Module 1
Introduction to empirical analysis
1 Introduction

The objective of this volume is to explain to readers how to carry out empirical analysis of the impact of trade on gender inequality. Module 1 introduces three elements that are necessary for such analysis: (a) empirical methodology; (b) data sources; and (c) statistical software to analyse the data. Modules 2–4 discuss in detail specific methodological approaches (microeconomic, macroeconomic, and sectoral) and include hands-on applications that enable readers to replicate analysis from published research papers.

Section 2 of this module introduces the methodologies and, in particular, discusses how to examine the link between trade and gender inequality. Section 3 lists the different data sources that can be used for trade, gender, and welfare analysis. It also provides a brief description of trade policy simulation models, such as the Global Trade Analysis Project (GTAP) and the World Integrated Trade Solution (WITS) simulation tools, and discusses how to use them. Section 4 contains a short description of Stata, one of the statistical software packages that can be used for data management and statistical analysis. The section describes the basic commands that serve to enter, explore, modify, manage, and analyse data, as well as the advanced commands that will be used in different estimation methods in Modules 2–4. It also provides references to open-source material that may help the reader learn Basic Stata, Advanced Stata, and Stata for Poverty Analysis. Section 5 puts forth a number of conclusions.

At the end of this module, students should be able to:

- Describe the three methodological approaches currently used in the literature to empirically assess the relationship between trade and gender, and compare the strengths and weaknesses of each of them;
- Identify which of the three methodological approaches best fits their particular research question;
- Understand the difference between *ex-ante* and *ex-post* studies;
- List relevant sources of trade data, household-level data, and macro-level data, as well as open-source trade models;
- Use the Stata statistical package and its basic commands for the manipulation and analysis of data.

2 Overview of empirical methodologies

The modules that follow discuss different quantitative methodologies that are useful to understand the relationship between international trade and gender outcomes – that is, the implications of trade for women’s economic and/or social status. However, it is important to note that in some cases qualitative methods of collecting observational data, such as focus groups, semi-structured interviews, and ethnographic studies, are equally or even more useful in capturing and understanding this relationship. The way trade affects gender outcomes often depends on the social dynamics in households, communities, and institutions where social norms, values, beliefs, personal experiences, and interests all play an important role. Uncovering and understanding this complex dynamic is not an easy task. Quantitative methods may be useful in this regard, but qualitative research methods can provide a very valuable and sometimes indispensable complement to the analysis. Qualitative methods can help shed light on the processes at work rather than just the “effects”, identify issues and questions for surveys, confirm the validity of proxy variables, establish the hypotheses that will be tested, and explain and interpret survey findings. The interested reader can find more about the use of qualitative methods in gender studies in Jarviliuoma *et al.* (2003), Metso and Le Feuvre (2006), and Warren (1998).

**Box 1**

**Definitions of trade in Volume 2**

As extensively explained in Volume 1 of this teaching material, we need to be specific about which aspect of “trade” we are referring to when we analyse the relationship between gender and trade. The aim here is to highlight the relevant definitions of trade that are used in Volume 2, referring the interested reader back to Module 1 of Volume 1 for a more detailed discussion.

In Module 2 of this volume, where we focus on the microeconomic approach, we use the concepts of trade policy and trade reform interchangeably. These terms make up a set of policy measures affecting international trade, including changes in tariff schedules, quotas, certifications, standards, and even subsidies (especially in agriculture). These changes can occur in the country’s own policies or the policies of its trading partners.
Introduction to empirical analysis

Module

2.1 Microeconomic approach

The microeconomic approach uses micro survey data to assess the distributional effects of trade policy in terms of gender by looking at the impact of those effects on individual agents such as consumers and producers. We will cover the details in Module 2.

This approach is useful for understanding the channels through which trade policy can affect the welfare of households or individuals. The methodology empirically explores two links: one that connects trade policies to prices of goods and factors of production, and a second that relates prices to household welfare. Results then can be aggregated by the relevant dimension — region, gender, income, etc. — so as to identify any subgroup that would gain or lose from trade policy measures or trade shocks. In this module, we focus on the gender dimension.

The elimination of tariffs on a certain good as a result of trade liberalization, for instance, may lead to a reduction of domestic prices of this good. The same trade policy measure may increase competition on the domestic market and thus pull down the wages (labour factor remuneration) in the production of this particular good. With respect to the link between prices and household welfare, changes in the price of the good will affect households as consumers via reduced or increased expenditure, while changes in wages will affect them as labour suppliers (workers) through reduced or increased income. As a consequence, households will adapt to new prices and wages through consumption and production decisions, affecting even the allocation of tasks within the household.

Moreover, in many countries, a large proportion of the poor do not necessarily work for wages — instead, they are self-employed or contributing family workers in household enterprises or farms. This is particularly true for women, who are often engaged in household production or informal activities. Therefore, price changes may also affect the income of households through reduced or increased sales of their products.

How does trade liberalization affect these different components of household income? And how does trade liberalization affect consumption? Do these effects differ depending on the gender of the household head? When attempting to answer these questions, consider that changes due to liberalization may affect different members of the household differently. Suppose for instance that trade liberalization has a negative impact on household income in total. This could mean that in addition to the responsibility for childcare, women would have to start to work outside the household. Trade liberalization may even affect crucial household investment decisions, such as whether or not to send children to school or to work. Further questions could be asked. Does trade liberalization reduce women’s welfare in the presence of imperfect labour markets? How can labour market reforms, education and training (skill development) policies, and trade reforms/policies contribute to reducing the gender-specific harmful effects of trade liberalization?

Definitions of trade in Volume 2

Trade policy and reform can lead either to a reduction or removal of tariffs and other trade restrictions (i.e. trade liberalization) or to an increase in tariffs and other trade barriers (i.e. trade protectionism). In Module 2, we will also refer to trade in terms of trade facilitation, which includes all the measures and practices that facilitate trade flows (e.g. reducing trade costs by improving transport infrastructure) and streamline and simplify trade procedures.

In Module 3, where we present the macroeconomic approach, the term trade takes on two different meanings: (a) trade openness in practice; and (b) trade orientation. While under definition (a) we are more interested in measuring the degree of the country’s integration into the world economy in terms of exports or imports, or both, under definition (b) we are more interested in measuring the direction of trade flows, i.e. with whom the country trades.

Finally, the term trade shock is used when a trade policy or event results in unanticipated effects on a country’s economy. An example of a trade shock is a sudden increase in export flows of a given product as a result of changes in trade policy or other factors, which may create more employment opportunities in the sector producing that product. A trade shock can also be an unexpected decline in trade volumes as a result of a country’s financial and economic collapse. This case will be examined in Module 4, where we look at the sectoral approach to study the effects of trade on gender.

