Note on the Trade Estimation

Overview

This note describes the procedures and decisions used to produce trade statistics for countries and areas that did not report for certain years, as well as for commodities considered under-reported or confidential. The procedure estimates data by trading partner for all 3-digit SITC, Rev.3, and, in addition, for several 4-digit codes (781.2, 784.1, 785.1, 785.2, and 785.31). Note that estimation is applied only to trade value (no quantity estimation).

The result of the estimation is called the "Full Trade Matrix," which indicates that the trade data includes all possible reporting countries, as well as under-reported and confidential commodities for both exports and imports. The matrix also includes high-level aggregated data, with World as both reporter and partner. It is used as an input table to calculate the following analytical outputs of tables 41, 42, and 43 of Monthly Bulletin of Statistics (also known as World Export Matrix Tables) and the full datasets are available at https://comtradeplus.un.org/TradeDataMatrix

Data Preparation

The following items are necessary as prerequisites for estimation:

- UN Comtrade data by trade flows for the past 5 years (for reported data)
- Total trade figures for not-yet-reported countries
- SITC, Rev. 3 data as reported or converted from original country data[1]

Estimation methodology

Overview of basic estimation principles

Extrapolation

If a country has reported data for a given year x, then estimates for years (x-1) and (x-2), as well as (x+1) and (x+2), are derived from the detailed data for year x and the total trade figures for the years being estimated. The estimates are produced using the ratio of total trade figures as the fixed multiplier for each commodity value by trading partner.

Inversion

Inversion is also known as Mirror Statistics, whereas estimated data are derived from trading partners. If a country did not report a complete data set within the time frame (x-2) to (x+2), the inversion of partner data will be applied. In this case, the total trade figures for the years are needed.

[1] As of reporting year 2008, only one country still reports in SITC classification.

The input to this estimation procedure is the reported data together with the estimated extrapolated data. By inverting those data, the reporter will become the partner and imports will become exports (or vice versa). The estimated trading partners will be limited to the reporting countries that either reported or were extrapolated for a given year. The inverted data are further multiplied by the ratio of total trade figures, used as the fixed multiplier.

Note that the mirror statistics may not fully reflect the data collected and compiled by the countries. The reasons can be traced back to the bilateral asymmetries, where trade flows reported by trading partners do not match the flows reported by compiling countries. See more at IMTS 2026 Chapter Bilateral Asymmetries.

Detailed Estimation Formula

As mentioned above, the basic principles of trade estimation are extrapolation and inversion. These principles can also be applied when estimating under-reported or confidential trade. Here are the detailed estimation methodologies along with their cases:

1. Extrapolation

Extrapolation is applied if reported data in adjacent years are available. There are three cases:

i. No partially reported data and no entries in the blocking table[2]

$$V(i,j,1) = V(i,j,0) * [T(i,1) / T(i,0)]$$

where i = Reporter i
j = commodity j
1 = target year
0 = reported year
V = commodity value
T = total value (either for imports or for exports)

ii. Commodity j exists in the blocking table

$$V(i,j,1) = B(i,j,1) * T(i,1)$$

where i = Reporter i
j = commodity j
1 = target year
0 = reported year
V = commodity value
B = multiplier from the blocking table
T = total value (either for imports or for exports)

iii. Reporter i exists in the blocking table (but not commodity j), or partially reported data is available

$$V(i,j,1) = V(i,j,0) * [T'(i,1) / T'(i,0)]$$

and $T'(i,1) = T(i,1) - SUM(V(i,r,1))$

and $T'(i,0) = T(i,0) - SUM(V(i,r,0))$

[2] The blocking table was used to adjust the multiplier of unlikely trade. It is no longer being used, but it is left here in the formula for completeness

where i = Reporter i
j = commodity j
1 = target year
V = commodity value
T = total value (either for imports or for exports)
T' = adjusted total
r = reported commodity or commodities in blocking table (this can be just one commodity, but it could be several as well)

2. Inversion

If there are no reported data in the adjacent year, inversion or mirror statistics will be applied. Here are the cases:

i. No partially reported data and no entries in the blocking table

$$V(i,j,1) = W(i,j,1) * [T(i,1) / Z(i,1)]$$

where i = "Reporter" i
j = commodity j
1 = target year
0 = reported year
V = commodity value
T = total value (either for imports or for exports)
W = sum of partner values for commodity j
Z = sum of partner values of all commodities

ii. Commodity j exists in the blocking table

$$V(i,j,1) = B(i,j,1) * T(i,1)$$

where i = "Reporter" i
j = commodity j
1 = target year
V = commodity value
B = multiplier from the blocking table
W = sum of partner values for commodity j
Z = sum of partner values of all commodities

iii. Reporter i exists in the blocking table (but not commodity j), or partially reported data is available

$$V(i,j,1) = W(i,j,1) * [T'(i,1) / Z'(i,1)]$$

and $T'(i,1) = T(i,1) - SUM(V(i,r,1))$

and $Z'(i,1) = Z(i,1) - SUM(V(i,r,1))$

where i = "Reporter" i
j = commodity j
1 = target year
V = commodity value
W = sum of partner values for commodity j
T = total value (either for imports or for exports)

Z = sum of partner values of all commodities
Z' = sum of partner values of all commodities minus sum of reported / blocking values
r = reported commodity or commodities in the blocking table (this can be just one commodity, but it could be several as well)

Note that input data for case 2(i), 2(ii) and 2(iii) are reported data and extrapolated data.

