digit Rev. 2 number, instead a 4-digit, 3-digit or more aggregate Rev. 2 number is used, with zero's appended onto it to achieve 5 digits. This problem was even more serious when going from Rev. 2 to Rev. 3, so in SITC2_3 only the 3-digit Rev. 3 numbers are used, with a similar convention for appended zero's.

All these files contain additional information on record layout in their first lines.

5N. NAFTA Tariff and Phase-Out Data (NAFTA_US.ASC, NAFTA_MX.ASC)

Contributed by: Carsten Kowalczyk,
The Fletcher School of Law and Diplomacy,
Tufts University,
Medford, MA 02155
E-mail: ckowalczyk@infonet.tufts.edu

This data should be referenced to the paper:

Carsten Kowalczyk and Donald Davis, 1996, "Tariff Phase-outs: Theory and Evidence from GATT and NAFTA," NBER Working Paper no. 5421.

The North American Free Trade Agreement (NAFTA) went into effect on January 1, 1994. Annex 302.2 of the Agreement identifies five general tariff phase-out categories specifying the number of equal-sized annual cuts to free trade (A: immediately, B: five stages, C: ten stages, C+: 15 stages, D: continued duty free) and some exceptional categories (B+: seven stages, B6 and B1: five stages with small initial reductions versus large initial reductions, C10: nine stages). The tariff schedules list products according to the Harmonized System and associate with each product a phase-out category and its 1991 base tariff.

The files NAFTA_US.ASC and NAFTA_MX.ASC list the imports into each of these countries (from each other, and from the world), along with the average 1991 *ad valorem* tariff rates for imports from the other country, and the average number of years for these tariffs to be phased-out under NAFTA. These are reported by selected SITC Rev. 3 categories at the 3-digit and 4-digit level. Both the *ad valorem* tariff rates, and the number of years for these to be phased-out, are computed as unweighted averages using sampled information from the detailed tariffs and phase-out schedules at more detailed levels, as follows.

Based on the 1991 United Nations Commodity Trade Statistics for Mexico and the United States, imports from each other and rest of world (and corresponding exports) are sampled at the 5-digit SITC level. Products accounting for relatively large shares of total trade within 2-digit categories were sampled first (>20% of 2-digit trade, >10% for large value categories SITC 7 and 8). Then 5-digit commodities were sampled within the omitted 2-digit categories that were not represented among the first set of products. For each 5-digit SITC code the corresponding six and eight digit Harmonized Codes were found using the UN (1986) classification. *Ad valorem*

tariffs and staging categories were gathered from NAFTA (1993) Annex, the latter as the number of years equal to the number of tariff reductions. Products with specific tariffs or tariff-quotas were not included. 5-digit SITC values of base rates and phase-outs were found by unweighted averaging across all relevant 6- and 8-digit Harmonized categories. The procedure results in 148 5-digit product lines for the United States and 685 lines for Mexico, with 56 common product categories. These commodities account for 34.6
percent of U.S. imports from Mexico and 15.4 percent of its imports from the world, and 38.5 and 40.1 percent of Mexico's imports from the U.S. and the world, respectively.

Record Layout:

The files NAFTA_US.ASC and NAFTA_MX.ASC have the following fields:

columns 1-23 - Harmonized System codes that were sampled from
columns 25-36 - Corresponding SITC Rev. 3 codes
columns 38-94 - SITC Rev. 3 commodity names
columns 96-105 - Imports from partner country (US$ 1,000)
In NAFTA_US this is imports from Mexico
In NAFTA_MX this is imports from the U.S.
columns 108-118 - Imports from the world
columns 121-125 - Average 1991 tariff (percentage)
columns 129-137 - Phase-out schedule
A = phase-out immediately,
B = phase-out in five years,
C = phase-out in ten years,
D = continued duty free.
When four numbers are given, the first is the number of sampled A cases; the second is B cases; the third is C cases; and the fourth is D cases.
columns 141-145 - Average number of years to tariff phase-out
This is computed as an unweighted average of the cases in the preceding columns, e.g. (0*A+5*B+10*C+0*D)/(A+B+C+D).

Special Considerations:

For Mexico, the import data in NAFTA_MX.ASC is stated freight-on-board (FOB), while for the U.S. the import data in NAFTA_US.ASC is cost-including-freight (CIF). In 1991, Mexico excluded maquiladoras trade from its merchandise trade and instead tabulated it as services trade, so that the import data in NAFTA_MX.ASC excludes maquiladora trade. In

contrast, U.S. merchandise data include trade with the maquiladoras: Mexico listed 1991 merchandise imports from the U.S. to be $25 billion while the U.S. listed 1991 merchandise exports to Mexico as $32 billion.

Sources:

North American Free Trade Agreement. 1993. Washington, D.C.: United States Government Printing Office.

United Nations Statistical Office. 1986. Standard International Trade Classification Revision 3. Statistical Papers Series M, No. 34/Rev. 3.

United Nations Statistical Office. 1992. Commodity Trade Statistics 1990-91. New York: United Nations.

5O. Data from the U.S. Foreign Trade Subzones Program (ZONES.TXT)

Contributed by: Deborah Swenson
Dept. of Economics
Univ. of California
Davis, CA 95616
E-mail: deswenson@ucdavis.edu

Source: Annual Reports of the Foreign Trade Zones Board to the Congress of the United States

This data should be referenced to the paper:

Deborah Swenson, 1997, "Explaining Domestic Content: Evidence from Japanese and U.S. Auto Production in the U.S.," in Robert C. Feenstra, ed. The Effects of U.S. Trade Protection and Promotion Policies. Univ. of Chicago Press and NBER.