Source: UNCTAD Secretariat.
trade liberalization? We could look at the impact on female labour market participation by sector and see whether new trade opportunities in the expanding export sectors employ women and whether this would have any consequences for family planning.

There are several ways in which trade liberalization can affect households. First, prices are one channel through which trade liberalization may affect household members. Both the price of consumption goods and the remuneration of factors (e.g. wages) may be affected by trade reforms. These effects may have a gender bias. Therefore, we need to ask to what extent these price changes are transmitted to and within households and whether there are barriers to price transmission. Second, competition and price changes induced by trade liberalization may sometimes be so intense that certain markets upon which the poor rely for income or consumption disappear. On the other hand, trade liberalization may create opportunities for new markets to emerge (for goods not traded before, or for new consumer goods), which in turn may benefit poor households and women. Third, trade policy and trade shocks often also have an impact on government revenues, affecting transfers and social programmes that may target more vulnerable groups, including women. Finally, trade liberalization may affect vulnerability and food security. Trade directly and indirectly affects the four components of food security as defined by the Food and Agriculture Organization of the United Nations (FAO): physical availability of food and food production, economic access to food (through the income effects of trade), stability of access to food (through control of the volatility of world food prices), and access to healthy and safe food (Diaz-Bonilla and Ron, 2010).

In this framework, the questions to answer are the following: How do households respond to price changes induced by trade shocks and can they adjust to these changes? How well can households protect themselves against the adverse effects induced by changes in markets and prices?

There are two types of microeconomic studies:

- **Ex-ante studies.** These studies analyse the welfare effects of trade reforms *ex ante*, i.e. before the reform takes place. They use simulation methodologies, such as partial equilibrium models or general equilibrium models, which can be combined with microsimulation models to conduct the analysis using household survey data. The GTAP and WITS models discussed later in this module are good examples of models employed in this type of analysis.

- **Ex-post studies.** These studies analyse the situation before and after a trade policy reform, and try to identify the reform’s effects on key outcome variables such as poverty and welfare. To carry out such an analysis, it is critical to have data about the situation before and after the trade reform episode and also to be able to unequivocally link the change in the outcomes of interest with the reform.

Despite its popularity, the microeconomic approach presented in the next module is far from perfect. As discussed in Section 3.3 below, the main problem is the dearth of reliable micro-level data, as well as a number of related issues that are highlighted in the rest of this section.

First, the microeconomic approach is limited by the fact that it is difficult to find data on capital investment and government transfers, especially for developing countries. Consequently, studies adopting the microeconomic approach often assume that the only sources of income for households are the wages of each of their members and/or the income from selling the household’s agricultural production. However, additional sources of income may derive from capital investment and government transfers. While we can assume that the amount of capital investment is negligible for households living in developing countries, government transfers also represent an indirect channel of influence of trade on household welfare. This implies that studies based on the microeconomic approach might be excluding an important and interesting piece of information from the analysis.

Second, studies using the microeconomic approach are often unable to capture substitution effects. According to Friedman et al. (2002), households may substitute goods that become more expensive with goods that are cheaper after trade liberalization. To capture this effect, a model should include cross-price elasticities, especially for goods that may substitute one another (e.g. wheat and rice), but data are not always available for their exact calculation.

It is also worth noting that the first applications of the microeconomic approach (see e.g. Porto, 2006) assume that changes in domestic prices of imported goods perfectly reflect changes in import tariff rates; in other words, the elasticity of the prices of imported goods with respect to import tariff rates is assumed to equal one. This,
however, is not always the case and a few authors have tried to address this problem. For instance, Nicita (2009) and Borraz et al. (2012) find that trade costs (proxied by distance from the border), domestic production prices, and exchange rates significantly affect domestic prices of imported goods. Trade costs are particularly relevant because the extent to which tariffs influence the prices of imported goods may depend on the region where the household is located. For example, Nicita (2009) finds that Mexican households living next to the border are more affected by changes in import tariff rates than households living in remote areas of the country. In this sense, import tariffs imperfectly pass through on prices; assuming a one-to-one relationship between tariffs and prices without appropriate controls (if the data are available) may yield misleading results about the effects of international trade on household welfare.

2.2 Macroeconomic approach

The macroeconomic approach focuses on the interconnections between trade policy and its outcomes at the economy-wide or macroeconomic level using aggregate data. Trade policy may trigger structural transformations in the economy as well as shifts in the level and growth of employment and income, which are the subjects analysed by macroeconomic studies of gender and trade. This approach is covered in detail in Module 3.

In macroeconomic studies, the basic empirical specification used by authors to estimate the relationship between trade and gender is:

\[ y_{ct} = \alpha_c + \alpha_{Tradect} + \alpha_{Xct} + \eta_t + \gamma + \nu_{ct} \]  

(1)

where \( y_{ct} \) stands for the gender outcome variable of interest (for instance, the gender wage gap or women’s empowerment) in period \( t \) of country \( c \), \( Tradect \) stands for a measure of openness (e.g. trade share), \( X_{ct} \) is a vector of other control variables (e.g. gross domestic product – GDP – per capita), and \( \eta_t + \gamma \) are country- and time-fixed effects. Alternatively, the estimation can be done in two stages. First, we can assess the relation between trade and the aggregate macro variables of interest (e.g. growth, employment creation, etc.) and then use a similar equation to study how those changes in the aggregate variables are transmitted to or distributed among the population, with a particular emphasis on gender outcomes. As far as growth is concerned, the literature on the existence of pro-growth effects of trade is inconclusive (Rodriguez and Rodrik, 2001). Also, the evidence that the poor, particularly women, benefit from growth is rather scant. Despite the advantages of the macroeconomic approach compared to the microeconomic approach – most notably the higher reliability of macro-level data as well as its more frequent collection – there may also be some issues with regard to the macroeconomic analysis.

