

Frequently asked questions about NBER-UN data

1. Are the data in nominal dollars or real dollars?

The data are in nominal thousands of US dollars (\$1,000).

2. Why does the same SITC4 code appears multiple times in some cases?

Example:

For 2000, Canadian exports to World for SITC4 product 0011, there are 3 entries; for product 0012, there are two entries; and for product 0013, there are three entries, as shown below:

<i>year</i>	<i>icode</i>	<i>importer</i>	<i>ecode</i>	<i>exporter</i>	<i>sitc4</i>	<i>unit</i>	<i>dot</i>	<i>value</i>	<i>quantity</i>
2000	100000	World	211240	Canada	0011			763827	
2000	100000	World	211240	Canada	0011	N		2264	433
2000	100000	World	211240	Canada	0011	W		1468	143
2000	100000	World	211240	Canada	0012	N		5793	52937
2000	100000	World	211240	Canada	0012	W		282	42
2000	100000	World	211240	Canada	0013			154	
2000	100000	World	211240	Canada	0013	N		443	319
2000	100000	World	211240	Canada	0013	W		303071	295830

Interpretation:

Within the SITC4 product 0011, there were 433 units valued at \$2,264,000 with the unit N (i.e. number) exported from Canada to the World. There were 143 metric tons valued at \$1,468,000 with the unit W (i.e. weight) exported. There was another \$763,827,000 with no units (and no quantity) exported from Canada to the World.

In order to get the total exports of Canada to the World in product 0011 you have to add up the values with and without quantities. That gives total exports of \$767,559,000:

<i>year</i>	<i>icode</i>	<i>importer</i>	<i>ecode</i>	<i>exporter</i>	<i>sitc4</i>	<i>unit</i>	<i>dot</i>	<i>value</i>	<i>quantity</i>
2000	100000	World	211240	Canada	0011			763827	
2000	100000	World	211240	Canada	0011	N		2264	433
2000	100000	World	211240	Canada	0011	W		<u>1468</u>	143

Total = 767559

The purpose of separating the values according to Units was to permit users to calculate unit values if they wished. In the case of product 0013, most of the exports are reported by weight, and a unit value might be calculated from that.

3. Why do some of the SITC4 codes end in A, X?

4-digit SITC codes ending in A, such as 001A, is an “aggregate” of all the 4-digit SITC codes within that 3-digit code, i.e. within 001. 4-digit SITC codes ending in X, such as 001X, represents “extra” trade of some 4-digit SITC codes within that 3-digit code, i.e. within 001. So when SITC 001A is used for the bilateral trade between one country and another, then you should not see any other 4-digit codes within SITC 001; but when SITC 001X is used for the bilateral trade between one country and another, then there might be additional trade within codes such as 0011, 0012, 0013, etc. See section 4 (pp. 9-10) of the documentation (NBER working paper #11040, referenced below).

4. Why do some of the SITC4 codes end in zero?

4-digit SITC codes ending in zero were introduced into the data because we substituted the U.S. values of exports and imports in place of the UN values, whenever the U.S. was a partner. In the U.S. values, an SITC Rev. 2 code ending in zero has the same meaning as a code ending in A or X; that is, it represents trade within that 3-digit code that could not be accurately assigned to a 4-digit code. For example, trade within SITC 0220 really means trade within one of the SITC industries 0222, 0223, or 0224.

5. What does DOT mean?

It means Direction of Trade, 1=Data is from importer, 2=Data is from exporter. This variable keeps track of whether the importer or exporter reported the trade flow. Note that the Direction of Trade (DOT) variable is missing for exports to the World or imports from the World. That variable only appears in bilateral (country to country) trade flows.

6. Are the data CIF (cost, insurance, freight) or FOB (free on board)?

The data reported by the importer (DOT=1) is CIF, whereas the data reported by the exporter (DOT=2) is FOB. As explained in the documentation, we give primacy to the importers’ reports, whenever they are available, assuming that these are more accurate than reports by the exporter.

7. Where are the Units reported?

See the NBER working paper #11040, p. 48:

Quantity Unit Codes (Blank indicates quantity unavailable or less than one-half a unit)

A	Area (1,000 square meters)	H	Energy (1,000 kilowatt hours)
K	Weight (kilograms)	L	Length (1,000 meters)
N	Units (number of items)	P	Pairs (number of pairs)
V	Volume (cubic meters)	W	Weight (metric tons)

8. The text of the working paper accompanying the NBER data set states that the data were mainly compiled on the basis of trade flows at the 4-digit SITC Rev.2 level, in excess of \$100,000 per year. Why do I find some individual bilateral trade flows at the 4-digit level, which are as low as \$1,000 per year in the data set?

The data that purchased from the United Nations for 1984-2000 only had values in excess of 100,000, for each bilateral flow. But because we revised these data, as described in the NBER working paper, then in a number of cases we introduced smaller-valued trade flows into the dataset. Those smaller value will be clustered on certain countries (i.e. those where we revised the data) or clustered on SITC codes ending in X (which we introduced into the dataset). That explains why we ended up with some values less than \$100,000 even though we did not purchase these records from the UN originally. (The situation before 1984 is different, and there should be more small-valued observations there). In 2000, for example, a quick count shows that there are 7,443 observations with the value < \$100,000, out of about 733,000 observations in total. So there are only about 1% of the observations with value < \$100,000. A substantial fraction of these have SITC ending in X.

9. Why does a country sometimes import from itself, or export to itself?

This occurs because the United Nations keeps track of countries that are smaller than those reported in the NBER-UN dataset, where we have aggregated as described in Appendix A. For example, Puerto Rico and the U.S. Virgin Islands are aggregated within the United States, so the U.S. will shows trade with itself when the mainland trades with those islands. Another example is Italy, which includes San Marino and the Holy See (Vatican).

