16.D

Tariff Data

Antoine Bouët, Yvan Decreux, Lionel Fontagné, Sébastien Jean, and David Laborde

16.D.1 Introduction

Simple questions such as the comparison of the level of protection across countries and industries are hardly satisfactorily answered at the worldwide level. The growing complexity of trade policies has left negotiators, as well as economists and the public debate, without suitable information about the present state of trade policies. This chapter presents the applied protection data used in the GTAP 6 Data Base which originates from the Market Access Map (MaMap) data base, the result of a joint effort by the International Trade Centre (ITC) (United Nations Conference on Trade and Development and World Trade Organization) and the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) to systematically collect detailed and exhaustive information on the level of applied trade barriers.

The source information on border protection come from national customs agencies. It is defined at the tariff-line level. The definition of tariff lines vary widely across countries but it is always based on the six-digit level of the Harmonized System classification (hereafter, HS-6 level). This non-harmonized information is not a suitable basis for a wide-ranging analysis of border protection across the world. UNCTAD's TRade Analysis and INformation System (TRAiNS) has played a leading role in trying to gather the relevant information. As a result from the collection by UNCTAD of information from national custom schedules, it provides with data at the tariff-line level about applied tariffs (ad valorem and specific) and Tariff Rate Quotas (TRQs), as well as import flows by origin for more than 140 countries. However, at least until recently, TRAiNS suffered from incomplete coverage of preferential agreements, and did not include ad valorem equivalent (AVE) calculations. For years, TRAiNS has been the main source of international information on applied border protection, and the only one allowing for a worldwide coverage. WTO's Integrated Database (IDB) is now an alternative source, although it only includes applied...
MFN tariffs, and does not reach a comparable coverage. Other efforts have been made to gather wide-ranging, harmonized data on border protection. Some databases have a more limited coverage in terms of products and/or countries (Agricultural Market Access Database (AMAD) or the Hemispheric Database). The Integrated Tariff Analysis System (ITAS) (see Fry et al., 2004) allows for a complete analysis of applied and bound tariffs and of the outcome of cutting bound protection, but for 17 countries and for industrial products only. Databases also differ from a methodological point of view, for example, in terms of aggregation procedure and estimation of AVEs.

Thus, although a lot of information existed, no well-suited, comprehensive assessment of AVE applied protection across the world was available. This resulted, in particular, in most assessment of the impact of multilateral trade liberalization being carried out without taking into account specific tariffs, nor trade preferences, even if the GTAP network has made considerable efforts in order to offer a consistent database. Gathering such information in a consistent and tractable way has been the first motivation of the MAcMap database. Beyond proper collection and harmonization of information, however, the development of MAcMap also aimed at dealing with the main methodological hurdles encountered when trying to produce tariff data well-suited for large-scale analysis, in particular as far as the calculation of the AVEs of specific duties and the aggregation procedures are concerned.

Basically, MAcMap is a set of files at the tariff-line level that can be used for several purposes, noticeably single-client studies and interactive web databases for the business community realized at the ITC. The dataset used in GTAP derives from one specific application of MAcMap, namely the construction and consolidation by the CEPII of a database at the HS-6 level, intended to provide a set of consistent and exhaustive AVEs of applied border protection across the world (165 reporting countries are covered, for 5,111 products, with 208 partners) in 2001, suitable for analytical purposes. MAcMap-HS6 is regularly improved and updated, and further information is available on the CEPII's website (www.cepii.fr).

16.D.2 Collecting and Harmonizing Information

For each importing country, information on the various instruments of protection at the border (ad valorem tariffs, special tariffs, quotas, etc.) is maintained in the MAcMap database at the most disaggregated level possible: the tariff-line level. In order to stick to the bilateral option, unilateral preferences and regional agreements are exhaustively documented. This is a rare feature among worldwide data bases on border protection.