Several authors have pointed out the many difficulties of using cross-section and panel data approaches involving countries that are very different in several dimensions and over long periods of time. Critics have focused on the controversial “empirical growth literature” for its lack of robustness (Florax et al., 2002). However, most of the problems in this strand of research are also relevant for other studies using panel data with data at a very aggregate level. Levine and Renelt (1992) examine growth regressions and find that the conclusions from existing studies are not robust to small changes in the set of explanatory variables; in other words, the results are sensitive to any change in the number and type of explanatory variables included in the original model. Mankiw et al. (1995) emphasize three problems with macroeconomic regression analysis: the simultaneity problem (entangled cause and effect), the multicollinearity problem (most of the potential determinants of growth are correlated with each other and are imperfectly measured, making it hard to figure out which is the true determinant), and the degrees-of-freedom problem (there are more plausible hypotheses than data points). Additionally, Harrison and Hanson (1999) point out that studies linking trade reform and growth are fragile because of problems associated with identifying the links between policies and economic performance, namely (a) endogeneity problems associated with the relation between trade policies and growth, and (b) problems with correctly interpreting the proxies for trade orientation, and difficulties with measuring trade openness. Winters et al. (2004) list the following problems with empirical macro-level research on openness and growth: (a) difficulties in measuring trade openness accurately; (b) the problem of causality (trade may stimulate growth, but countries may also only open up to trade, or may trade more, once
their growth rates are higher); (c) general problems with cross-country regressions (trade is assumed to affect growth similarly in poor and rich countries); and (d) the need for supportive policies and institutions that are required for trade to have long-term, permanent effects on growth.

It is not easy to empirically disentangle the role of trade volumes and supporting economic policies from each other in cross-country analyses of economic growth. The same caveats of the trade and growth literature also apply to cross-country regressions attempting to link trade and growth to gender outcomes. In particular, as shown in Volume 1 of this teaching material, the relationship between trade and gender outcomes involves direct and indirect transmission mechanisms going both ways – i.e. trade affects gender but at the same time gender bias affects trade – that may be captured by employing sophisticated econometric techniques.  

2.3 Sectoral approach

This last methodology assesses the relationship between trade and gender by analysing changes within specific sectors or industries of the economy. The value chain analysis examines the entire process chain of procuring raw material, production, and distribution of goods within a particular sector or industry (see Module 4). Sectoral studies may either use macroeconomic or micro-economic data. Furthermore, while the empirical strategies in the previous two approaches were more or less defined, sectoral studies often do not share a common strategy. For instance, Nicita and Razzaz (2003) study the effect of a boom in the textile sector on wage differentials in Madagascar first using a propensity score model to identify the individuals who would likely switch to the textile industry, and then estimating the wage premium this industry commands. When addressing the gender impact of agricultural exports in Ecuador, Newman (2002) uses a different approach, namely quasi-experimental data where the “treatment” group is in a geographical area where the cut flower industry, which has a high demand for female labour, is located. The “control” group is in a culturally similar but economically more traditional valley that does not produce flowers for the export market. This approach addresses the problem of endogeneity that arises when measuring the effects of contemporaneous household labour supply decisions. Furthermore, Depetris et al. (2011) look at the cash crop sector in Africa using a methodology similar to the microeconomic approach, but instead of analysing the effects of trade policy on prices, they examine the internal marketing arrangements and assess how changes in the level of competition in one of the layers of the value chain affect farmers, with a particular emphasis on female-headed and poor households.

Although the sectoral approach allows us to focus on the extent to which trade-related changes in the structural composition of an economy translate into more economic opportunities for women, it has some shortcomings that are worth discussing here.

The first issue, as is often the case with empirical research on developing countries, concerns the availability of data. Sector-level data are usually derived from the aggregation of lower-level data that are not always available, especially for developing countries. In some cases, we are able to retrieve information on, for example, output, value added, wages, and number of employees available at a very aggregated level (such as the one-digit level of the International Standard Industrial Classification – ISIC) for agriculture, manufacturing, and services. However, gender-disaggregated information is only available for some countries, which makes it difficult to analyse the distribution of employment in terms of gender by sector and to understand the gender repercussions of trade at the sectoral level.

It should be noted that there have been recent improvements in the collection of gender-disaggregated data. However, assuming that this information is available and that we are able to carry out our research on trade and gender using the sectoral approach, there is a second issue that emerges. By looking at only one sector of the economy, this methodology cannot grasp the effects of trade on gender in the economy as a whole. Oversimplifying, assume that, for example, trade results in a shift of a country’s patterns of specialization from relatively higher female-labour-intensive products (e.g. textiles) to relatively higher male-labour-intensive products (e.g. machinery). On the one hand, this would adversely affect women employed in the textile sector who may lose their jobs. On the other hand, it would favour men engaged in the machinery sector who may experience an increase in their wages. If we only looked at the textile sector, our conclusion would be that trade has hurt women workers. But this might not be the end of the story. In the best-case scenario, women in the textile industry have received training, which has contributed to their skill development and their ability to relocate to better-paid positions in other sectors of the economy. The sectoral approach might fail to capture this effect and therefore yield misleading results.
It is important to note that there is no best or worst methodology as such. There is, however, the most or least appropriate approach to employ for the purpose of your particular study. Literature published on the topic in academic journals can be a useful guide in terms of ascertaining the empirical and theoretical approaches that would be appropriate for your study.

3 Data sources

3.1 Trade data

Access to detailed trade data is useful for the type of analysis we will pursue in this teaching material. The main source of information and data on a country’s trade policies, regulations, and flows is the country itself. Most countries have an official statistical agency, although sometimes trade statistics and information about trade policy are recorded by the Ministry of Foreign Affairs or the Ministry of Trade.

Besides the national statistical agency or data collection agency, there are a number of international organizations, research centres, and think tanks that systematically gather trade data and information. These organizations are a good source of data and information for both the quantitative and qualitative analysis, in particular when such analysis involves multiple countries. The main sources of this type of trade data and information are described below.

3.1.1 World Trade Organization

The World Trade Organization (WTO) is an international organization with 160 member countries or economies (as of June 2014) that have signed WTO agreements containing global rules for trade between countries. The following sources of data and information can be accessed on the WTO website (http://www.wto.org):

- **International Trade Statistics** (http://www.wto.org/english/res_e/statis_e/its_e.htm) is an annual publication that analyses trends in trade, including at the product level. In addition, the WTO statistics database (http://stat.wto.org/Home/WSDBHome.aspx) allows for interactive data retrieval of international trade statistics.

- **The Integrated Trade Intelligence Portal (I-TIP)** (http://www.wto.org/english/res_e/statis_e/itip_e.htm) provides a single entry point for information compiled by the WTO on trade policy measures. It covers both tariff and non-tariff measures affecting trade in goods, as well as information on trade in services, regional trade agreements, and the accession commitments of WTO members.

- **The World Trade Report** (http://www.wto.org/english/res_e/reser_e/wtr_e.htm) is an annual publication that aims to enhance understanding about trends in trade, trade policy issues, and the multilateral trading system.

