# Construction of the Trade Data for the GTAP Data Base

**Mark Gehlhar, Angel Aguiar, Robert McDougall,**

**and Badri Narayanan**

Paper prepared for Organized Session on Different Approaches to Reconcile World Trade Asymmetries at the 22nd International Input-Output Conference, Lisbon Portugal, July 2014

## *Introduction*

Reconciling world trade asymmetries poses many challenges for developing an analytical database. In the GTAP Data Base, countries are connected through bilateral commodity trade. The construction of the trade data for the GTAP Data Base brings together trade statistics for merchandise and services trade data from different sources. In the GTAP model, an accounting identity exists whereby the value of goods imported at *cif* prices minus the value of transportation services equals the value of exports of exports at *fob* prices. The problem is that trade data as reported by partner pairs rarely if ever satisfies the basic accounting conditions. For nearly every bilateral transaction, the reported import value differs either substantially more or less than the reported export value. Thus, reported trade statistics "as reported" are unsuitable for the GTAP data base. What is done to the data to satisfy the accounting identities is central to this paper.

Uncertainty associated with any component of the base data is troubling given that it can influence an outcome for policy analysis. However, the fact that large discrepancies exist in the reported trade statistics does not alone suggests the data is not credible. There is no doubt there is erroneous trade data reporting. But errors are made known only by having trade reported by both exporter and importer. In many cases we can be highly confident in the data upon understanding the source of errors and taking corrective action.

There are many choices in how to deal with trade asymmetries. The decision largely depends on the end goal. A guiding principle for reconciling discrepant trade flows for GTAP is to preserve as much factual information as possible. This becomes important because the reliability of trade data affects credibility of model results. Much of the challenge lies in distinguishing factual information from fiction in partner trade flows. The method to reconcile bilateral merchandise trade data for the GTAP Data Base Trade reconciliation primarily involves is a decision to accept or reject reported trade flows. As a last resort it may involve a compromise by adjusting data using a weighting scheme. We outline the specifics of our approach in this paper.

The GTAP Data Base is intended for modeling all economic activity including production, trade, and consumption of all goods and services. The main source for merchandise bilateral trade data is the United Nations COMTRADE data. However this covers only trade in goods not services. Accordingly, we need another data source for services. In GTAP 8, we use UN service trade data and EUROSTAT's international trade in services (Narayanan, et al., 2012).

Starting in GTAP version 7, the reconciliation procedure used in previous versions of the GTAP Data Base (Gehlhar, 1996) was enhanced by adding an optimization procedure to obtain more accurate trade results for China and Hong Kong.

The largest discrepancies in bilateral trade are the result of re-export activity. A large part of

China's trade passes through Hong Kong, which earns substantial revenue from the difference between import and re-export prices. We account for this revenue as an export of trade services from Hong Kong to the countries of destination of the merchandise. In GTAP 8, we also account for re-exports for the Netherlands.

The GTAP Data Base also contains data on international trade margins, that is, the services used or costs incurred in moving goods from point of export to point of import. Margin services are considered exports of the country that supplies the service, and imports of the country that receives the merchandise to which they are applied to. Accordingly, they are included in the services trade statistics. Another special case is travelers' expenditures. The services trade statistics treat travelers' expenditures as a distinct commodity, but in the GTAP data structure, they are counted as trade in the goods and services actually purchased. Purchases in one country by residents of another country are considered exports from the first country to the second. This includes tourism, but also such things as expenditures incurred in short-term employment overseas. And finally the result of this construction process is a reconciled trade data set that can be used for economic analysis in a general equilibrium type of model.

## *Causes of Asymmetry in World Trade*

Non-reporting is one of the simplest reasons for asymmetry in world trade flows. The fact that not all countries report trade data or that meet the international standards for reporting requirements set forth by the United Nations Statistics Division (UNSD) does not pose a serious problem for constructing the global merchandise trade data for GTAP. The number of non-reporting countries using the Harmonized System of classification has ranged from 30-50 in recent years (see Appendix tables). Most countries are relatively minor trading partners in global trade. Estimating missing trade as a result of non-reported trade is not discussed in this paper.

A more serious problem is the suppression of bilateral commodity detail by reporting countries. This takes the form of a reporting country intentionally withholding partner flow data. This is common for energy-based commodities. For example, Saudi Arabia's export destinations for petroleum oil are often suppressed. Similarly, India's imports for many energy commodities are not identified by exporting partners (see Appendix tables for details).