Protection data in the MAcMap database originates from the source files of TRAINS, from countries notifications to the WTO, from AMAD, and from national custom information (reported to UNCTAD, or directly to ITC). Occasionally, this information is complemented by data from other relevant sources, such as: administrations, statistical institutes, international organizations, and

---

4None of these two aspects were accounted for in the tariff data included in the GTAP 5 Data Base (see Dimaranan and McDougall, 2002), which has been the workhorse for empirical assessments of the impact of multilateral liberalization.
websites of regional agreements. Information on antidumping duties are drawn from the notifications of member countries to the WTO. This combined information characterizes the trade policy applied by 165 countries to 208 exporting partners. It includes tariffs (ad valorem, specific, mixed, compound and antidumping duties), and tariff quotas\(^5\).

The protection data is generally available at the tariff-line level but no international harmonization of classifications exists beyond the HS6 level. Since HS6 is also, for most countries, the most detailed classification for which foreign trade statistics are available, this classification (more specifically, its revision 1 of 1996) has been adopted to put together all data into a database tractable for analytical purposes.

Trade data are sourced from the BACI database\(^6\). Based on COMTRADE, BACI includes a harmonization of classifications, a reconciliation of mirror declarations, and a treatment of unit values, in order to make them comparable and to check their consistency. This results in a harmonized trade data base at the HS6 level.

### 16.D.3 Calculating AVEs

The source information includes various instruments which cannot be directly compared or summed and which are not all readily usable in large-scale modeling exercises. The natural solution to overcome these problems and to make the data base fully operational for analytical purposes is to compute AVEs of each instrument.

#### 16.D.3.1 Specific Tariffs

Specific tariffs were converted to AVEs by dividing the duty by a unit value (UV).\(^7\) The whole problem lies in the choice of the UV, and this raises important issues, both from a statistical and from a theoretical point of view. This is not a secondary issue, given that vertical specialization of countries along the quality ladder has been shown to be widespread, and that poorer countries tend to specialize in lower quality goods. It means that the restrictive impact of specific tariffs may vary substantially across trade partners, depending on their quality specialization, and that their impact is likely to be systematically stronger on imports from developing countries.

From a statistical point of view, using bilateral UVs (BUVs) might seem appealing because it is fully consistent with the amount of tariff receipts collected, and because it allows the quality specialization of the corresponding trade flow to be taken into account. However, it is flawed with

\(^{5}\) *Ad valorem* equivalents of prohibitions were also taken into account in previous releases. Given their very limited number, however, their omission now does not make significant changes.

\(^{6}\) BACI is the French acronym for *Base de données pour l’Analyse du Commerce International*. For further details, see: http://www.cepii.fr/anglaisgraph/bdd/baci/baci.pdf.

\(^{7}\) Alternative methods include an estimation based on price wedges; that is hardly tractable at the level of detail and coverage of this data base. The “revenue” method consists of dividing tariff revenues by the value of imports. In addition to being difficult to implement, it is clearly not suitable in the presence of preferential agreements (see e.g. WTO, 2003, for a discussion).
lack of robustness, thus introducing significant variance across AVE protection faced by different partners in the same market, often to a surprising extent.\(^8\) It is due in many instances to measurement errors, or even reporting errors (errors in the physical units reported, for instance). It also stems from the fact that small bilateral flows are not always representative, and are more prone to exhibit out-of-range UVs. In addition, it can only be implemented as such when imports did take place.\(^9\)

Computing AVEs based on a worldwide import average, like for instance Gibson et al. (2001), is appealing in terms of robustness, but it completely disregards the question of quality differences.

Hence the choice of UVs must satisfy two purposes: to reflect the different restrictive impacts of a specific tariff according to the vertical specialization of the trade partners, and to exhibit moderate volatility. This need for an intermediate approach has led to the use of the median unit value of worldwide exports from an exporter's reference group in the AVE calculations in MAcMap. Each country is grouped with a reference group of similar countries identified using a hierarchical clustering analysis based on GDP per capita (in terms of PPP) and trade openness.\(^10\) The five groups determined as a result of this procedure can be loosely labeled as follows: (1) richest countries; (2) high openness, middle income countries; (3) low openness, middle income countries; (4) high openness, low income countries; (5) low openness, low income countries. The full set of countries and reference groups is provided on the CEPII website (www.cepii.fr).