- **The Trade Policy Reviews** (http://www.wto.org/english/tratop_e/tp/e.htm) contain, among others, policy statements from governments about changes in their countries’ trade regulations, policies, or practices, as well as reviews written by the WTO Secretariat on specific countries.

- **Regional trade agreements (RTA) and preferential trade agreement (PTA) databases** (http://www.wto.org/english/tratop_e/region_e/rtas_e.htm) contain WTO records on notified RTAs and PTAs.

- **The trade monitoring database** (http://tmdb.wto.org) provides detailed information on trade measures implemented by WTO members and observers since October 2008.

3.1.2 United Nations

- **The United Nations Commodity Trade Statistics Database** (http://comtrade.un.org/db) is the most comprehensive trade database available (over 1 billion records), and it is continuously updated. The database has import and export data reported by almost 200 countries since 1962, and standardized by the United Nations. Data can be searched by products, which are classified according to the Harmonized Commodity Description and Coding System (HS) product classification at the 6-digit level.

- The United Nations also publishes the **International Trade Statistics Yearbook** (http://comtrade.un.org/pb/first.aspx), which provides information on the international trade performance of some 180 countries or regions and, in particular, on world trade flows of selected commodities (at the 3-digit level of the Standard International Trade Classification – SITC – revision 3). The publication is composed of two volumes. Volume 1 contains information on: (a) trade flows of individual countries in terms of values and, if available; (b) quantities of the key commodities traded by individual countries (for the latest four
years); (c) the countries’ trade with their main trading partners and regions (for the latest five years); (d) imports by Broad Economic Categories (BEC); and (e) the percentage share of countries’ trade with each region of the Millennium Development Goals’ (MDG) regional groupings. Volume 2 contains tables showing total trade for selected commodities (at the 3-digit level of the SITC, revision 3) for each MDG regional grouping and main trading countries. It also provides analytical data on total trade, exchange rate conversion factors, trade indices, and import and export flows.

- The International Trade Centre’s (ITC)\textsuperscript{19} Trade Map (http://www.trademap.org) is a good source for indicators on export performance, foreign demand, and alternative and competitive markets. All the information is organized in tables, graphs, and maps covering 220 countries and territories and 5,300 products of the HS product classification. The data about monthly, quarterly, and yearly import and export flows are available from the most aggregated level to the tariff line level.

- The ITC’s Market Access Map (http://www.macmap.org) is a market analysis tool that provides information on market entry requirements of a particular country. It contains data on tariffs, non-tariff measures, and trade flows, as well as information on trade agreements and rules of origin by country of interest.

3.1.3 International Monetary Fund

- The International Monetary Fund’s (IMF) Direction of Trade Statistics (DOTS) (http://elibrary-data.imf.org/findDataReports.aspx?id=33061&ei=170921) records countries’ exports and imports and their area of distribution by trading partner. The DOTS yearbook covers seven years of data for about 187 countries. The DOTS quarterly issue provides data for the most recent six quarters and the latest year for about 156 countries, as well as data for the most recent 10 quarters and the latest five years for the world and selected regions of the world.

3.1.4 Other sources

- The Global Trade Alert (http://www.globaltradealert.org) provides real-time information on government measures that are likely to affect foreign trade. It goes beyond other monitoring initiatives by identifying the trading partners that are likely to be harmed by these measures.

- GTAP\textsuperscript{20} (https://www.gtap.agecon.purdue.edu) produces a global database that describes patterns of bilateral trade, production, consumption, and intermediate use of commodities and services. Access to this database is not free in most cases, but the dedicated web page also contains links to open-source data.

3.2 Trade models

A trade simulation model is needed to simulate the welfare effects of trade policies before they are implemented. Below is a list of open-source software that will allow you to conduct your simulation according to the purpose of your analysis.

- The standard GTAP model (https://www.gtap.agecon.purdue.edu/models/current.asp) is a computable general equilibrium (CGE) model covering different regions and sectors. The model is based on perfect competition with constant returns to scale. The GTAP model enables users to choose between different closure options, including unemployment, tax revenue replacement, and fixed trade balance closures, as well as a selection of partial equilibrium closures (which facilitate comparison with results obtained using partial equilibrium assumptions). The website also contains a list of books explaining the basic functioning of the model. The model’s code is downloadable.

- WITS (http://wits.worldbank.org/wits) is a software developed by the World Bank, in close collaboration and consultation with various international organizations, including UNCTAD, the ITC, the United Nations Statistical Division, and the WTO.\textsuperscript{21} The WITS Global Tariff Cuts and Trade Simulator enables you to:
  - Simulate tariff cuts by cutting the applied tariff rates according to already available formulas. In particular, users can implement a new or new maximum tariff rate, apply a linear percentage cut, or apply the Swiss formula.\textsuperscript{22} Tariff cuts can also be simulated with different formulas (or the same formula with different parameters) applied to different products and countries: in this case, both pre- and post-tariff cut rates are shown for each HS 6-digit level product and each combination of importer-exporter.

- Carry out a global simulation using the Global Simulation Model (GSIM), a partial equilibrium model developed by Francois and Hall.
(2002). By employing national product differentiation, the model aims to analyse global trade policy changes at the industry (product) level on a global, regional, and national scale. Results from the global simulation reveal the distributional effects of tariff revenues, exporter (producer) surplus, and importer (consumer) surplus. More information on the GSIM methodology is available at: http://wits.worldbank.org/data/public/GSIMMethodology.pdf.

- Use the System of Market Analysis and Restrictions on Trade (SMART) with the users’ own data. This module enables users to export the available template and run a simulation with their own data. The package allows you to download the missing data from WITS. More information on the SMART simulation methodology is available at: http://wits.worldbank.org/data/public/SMARTMethodology.pdf.

- The ITC’s Market Access Map (see above) also provides a tool for simulating tariff reductions. This can be used to prepare for trade negotiations and study the welfare effects of trade policy changes.

- Simple Excel models (https://sites.google.com/site/jgilberteconomics/Home/excel) bring together a number of general and partial equilibrium numerical simulation models on various aspects of international trade theory and policy, all built in Excel and using both tabular and graphical presentations.

3.3 Micro-level data

Some of the best microeconomic data for analysis of the effects of trade on gender outcomes are those coming from household surveys and labour force surveys. Module 2 discusses how to use microeconomic data to analyse the effects of trade policy and other shocks on gender outcomes (the microeconomic approach). There are nearly 2,500 types of household survey questionnaires in the world. They vary in design and selection of variables, and therefore some familiarization with their structure is required prior to potential use. The International Household Survey Network (IHSN) houses a catalogue of more than 4,000 household surveys that include economic and social variables from most countries in the world. A subset of 266 of these surveys provides data on income and expenditure. The IHSN does not have ownership rights to the data and is not mandated to disseminate country microdata. However, the network maintains a central survey catalogue and provides links to national or international databases from which the survey data can often be retrieved.