10. How are the World totals calculated?

The World totals are calculated by summing over all values in the dataset. For example, consider Canadian exports to the World within SITC product 0011:

<i>year</i>	<i>icode</i>	<i>importer</i>	<i>ecode</i>	<i>exporter</i>	<i>sitc4</i>	<i>unit</i>	<i>dot</i>	<i>value</i>	<i>quantity</i>
2000	100000	World	211240	Canada	0011			763827	
2000	100000	World	211240	Canada	0011	N		2264	433
2000	100000	World	211240	Canada	0011	W		1468	143

The first value of \$763,827,000 represents Canadian exports to all countries for which we have a value in the dataset, for SITC product 0011 and no Units. The second value of \$2,264,000 represents Canadian exports to all countries for which we have a value in the dataset, for SITC product 0011 and the Unit N (i.e. number).

11. If I add up all the values in the dataset, will I get total world trade?

No! If you add up all the values in the dataset you will get *twice* the value of world trade, because the World values are included. If you add up all the values *not including* cases where World appears as a partner, then you will get the total value of world trade. Similarly, if you add up all the cases where World appears as a partner, then you will get the same value of world trade. However, this value for total world trade is only our estimate, obtained by using the reported import data whenever possible, and otherwise using the reported export data. Because the import reports and the export reports differ, there is no unique value for world trade.

12. A complicated question about A and X codes:

I have a question about the SITC categories. In section 4 of the working paper, “Adjusting the UN Data at Differing Levels of Aggregation: A and X codes for 1984-2000,” it says that:

“The second set of codes deals with the case in which there are no imports or exports reported at the disaggregate SITC level, but there is such a value at the higher aggregate SITC level. As an example, suppose the value of imports is reported as \$200 million for SITC 444, but there are no corresponding four-digit SITC with leading numbers of 444. To deal with such a case, an additional SITC is created which combines the beginning of the 3-digit SITC with an ending of A (444A) and is given a value equal to the value of import at the 3-digit SITC level (\$200). This residual category represents *aggregate* imports or exports in SITC 444.”

But as I checked the data, say U.S. exports to the World, I get the following for SITC 014:

<i>year</i>	<i>icode</i>	<i>importer</i>	<i>ecode</i>	<i>exporter</i>	<i>sitc4</i>	<i>unit</i>	<i>dot</i>	<i>value</i>	<i>quantity</i>
1995	100000	World	218400	USA	0141			184	
1995	100000	World	218400	USA	0141	W		7062	1835
1995	100000	World	218400	USA	0142			2169	
1995	100000	World	218400	USA	0142	W		102385	35455
1995	100000	World	218400	USA	0149			11712	
1995	100000	World	218400	USA	0149	W		363491	136861
1995	100000	World	218400	USA	014A			234	
1995	100000	World	218400	USA	014X			<u>41</u>	

Total = 487278

The second-last row has SITC4 014A but that value isn't the aggregate of the rest of the 0141, 0142, etc; also, given the existence of 0141, 0142, etc, there shouldn't be a 014A according to the documentation. Why do the 014A and 014X codes appear, and what do they mean?

Answer:

The category 014A will be the aggregate for *some* trade partner of the U.S., when the underlying 4-digit data were not available for that partner. But when adding up to World exports, then 014A becomes just an extra category of exports, similar to 014X, and is not the aggregate of anything shown here. Total U.S. exports to the World within 014 for 1995 would be obtained by adding up over all the lines that you show, obtaining \$487,278,000.

13. An issue concerning African countries in early years, from Johannes Moenius

<Johannes_Moenius@redlands.edu>

While working on the UN-NBER data, I noticed some more irregularities / inconsistencies of the data with the documentation. I describe below how I think it should be dealt with

The codes:

168350 Frm Tangayika (until 1964)
 168360 Frm Zanzibar-Pemb. (until 1964)
 167170 Frm Rhodesia NYAS (until 1964)

contain only missing values.

Tangayika and Zanzibar seemed to have already been included in the Tanzania data, since

168340 Tanzania

already has data for 62-64, which it should not have. Moreover,

167160 Zimbabwe, since 65
 164540 Malawi, also since 65

have data for 63 and 64, which they should not and

168940 Zambia

has also data for 62-64 (which it should not have). Adding the last three codes for 62-64 seems to give quite reasonable figures for the former Rhodesia NYAS.

I would suggest to keep Tanzania from 62 onward, use Zimbabwe and Malawi from 65 onward and Zambia from 62 onward (without any changes to the data).

14. A question about the treatment of Taiwan:

I found a puzzle which appears on p. 6, p. 55 and p. 59 of NBER Working Paper no. 11040 about the country code for Taiwan.

On p. 6. the statement is " ...NBER-UN country code for Taiwan is '454900'. This code corresponds to UN country code of '458960'.

On both pages 55 and 59 on Appendices A and B, the code for Taiwan were just the opposite. That is NBER-UN code for Taiwan was 458960.

Answer:

We checked the NBER-UN dataset and UN Comtrade website and found that you are right. In NBER-UN dataset, the country code for Taiwan is 458960, while the UN country code for Taiwan, or "Other Asia, nes", is 454900. In NBER-UN dataset, the code 454900 is non-existent. Thus it is a mistake on page 6 of NBER WP 11040, as you point it out.

Documentation:

Feenstra, Robert C., Robert E. Lipsey, Haiyan Deng, Alyson C. Ma, and Hengyong Mo, "World Trade Flows: 1962-2000," NBER Working Paper no. 11040, 2004.