The data in ZONES.TXT record the activities conducted within foreign trade subzones in the U.S. between 1984 and 1993 Each subzone operation involves the activity of a single firm operator. The data that follow track each firm's shipments to the U.S. as well as shipments for direct export. The data also indicate the quantities of domestic and foreign sourced items that were brought into the subzones by the zone operators.

The data are organized in three panels. In order to link the three data panels, the data need to be matched on the basis of the “ZONE IDENTIFIER” which is contained in each panel. The zone identifier contains a number and a letter. The number indicates the general purpose trade zone to which the subzone operator is attached. Since many general purpose trade zones have multiple subzone operators, the letter is used to distinguish each individual firm's operations.

Structure of the Data in ZONES.TXT:

PANEL 1 includes data on the foreign interactions of each subzone.

Receipts measure the inflows of “foreign status” merchandise to the subzone. (19 CFR 146.41/42) These are all foreign items for which customs entry has not been made.

Shipments indicate all outflows from the zone directly going to export.

PANEL 2 includes data on the domestic activities of each subzone.

Receipts in this panel indicate flows to the subzone of “domestic status” merchandise (19 CFR 146.43) This encompasses items of domestic origin as well as duty paid and duty free items.

In panel 2 shipments measures all outflows from the subzone that travel to a U.S. location.

PANEL 3 provides the name of the firm in each subzone. Panel 3 also includes the industry activity conducted within the subzone.

Notes:

All values are millions of dollars.

Non-existing values are indicated by a blank space.

5P. Materials Used by U.S. Manufacturing, 1972-1992 (MAT7292.ASC)

Contributed by: Gordon H. Hanson
Department of Economics
University of Texas
Austin, TX 78712
E-mail: hanson@mundo.eco.utexas.edu

This data should be referenced to the paper:

Feenstra, R.C. and G.H. Hanson, 1996, "Globalization, Outsourcing, and Wage Inequality," American Economic Review, May, 240-245.

The file MAT7292.ASC contains data on materials consumption by four-digit SIC manufacturing industries in the United States for the period 1972 to 1992. The data are from the "Materials Consumed by Kind" table (Table 7) in the 1972, 1977, 1982, 1987, and 1992 *U.S. Census of Manufactures* (data from Censuses after 1972 have been concorded to the 1972-basis SIC code). The data contained in the file show how much each four-digit manufacturing industry purchases from other manufacturing industries and the value of materials each industry purchases in total (the sum of manufacturing and nonmanufacturing materials) for each Census year.

The Census collects data on the purchase of about 1,000 different materials (these are the raw data that the BEA uses to construct input-output tables). It assigns each material a six- digit code, the first four digits of which correspond to the SIC industry in which the good was produced (see variable definitions below). Of course, not all industries purchase appreciable quantities of all materials. For each industry, the Census collects data on the materials that account for most material purchases by that industry (for most industries this means purchases in excess of $10,000). The Census uses surveys and other means to determine which materials belong in this set. For the typical four-digit industry, the Census shows the value of purchases for 10 to 15 different materials. The listing of material purchases has become more complete over time, so that in later Census years there are more observations per purchasing industry than in earlier Census years. For a given industry, then, the list of materials on which the Census collects data changes over time. This means that in some cases it is not possible to track the purchase of a specific material by a specific industry between Census years.

While the Census shows purchases of all types of materials -- manufacturing and nonmanufacturing -- the file MAT7292.ASC contains data on (1) the purchase of manufacturing materials, and (2) the value of total material purchases by each industry (the file excludes a detailed listing of nonmanufacturing materials). See the notes accompanying "Materials Consumed by Kind" in *Census of Manufacturers* for more details on the materials data.

Record Layout:

The file MAT7292.ASC contains the following variables:

columns 1-4 - *year*	Indicates the year on which the observation is taken.
columns 7-10 - *sic72*	Indicates the purchasing industry by 1972 SIC code. Data for Census years following 1972 have been concorded to the 1972 SIC code.
columns 13-16 - *mat72*	Indicates the source industry (the industry from which the material is purchased) by 1972 SIC code. The Census uses a six-digit materials code to identify the material being purchased. For most cases, the materials code corresponds to a single four-digit SIC industry. For some cases, however, the materials code groups together materials from several four digit-industries in the same two or three-digit SIC industry. In cases of the second type, the industry indicated in *mat72* is the corresponding two or three-digit SIC industry.
columns 18-28 - *matcost*	Indicates the nominal value of materials purchased by the industry given in *sic72* from the industry given in *mat72* in the given year.
columns 30-40 - *tmatcost*	Indicates the total nominal value of materials purchased (from manufacturing and nonmanufacturing industries) by the industry given in *sic72* in the given year.

Special Considerations:

For a few industries in each year, no detailed information on material purchases is available. In most of these cases, information is available on total material purchases. For these industries the entries for the variables *mat72* and *matcost* are both zero.

Size: MAT7292.ASC contains 27,067 records.

5Q. U.S. Domestic and Imported Auto Data, 1971-1990 (AUTOS.ASC)

Contributed by: James Levinsohn
Department of Economics
University of Michigan
Ann Arbor, MI 48109
E-mail: jamesl@econ.lsa.umich.edu

The data in AUTO.ASC was collected from the annual issues of Automotive News, and from various issues of Consumer Reports. It can be referenced to the article:

Berry, Steven, James Levinsohn and Ariel Pakes, 1995, "Automobile Prices in Market Equilibrium," Econometrica, 63(4), 841-890