The main advantage of household survey data is that they allow the researcher to examine the effects of policies at the household level, while controlling for various household characteristics. For instance, we can assess if a policy measure has a gender bias, if it is pro-poor, or if it benefits urban more than rural areas. The level of detail in household data can help policymakers devise better ways to implement policies. However, the analysis of trade policies using household surveys is complicated for several reasons:

- Household surveys can be costly, and countries may reduce survey frequency, sample size and content in order to cut costs. The quality and frequency of household surveys vary significantly between and within countries. A country may change the thematic emphasis from one survey to another, depending on its policy design needs. Analysts often have to adjust their estimation methods, depending on questionnaire and sample variability between different waves of surveys.

- Another issue with household surveys is the reliability of data. For instance, individuals may not declare their true level of income or if they are involved in informal work.

- Household survey methods have evolved over the years, and redundant questions are often included in surveys in an effort to improve cross-checking and accuracy.

- Despite the decreasing costs of computing, data management and processing have tended to grow more complex.

- There is wide variation in the definition of key variables.

Notwithstanding these difficulties, household surveys are useful for the analysis of the effects of trade policies. The volume and quality of information at the household level tend to be good enough to provide accurate estimations of poverty impact, which may help in developing strategies to reduce poverty in line with the MDGs. Household expenditure, level of education, gender, household location, and other variables can have a significant influence on the effect of a given trade shock.

Three surveys have made particular efforts to address some fundamental issues related to
data collection such as solving methodological and statistical problems, documenting the preparation, implementation, and analysis of surveys, and publicizing and publishing the survey results. These surveys are the World Bank’s Living Standards Measurement Study (LSMS); the United States Agency for International Development’s (USAID) Demographic and Health Survey (DHS), and the United Nations Children’s Fund’s (UNICEF) Multiple Indicator Cluster Survey (MICS). These sponsoring institutions have collaborated with many countries to implement household surveys. Such collaboration has contributed to strengthening national capacity, which was another important goal of the sponsoring institutions.

### Table 1

<table>
<thead>
<tr>
<th>Module Information</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household composition</td>
<td>Household roster, demographic data, information on parents of all household members</td>
</tr>
<tr>
<td>Food expenditures</td>
<td>Food expenditures in the past two weeks and past year; consumption of home production in the past year</td>
</tr>
<tr>
<td>Non-food expenditures</td>
<td>Non-food expenditures in the past two weeks and past year; remittances to other households in the past week and past year</td>
</tr>
<tr>
<td>Housing</td>
<td>Type of dwelling, housing and utilities expenditures over the week and year of the interview</td>
</tr>
<tr>
<td>Durable goods</td>
<td>Inventory of durable goods and their characteristics</td>
</tr>
<tr>
<td>Economic and production activities and assets</td>
<td>Non-farm employment, agro-pastoral production, land, livestock, and equipment owned in the past week and past year</td>
</tr>
<tr>
<td>Savings</td>
<td>Savings and debts</td>
</tr>
<tr>
<td>Education</td>
<td>Completed schooling and schooling expenditure for all household members; attendance and non-attendance information</td>
</tr>
<tr>
<td>Health</td>
<td>Health expenditure of all household members and use of health services in the past four weeks</td>
</tr>
<tr>
<td>Migration</td>
<td>Place of birth, length of stay at current residency</td>
</tr>
<tr>
<td>Fertility</td>
<td>Subsample with data on birth history, use of maternity services, and duration of breastfeeding</td>
</tr>
<tr>
<td>Anthropometrics</td>
<td>Height and weight measurements of all household members</td>
</tr>
</tbody>
</table>

Source: IHSN.

### Table 2

<table>
<thead>
<tr>
<th>Topic Information</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household composition</td>
<td>Name, age, sex, marital status</td>
</tr>
<tr>
<td>Education</td>
<td>School attendance and attainment, literacy from birth to age 24, literacy test for people older than age 7</td>
</tr>
<tr>
<td>Characteristic of the dwelling</td>
<td>Water, sanitation, second-hand smoke, construction materials, electricity, mosquito netting, inventory of possessions (durable goods, livestock)</td>
</tr>
<tr>
<td>Anthropomorphic measurements</td>
<td>Measurements for each household member, including hemoglobin and HIV tests</td>
</tr>
<tr>
<td>Reproductive health</td>
<td>Contraception, pregnancies, and birth outcomes, pre- and post-natal care</td>
</tr>
<tr>
<td>Child immunization, health, and nutrition</td>
<td>Vaccination records for all children, types of food given to infants</td>
</tr>
<tr>
<td>Marriage and sexual activity</td>
<td>Data on sexual partners, fertility preference</td>
</tr>
<tr>
<td>Work and work decisions</td>
<td>Employment and work decisions by men and women</td>
</tr>
<tr>
<td>Human immunodeficiency virus (HIV)</td>
<td>Knowledge, behaviour</td>
</tr>
</tbody>
</table>

Source: IHSN.
Introduction to empirical analysis

Module

Source: UNICEF.

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household questionnaire</td>
<td>Household information panel, list of household members, education, child labour, child discipline, household characteristics, insecticide treated nets, indoor residual spraying, water and sanitation, hand washing, salt iodisation</td>
</tr>
<tr>
<td>Individual questionnaire for women</td>
<td>Woman’s information panel, woman’s background, access to mass media and use of ICT, fertility (mortality) with optional birth history, desire for last birth, maternal and newborn health, post-natal health checks, illness symptoms; contraception, unmet need, female genital mutilation/cutting, attitudes towards domestic violence, marriage/union, sexual behavior; HIV/acquired immunodeficiency syndrome (AIDS), maternal mortality, tobacco and alcohol use, life satisfaction</td>
</tr>
<tr>
<td>Questionnaire for children under five</td>
<td>Under-five child information panel, age; birth registration; early childhood development; breastfeeding and dietary intake; immunization; care of illness; anthropometry</td>
</tr>
<tr>
<td>Individual questionnaire for men</td>
<td>Man’s information panel, man’s background, access to media and use of ICT, fertility; attitudes towards domestic violence; marriage/union; sexual behaviour; HIV/AIDS; circumcision; tobacco and alcohol use, life satisfaction</td>
</tr>
</tbody>
</table>

Aside from the central survey catalogue and the links to national and international datasets maintained by the IHSN, there are other sources of micro datasets for development economists and others:

- OpenMicroData ([http://openmicrodata.wordpress.com](http://openmicrodata.wordpress.com));
- Micro Data for Development Economics ([https://sites.google.com/site/medevecon/development-economics/devecondata/micro](https://sites.google.com/site/medevecon/development-economics/devecondata/micro)).