More specifically, this ERGUV (Exporter's Reference Group Unit Value) is calculated using "weighted" medians, obtained by assuming that each UV is repeated as many times as the underlying trade flow contains dollars.\(^11\) For the sake of robustness, UVs are computed based on three-year-average trade flows (across the 2000-2002 period). In addition, a filter rule is used. Any ratio of ERGUVs to the world median unit value outside the bracket \([1/3 ; 3]\) is truncated to the top or bottom limit. Moreover, a sequential procedure is used to fill missing values for reference groups: any blank is substituted by the value of the closer reference group.

Using this ERGUV offers four advantages:

(i) the differences in unit values across countries, linked in particular to different products quality, is accounted for;

(ii) the endogeneity bias (there is an incentive to alter product quality in response to a specific tariff) is lessened compared to a bilateral unit value, since the value is calculated based on worldwide exports;

\(^8\)This method has been used in previous releases of MacMap, and its use for analytical purposes proved to suffer from insufficient robustness.

\(^9\)Although this does not exclude the possibility of choosing an alternative rule when no imports took place.

\(^10\)This is a major change from the first release of MAcMap, in which reference groups were based on an arbitrary classification (GDP PPP per capita).

\(^11\)Using traded quantities as weights instead of trade flows could appear more consistent, but from a statistical point of view, the distribution of trade flows is more robust and less sensitive to measurement errors.
(iii) more importantly, ERGUVs are more robust to measurement errors than BUVs: due to the use of median, outliers do not influence strongly the result, in contrast to a calculation based on the average.

On the one hand, one drawback with the use of ERGUV is that some information is lost relative to differences in product quality at the bilateral level. On the other hand, this quality information is only useful insofar as it is structural in nature, in which case we would expect it to remain relatively constant across reference groups (if not, it is likely that these differences are endogenous to protection). In particular, if unit values differ because of an exporter's cost-competitiveness, and not because of quality-related factors, then there is no point in computing a different AVE.

**16.D.3.2 Mixed and Compound Tariffs**

Mixed tariffs, i.e. tariffs involving a choice (a maximum or a minimum operator) between various terms raise the question of the term to be chosen. This choice is made without any calculation, using the following rules:

— when the tariff is defined as an *ad valorem* base tariff, with in addition a cap and a floor (which are defined in specific terms), the base tariff is retained. If the base tariff is in specific terms and the cap and the floor are *ad valorem*, a simple average of the two bounds is retained. This prevents from adding any additional noise through AVE calculation;

— when the tariff involves choosing between two terms, priority is given to *ad valorem* tariffs\(^ {12}\) for the same reason as previously;

— when the tariff is a choice between two compound tariffs, the second one is ignored. Additional elements are ignored.

**16.D.3.3 Tariff Rate Quotas**

When trying to summarize and harmonize the information about border protection, TRQs originate a number of problems. Due to their intrinsic nature, they cannot be perfectly summarized through an AVE, hence the question about the right way to handle them in multi-country, multi-product databases and models (see e.g. IATRC, 2001, Liapis and Blitz, 2001). Since TRQs are defined at very different levels (from HS-4 to tariff line), their treatment must include a consistent aggregation procedure. Keeping the basic information about Inside Quota Tariff Rate (IQTR), Outside Quota Tariff Rate (OQTR) and quota as such, for instance, is not consistent with aggregation: the average IQTR and OQTR, combined with the sum of quotas, do not describe consistently the aggregate impact of various TRQs. Finally, another requirement is that the information produced could be

---

\(^{12}\)This means that if the tariff is the maximum between a compound tariff (the sum of a specific and an *ad valorem* tariff) and an *ad valorem* tariff, only the second term, i.e. the *ad valorem* tariff is taken into account.
conveniently combined with the tariff data, both for descriptive and statistical matters, as well as for incorporation in CGE analyses.