Record Layout: Each record of AUTO.ASC contains the following variables, with the variables names included on the first record:

columns 1-6	-	*name*	Model name (abbreviation)
columns 9-12	-	*id*	Identification number for model
columns 15-16	-	*yr*	Year
column 19	-	*cy*	Number of cylinders
column 22	-	*dr*	2=Two wheel drive, 4=Four wheel drive
column 25	-	*at*	1=Automatic transmission, 0=otherwise
column 28	-	*ps*	1=Power steering, 0=otherwise
column 31	-	*air*	1=Air conditioning, 0=otherwise
columns 34	-	*drv*	1=Front wheel drive, 0=otherwise (the 1971-1974 data is missing)
columns 37-41	-	*price*	Price (dollars)
columns 44-47	-	*wght*	Weight (pounds)
column 50	-	*dom*	1=U.S. built, 0=imported
columns 53-57	-	*disp*	Displacement (cubic inches; liters in 1990)
columns 60-63	-	*hp*	Horsepower
columns 65-69	-	*lngth*	Length (inches)
columns 72-75	-	*wdth*	Width (inches)
columns 78-82	-	*wb*	Wheel base (inches)
columns 85-89	-	*mpg*	Miles per gallon
columns 92-97	-	*sales*	Sales
columns 100-101	-	*fid*	Firm identification number
column 104	-	*eu*	1=European model, 0=otherwise
column 107	-	*re*	Reliability index (from Consumer Reports)
column 110	-	*dfi*	1=Transplant production in the U.S. 0=otherwise
columns 113-118	-	*markup*	Estimated markup (thousands of dollars)

Special Considerations:

The displacement variable switched from cubic inches to liters in 1990. To convert these back to cubic inches, the displacement can be multiplied by 61.02.

5R. Antidumping Cases (DUMP1.ASC, DUMP2.ASC)

Contributed by: Robert Staiger
Dept. of Economics
Univ. of Wisconsin
Madison, WI 53706
E-mail: rstaiger@facstaff.wisc.edu

Frank Wolak
Dept. of Economics
Stanford University
Stanford, CA 94305-6072
wolak@zia.stanford.edu

This data should be referenced to:

Robert W. Staiger and Frank A. Wolak, "Measuring Industry-Specific Protection: Antidumping in the United States, Brookings Papers on Economics Activity: Microeconomics, 1994, 51-118.

The files DUMP1.ASC and DUMP2.ASC contain various information on antidumping cases in the United States, filed between 1980 and 1986. Each observation corresponds to an antidumping suit petition. Each petition is filed against one or several commodities as identified by the Tariff Schedule of the United States Annotated (TSUSA). Disk 1 of the NBER Trade Database contains complete information on U.S. imports according to TSUSA categories, so the dumping information in these files can be matched with the TSUSA data on Disk 1.

Record Layout for DUMP1.ASC:

columns 1-3 - *petno* = petition number (petitions are numbered in increasing order from past to present)
columns 5-7 - *lineno* = line number (line used to match electronic data to original hard copy)
columns 9-10 - *filemo* = month petition was filed in
columns 12-13 - *fileday* = day of month petition was filed in
columns 15-16 - *fileyr* = year petition was filed in
columns 18-19 - *lfvmo* = month preliminary less than fair value decision was made
columns 21-22 - *lfvday* = day preliminary less than fair value decision was made
columns 24-25 - *lfvyr* = year preliminary less than fair value decision was made
columns 27-28 - *decmo* = month of final decision (duties, negative finding, suspension, or withdrawal)
columns 30-31 - *decday* = day of final decision (duties, negative finding, suspension, or withdrawal)
columns 33-34 - *decyr* = year of final decision (duties, negative finding, suspension, or withdrawal)
columns 36-39 - *year* = year petition filed
columns 41-44 - *sic* = sic code of domestic competing product
columns 46-48 - *petrecno* = petition number
column 50 - *prelinj* = preliminary injury (0 = negative, 1 = affirmative)

column 52 - *tpm* = trigger price mechanism (1 = if triggered, 0 = if not triggered or not relevant)
column 54 - *suspend* = petition is suspended (0 = negative, 1 = affirmative)
column 56 - *dduty* = duties imposed (0 = negative, 1 = affirmative)
column 58 - *revoke* = if dumping duties revoked within six months
column 60 - *reinit* = current suit is re-initiation of previous suit (0 = false, 1 = true)
column 62 - *japan* = suit initiated against producers from Japan
column 64 - *canada* = suit initiated against producers from Canada
column 66 - *mex* = suit initiated against producers from Mexico
column 68 - *latam* = suit initiated against producers from Latin America (except Mexico)
column 70 - *eur* = suit initiated against producers from Western Europe
column 72 - *eeur* = suit initiated against producers from Eastern Europe (including former USSR)
column 74 - *easia* = suit initiated against producers from East Asia (Korea, Taiwan, Hong Kong, Singapore)
column 76 - *seasia* = suit initiated against producers from Southeast Asia (Philippines, Thailand, China)
column 78 - *safr* = suit initiated against producers from South Africa
column 80 - *mide* = suit initiated against producers from Mideast (Turkey, India, Israel)
column 82 - *ausnz* = suit initiated against producers from Australia and New Zealand
column 84 - *pnlfv* = preliminary less than fair value determination is negative (1 = negative, 0 = affirmative)
columns 86-87 - *tsusnum* = number of TSUSA codes petition filed against
column 89 - *pwct* = petition withdrawn, case terminated (1 = yes, 0 = no)

Size: DUMP1.ASC has 320 records.

Record Layout for DUMP2.ASC:

This file lists petition numbers followed by TSUSA codes associated with that petition. The format of the file is one petition number/TSUSA code combination per line. Note that the year of 2000, with SIC and TSUSA of -1, were used for petitions that did not match with any known TSUSA. These petitions were excluded from further analysis. The record layout for the file is:

columns 1-3 - Petition Number
columns 6-9 - Year
columns 11-14 - SIC Code
columns 16-22 - TSUSA Number

Size: DUMP2.ASC has 2,141 records.

5S. State-Level Exports (STATE*.ASC)

The files STATE*.ASC contain export data for U.S. states, by 2-digit SIC codes, for the years 1991-1994. These data *include* re-exports, so that the total exports over all states match the totals listed for these years in column 3 of Table 1 (page 4).