### 3.4 Macro-level data

This section presents a non-exhaustive list of sources of macroeconomic data that will be useful when we present the macroeconomic approach in Module 3. Such data will allow you to analyse the interaction between trade and gender at a higher level of aggregation (units of analysis are usually countries, regions, or groups of regions), where it is easier to collect the data. Most of the shortcomings of micro-level data – notably the reliability of the data itself – do not apply in this case. However, one needs to take into account that micro and macro variables serve different purposes and complement rather than substitute for each other. Most international organizations keep a record of macroeconomic variables (GDP, female labour force participation, etc.) and make this information available on their websites. Most national statistical offices also collect these data, and their websites can be a useful source of this type of information. Sources of macro-level data include:

- World Development Indicators (WDI) ([http://data.worldbank.org/data-catalog/world-development-indicators](http://data.worldbank.org/data-catalog/world-development-indicators)), the World Bank’s primary database of development indicators collected by officially recognized international sources. By providing over 800 indicators covering more than 150 countries, WDI is the most updated and accurate open-source global development database available. Users can access a selection of WDI data online and browse the information by country, indicators, and topic.
- Penn World Table (PWT) ([http://www.rug.nl/research/ggdc/data/penn-world-table](http://www.rug.nl/research/ggdc/data/penn-world-table)), developed and maintained by researchers at the University of California, Davis, and the Groningen Growth Development Centre of the University of Groningen in the Netherlands. The PWT contains information on national accounts, capital, productivity, employment, and population, and is currently covering 167 countries over the period 1950–2011. The information is constantly being updated. Compared to other databases, such as the World Bank’s WDI,
the PWT allows for more time coverage as well as more data to compare productivity across countries and over time. However, it does not contain gender-specific variables.

- UNCTAD’s Foreign Direct Investment (FDI) Statistics database (http://unctad.org/en/Pages/DAIE/FDI%20Statistics/Interactive-database.aspx), which contains annual FDI data (flows and stocks) for 189 countries starting in 1970, as well as data on mergers and acquisitions (sales and purchases) starting in 1987. The database is the main source of data for UNCTAD’s World Investment Report.

- UN Data (http://data.un.org), an Internet search engine that allows for retrieving and downloading data series provided by the United Nations system on a number of different topics, such as education, employment, energy, environment, food and agriculture, health, and human development.

- The Organisation for Economic Co-operation and Development (OECD) (http://www.oecd.org/statistics), which publishes comparable statistics for OECD member countries on a wide range of topics. The statistics are available in several forms: (a) as interactive databases on iLibrary; (b) as static files or dynamic database views on the OECD Statistics portal; and (c) as StatLinks (in most OECD books, there is a URL that links to the underlying data). OECD statistics represent a good source of data on such areas as gender differences in employment outcomes (i.e. labour force participation, hours spent on paid and unpaid work, and employment conditions), gender equality, gender wage gap, and mean scores and gender differences in the Programme for International Student Assessment (PISA) rankings.

- The ILO’s Labour Statistics (http://laborsta.ilo.org). The ILO collects and provides access to a series of labour market statistics that are often useful to study gender outcomes. Among these datasets are:
  - ILOSTAT: Provides annual, monthly, quarterly, and semi-annual labour market statistics for over 100 indicators and 230 countries, areas, and territories;
  - Key Indicators of the Labour Market (KILM): Published every two years since 1999, the KILM is a collection of 20 key indicators of the labour market, ranging from employment and variables relating to employment (status, sector, hours, etc.) to education, wages and compensation costs, labour productivity, and working poverty.

- Labour force surveys: These surveys compile websites that contain data from national statistical agencies, the ILO, and other sources. The surveys include links to source websites, and references to print publications available in the ILO library;

- NATLEX: Provides abstracts of legislation and relevant citation information, indexed by keywords and by subject classifications;

- NORMLEX: Comprehensive database providing open access to the latest information on ILO international labour standards as well as national labour and social security laws.

- Other sources such as the Harvard Dataverse Network (http://dvn.iq.harvard.edu/dvn), the Cingranelli-Richards Human Rights Dataset (http://www.humanrightsdata.com/p/data-documentation.html), the Socio-Economic Database for Latin America and the Caribbean (http://sedlac.econo.unlp.edu.ar/eng), and the Harvard University Centre for International Development Data (http://www.cid.harvard.edu/ciddata/ciddata.html).

4 The Stata statistical package

4.1 Online references to learn Stata

We will use the Stata statistical package for estimation and data analysis. Stata is a modern command-line driven package for statistical analyses, data management, and graphics. It provides commands for the analysis of panel, cross-sectional, time-series, survival-time, and cohort study data, as well as other data. Stata is user-friendly and provides a large set of references with online access. The software package also has networking capabilities that facilitate installing new commands or updating installed packages.

Using this teaching material does not require extensive previous knowledge of Stata because we will discuss many commands in detail. However, the interested reader can gain further knowledge by referring to the following learning material:
Stata for beginners

- A good place to start learning Stata is the Stata Starter Kit (http://www.ats.ucla.edu/stat/stata/sk/default.htm), which contains class notes, learning modules, and other useful resources;

- The LSE Research Laboratory (http://rlab.lse.ac.uk) has a good Stata manual that includes an introduction to Stata (http://rlab.lse.ac.uk/it/it_docs/introduction_to_stata.pdf) and training and practical notes, which are class notes from the London School of Economics;

- The Stata tutorial from Princeton University is also a good introductory resource for Stata (http://data.princeton.edu/stata).

Stata for advanced users

- The UCLA Stata website (http://www.ats.ucla.edu/stat/stata) is the most complete Stata resource on the Internet, featuring textbook examples, class notes, tutorials, and other useful links;

- The Stata Journal (http://www.stata-journal.com/archives) posts the latest news on Stata commands, data analysis, statistics and econometrics techniques, and programming tips;

- The Stata Users Group (http://ideas.repec.org/g/stataus.html) provides useful resources such as working papers, journal articles, books, and Stata packages;

- Ben Jann’s website (http://www.soz.unibe.ch/content/ueber_uns/jann/stata_packages) is an excellent source for Stata packages used in data analysis, graphics, statistics, programming, and matrix manipulation. However, the content of his website goes beyond what we will need for the modules in this teaching material.