The methodology has been radically changed in MAcMap-HS6 version 1, based on the idea that the calculation should reflect the marginal level of protection, as well as the rents involved, on a bilateral basis (figure 16.D.1).\textsuperscript{13}

Data on tariff quotas mainly comes from the AMAD data base.\textsuperscript{14} The available information is: reporter, quota identification, product coverage, quota, imports, allocation of quotas, inside tariff, outside tariff.\textsuperscript{15}

Figure 16.D.1: Tariff rate quotas treatment procedure

A fill rate is first computed for each TRQ, as the ratio of imports to the size of the quota itself.\textsuperscript{16} As soon as a quota involves a specified bilateral allocation, this fill rate is computed separately for each (group of) partner(s) which is allocated a separate quota. A shadow tariff is then defined as the \textit{ad valorem} tariff that would lead to the same level of imports as is observed under the tariff rate quota. Three market regimes are considered, depending on the level of the fill rate:

0 – If the fill rate is less than 90\% (quota not binding), the inside quota tariff rate is chosen as the shadow rate.

1 – In the (90\%,99\%) range (quota assumed to be binding\textsuperscript{17}, OQTR prohibitive), a simple arithmetic average is used (except in a few cases where external information allows a better guess to be made).

\textsuperscript{13}In previous releases of MAcMap, the AVE of tariff rate quotas (TRQs) was calculated as a trade-weighted average of inside- and outside-quota tariff rates. This method proved to understate significantly the real level of protection implied by TRQs. When out-of-quota imports are limited, in particular, it results in an AVE close to the inside-quota tariff rate (IQTR), while in fact this reflects the very high protection resulting from a high outside-quota tariff rate (OQTR).

\textsuperscript{14}Thanks to Paul Gibson for making these data available.

\textsuperscript{15}Ad valorem equivalents of specific tariffs and mixed tariffs (inside or outside the quotas) are calculated according to the same methodology as tariff barriers.

\textsuperscript{16}This fill rate is thus not restricted to being inferior or equal to one.

\textsuperscript{17}The quota is assumed to be binding as soon as the fill rate exceeds 90\%, since administration methods or other reasons might well prevent a small share of the quota from being used, even though the quota is actually binding.
2 – If it is higher than 99% (quota binding, OQTR not prohibitive), the shadow rate is equal to the outside quota tariff rate (OQTR).¹⁸

In the case of missing data about the quota level and its fill rate, regime 2 is assumed but the rent is set to zero¹⁹.

This data base is then merged with MAcMap, keeping information on AVE of tariff quota (ad valorem and specific components), IQTR, market regime and allocated quota (for every reporter-partner-HS6 relation).

The quota rent (rent) is computed as follows:

\[
rent = \text{Min}\left( uv \times q \times \frac{SR - IQTR}{1 + IQTR}, tradev \times \frac{SR - IQTR}{1 + SR} \right)
\]

Where \( uv \) refers to the unit value, \( q \) to the quota allocated to the line, \( tradev \) to the trade value, \( SR \) to the shadow tariff rate, and \( IQTR \) to the inside quota tariff rate.

The shadow tariff is also used to compute a new applied tariff when it is inferior to the original applied tariff in regimes 0 and 1, or in some cases when the applied tariff is missing or incorrect. However it is ignored when it originates from a quota defined at the HS4 level.

**16.D.3.4 Prohibitive Duties and "Water in the Tariff"**

The presence of prohibitive tariffs is problematic when calculating AVEs with a view of using them for CGE analysis. Indeed, prohibitive tariffs are seldom exactly equal to the lowest tariff that would drive import demand down to zero. Generally, there is "water in the tariff", meaning that actual tariffs exceed this lowest level. This implies that a small tariff cut would not lead to any change in the corresponding trade flow. This is misleading for analytical purposes, since the study of the consequences of trade liberalization generally relies on the assumption that cutting tariffs systematically increases import demand. When the tariff is really prohibitive (i.e., initial imports are zero), the problem is of a specific nature, in the sense that a standard CGE model cannot give any insight about the impact of liberalization: it needs to use initial imports as a basis for calibrating the import demand function (or the share coefficient of imports in the utility function.)

¹⁸ An apparently more logical approach would lead to chose the minimum of the outside rate and the applied rate. Nevertheless, due to the occurrence of tariff rate quotas on lines with tariff peaks within the HS6 position for which the applied tariff is defined, the difference between the inside rate and the minimum of the applied rate (at HS6 level) and the outside rate will poorly reflect the tariff gap that generates the rents. As outside rates and inside ones are defined at the same level, the chosen solution appears to be better.