The state-level exports are collected by the U.S. Department of Commerce, Bureau of the Census, using information provided on the Shipper's Export Declaration (SED). There are two places on this declaration where geographical information in provided: (i) in the 2-letter postal service code reported in item 6 of the SED, the point (state) of origin; (ii) in the 5-digit ZIP code reported in item 1a of the SED, the name and address of the exporter. Prior to 1993, the Census used source (i) as a means of determining the state origin of the exports, though this information does not necessarily correspond with where the product is manufactured. In particular, the export data may not correlate with the "Origin of Exports of Manufactured Products" publication, which is based on data from the Census Bureau's "Annual Survey of Manufactures," conducted by the Bureau's Industry Division. These cautions apply to the state-level data for 1991 and 1992, which are called the "Origin of Movement (OM)" series by the Census.

Beginning in 1993, the Census began to use source (ii) as a means of determining the state origin of exports. It was felt that the ZIP code data more accurately reflected the physical location of the exporter, though not necessarily the producer. Studies by the Census showed that the ZIP code reported by wholesalers and other intermediaries reflects their own location and not the location of the manufacturing plant, farm, or quarry where the merchandise that they exported was made, grown or mined. The ZIP code reported by manufacturers, wholesalers and other intermediaries reflects the location of the plant or establishments that exported directly to a foreign country, and not the headquarters of the firm. The data for 1993 and 1994 included on the CD-ROM reflect this method for determining state location, and is called the "State Location of the Exporter (LE)" series by the Census.

Record Layout for STATE.ASC:*

columns 1-2	-	State (2-digit postal code)
columns 2-5	-	SIC code (2-digit)
columns 7-14	-	Name of destination country
columns 16-21	-	Country code United Nations codes are used (see Appendix A)
columns 23-34	-	Value of exports (dollars)
columns 36-37	-	Year

Size: Each file STATE*.ASC is about 3 megabytes.

5T. Foreign Affiliates Employment in the U.S. (FDIEMP1.TXT, FDIEMP2.TXT)

Contributed by: Deborah Swenson
Dept. of Economics
Univ. of California
Davis, CA 95616
E-mail: deswenson@ucdavis.edu

Source: Survey of Current Business [United States Department of Commerce / Bureau of Economic Analysis]
(Various Issues: "U.S. Affiliates of Foreign Companies: Operations in 19XX")

This data should be referenced to the paper:

Deborah Swenson, 1997, "The Effect of U.S. State Tax and Investment Promotion Policy on the Distribution of Inward Direct Investment," University of California, Davis.

The files FDIEMP1.TXT and FDIEMP2.TXT contain data on the employment of foreign affiliates in the U.S., by state: the first is for all non-bank affiliates, 1981-1994; and the second in for manufacturing affiliates only, 1988-1994. The layout of these files is self-explanatory.

Notes for FDIEMP1.TXT:

1. This file measures employment in 1000's.
2. The data are for all non-banking foreign affiliates.
3. 'Other U.S. Areas' = U.S. Virgin Islands, Guam, American Samoa and all other outlying US areas.
4. 'Foreign' = employees of US affiliates working abroad.

Notes for FDIEMP2.TXT:

1. This file measures manufacturing employment in 1000's.
2. The data are for non-banking foreign affiliates only.
3. 'Other U.S. Areas' = U.S. Virgin Islands, Guam, American Samoa and all other outlying US areas.
4. 'Foreign' = employees of US affiliates working abroad.

5U. Outward Foreign Investment from the U.S., 1983-1994 (FDIOUT*.GZ)

Contributed by: Matthew Slaughter
Dept. of Economics
Dartmouth College
Hanover, NH 03755
E-mail: slaughter@dartmouth.edu

Source: U.S. Direct Investment Abroad: Operations of U.S. Parent Companies and Their Foreign Affiliates, Bureau of Economic Analysis, Department of Commerce.

The files FDIOUT83.ASC,...,FDIOUT94.ASC contain all tables from the source publications listed above. The extension GZ refers to the GNU Zip program, which is a public-access program used to create zipped files on a UNIX platform. These files can be unzipped with several programs, including the public-access Stuffit Expander, which is available from Aladdin Systems via anonymous FTP at the site **ftp://ftp.aladdinsys.com/.** It is recommended that you download the Stuffit Expander if you have any difficulty in decompressing these files. For each .GZ file, there is a like-named file with the extension .TXT that contains details about the tables.

The U.S. Direct Investment Abroad surveys contain data on the financial structure and operations of a sample of nonbank U.S. companies and their nonbank foreign affiliates for each fiscal year. A report is required from every nonbank U.S. person having a nonbank foreign affiliate at the end of each fiscal year with assets, sales or net income exceeding $10 million. The estimates shown in the tables were obtained by expanding the sample data to universe totals; that is, data for U.S. parent companies and foreign affiliates not in the sample are estimated and then added to the reported data for parents and affiliates in the sample to yield universe totals. For 1989 and 1994, the Tables are taken from larger benchmark surveys. The 1989 survey covered the universe of U.S. direct investment abroad, which consists of all foreign business enterprises owned 10 percent or more, directly or indirectly, by a U.S. person. The 1994 survey covered all foreign affiliates of U.S. direct investors (foreign companies owned 10 percent or more by a U.S. person) that had assets, sales, or net income of more than $3 million.

The tables contained in these files were taken directly from disks sold by the U.S. Department of Commerce. For further information you can consult the hard copy of these publications, and the following person at the Bureau of Economic Analysis: Raymond Mataloni, (202) 606-9867, raymond.mataloni@bea.doc.gov.