The most common cause of for asymmetry in world trade data mistaken trade flows from miss-classification, miss-identified partners, intentionally over or under-invoicing the value of goods, or improper valuation of goods due to currency conversion mistakes. All of these factors lead to discrepant bilateral trade flows.

As a result there is no known means to meaningfully adjust trade flows using systematic reporting errors or upward or downward reporting biases. What can be measure in a systematic manner is that certain reporters are more prone to reporting errant data than others. And each reporter does not exhibit the same reporting behavior across all countries. An exporter may demonstrate competence in reporting iron and steel but fail to show any credibility reporting electronic equipment. It is for that reason we cannot pre-judge any country's trade data based only on a single industry. Often a country's government may have little influence in the actual reporting of trade data but rather it is private entities and the shipper's custom declaration that largely determines data quality and truthfulness in the data.

## *Developing Trade Reliability Indexes*

Determining what trade flows are more likely credible from those that are errant is key to our reconciliation approach. To do this we must devise an indicator of the reporter's commodity specific reliability both as an exporter and importer.

As a first step for determining the size of the bilateral discrepancies we must ensure reported exports and reported imports are comparable using the same value basis. Typically the reported import value includes freight and insurance cost. The conversion of the import value is made to eliminate the transport margin when it is inclusive. We denote $M^{cif}_{(i,r,s)}$ as reported imports by importer *s*, from exporter r, for commodity *i*, valued on a *cif* basis. We denote $T^{c}_{(i,r,s)}$ as the cost to carry commodity *i*, from exporter *r* to importer *s*. The reported import value minus the transport cost becomes the import value free of transport cost, denoted here as $M^{fob}_{(i,r,s)}$.

$$M^{fob}_{(i,r,s)} = M^{cif}_{(i,r,s)} - T^{c}_{(i,r,s)}$$

For each trade flow having counterpart trade reported we measure the relative size of the discrepancy. This difference is compared with an established allowable threshold factor or accuracy level, denoted as $Ae$. Whether the export value is greater or less than the import value is immaterial for our purposes. Whether the exporter or importer grossly under or over-reports is not important. We are primarily interested in whether or not the transaction is accurate or not.

In this approach counterpart trade flows are either deemed "accurate" or "errant" based on the following conditions:

If $\frac{|\ M^{fob}_{(i,r,s)} - X^{R}_{(i,r,s)}\ |}{M^{fob}_{(i,r,s)}} \leq Ae,$

then the transaction is deemed accurate because the relative size of the discrepancy does not exceed the allowable error factor. In that case, both the exporter and importer exhibit credibility for reporting commodity *i*. When this condition is met both the exporter and the importer deserve credit for the reporting this transaction. By giving credit, we add up all reported value deemed accurate for each respective reporter as shown further in the paper to develop an index.

If however, $\frac{|\ M^{fob}_{(i,r,s)} - X^{R}_{(i,r,s)}\ |}{M^{fob}_{(i,r,s)}} \geq Ae,$

then the transaction is deemed errant. As a reporting pair both fail in this instance to exhibit credibility in reporting commodity *i.* However, this alone does not in any way suggest which reporter is errant in its reporting. Nevertheless we cannot assign credit to either exporter or importer. How this transaction will be reconciled has yet to be determined.

All that is known from this information is that as a reporting pair, they fail to show creditable thereby neither deserves credit for what they have reported. Judging individual commodity flows is of little use for determining a reporter's overall credibility or reputation as a reporter. We examine each reporter in a comprehensive manner. To do that it involves evaluating the transactions with all trading partners.

Once this is done we retain an accumulative total for both the accurate and errant trade flows on a commodity-specific basis. We denote $M^{A}_{(i,r,s)}$ as only those import flows deemed accurate. We denote $M^{TA}_{i,s}$ as the sum of all trade reported by importer *s* for commodity *i* . As a reporting pair these transactions are determined to be sufficiently accurate because of the relatively small reporting discrepancy.

$$M^{Ta}_{i,s} = \sum_{r} M^{a}_{(i,r,s)}$$

Likewise we denote $X^{A}_{(i,r,s)}$ as those reported export flows deemed as accurate transactions. We

denote $X_{i,r}^{TA}$ as the sum of all trade reported by exporter *r* for commodity *i* where as a reporting pair the transactions are determined to be sufficiently accurate.