¹⁹ In a few cases, like the Japanese tariff rate quotas on rice, a special treatment is made for dealing with “water in tariffs” and with the capture of the rents by national agents: regime 1 is assured and rents are set to zero.
But initial demand is generally not zero in such analysis, for two reasons. Firstly, while in theory, by definition, there is no import for a product protected by a prohibitive tariff, in practice, non-zero trade flows are frequently recorded, even in presence of a tariff set at a prohibitive level. It results from the existence of very specific niches (within a given tariff line), or of the behavior of very specific consumers. However, these imports are generally very limited, they are not representative of the whole demand for the product, and their price responsiveness is almost zero (otherwise, this demand would not exist, given the initial level of the tariff.) Secondly, CGE analysis did not use to be carried out at a very detailed level. Every sector aggregates a number of products. If one product is protected by a prohibitive tariff, then it will probably be aggregated with other products, with lower protection and significant initial imports. The AVE tariff duty calculated for the prohibitive tariff will then enter the average tariff computed for the sector if, as it is the case in MACMap, this average is not import-weighted.

In both cases, this means that the presence of “water in the tariff” would lead to overstate the impact of a tariff cut. Properly correcting this bias would require removing the water in the tariff, i.e. replacing any prohibitive tariff by the lowest prohibitive tariff duty for the product concerned. This is far beyond the scope of this database. Therefore, we take the approach of establishing an upper limit to the AVE in the model starting at the HS6 level. This upper limit is set to 1,000% for the sum of all instruments. Practically, this upper bound is not applied in the source data base, but it is used as a preliminary stage for any aggregation intended to feed a CGE model. In particular, it is applied when building the MACMap for GTAP data base.

### 16.D.4 Aggregation Procedures

While the corresponding questions have been already widely discussed (see e.g. Balassa, 1965; Laird, 1996), there is still no consensus about how to acknowledge the respective importance of products (as well as exporters and importers), without introducing very large biases. In building the MACMap data base, a specific aggregation procedure has been developed using a weighting scheme based on a clustering of countries into reference groups (see Bouët et al., 2004, for details). This procedure intends to minimize the extent of the well-known endogeneity bias, and a dataset using this aggregation procedure is available as a separate table on GTAP website (http://www.cepii.fr/anglaisgraph/bdd/macmap.htm). In order to ease comparability with previous releases and to ensure consistency with tariff receipt data, the protection data used in GTAP 6 are aggregated using import-weighted averages. As a general rule, given the existence of these two competing datasets (aggregation based on import-weighted average, the default, and aggregation based on reference-group clustering), we think that the downloadable dataset should be preferred for

---

20 This methodology is based on a clustering of countries into the same reference groups as for estimating AVE of specific tariffs. For each hs6 – partner – reporter triplet, the following weight is used:

Where \( M_{hs6,partner,reporter} \) refers to the value of product “hs6” imported by country “reporter” from country “partner”, “RefGroup(reporter)” refers to the reference group the country “reporter” belongs to, and “." refers to the total, so that \( M_{.,reporter} \) refers to the total value of “reporter’s” imports. The aggregation SAS program is given on the CEPII website (www.cepii.fr).
uses dealing with trade liberalization, since the measure of protection then focuses on trade restrictiveness. For other uses, the default dataset is preferable, since it allows consistency with public finance data, as well as with the user cost of intermediate inputs.

During the aggregation procedure, an infinitesimal trade flow is added when there is no trade between two GTAP zones for a given GTAP sector.

It is noteworthy that for all aggregations, the CEPII relies on the BACI trade data base instead of the GTAP trade vector, so that minor differences may appear between trade-weighted averages presented in the CEPII's documentation (MAcMap technical note by example, or the Table 16.D.1) and those obtained with the GTAP aggregation procedure21.