6. Programs to Construct and Update the Export and Import Data

Included on the CD-ROM are a number of SAS (and BASIC) programs that can be used to construct the 4-digit SIC import and export data. These programs are intended to serve as a guide for researchers who would like to update this data using additional years. However, they are *not* intended to work automatically; rather, they will need to be modified to fit the local platform that you are using, and the form in which you receive additional years of data. These programs can also be used to construct 4-digit SIC level aggregates over other variables, such as tariff revenue, or any of the country subcodes listed in the import data.

Constructing the SIC import data (Disk 1 of the NBER Trade Database):

1clist94.sas - Converts the country numbers from those used in the raw Census data, to the UN codes used on the CD-ROM. This program would use the raw 1994 Census data as input.

1imp88.sas - Reads in the IMP*.ASC files from the CD-ROM for 1988, for later use.

1imp89.sas - Reads in the IMP*.ASC files from the CD-ROM for 1989-1992 and sums the data according to 1987-basis XSIC codes (4-digit).
Output: SIC89,…, SIC92

1sumsic.sas - Reads in SIC* and adds the observations for WORLD.
Output: MSIC89,…,MSIC92

1sic72_4.sas - Creates the weighting matrix MSIC8772 using the 1988 data, and converts the 1987-basis MSIC data for 1989-1992 to the 1972-basis MSIC.
Output: M72_89,…,M72_92

1country.bas - Basic program that creates the cross-product of the files SIC_MSIC.TMP and COUNTRY1.ASC, where these are identical to the files SIC_MSIC.ASC and COUNTRY.ASC on Disk 1, respectively, but with the last column deleted. The user needs to specify the number of rows in these files within the program.
Output: SICMSICC.ASC

1insic1.sas - Preliminary program for reading in certain data for the SIC calculation.

1insic2.sas - Main program to compute the domestic-based SIC imports. This program must be run separately for each year.

1insic3.sas - Merge the import data with the earlier years from Abowd, and print out.

Constructing the SIC export data (Disk 3 of the NBER Trade Database):

3clist94.sas - Converts the country numbers from those used in the raw Census data, to the UN codes used on the CD-ROM. This program was used to convert the Census country codes within the state-level data to UN codes, but can also be used to convert the county codes in the raw Census export data.

3exp89.sas - Prints out the EXP*.ASC files for 1989-1994, while making certain changes to the 1987-basis XSIC codes for some HS numbers; also sums the data according the 1987-basis XSIC and also SITC
Output: SIC89,…,SIC92

3sumsic.sas - Reads in SIC* and adds the observations for WORLD.
Output: XSIC89,…,XSIC92

3sic72_4.sas - Creates the weighting matrix XSIC8772 using the 1988 data (assuming it is already read in) and converts the 1987-basis XSIC data for 1989-1992 to the 1972-basis XSIC.
Output: X72_89,…,X72_92

3country.bas - Basic program that creates the cross-product of the files SIC_XSIC.TMP and COUNTRY3.ASC, where these are identical to the files SIC_XSIC.ASC and COUNTRY.ASC on Disk 3, respectively, but with the last column deleted. The user needs to specify the number of rows in these files within the program.
Output: SICXSICC.ASC

3insic1.sas - Preliminary program for reading in various data for the SIC calculation.

3insic2.sas - Main program to compute the domestic-based SIC exports. This program must be run separately for each year.

3insic3.sas - Merge the export data with the earlier years from Abowd, and print out.

References

Abowd, John M. (1991) "Appendix: The NBER Immigration, Trade, and Labor Market Files," in John M. Abowd and Richard B. Freeman, eds. Immigration, Trade, and the Labor Market. Chicago: Univ. of Chicago Press and NBER, 407-422.

Abowd, John M. and Richard B. Freeman, eds. Immigration, Trade, and the Labor Market. Chicago: Univ. of Chicago Press and NBER.

Alterman, William, "Price Trends in U.S. Trade: New Data, New Insights," in Peter Hooper and J. David Richardson, eds. *International Economic Transactions*, Chicago: Univ. of Chicago Press and NBER, 1991, pp. 109-139.

Kowalczyk, Carsten and Donald Davis, 1996, "Tariff Phase-outs: Theory and Evidence from GATT and NAFTA," NBER Working Paper no. 5421.

Shiells, Clinton R. (1991) "Errors in Import-Demand Estimates Based Upon Unit-Value Indexes," Review of Economics and Statistics, 55, 456-464.

Census Publications

[1] U.S. Bureau of the Census, U.S. Exports, Schedule B, Commodity by Country, FT410, Washington, D.C.: Department of Commerce, annual, 1972-1977.

[2] U.S. Bureau of the Census, U.S. Exports, Schedule B, Commodity by Country, FT446, Washington, D.C.: Department of Commerce, annual, 1978-1988.

[3] U.S. Bureau of the Census, U.S. Exports, Harmonized System, Commodity by Country, FT447, Washington, D.C.: Department of Commerce, annual, 1989-1994.

[4] U.S. Bureau of the Census, U.S. Foreign Trade Statistics. Classifications and Cross-Classifications, 1974. Washington, D.C.: Department of Commerce, December, 1975.

[5] U.S. Bureau of the Census, U.S. Commodity Exports and Imports as Related to Output, Washington, D.C.: Department of Commerce, annual, various years.

[6] U.S. Bureau of the Census, U.S. Exports. Domestic Merchandise, SIC-Based Products by World Area, FT210, Washington, D.C.: Department of Commerce, 1981-1984.

Appendix A: Country Codes and Names (COUNTRY.ASC)

The U.S. export data collected by the Bureau of the Census keeps track of the source country by certain Census codes. For the export database, the United Nations (UN) country codes and names are used instead. The file COUNTRY.ASC gives a complete list of the UN codes, UN country abbreviations, the corresponding Census codes for 1994, and the full name of the Census country. This file is printed on the following two pages.

