VIRTUAL INSTITUTE TEACHING MATERIAL ON
TRADE AND GENDER
VOLUME 2: EMPIRICAL ANALYSIS OF THE TRADE AND GENDER LINKS

UNITED NATIONS

VIRTUAL INSTITUTE TEACHING MATERIAL ON
TRADE AND GENDER

VOLUME 2
Empirical Analysis of the Trade and Gender Links

NOTE

The views expressed in this volume are those of the author and do not necessarily reflect those of the United Nations.

The designations employed and the presentation of the material do not imply the expression of any opinion on the part of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Material in this publication may be freely quoted or reprinted, but acknowledgement of the UNCTAD Virtual Institute is requested, together with a reference to the document number. A copy of the publication containing the quotation or reprint should be sent to the UNCTAD Virtual Institute, Division on Globalization and Development Strategies, Palais des Nations, 1211 Geneva 10, Switzerland.

This publication has been edited externally.

<table>
<tr>
<td>

The UNCTAD Virtual Institute is a capacity-building and networking programme that aims to strengthen teaching and research of international trade and development issues at academic institutions in developing countries and countries with economies in transition, and to foster links between research and policymaking.

For further information about the UNCTAD Virtual Institute, please contact:

Ms. Vlasta Macku
Chief, Virtual Institute
Division on Globalization and Development Strategies
E-mail: *vlasta.macku@unctad.org*
Website: *http://vi.unctad.org*

</td>
<td>

The Trade, Gender and Development Section carries out UNCTAD's work programme on trade and gender with the aim of making trade policy a force for inclusive development.

For further information about the UNCTAD work programme on trade, gender and development, please contact:

Ms. Simonetta Zarrilli
Chief, Trade, Gender and Development Section
Division on International Trade in Goods and Services, and Commodities
E-mail: *simonetta.zarrilli@unctad.org*
Website: *http://unctad.org/gender*

</td>
</tr>
</table>

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF BOXES

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIDS	ACQUIRED IMMUNODEFICIENCY SYNDROME
BEC	BROAD ECONOMIC CATEGORIES
BREAD	BUREAU FOR RESEARCH AND ECONOMIC ANALYSIS OF DEVELOPMENT
CIRI	CINGRANELLI-RICHARDS
CLAD	CENSORED LEAST ABSOLUTE DEVIATIONS
DHS	DEMOGRAPHIC AND HEALTH SURVEY
DOTS	DIRECTION OF TRADE STATISTICS
FAO	FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
FDI	FOREIGN DIRECT INVESTMENT
GDI	GENDER DEVELOPMENT INDEX
GDP	GROSS DOMESTIC PRODUCT
GEE	GENERALIZED ESTIMATION EQUATION
GEM	GENDER EQUALITY MEASURE
GMM	GENERALIZED METHOD OF MOMENTS
GTAP	GLOBAL TRADE ANALYSIS PROJECT
HIV	HUMAN IMMUNODEFICIENCY VIRUS
HS	HARMONIZED COMMODITY DESCRIPTION AND CODING SYSTEM
IHSN	INTERNATIONAL HOUSEHOLD SURVEY NETWORK
IDRF	HOUSEHOLD EXPENDITURE AND INCOME SURVEY (CAPE VERDE)
ILO	INTERNATIONAL LABOUR ORGANIZATION
IMF	INTERNATIONAL MONETARY FUND
ISIC	INTERNATIONAL STANDARD INDUSTRIAL CLASSIFICATION
ITC	INTERNATIONAL TRADE CENTRE
I-TIP	INTEGRATED TRADE INTELLIGENCE PORTAL
ITS	INTERNATIONAL TRADE STATISTICS
IV	INSTRUMENTAL VARIABLE
KILM	KEY INDICATORS OF THE LABOUR MARKET
LSMS	LIVING STANDARDS MEASUREMENT SURVEY
MDG	MILLENNIUM DEVELOPMENT GOAL
MERCOSUR	SOUTHERN CONE COMMON MARKET
MICS	MULTIPLE INDICATOR CLUSTER SURVEY
OECD	ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
PISA	PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT
PWI	PENN WORLD TABLES
RTA	REGIONAL TRADE AGREEMENT
SAP	STRUCTURAL ADJUSTMENT PROGRAMME
SITC	STANDARD INTERNATIONAL TRADE CLASSIFICATION
SMART	SYSTEM OF MARKET ANALYSIS AND RESTRICTIONS ON TRADE
UN	UNITED NATIONS
UNCOMTRADE	UNITED NATIONS COMMODITY TRADE STATISTICS DATABASE
UNCTAD	UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT
UNDP	UNITED NATIONS DEVELOPMENT PROGRAMME
UNICEF	UNITED NATIONS CHILDREN'S FUND
USAID	UNITED STATES AGENCY FOR INTERNATIONAL DEVELOPMENT
WCO	WORLD CUSTOMS ORGANIZATION
WDI	WORLD DEVELOPMENT INDICATORS
WEF	WORLD ECONOMIC FORUM
WITS	WORLD INTEGRATED TRADE SOLUTION
WTO	WORLD TRADE ORGANIZATION

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,[1] 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 Järviluoma *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.

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.

