

Trade, Production, and Protection Database, 1976–2004

Alessandro Nicita and Marcelo Olarreaga

The database described in this article provides researchers with a broad set of data on trade, production, and protection for 28 manufacturing sectors at the three-digit level of the International Standard Industrial Classification, Revision 2. The database covers up to 100 developing and developed countries over the period 1976–2004, but data availability varies by country and year. The trade, production, and protection database is available online and can be freely accessed through the World Bank trade website. JEL code: C8.

The trade, production, and protection database includes annual data on trade flows (exports and imports), domestic production (output, value added, employment), and trade protection (tariffs and nontariff barriers) for up to 100 countries over the period 1976–2004. The main contribution of this database is that it merges data from different sources in a common industry classification. The data are disaggregated into 28 manufacturing sectors, corresponding to the three-digit level of the International Standard Industrial Classification, Revision 2 (ISIC Rev. 2). The database is available for download free of charge on the World Bank trade website (www.worldbank.org/trade; click on the ‘Data & Statistics’ tab).

The database updates the earlier release made available in Nicita and Olarreaga (2001). Besides the longer time coverage, the database has been improved in a number of ways. The coverage has increased to 100 countries. The concordance table between Standard International Trade Classification, Revision 2 (SITC Rev. 2) and ISIC Rev. 2 has been updated, and more

Alessandro Nicita (corresponding author) is a consultant in the Development Economics Research Group (Trade) at the World Bank; his email address is anicita@worldbank.org. Marcelo Olarreaga is a senior economist in the Office of the Chief Economist for Latin America at the World Bank; his email address is molarreaga@worldbank.org. The authors are grateful to three anonymous referees, the editor of the World Bank Economic Review, and many users of the earlier release of the database for their comments and suggestions. They also thank the World Bank’s Research Support Budget for funding. A supplemental appendix to this article is available at <http://wber.oxfordjournals.org/>.

THE WORLD BANK ECONOMIC REVIEW, VOL. 21, NO. 1, pp. 165–171
Advance Access Publication 31 January 2007

doi:10.1093/wber/lhl012

© The Author 2007. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development / THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org

variables have been added (nontariff barriers and their ad valorem equivalents, elasticities of import demand, import price indices, import-weighted tariffs and their standard deviation). The database has also been made more compact by not including the ISIC four-digit classification. A supplemental appendix to this article (available at <http://wber.oxfordjournals.org>) describes the different dimensions of the database, variable definitions, and data availability in more detail.

Researchers may wish to use this database jointly with others available on the web (see online resources in the reference list for details). These include the trade-related data provided by Jon Haveman and Raymond Robertson; data provided by the Centre d'études prospectives et d'informations internationales (CEPII), in particular the extension by Mayer and Zignago (2005) of the earlier release of the trade and production database; and another wide collection of trade and protection data provided by the University of California at Davis, Center for International Data.

I. PRODUCTION

The source of domestic production-related data is the United Nations Industrial Development Organization (UNIDO), which collects annual data from member countries. These data are published annually in the *International Yearbook of Industrial Statistics* using the three-digit level of ISIC Rev. 2. The production-related data of this database include information on output, value added, gross fixed capital formation, index of real output, wage bill, number of establishments, number of employees, and number of female employees for each of the 28 manufacturing sectors.

Both country and time coverage of the production-related data are limited. Some countries, especially in the developing world, report data only sporadically and do not report on all the variables mentioned above. UNIDO makes a great effort to standardize the data so that they are comparable across countries and years, but some problems persist. Yamada (2005) provides detailed information on the UNIDO data and issues related to its use in research projects.

II. TRADE

The source of trade data is the commodity trade statistics database (Comtrade) kept by the United Nations Statistic Division. Comtrade provides trade data using the SITC Rev. 2. These data are converted into ISIC Rev. 2 using a concordance table. Trade data are reported at both the aggregate and the bilateral level for exports and imports. Country and time coverage are both very complete. However, there are some missing observations, especially among developing countries. To fill missing observations, researchers often resort to mirrored data (see, for example, Feenstra and others 2005). For example, the

missing export data of country C are calculated using imports from country C reported by its trading partners. The database provides mirrored data for both imports and exports. Individual researchers can determine whether and how to use mirrored data.