Record Layout:

columns 1-6 - United Nations (UN) code
columns 8-15 - Abbreviated UN country name
columns 17-20 - Census country code for 1994
columns 22-50 - Full Census country name

The records are sorted by the six-digit UN codes. The first two-digits of that code are a regional identifier, the next three-digits are a specific country code, and the last digit is a special modifier than equals zero in nearly all cases.

Missing values: In a few cases, the Census country code for 1994 is missing, as indicated by a blank. This means that the country in question did not appear in the Census export data for 1994, but did appear in some earlier year; an example is East Germany.

Special considerations: There are more Census country codes than UN codes. This means that a given UN code may appear on several subsequent records, followed by the same abbreviated UN country name; on each of these records, a different Census country code and Census county name will appear. For example, South Africa is treated as one country in the UN codes, but is broken down into several smaller regions in the Census codes and names.

Size: COUNTRY.ASC has 242 records.

Sources: The UN codes are the six-digit Standard Classification of Customs Areas and Territories, and are nearly the same as those used by Statistics Canada in their World Trade Database. The Census codes and country names are taken from the file COUNTRY.DBF contained on:

U.S. Exports of Merchandise on CD-ROM [machine-readable data file] / prepared by the Bureau of the Census. - Washington: The Bureau [producer and distributor], 1994.

```
117100 S_AFRICA 7910 REPUBLIC OF SOUTH
                     AFRICA
117100 S_AFRICA 7920 NAMIBIA
117100 S_AFRICA 7930 BOTSWANA
117100 S_AFRICA 7950 SWAZILAND
117100 S_AFRICA 7990 LESOTHO
130120 ALGERIA  7210 ALGERIA
134340 LIBYA    7250 LIBYA
135040 MOROCCO  7140 MOROCCO
135040 MOROCCO  7370 WESTERN SAHARA
137360 SUDAN    7320 SUDAN
137880 TUNISIA  7230 TUNISIA
138180 EGYPT    7290 EGYPT
141200 CAMEROON 7420 CAMEROON
141400 C_AFRICA 7540 CENTRAL AFRICAN
                     REPUBLIC
141480 CHAD     7560 CHAD
141780 CONGO    7630 CONGO
142660 GABON    7550 GABON
160240 ANGOLA   7620 ANGOLA
161080 BURUNDI  7670 BURUNDI
161800 ZAIRE    7660 ZAIRE
162040 BENIN    7610 BENIN
162260 EQ_GNEA  7380 EQUATORIAL GUINEA
162300 ETHIOPIA 7740 ETHIOPIA
162300 ETHIOPIA 7741 ERITREA
162300 ETHIOPIA 7749 ETHIOPIA
162620 DJIBOUTI 7770 DJIBOUTI
162700 GAMBIA   7500 THE GAMBIA
162880 GHANA    7490 GHANA
163240 GUINEA   7460 GUINEA
163840 IVY_CST  7480 IVORY COAST
164040 KENYA    7790 KENYA
164300 LIBERIA  7650 LIBERIA
164500 MADAGAS  7880 MADAGASCAR
164540 MALAWI   7970 MALAWI
164660 MALI     7450 MALI
164780 MAURITN  7410 MAURITANIA
164800 MRITIUS  7850 MAURITIUS
165080 MOZAMBQ  7870 MOZAMBIQUE
165620 NIGER    7510 NIGER
165660 NIGERIA  7530 NIGERIA
166240 G_BISAU  7642 GUINEA-BISSAU
166240 G_BISAU  7643 CAPE VERDE
166240 G_BISAU  7644 SAO TOME AND
                     PRINCIPE
166380 FR_IND_O 7890 COMOROS
166380 FR_IND_O 7904 REUNION
166380 FR_IND_O 7905 FR SOUTHERN-
                          ANTARTIC
LANDS
166460 RWANDA   7690 RWANDA
166540 S_HELNA  7580 ST. HELENA
166860 SENEGAL  7440 SENEGAL
166900 SEYCHEL  7800 SEYCHELLES
166900 SEYCHEL  7810 BRITISH INDIAN
                     OCEAN TERR.
166940 SIER_LN  7470 SIERRA LEONE
167060 SOMALIA  7700 SOMALIA
167160 ZIMBABWE 7960 ZIMBABWE
167680 TOGO     7520 TOGO
168000 UGANDA   7780 UGANDA
168340 TANZANIA 7830 TANZANIA
168540 BURKINA  7600 BURKINA
168940 ZAMBIA   7940 ZAMBIA
211240 CANADA   1220 CANADA
220600 BERMUDA  2320 BERMUDA
223040 GREENLD  1010 GREENLAND
226660 SP_MQEL  1610 ST. PIERRE AND
                     MIQUELON
330320 ARGENT   3570 ARGENTINA
330680 BOLIVIA  3350 BOLIVIA
330760 BRAZIL   3510 BRAZIL
331520 CHILE    3370 CHILE
331700 COLOMBIA 3010 COLOMBIA
332180 ECUADOR  3310 ECUADOR
334840 MEXICO   2010 MEXICO
336000 PARAGUA  3530 PARAGUAY
336040 PERU     3330 PERU
338580 URUGUAY  3550 URUGUAY
338620 VENEZ    3070 VENEZUELA
341880 COS_RICA 2230 COSTA RICA
342220 SALVADR  2110 EL SALVADOR
343200 GUATMALA 2050 GUATEMALA
343400 HONDURA  2150 HONDURAS
345580 NICARAGA 2190 NICARAGUA
350440 BAHAMAS  2360 BAHAMAS
350520 BARBADO  2720 BARBADOS
351920 CUBA     2390 CUBA
352140 DOM_REP  2470 DOMINICAN REPUBLIC
353120 GUADLPE  2831 GUADELOUPE
353120 GUADLPE  2839 MARTINIQUE
353320 HAITI    2450 HAITI
353880 JAMAICA  2410 JAMAICA
353880 JAMAICA  2430 TURKS AND CAICOS
                     ISLANDS
353880 JAMAICA  2440 CAYMAN ISLANDS
355320 N_ANTIL  2771 NETHERLANDS
                           ANTILLES
355320 N_ANTIL  2779 ARUBA
356580 ST_K_NEV 2481 ANGUILLA
356580 ST_K_NEV 2482 BRITISH VIRGIN
                     ISLANDS
356580 ST_K_NEV 2483 ST. KITTS-NEVIS
356580 ST_K_NEV 2484 ANTIGUA
356580 ST_K_NEV 2485 MONTSERRAT
356580 ST_K_NEV 2486 DOMINICA
356580 ST_K_NEV 2487 ST. LUCIA
356580 ST_K_NEV 2488 ST. VINCENT