Distribution of non-specified partners

In some cases, countries reported data with very high values for unspecified partners (which are: Antarctica, Bunkers, Free Zones, confidential partner, and partner not-elsewhere-specified). This is particularly true if a country intended to suppress the detailed partner breakdown for a specific commodity (such as crude oils). To take advantage of the inversion principle, the "real" detailed partner breakdown can be estimated.

In order for the inversion to be used for estimation, the following criteria must be met:

- The total value of non-specified partners is more than 50% than the value of the sum of all partners
- The total value from mirror statistics is more than 25% than the value of the sum of all partners
- The number of actual reporters is more than or equal to 100 countries for the estimation year

The new partner breakdown will be derived from mirror statistics and then multiplied by the ratio of the total value across all partners to the total value from mirror statistics.

$$P(i,j) = P'(i,j) * [T(i,j) / T'(i,j)]$$

where i = Reporter i
j = commodity j
P = partner value for commodity j
P' = partner value from mirror statistics for commodity j
T = sum of all partners for commodity j
T' = sum of all partners from mirror statistics for commodity j

Manual Estimation of Under-Reported and/or Confidential Trade

In some cases, reported data is incomplete (missing major commodities, such as oil exports). This can lead to underreported trade for certain commodities, as missing commodities are not estimated by default. To overcome this limitation, estimation can be forced for certain commodities even when the data are reported by country.

Disclaimer:
The manual estimation is based on the best judgment of the staff at the Business and Trade Statistics Section (BTS) of the United Nations Statistics Division (UNSD).

Aggregation to commodity TOTAL

With the introduction of new under-reported and/or confidential data, the aggregation procedure must be adjusted to account for the new estimated data and for data in commodity code 931 (commodity not-elsewhere-specified). Without this adjustment, the value of totals trade might

contain duplicates, because, for some countries, confidential data are lumped in commodity code 931. Therefore, they should be taken into account and removed during data aggregation.

Here are step-by-step procedures:

i. Aggregate data without under-reported and/or confidential data
ii. Aggregate data with under-reported and/or confidential data, but without commodity code 931
iii. Compare data for each commodity and partner combination from step (i) and step (ii), then combine them by taking the higher trade value
iv. Create new data for commodity code 931 by "subtracting" the result from step (iii) and step (ii)
v. Combine the result from step (iv) and step (ii). This is a new detailed data that contains reported, estimation of under-reported, and/or confidential data and adjusted commodity code 931 data
vi. Aggregate the result from step (v)

The final aggregated data usually have slightly higher totals compared to the original aggregated data. The differences can be attributed to the under-reported and/or confidential data.

Data Layout

The final result, called the Full Trade Matrix, has a data layout similar to UN Comtrade. In addition, an estimation flag has been added to indicate whether the value is estimated. All quantity information that was available in the original reported data has been removed. The data items are described in the table below:

Data Item	Variable Name	Description
Reporter	reporterCode	Reporting countries, includes reporter World
Year	period	Year
Trade Flow	flowCode	Import or Export
Commodity	cmdCode	TOTAL, all 1,2 and 3-digits SITC Rev.3 or Rev.4
Partner	partnerCode	Trading partners, includes partner World
Trade Value	primaryValue	Value of trade in USD
Estimation Flag	isReported	'1' indicates the records were estimated

Annex I: Example on Selecting Estimation Techniques

When to use the extrapolation or inversion technique? It depends on whether reported data are available for adjacent years. If they are, extrapolation can be applied to year-2 and year+2. Otherwise, inversion must be applied.

Take a look at an example of data availability below:

	2004	2005	2006	2007	2008
Algeria	R	R	R	-	-
Australia	R	R	R	R	R
Congo	R	R	-	-	-
Libya	-	R	R	-	-
Tuvalu	-	-	-	-	-

Legend: *R* denotes Reported Data

Algeria did not report data for 2007 and 2008. Fortunately, 2006 data is available, so missing data for 2007 and 2008 can be extrapolated from it. Congo has a similar case for 2005 and 2006, whereas missing data is extrapolated from 2004 data. However, data for 2008 must be inverted because the closest reported data was in 2005 (year - 3). For Libya, 2004 will be extrapolated from 2005 data, and 2007 and 2008 will be extrapolated from 2006 data. No data are available for Tuvalu; therefore, the only applicable estimation technique is inversion. Finally, no estimation is needed for Australia because all data are available.

The final decision table:

	2004	2005	2006	2007	2008
Algeria	R	R	R	X	X
Australia	R	R	R	R	R
Congo	R	R	X	X	I
Libya	X	R	R	X	X
Tuvalu	I	I	I	I	I

Legend: *R* denotes Reported Data, *X* denotes eXtrapolation, and *I* denotes Inversion.