The trade data include information on the value of shipments (in thousands of U.S. dollars) and physical quantities (in kilograms). The database also contains unit values, measured in dollars per kilogram, calculated as the ratio of the value of shipments and physical quantities at the three-digit level of ISIC Rev. 2.¹ They are provided for exports and imports at the aggregate and bilateral level. Important cautionary notes regarding the use of mirrored data and unit values are provided in section V.

III. PROTECTION

The protection data in this database include data on tariffs and nontariff barriers. The main source of protection data is the United Nations Conference on Trade and Development (UNCTAD) Trade Analysis and Information System (TRAINS). The authors also collected additional data from national statistical documents and websites. Protection data availability starts with 1988. These data are received at the Harmonized System (HS) six-digit level and converted to the three-digit level of ISIC Rev. 2 using a concordance table. Both country and time coverage of the protection data are far from complete, especially for the data on nontariff barriers.

The tariff data contain simple and import-weighted average tariffs for the 28 manufacturing sectors. Standard deviation and maximum and minimum values at the six-digit level of the HS are also reported within each ISIC code for applied tariffs and most favored nation tariffs. Applied tariffs take into consideration the available data on preferential schemes and are therefore calculated at the aggregate level using bilateral tariff and import data.² Most favored nation tariffs are the rates granted to all World Trade Organization members to which no preferential access is granted.

Data on nontariff barriers are reported as a single category core nontariff barrier (Core NTB), which includes price-control measures, finance-control measures, and quantity-control measures. Nontariff barrier data are reported using coverage ratios (the percentage of imports subjected to nontariff barriers)

1. For about 5 percent of products at the SITC five-digit level, quantities are not reported at all or are reported in number of units rather than metric weight. The aggregate quantities at the ISIC level do not take these products into account.

2. Bilateral protection data cover most preferential trade agreements, but not all. Researchers interested in bilateral tariffs should refer directly to the UNCTAD TRAINS database, available through World Integrated Trade Solutions (WITS) (wits.worldbank.org). An alternative source of bilateral tariff data that includes a careful calculation of ad valorem equivalents of specific tariffs is Bouet and others (2004).

and frequency ratios (the percentage of tariff lines subjected to nontariff barriers). This database also provides simple and import-weighted averages of nontariff barriers at the three-digit level of the ISIC. The methodology to obtain the averages of nontariff barriers is described in Kee, Nicita, and Olarreaga (2006).

IV. OTHER DATA

Trade, production, and protection data are often used in gravity models. To facilitate work with these types of models, this database includes a number of gravity-type variables: geodesic distance between national capitals, language, GDP, GDP per capita (adjusted for purchasing power parity), and dummy variables for common language, shared border, being landlocked, and being an island. Some of these data come from the World Bank's *World Development Indicators*, and some have been collected and constructed by the authors. Researchers working with gravity models may find additional data at CEPII's website: <http://www.cepii.fr/anglaisgraph/bdd/distances.htm> and at Andrew Rose's website: <http://faculty.haas.berkeley.edu/arose/RecRes.htm#Software>.

The trade, production, and protection database also includes import demand elasticities for each country and each of the ISIC codes at one point in time. These data are particularly useful for simulation exercises. The methodology used to estimate the import demand elasticities is described in detail in Kee, Nicita, and Olarreaga (2004). Import demand elasticities are provided in a separate file.

The database also contains information on input–output tables from the Global Trade Analysis Project (GTAP) database version 4, which is based on data from the early 1990s.³ Because the GTAP 4 database aggregates some countries by region, countries in the same GTAP region will have the same input–output table. To give more flexibility to users of these tables, the data are split into two tables.⁴ One table reports the share of each manufacturing sector output that is sold as an input to the production of each sector. Another table reports the amount of all of the other sectors' output necessary to produce one unit of final output in each sector. Because the GTAP industry disaggregation does not exactly match the ISIC Rev. 2 three-digit level industry disaggregation, the input–output data are provided at a higher level of aggregation.