356580 ST_K_NEV 2489 GRENADA
357800 TRINIDAD 2740 TRINIDAD AND
TOBAGO
360840 BELIZE   2080 BELIZE
362380 FALK_IS  3720 FALKLAND ISLANDS
362540 FR_GUIAN 3170 FRENCH GUIANA
363280 GUYANA   3120 GUYANA
365900 PANAMA   2250 PANAMA
367400 SURINAM  3150 SURINAME
368960 US_NES   6810 MARSHALL ISLANDS
368960 US_NES   6820 FEDRATED STATES OF
                     MICRONESIA
368960 US_NES   6830 PALAU
368960 US_NES   9350 GUAM
368960 US_NES   9800 U.S. OUTLYING
                     ISLANDS
368961 PRT_RICO 9000 PUERTO RICO
368961 PRT_RICO 9030 PUERTO RICO
368962 VGN_ISL  9110 VIRGIN ISLANDS
413760 ISRAEL   5081 ISRAEL
413760 ISRAEL   5082 GAZA STRIP ADMNSTD
                     BY ISRAEL
413760 ISRAEL   5083 WEST BANK  ADMNSTD
                     BY ISRAEL
413920 JAPAN    5880 JAPAN
```

```
440480 BAHRAIN  5250 BAHRAIN
441960 CYPRUS   4910 CYPRUS
443640 IRAN     5070 IRAN
443680 IRAQ     5050 IRAQ
444000 JORDON   5110 JORDAN
444140 KUWAIT   5130 KUWAIT
444220 LEBANON  5040 LEBANON
445120 OMAN     5230 OMAN
446340 QATAR    5180 QATAR
446820 SD_ARAB  5170 SAUDI ARABIA
447200 YEMEN_S  5220 SOUTH YEMEN
447600 SYRIA    5020 SYRIAN ARAB
                     REPUBLIC
447840 ARAB_EM  5200 UNITED ARAB
                     EMIRATES
447920 TURKEY   4890 TURKEY
448860 YEMEN_N  5210 YEMEN ARAB
REPUBLIC
450000 ASIA_NES 5610 BRUNEI
450000 ASIA_NES 5682 BHUTAN
450040 AFGHAN   5310 AFGANISTAN
450500 BNGLDSH  5380 BANGLADESH
451040 BURMA    5460 BURMA
451160 CAMBOD   5550 CAMBODIA
451440 SRI_LKA  5420 SRI LANKA
453440 HONGKONG 5820 HONG KONG
453560 INDIA    5330 INDIA
453600 INDONES  5600 INDONESIA
453600 INDONES  5683 MALDIVE ISLANDS
454100 KOREA_S  5800 KOREA, REPUBLIC OF
454180 LAO      5530 LAOS
454460 MACAU    5660 MACAO
454580 MALAYSIA 5570 MALAYSIA
455240 NEPAL    5360 NEPAL
455860 PAKISTAN 5350 PAKISTAN
456080 PHIL     5650 PHILIPPINES
457020 SINGAPR  5590 SINGAPORE
457640 THAILAND 5490 THAILAND
458960 TAIWAN   5830 CHINA (TAIWAN)
481560 CHINA    5700 CHINA (MAINLAND)
484080 KOREA_N  5790 NORTH KOREA
484960 MONGOLA  5740 MONGOLIA
487040 VIETNAM  5520 VIETNAM
530560 BEL_LUX  4231 BELGIUM
530560 BEL_LUX  4239 LUXEMBOURG
532080 DENMARK  4091 FAROE ISLAND
532080 DENMARK  4099 DENMARK
532500 FRANCE   4271 ANDORRA
532500 FRANCE   4272 MONACO
532500 FRANCE   4279 FRANCE
532800 GERMAN   4280 GERMANY
533000 GREECE   4840 GREECE
533720 IRELAND  4190 IRELAND
533800 ITALY    4751 SAN MARINO
533800 ITALY    4752 VATICAN CITY
533800 ITALY    4759 ITALY
535280 NETHLDS  4210 NETHERLANDS
536200 PORTUGAL 4710 PORTUGAL
537240 SPAIN    4700 SPAIN
538260 UKINGDOM 4120 UNITED KINGDOM
550400 AUSTRIA  4330 AUSTRIA
552460 FINLAND  4050 FINLAND
553520 ICELAND  4000 ICELAND
555780 NORWAY   4031 SVALBARD, JAN
MAYEN                ISLAND
555780 NORWAY   4039 NORWAY
557520 SWEDEN   4010 SWEDEN
557560 SWITZLD  4411 LIECHTENSTEIN
557560 SWITZLD  4419 SWITZERLAND
572920 GILBRALT 4720 GIBRALTAR
574700 MALTA    4730 MALTA AND GOZO
580080 ALBANIA  4810 ALBANIA
581000 BULGARIA 4870 BULGARIA
582000 CZECHO   4351 THE CZECH REPUBLIC
582000 CZECHO   4359 SLOVAKIA
582780 GERMAN_E      EAST GERMANY
583480 HUNGARY  4370 HUNGARY
586160 POLAND   4550 POLAND
586420 ROMANIA  4850 ROMANIA
598900 YUGOSLAV 4790 YUGOSLAVIA
598900 YUGOSLAV 4791 CROATIA
598900 YUGOSLAV 4792 SLOVENIA
598900 YUGOSLAV 4793 BOSNIA-HERCEGOVINA
598900 YUGOSLAV 4794 MACEDONIA
598900 YUGOSLAV 4799 YUGOSLAVIA
688100 USSR     4470 ESTONIA
688100 USSR     4490 LATVIA
688100 USSR     4510 LITHUANIA
688100 USSR     4621 RUSSIA
688100 USSR     4622 BELARUS
688100 USSR     4623 UKRAINE
688100 USSR     4631 ARMENIA
688100 USSR     4632 AZERBAIJAN
688100 USSR     4633 GEORGIA
688100 USSR     4634 KAZAKHSTAN
688100 USSR     4635 KYRGYZSTAN
688100 USSR     4641 MOLDOVA
688100 USSR     4642 TAJIKISTAN
688100 USSR     4643 TURKMENISTAN
688100 USSR     4644 UZBEKISTAN
710360 AUSTRAL  6021 AUSTRALIA
710360 AUSTRAL  6022 NORFOLK ISLAND
710360 AUSTRAL  6023 COCOS ISLANDS
710360 AUSTRAL  6024 CHRISTMAS ISLAND
710360 AUSTRAL  6029 HEARD AND MCDONALD
                     ISLANDS
715540 NEW_ZEAL 6141 NEW ZEALAND
715540 NEW_ZEAL 6142 COOK ISLANDS
715540 NEW_ZEAL 6143 TOKELAU ISLANDS
715540 NEW_ZEAL 6144 NIUE
722420 FIJI     6862 NAURU
722420 FIJI     6863 FIJI
722420 FIJI     6864 TONGA
722960 KIRIBATI 6223 SOLOMON ISLANDS
722960 KIRIBATI 6224 VANUATU
722960 KIRIBATI 6225 PITCAIRN ISLAND
722960 KIRIBATI 6226 KIRIBATI
722960 KIRIBATI 6227 TUVALU ISLANDS
725400 NEW_CALE 6412 NEW CALEDONIA
725400 NEW_CALE 6413 WALLIS AND FUTUNA
725400 NEW_CALE 6414 FRENCH POLYNESIA
725980 NEW_GUIN 6040 PAPUA NEW GUINEA
728882 SAMOA    6150 WESTERN SAMOA
728882 SAMOA    9510 AMERICAN SAMOA
999990 INT_ORG  8500 INTERNATIONAL
                     ORGANIZATIONS
999991 ALL_CTY  8990 ALL COUNTRIES
999999 UNKNOWN  9610 NORTHERN MARIANA
                     ISLANDS
999999 UNKNOWN  8220 UNIDENTIFIED
999999 UNKNOWN  9980 UNIDENTIFIED
```