It is important to keep in mind that this material will cover only a small portion of the possible methodologies used in gender analysis. Our focus will remain on the effects of trade on gender outcomes. However, there are a number of methodologies and approaches related to poverty, inequality, and gender measurement and diagnosis, most of which are implemented in Stata. Below are relevant links to these methodologies.

- The World Bank’s Introduction to Poverty Analysis manual (http://siteresources.worldbank.org/PGLP/Resources/PovertyManual.pdf) provides an overview of the basic methods related to poverty measurement and diagnosis, and also shows you how to apply these methods using household survey data in Stata.

- The International Food and Policy Research Institute’s manual titled Using Stata for Survey Data Analysis (http://www.ifpri.org/publication/using-stata-survey-data-analysis) is another good source that is freely available.


4.2 How to use Stata

4.2.1 The Stata interface

When you start Stata you will see five docked windows, initially arranged as shown in Figure 1. The windows are as follows:

- Results window – all outputs, except graphs, appear in this window;
- Command window – where you type your commands to execute them;
- Variables window – all variables in the currently-open dataset will appear here;
- Review window – previously used commands are listed here and can be transferred to the command window by clicking on them;
- Properties window – introduced in version 12 of Stata, this window displays properties of your variables and dataset.
At the top of the screen, you will find the menu bar and the toolbar. The most important functions on the toolbar are:

- **Open** (equivalent to command `use`) – opens a new data file
- **Save** – saves the current data file
- **Print Results** – prints the content of the results window
- **New Viewer** – opens a new viewer window
- **New Do-file Editor** – opens a new do-file editor
- **Data Editor** – opens the data editor window
- **Data Browser** – opens the data browser
- **Break** – allows for cancelling currently running calculations

Commands can be called from the menu bar, but using them will slow you down, so we always recommend working with the command window or writing do-files.

Stata syntax mostly follows the basic structure in which square brackets denote optional qualifiers (see `help language`):

```
[by varlist:] command [varlist] [=exp] [if] [in]
[using filename], options
```

Example: `bysort gender: tabulate age if weight < 50, nolabel`

### 4.2.2 Do-files

A do-file is a set of Stata commands typed in a plain text file that allows you to run your commands repeatedly and not lose them once Stata is closed. When the do-file is run using the do-file editor, all commands are executed automatically in the same order as in the do-file. If all steps of a project have been documented in one or more do-files, all analyses and results can be reproduced by other users at a later stage. To access Stata’s do-file editor, use Ctrl-9 in versions 12 and 13 (Ctrl-8 in earlier versions) or select “Window > Do-file Editor > New Do-file Editor” in the menu bar.

### 4.2.3 General commands

Most Stata commands can be abbreviated. For example, instead of typing `generate`, Stata will also accept `gen`. The help screen demonstrates how each command can be abbreviated by showing underlined letters in the syntax section.

- `cd`: Stata uses a working directory where datasets are saved if no path has been entered. The current working directory is displayed on the status bar at the bottom of the user interface. You can also display the current directory in the results window by using the command `pwd`. The working directory can be changed by using the command `cd` (change directory). If a directory name contains spaces, the whole path has to be entered with quotation marks, e.g., `cd "C:/Documents and Settings/Admin/My Documents/gender data"`.

- `help`: The help screen for any command can be displayed in a separate window with the command `help`. The syntax is `help commandname`. 

Source: Stata 12.1.
findit: The command findit is the best way to search for information on a topic across all sources, including online help, FAQs on the Stata website, the Stata Journal, and all other Stata-related Internet sources. The syntax is: findit word.

You can look up the meaning of error messages by either clicking on the return code or by using findit rc#, where # stands for the number of the return code.

set memory: Stata reads the whole dataset into the working memory; thus, sufficient memory has to be available or an error message will be displayed. Therefore, you should set the size of the working memory reserved for Stata before loading a (big) dataset with the command set memory. Example: set memory 100m.

4.2.4 Data input and saving

insheet: If the data come from an external source (Excel, Access, SPSS, etc.), they first have to be read into Stata. The data should be exported as tab-separated, comma-separated, or semicolon-separated text (ASCII) files. This option can generally be found in the file menu under “Save as” or “Export”. Other methods for reading non-Stata data are described in help infiling. In Stata, this text file is then read with the command insheet: insheet using filename [, options]. It can be specified in the options if the external data file is tab-separated or otherwise (see help insheet). A check on the raw data then needs to be performed to determine whether these data are complete or correctly imported, and if further data management tasks are required. Common data management tasks include renaming variables, changing string variables to numerical or date format, replacing a comma as decimal separator with a period, and labelling. Vice versa, data can be exported from Stata to a tab-separated text file with: outsheet using filename.

use: Datasets with the Stata specific ending .dta can be opened with the command use. The syntax is: use filename.dta or use "c:/.../gender_paper.dta" for a file from a parent directory.

edit: Data can also be manually entered or changed using the data editor with edit.

save: The data are saved with the command save. The syntax is: save filename.dta [, options]. Take into account that old versions of Stata may not be able to open data saved in new versions of Stata. If you want to keep compatibility, you should save your data with the command saveold.

In do-files, you would use the option replace for the command save most of the time, as datasets are overwritten every time the do-file is run. This means that you should pay attention when using this command, as you might lose the file with the original dataset if you do not save the edits made under a different name.

4.2.5 Data management

by: The by qualifier tells Stata to execute the subsequent command repeatedly along with the different values of a given varlist. Note that not all commands support this feature. To use the command by, data have to be first sorted by varlist. Using bysort instead of by makes previous sorting redundant. For example, we could summarize the variable educational level (edlevel) by gender (sex), as shown in Figure 2.

Example of data management using the by-qualifier

. bysort sex: summarize edlevel
-> sex = Male
Variable |       Obs.      Mean    Std. Dev.     Min       Max
-------------+-----------------------------------------------------------
edlevel |       962 5.087318    1.296223      1         9
-------------+-----------------------------------------------------------
edlevel |      1161 4.828596    1.593329       1         8

Source: UNCTAD’s estimations, based on data from Newman (2002).
4.2.6 Data manipulation

generate: New variables are generated with the command generate (can be abbreviated as gen). The syntax is: `generate newvar = exp [if] [in]`, where `exp` can be either an algebraic or a string expression. An empty algebraic variable can be created with `generate varname = .`, while an empty string variable can be created with `generate varname = "".