3. The GTAP database is now in its sixth release. The use of newer releases of the GTAP database permits building updated input–output tables. For information on how to access the GTAP databases, visit www.gtap.org.

4. The supplemental appendix provides more detail on the construction of these tables.

V. SPECIAL CONSIDERATIONS

The data in the trade, production and protection database are organized to facilitate its use for many different purposes. The objective of the database is not to produce quick answers for researchers, but rather to ease the lengthy and cumbersome exercise of collecting and organizing data into a common classification. To use the database meaningfully, it is important that researchers be aware of its limitations.

The database is an unbalanced panel, containing many missing observations. Thus comparisons must be made with care. Industry or country averages may not be very meaningful, for example, if they correspond to different time periods or contain different countries in different years. While missing data may be interpolated, the decision on whether and how to do so is left to the researcher.

A second issue relates to the use of mirrored data. In theory, export data are recorded as free on board (f.o.b.), while import data are recorded as cost, insurance, and freight (c.i.f.). That makes it appealing to use the difference between the bilateral import value and its corresponding export value to impute trade costs. In practice, however, the differences between a trade flow (whether imports or exports) and its mirrored counterpart should not be considered a good measure of trade costs, especially for countries with weak customs capacity. In many cases the discrepancies between the two values are attributable to many reasons besides trade costs, such as customs corruption, underinvoicing, weak accounting methods, existence of entrepôts, and different product classifications. And for trade aggregates discrepancies may be due to missing reporting partners in the mirrored data (not all trade partners report trade to Comtrade).

Consider the existence of entrepôts (countries that are neither the origin nor the final destination of trade but through which transits of trade occur), that may create accounting discrepancies between reported data and mirrored data (Hanson and Feenstra, 2001). In some cases the country of origin mistakenly reports the entrepôt as the destination. Meanwhile, the entrepôt country does not report the import, and the final importer reports the original exporter as the country of origin. This creates discrepancies when bilateral imports and exports are compared. The researcher should keep this in mind, especially when analyzing bilateral trade flows that may involve such entrepôts as Hong Kong, China; Macao, China; Singapore; and the Netherlands.

In some cases, however, mirrored data may be considered of better quality than reported data, such as when the partner has much better customs administration than the reporter. See Yeats (1995) for a discussion.

Another important consideration relates to a few cases (about 1 percent of observations) in which the value of exports is larger than the sum of output plus imports. This could arise for several reasons. First, there could be discrepancies between the year of production and trade flows if goods produced one

year are exported the next year. Second, production data may be misallocated across ISIC categories. Third, for some countries reported production data may exclude a significant portion of industrial activity because coverage of small-scale establishments is incomplete or because the data refer only to a certain area of the country or only to part of the manufacturing sector (excluding the informal sector, for example). Thus, researchers should be attentive to the possibility of measurement error.

Protection data also raise some issues. First, while the applied tariff data take into account the preferential access schemes of developed countries, some of the smaller agreements between developing countries may not be included. This issue is usually more relevant for data for early years. Second, there is no systematic information on preference scheme utilization rates. The data assume that the schemes are fully utilized, but some are not, often because of an inability to meet origin requirements. Third, tariff data include only the ad valorem component of tariff schedules, with no ad valorem equivalents for specific duties. This is not a major omission, however, because the database focuses on manufacturing, where specific duties are rare. Fourth, the database provides simple and import-weighted average tariffs. Neither of these has a sound theoretical basis as a measure of trade restrictiveness. For example, in calculations of import-weighted average tariffs, goods subject to prohibitively high tariffs have zero weight, underestimating trade restrictiveness. Similarly, very low tariffs on economically meaningless goods downwardly bias simple average tariffs as a measure of trade restrictiveness. For a detailed discussion and some solutions to these problems, see Kee, Nicita, and Olarreaga (2006).