Appendix B: Units for Quantity

The following units for quantity are used in EXP*.ASC for 1972-1988:

AOZ	Avoirdupois ounce
BBL	Barrel
BFT	Board feet
BU	Bushel
C	One hundred
CAR	Carat
CD	Cord
CFT	Cubic feet
CGM	Content gram
CLB	Content pound
CRT	Crate
CST	Content short ton
CTN	Content ton
CTO	Content troy ounce
CUR	Curie
CWT	Hundredweight (100 pounds)
CYD	Cubic Yard
DOZ	Dozen
DPC	Dozen Pieces
DPR	Dozen Pair
GAL	Gallon
GBX	Gross boxes
GR	Gross
GRL	Gross lines
GM	Gram
GTN	Gross ton
JWL	Jewels
LB	Pound
LF	Leaf
LFT	Linear Foot
LTN	Long ton (2,240 pounds)
LYD	Long yard
M	Thousand (1,000)
MBF	Thousand board feet
MC	Milli-Curie
MCF	Thousand cubic feet
MFT	Thousand feet
MLF	Thousand linear feet
MSF	Thousand square feet
MUN	Million units
NO	Number
PC	Piece
PFG	Proof gallon
PGR	Pencil gross
PK	Pack
PR	Pair
RBA	Running bale
RM	Ream
SET	Set
SFT	Square foot
SQ	Square
SQI	Square inch
STN	Short ton (2,000 pounds)
SUF	Superficial foot
SYD	Square yard
TON	Long Ton (2,240 pounds)
TOZ	Troy ounce
WG	Wine gallon
YD	Yard

The following units for quantity are used in EXP*.ASC for 1989-1994:

BBL	Barrel
CAR	Carat
CBM	Cubic Meters
CKG	Content Kilogram
CTN	Content Metric Ton
CUR	Curie
CYK	Clean Yield Kilogram
DOZ	Dozen
DPC	Dozen Pieces
DPR	Dozen Pair
FBM	Fiber Meter
GCN	Gross Containers
GKG	Kilogram (gross)
GM	Gram
GRS	Gross
HUN	Hundred
KG	Kilograms
KGS	Kilogram Total Sugars
KWH	Kilowatt-hours
LTR	Liters
M2	Square Meters
MC	Milli-Curie
MCU	Micro-Curie
MTR	Meter
NO	Number
ODE	Ozone Depletion Equivalent
PCS	Pieces
PFL	Proof Liter
PKS	Packs
PRS	Pairs
RBA	Running Bales
SCM	Square Centimeters
SET	Sets
SQ	Square
SQM	Square meters
TBE	Thousand Standard Brick Equivalent
TCM	Thousand Cubic Meter
THS	Thousands

TON Metric Ton