For an overview of functions that can be used in expressions, type help functions. Mathematical operators are: + addition, − subtraction, * multiplication, / division, and ^ power. Important mathematical functions are: abs(x) absolute value, sqrt(x) square root, ln(x) natural logarithm, and round(x) round to nearest whole number. Example: `gen age_sq = age^2`.

gen: Extensions to generate can be found in the command gen which offers a set of algebraic or string functions that are sometimes needed for data management tasks (see help gen for an overview of available functions). The syntax for the gen function is: `gen newvar = fcn(arguments) [if] [in] [options]`. For example, `gen averageincome = mean(income)` creates a variable with the average value of the variable income.

replace: The values of existing variables can be changed with the command replace. It works similarly to the command generate, accepting expressions and allowing for in and if qualifiers. The syntax for replace is: `replace oldvar =exp [if] [in]`. Example: `replace income = income/100`.

drop: Variables or observations can be deleted using the command drop. Variables are deleted using the following version of drop: `drop varlist`.

Observations are deleted by applying another version of drop, and the syntax is: `drop if exp or drop in range [if exp]`. Example: `drop if gender == "Male"`.

keep: This command is an opposite of drop as it keeps variables or observations instead of deleting them. Keeping variables is done with: `keep varlist`.

For keeping observations, you use: `keep if exp or keep in range [if exp]`.

Remember that if you modify or drop variables and save the file, you will lose your original dataset.

4.2.7 Data formatting

rename: A variable can be renamed with the command rename. The syntax is: `rename old_varname new_varname`.

label: There are two ways to label variables. The first one is to label the variable itself; in this case, the syntax is: `label varname "label"`. Example: `label variable hh_income "Household income"`.

The second option is to assign labels to the values of categorical variables. This is done in two steps. First, a value label has to be defined. For this first step, the syntax is: `label define lbname # "label" [ # "label" ...]`. Example: `label define city_label # "Buenos Aires" 2 "Rio de Janeiro"`. Second, this value label is assigned to the respective variable. The syntax is: `label values varname [lbname]`. Example: `label values city city_label`.

4.2.8 Data description

describe: General information about the dataset can be retrieved using the command describe. This command displays the number of observations, number of variables, and the size of the dataset, and it lists all variables together with basic information.
codebook: The command codebook delivers information about one or more variables, such as storage type, range, number of unique values, and number of missing values. The command offers further interesting features that can be seen with help codebook. The syntax for this command is: codebook [varlist] [if] [in] [, options].

sort: Data can be sorted in ascending order with the command sort. The syntax is: sort varlist. Example: sort gender age income.

codebook: The command codebook delivers information about one or more variables, such as storage type, range, number of unique values, and number of missing values. The command offers further interesting features that can be seen with help codebook. The syntax for this command is: codebook [varlist] [if] [in] [, options].

sort: Data can be sorted in ascending order with the command sort. The syntax is: sort varlist. Example: sort gender age income.

browse: The data browser can be opened with the command browse. Example: browse age income. The browser does not allow data manipulation (as does edit), but data can be sorted using the sort button.

list: Similar to the data browser, values of variables can be listed in the results window with the command list. Here, if and in qualifiers are often useful. The syntax is: list [varlist] [if] [in] [, options].

summarize: The most important descriptive statistics for numerical variables are delivered with the command summarize. The syntax for summarizing descriptive statistics is: summarize [varlist] [if] [in] [weight] [, options].

This command displays the number of (non-missing) observations, mean, standard deviation, minimum, and maximum. Additionally, summarize varlist, detail shows certain percentiles, skewness, and kurtosis. Tables of summary statistics can be drawn with table.

tabulate: One-way and two-way frequency tables for categorical variables can be drawn with the command tabulate. The syntax is: tabulate varname [if] [in] [weight] [, options].

4.2.9 Data merging

append: A second dataset can be appended to the end of the one currently used by using the command append. Note that Stata will automatically match variables that are common to both datasets, provided they have the same label and are of the same type (e.g. numeric or string) and will keep any other additional variable from the two datasets. The syntax is: append using filename.dta [, options].

merge: Datasets sharing the same kind of observations, but having different variables, can be joined with the command merge. The currently used dataset (“master” dataset) is merged with the corresponding observations from one or more other files (“using” datasets). The “master” and “using” datasets need to share at least one common variable, the so-called “primary key”, in order to make the match possible. The match variable(s) is (are) defined in varlist. After merging, Stata automatically generates a variable that contains information about the matching of the data:

_merge == 1 Observations only from “master” dataset
_merge == 2 Observations only from “using” dataset(s)
_merge == 3 Observations from “master” and “using” dataset(s)
_merge == 4 Observations from both, missing values updated
_merge == 5 Observations from both, conflicting non-missing values

4.2.10 Graphs

One of the advantages of Stata is its vast graphics capabilities. However, commands for comprehensive graphs can get quite long, and it takes some time to get used to the code structure. Using dialogue boxes might have an advantage in certain cases. The starting point for learning about graphs is help graph. In addition, the Stata command help offers a separate tutorial for basic graphs that can be accessed with help graph_intro. An example of a simple bar graph of two variables would be: graph bar variable_1 variable_2. A basic histogram of the variable age would be: histogram age. If you want to show the relationship between two variables, you could use the command twoway. You can use it, for example, with a scatter (twoway scatter) or a line (twoway line).

5 Conclusions

In this introductory module, we have set the stage for the subsequent modules where we describe in more detail how the relationship between trade and gender can be analysed using the microeconomic, macroeconomic, or sectoral approach. We have seen that there are numerous sources of trade, micro- and macro-level data at your disposal; identifying the source that is most appropriate for your purposes depends on the objective of your study. You will also need to know whether you wish to conduct an ex-post or ex-ante analysis of the impact of trade on gender. In the latter case, there are open-source trade models at your disposal that allow you to simulate the welfare effects of a trade policy or reform of a particular country.
At a later stage, when the aim of your study is clarified and you have collected the necessary data and information for your research, you will need to manage and analyse the data to test your hypothesis. As a reference for this exercise, this module has provided an introduction to the statistical software package Stata that is the most widely used in academic circles. We will use Stata to replicate the studies reviewed in the following modules. This module has therefore presented the key commands needed to perform those replications. It is important that you master these commands, since you will also need them to work on the data for your own empirical analysis.
REFERENCES


UNCTAD (2012). Virtual Institute teaching material for the online course on trade and poverty. UNCTAD Virtual Institute, Geneva.