There is also an important caveat on the use of unit values. Unit values are useful for analyzing many international trade issues, in particular price competitiveness. However, they should be used with caution as they are a noisy proxy for prices.⁵ This is particularly the case for large product aggregates, such as those in this database. The main reason is that changes in product quality or product mix can affect average unit values. There are no straightforward solutions to these problems but rather a need for awareness and caution when using the data. In econometric work, one way out of some of these issues is to instrument unit prices to get rid of measurement error.

VI. TECHNICAL INFORMATION

Most of the data are stored as ASCII files and can be read with any text editor or statistical software. A supplemental appendix to this article describes the different dimensions of the database, variable definitions, and data availability in more detail. It can be found at <http://wber.oxfordjournals.org> or on the World Bank Trade website (www.worldbank.org/trade).

5. In practice, sudden jumps in the time series of unit values, when not substantiated by other data, are usually an indication of a change in the accounting or recording methods.

REFERENCES

- Bouet, Antoine, Yvan Decreux, Lionel Fontagné, Sébastien Jean, and David Laborde. 2004. "A Consistent, Ad-valorem Equivalent Measure of Applied Protection across the World: The MacMap-HS6 Database." CEPII Working Paper 2004-22. Paris: Centre d'études prospectives et d'informations internationales.
- Centre d'études prospectives et d'informations international. *Additional data for gravity models*. <http://www.cepii.fr/anglaisgraph/bdd/distances.htm>.
- Centre d'études prospectives et d'informations internationales. *Extension by Mayer and Zignago (2005) of the earlier release of the trade and production database*. www.cepii.fr/anglaisgraph/bdd/TradeProd.htm.
- Feenstra, Robert, Robert E. Lipsey, Haiyan Deng, Alyson C. Ma, and Hengyong Mo. 2005. "World Trade Flows: 1962–2000". NBER Working Paper 11040. Cambridge, Mass.: National Bureau of Economic Research.
- Hanson, Gordon H., and Robert C. Feenstra. 2001. "Intermediaries in Entrepôts Trade: Hong Kong Re-exports of Chinese Goods". NBER Working Paper 8088. Cambridge, Mass.: National Bureau of Economic Research.
- Haveman, Jon, and Raymond Robertson. *Additional trade-related data*. www.maclester.edu/research/economics/PAGE/HAVEMAN/Trade.Resources/TradeData.html.
- Kee, Hiau Looi, Alessandro Nicita, and Marcelo Olarreaga. 2004. "Import Demand Elasticities and Trade Distortions." Policy Research Working Paper 3452. Washington, D.C: World Bank.
- . 2006. "Estimating Trade Restrictiveness Indices." Policy Research Working Paper 3840. Washington, D.C: World Bank.
- Mayer, Thierry, and Soledad Zignago. 2005. "Market Access in Global and Regional Trade." CEPII Working Paper 2005-02. Paris: Centre d'études prospectives et d'informations internationales.
- Rose, Andrew. *Additional data for gravity models*. <http://faculty.haas.berkeley.edu/arose/RecRes.htm#Software>.
- Nicita, Alessandro, and Marcelo Olarreaga. 2001. "Trade and Production: 1976–1999." Policy Research Working Paper 2701. Washington, D.C: World Bank.
- United Nations Conference on Trade and Development (UNCTAD). *TRAINS database. Data on bilateral tariffs*. www.worldbank.org.
- University of California at Davis, Center for International Data. *Wide collection of trade and protection data*. cid.econ.ucdavis.edu.
- World Bank. Various years. *World Development Indicators*. Washington, D.C.
- Yamada, Tetsuo. 2005. "Relevance and Applicability of the UNIDO Industrial Statistics Database for Research Purposes." UNIDO ESA/STAT/AC.105/21. Vienna.
- Yeats, Alexander. 1995. "Are Partner Country Statistics Useful for Estimating 'Missing' Trade Data?" Policy Research Working Paper 1501. Washington, D.C: World Bank.