$$X_{i,r}^{Ta} = \sum_{s} X_{(i,r,s)}^{a}$$

As with accurate transactions we retain an accumulative value of trade for all errant transactions on a commodity-specific basis. Here we denote $M_{(i,r,s)}^{e}$ as those reported import flows deemed as errant transactions. We denote $M_{i,s}^{Te}$ as the sum of all trade reported by importer *s* for commodity *i* . As partner pairs these transactions are insufficient to qualify as accurate because of relatively large reporting discrepancy.

$$M_{i,s}^{Te} = \sum_{r} M_{(i,r,s)}^{e}$$

Likewise we denote $X_{(i,r,s)}^{e}$ as those reported export flows deemed as accurate transactions. We denote $X_{i,r}^{Te}$ as the sum of all trade reported by exporter *r* for commodity *i* where transactions are determined to be errant.

$$X_{i,r}^{Te} = \sum_{s} X_{(i,r,s)}^{e}$$

For developing a meaningful index of a reporter's overall reliability on a commodity-specific basis, we use the accumulative value described above expressed is a ratio. We denote $RIM_{(i,s)}$ as the importer reliability index for the commodity *i* for importer *s*. The index is expressed as the ratio of accurately reported imports divided by the total of accurate and errant reported imports.

$$RIM_{i,s} = \frac{M^{Ta}_{(i,s)}}{M^{Ta}_{(i,s)} + M^{Te}_{(i,s)}} \quad ,$$

Similarly, we construct an index for the exporter on a commodity-specific basis. We denote $RIX_{(i,r)}$ as the commodity-specific reliability for the exporter.

$$RIX_{i,r} = \frac{X^{Ta}_{(i,r)}}{M^{Ta}_{(i,r)} + M^{Te}_{(i,r)}}$$

We note here that total trade reported by the importer is actually less than the sum of accurate and errant trade. This is because in addition to the trade involving in counter-part reports each reporter conducts trade with non-reporting partners. However, these transactions are not included in the above ratio. The reason is that including such trade would bias the index downward as a result of non-reporting and not because of lack of reliability.

## *A Numerical Example in Reconciling Bilateral Trade*

We now provide a numerical example applying the approach described above. The data sample here is illustrative of the problems encountered in global bilateral trade statistics. Although the reported import values shown in table 1 are free of transport costs, reported exports often exceed the import value. In fact bilateral discrepancies are rarely if ever explained by transport cost

alone playing a very minor factor is the differences reported.

Nearly all reporting countries conduct trade with one or more non-reporting countries. In this sample reporters would be discredited if the measure of reliability included trade with the non-reporting partner. Those reporters in particular which conduct considerable trade with non-reporters would be unfairly discredited reporters.

Table 1. Reported Imports

| | Importer | | | | |
|---|---|---|---|---|---|
| Exporter | 1 | 2 | 3 | 4 | Total |
| 1 | | 15 | 30 | | 45 |
| 2 | 145 | | 40 | | 185 |
| 3 | 55 | 100 | | | 155 |
| 4 | 12 | 3 | 15 | | 30 |
| Total | 212 | 118 | 85 | | 415 |

Table 2. Reported Exports

| | Importer | | | | |
|---|---|---|---|---|---|
| Exporter | 1 | 2 | 3 | 4 | Total |
| 1 | | 17 | 21 | 10 | 48 |
| 2 | 41 | | 143 | 2 | 184 |
| 3 | 47 | 98 | | 1 | 145 |
| 4 | | | | | 0 |
| Total | 88 | 115 | 164 | 13 | 377 |

Typically discrepancies appear sporadic and inexplicable. For example, there no apparent reason why there is 250 percent discrepancy between importer 1 and exporter 2 while only a 17 percent discrepancy between exporter 1 and exporter 3. At the same time importer 2 shows only a 2 percent discrepancy with exporter 3. It is for these reasons that there is no apparent solution for reconciling trade flows.