16.D.5 GTAP 6 vs. GTAP 5

GTAP 5 makes use of AMAD 1998 data for agriculture, and of TRAINS22 1997 data for non agricultural products. As outlined above, and in addition to the difference in the period of reference, methodological differences between these datasets and MAcMap abound. Still, it is interesting to make clear how GTAP 5 average duties differ from our assessment of average MFN AVE tariffs in MAcMap, at the country level.23

Comparing GTAP 5 and MAcMap average AVE duties by country does not highlight a systematic bias. Nevertheless, assessment of sector protection may hugely differ between data bases in a few cases, as illustrated for instance by agriculture in Estonia, Mozambique and Zimbabwe. More importantly, these two measures appear to be very poorly correlated across countries (the R-square of the bivariate regression is 57% for non-agricultural products, and only 16% for agricultural products). The change in protection between 1997 (or 1998 in agriculture) might explain this mismatch, but only for a small part. We argue that this poor correlation illustrates the value added brought about by the deep data and methodological improvements involved in constructing MAcMap.

---

21 Moreover, when figures are computed for an EU member, no intra-European trade flow is taken into account.
22 As available through WITS interface.
23 The comparison with GTAP 5 is made using MFN tariffs, since GTAP 5 does not take into account preferential agreements, except the EU, the EU-EFTA agreement, ANZCERTA, SACU and NAFTA. On these two figures, GTAP 5 country codes are used.
Table 16.D.1 Protection Data in GTAP 6 Using Two Difference Aggregation Procedure and Compared to GTAP 5 (Ad valorem rates)

<table>
<thead>
<tr>
<th>Source</th>
<th>GTAP 5</th>
<th>MAcMap, Applied Tariffs</th>
<th>MAcMap, MFN Tariffs</th>
</tr>
</thead>
<tbody>
<tr>
<td>World</td>
<td>4.5</td>
<td>3.6</td>
<td>5.3</td>
</tr>
<tr>
<td>Argentina</td>
<td>14.0</td>
<td>12.6</td>
<td>10.0</td>
</tr>
<tr>
<td>Australia</td>
<td>5.2</td>
<td>5.3</td>
<td>5.4</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>21.3</td>
<td>16.3</td>
<td>19.5</td>
</tr>
<tr>
<td>Brazil</td>
<td>14.0</td>
<td>11.9</td>
<td>9.9</td>
</tr>
<tr>
<td>Canada</td>
<td>4.7</td>
<td>2.5</td>
<td>0.8</td>
</tr>
<tr>
<td>China</td>
<td>13.2</td>
<td>13.1</td>
<td>11.6</td>
</tr>
<tr>
<td>Europe</td>
<td>4.4</td>
<td>2.3</td>
<td>1.5</td>
</tr>
<tr>
<td>India</td>
<td>22.0</td>
<td>29.8</td>
<td>27.4</td>
</tr>
<tr>
<td>Japan</td>
<td>1.8</td>
<td>1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>Korea</td>
<td>6.4</td>
<td>5.8</td>
<td>4.5</td>
</tr>
<tr>
<td>Madagascar</td>
<td>N.A.</td>
<td>4.2</td>
<td>4.1</td>
</tr>
<tr>
<td>Mexico</td>
<td>11.0</td>
<td>9.3</td>
<td>4.3</td>
</tr>
<tr>
<td>Morocco</td>
<td>16.6</td>
<td>19.1</td>
<td>20.5</td>
</tr>
<tr>
<td>Mozambique</td>
<td>14.7</td>
<td>9.3</td>
<td>9.3</td>
</tr>
<tr>
<td>Norway</td>
<td>N.A.</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>South Africa</td>
<td>24.4</td>
<td>7.3</td>
<td>5.7</td>
</tr>
<tr>
<td>Switzerland</td>
<td>1.6</td>
<td>1.4</td>
<td>2.1</td>
</tr>
<tr>
<td>Thailand</td>
<td>12.4</td>
<td>11.2</td>
<td>9.3</td>
</tr>
<tr>
<td>Tunisia</td>
<td>N.A.</td>
<td>17.3</td>
<td>12.6</td>
</tr>
<tr>
<td>Turkey</td>
<td>5.3</td>
<td>3.1</td>
<td>1.4</td>
</tr>
<tr>
<td>USA</td>
<td>3.5</td>
<td>2.1</td>
<td>1.6</td>
</tr>
<tr>
<td>Vietnam</td>
<td>17.5</td>
<td>12.8</td>
<td>14.9</td>
</tr>
</tbody>
</table>

Source: MAcMap and GTAP 5, authors’ calculations.

References