Table 3. Reconciled Trade Flows with reliability index

| Importer | Exporter | Reported Imports | Reported Exports | Discrepancy | RIM | RIX | Reconciled |
|---|---|---|---|---|---|---|---|
| 1 | 1 | | | | 0.28 | 0.64 | |
| 1 | 2 | 145 | 41 | 2.45 | 0.28 | 0.00 | 145 |
| 1 | 3 | 55 | 47 | 0.17 | 0.28 | 1.00 | 47 |
| 1 | 4 | 12 | | | 0.28 | na | 12 |
| 2 | 1 | 15 | 17 | 0.12 | 0.98 | 0.64 | 15 |
| 2 | 2 | | | | 0.98 | 0.00 | |
| 2 | 3 | 100 | 98 | 0.02 | 0.98 | 1.00 | 100 |
| 2 | 4 | 3 | | | 0.98 | na | 3 |
| 3 | 1 | 30 | 21 | 0.43 | 0.30 | 0.64 | 21 |
| 3 | 2 | 40 | 143 | 0.72 | 0.30 | 0.00 | 40 |
| 3 | 3 | | | | 0.30 | 1.00 | |
| 3 | 4 | 15 | | | 0.30 | na | 15 |
| 4 | 1 | | 10 | | na | 0.64 | 10 |
| 4 | 2 | | 2 | | na | 0.00 | 2 |
| 4 | 3 | | 1 | | na | 1.00 | 1 |
| 4 | 4 | | | | na | na | |

Using the data shown above reliability indices were calculated for each reporting importer and exporter. The allowable error factor in this case is 0.2, meaning the absolute percentage difference between reported exports and imports cannot exceed 20 percent to qualify as credible.

We either accept or rejected reported trade flows based solely on the reliability information. The reliability index as stated above is the ratio of accurately reported trade to accurately and errant trade. For example, the reliability index for importer 1 is 0.28. This is the ratio of accurately reported trade (55) to accurate plus errant trade (55 + 145=200).

Most noteworthy in our same here is the how the largest discrepancy is resolved for trade between exporter 2 and importer 1. In this case we disregard the reported trade for exporter 2 in favor of importer 1. Importer 1 has demonstrated credibility by accurately reporting its trade with exporter 3, having discrepancy (0.17) less than the allowable error. Exporter 2 failed to demonstrate bilateral credibility despite that fact that the total it reported (184) is nearly identical to what its reporting partners reported (185). Our goal therefore would be to preserve the row sum for exporter 2 of 185 because this sum is highly credible. However, there is evidence that exporter 2 is prone to miss-identifying its partners. As reported exporter 2 mistook importer 1

for importer 3 and as a result reliability is zero. Reporters prone to miss-identifying partners is a common problem but detectable one. Using an alternative method using a weighting scheme to adjust trade flows in this instance would have compromised the authentic data and change the row some for exporter 2 different than 185. This is because the errors reported by exporter 2 would have been used to adjust what is reported for exporter 1. This would be a mistake to alter the row sum.

Insert table 4 here

## *Further Extensions of Reliablity Indictors for Adjusting Re-exports*

One of the primary reasons for the discrepancies is the intermediary role of Hong Kong in China's external trade. A large share of China's trade with the world passes through Hong Kong, while current reporting practices in China and their trading partners do not fully reflect this role thus provide a misleading picture of the origin and final destination of Chinese exports and imports, leading to conflicting official bilateral trade balances. For example, China only started to identify the final destinations of its goods shipped through Hong Kong in1993, but the work is incomplete, in part because traders really do not know the final destinations when good left China. In these cases, they are recorded as exports to Hong Kong by the Chinese Customs. For this reason, it is not a surprise to see that in Chinese customs statistics, Hong Kong is one of China's largest export destinations, only second to the Unites States. In fact, Hong Kong re-exports most its imports from China to other countries.

We employed a mathematical programming model to estimate re-export markup and reconcile detailed bilateral trade statistics from China, Hong Kong and their trading partners. Five key steps to link the model with real world trade statistics are discussed in details. The model was applied to 2004 bilateral world trade data in GTAP sectoral and region classification to produce Hong Kong re-exports adjusted trade flows contributing to version 7 GTAP database. Preliminary result shows that the model is able to eliminate the statistical discrepancy efficiently. Hong Kong's re-export mark-up, each trading partner's transshipment via Hong Kong as percent of the country's total exports to and import from China, and adjusted bilateral balance of trade among China, Hong Kong and their partner countries by each covered commodity are all part of the model solution. Therefore, the model provides a flexible tool to reconcile trade statistics from China, Hong Kong and their trading partners simultaneously. The model is quite flexible in its data requirement and has desirable theoretical and empirical properties, therefore can be applied to reconcile direct and indirect trade in other transshipment settings. It not only provides a convenient tool for the preparation of global trade data in future versions of GTAP database, but also contribute to the methodological development on how to accurately estimate and reconcile

discrepancies in international trade statistics when transshipment and re-export activities heavily diminish the ability of a country identifying its correct partner countries.

## *References*

**Appendix Table A1**

**Appendix Table A2**

**Appendix Table A3**