2.1 Microeconomic approach[2]

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?[3] 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? 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,[4] 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.[5]

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[6] to estimate the relationship between trade and gender is:

$$y_{ct} = \alpha_o + \alpha_1 Trade_{ct} + \alpha_2 X_{ct} + \eta_c + \gamma_t + \nu_{ct} \qquad (1)$$

where y_{ct} stands for the gender outcome variable of interest (for instance, the gender wage gap[7] or women's empowerment) in period t of country c, $Trade_{ct}$ stands for a measure of openness (e.g. trade share),[8] X_{ct} is a vector of other control variables (e.g. gross domestic product – GDP – per capita), and $\eta_c + \gamma_t$ are country- and time-fixed effects.[9] 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.[10] 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, particu-

larly women, benefit from growth is rather scant. Regarding employment creation, greater openness to trade may cause significant shifts in the demand for low- and high-skilled labour and may have different effects on different groups. According to standard economic theory, for example, developing countries experience an increase in the demand for low-skilled labour as a result of trade liberalization, which theoretically should benefit women given their overrepresentation among low-skilled workers (Wood, 1995).[11]

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[12] 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.[13]

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 microeconomic 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[14] 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.[15] 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/tpr_e/tpr_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/rta_pta_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** (UNCOMTRADE) (*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.[16]

- 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).[17] 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);[18] 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)[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?d=33061&e=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[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.[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.[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[23]

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.[24] 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.[25] 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

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

Source: IHSN.

Table 2

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

Source: IHSN.

Table 3

Modules in UNICEF's Multiple Indicator Cluster Survey	
Questionnaire	Information
Household questionnaire	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
Individual questionnaire for women	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 immuno-deficiency syndrome (AIDS); maternal mortality; tobacco and alcohol use; life satisfaction
Questionnaire for children under five	Under-five child information panel; age; birth registration; early childhood development; breastfeeding and dietary intake; immunization; care of illness; anthropometry
Individual questionnaire for men	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

Source: UNICEF.

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:

- Bureau for Research and Economic Analysis of Development (BREAD) (*http://ipl.econ.duke.edu/dthomas/dev_data/index.html*);
- World Bank Central Microdata Catalog (*http://microdata.worldbank.org/index.php*);
- Datasets for Development Economists (*http://devecondata.blogspot.com.ar*);
- OpenMicroData (*http://openmicrodata.wordpress.com*);
- Micro Data for Development Economics (*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:

- The World Bank Gender Statistics database (*http://data.worldbank.org/data-catalog/gender-statistics*), which provides indicators on key gender topics. Themes included are demographics, education, health, labour force, and political participation. The data are available annually since 1960 and cover 213 economies.[26]

- World Development Indicators (WDI) (*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*), 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/DIAE/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.[27]

- 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;[28]

  - 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.

- The Word Bank's Poverty Analysis Toolkit (*http://econ.worldbank.org/external/default/main?print=Y&theSitePK=477894&contentMDK=20292195&menuPK=546578&pagePK=64168182&piPK=64168060*) contains Stata programmes for growth inequality decomposition, growth incidence curves, egen functions for inequality and poverty measurements, and more. You can also find there programmes for applied economic research.

- Useful Stata codes are available in Angus Deaton's (1997) *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy* (*http://go.worldbank.org/UWQRT802H0*). This book focuses on the analysis of household survey data from developing countries.

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.

Screenshot of Stata 12.1 user interface

Source: Stata 12.1.

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[29]

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 varlist1:] command [varlist2] [=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*.

`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 *varlist1*. Note that not all commands support this feature. To use the command `by`, data have to be first sorted by *varlist1*. 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.

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
-------------------------------------------------------------------------

-> sex = Female
    Variable |       Obs.       Mean     Std. Dev.       Min         Max
-------------+-----------------------------------------------------------
     edlevel |       1161   4.828596    1.593329          1           8
```

Source: UNCTAD's estimations, based on data from Newman (2002).

`if:` `if` can be put at the end of a command in order to use only the data specified. It is allowed with most Stata commands. Example: `summarize edlevel if sex == "Male"` would generate only the top panel in the figure above, as observations recorded as female will not be considered. Several `if` qualifiers can be used to define the range of the data. For example: `summarize edlevel if (age > 45 & edlevel >= 5)`. The `if` qualifiers are connected with logical operators and are used with relational operators. Logical operators are: & for and, | for or, and ! for not. Relational operators are: > greater than, < less than, >= greater than or equal to, <= less than or equal to, == equal to, != not equal to. Note that string values have to be put in quotation marks (example `!="Male"` is a female). Note also that Stata marks a missing value for numerical variables as `.` (period) and interprets it as infinite. This is very important when referring to "greater than" and when you do not want to include missing values (so you could use `">100 & !=."`).

`in:` The qualifier `in` at the end of a command means the command should only use the specified observations. Example: `summarize edlevel in 1/100` summarizes the educational level of individuals in observations 1 through 100.

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`. Arithmetical 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`.

`egen:` Extensions to `generate` can be found in the command `egen` which offers a set of algebraic or string functions that are sometimes needed for data management tasks (see `help egen` for an overview of available functions). The syntax for the egen function is: `egen` *newvar = fcn(arguments) [if] [in] [, options]*. For example, `egen 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 variable` *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` *lblname # "label" [# "label" ...]*. Example: `label define city_label 1 "Buenos Aires" 2 "Rio de Janeiro"`. Second, this value label is assigned to the respective variable. The syntax is: `label values` *varname [lblname]*. 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.

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

Borraz F, Rossi M and Ferres D (2012). Distributive effects of regional trade agreements on the 'small trading partners': Mercosur and the case of Uruguay and Paraguay. *Journal of Development Studies*. 48 (12): 1828–1843.

Deaton A (1997). *The Analysis of Household Surveys. A Microeconometric Approach to Development Policy*. John Hopkins University Press for the World Bank, Baltimore.

Depetris Chauvin N and Porto G (2011). Market competition in export cash crops and farm income in Africa. Available at: *http://acetforafrica.org/wp-content/uploads/2011/09/paper-depetris-porto-acet-web.pdf.*

Díaz-Bonilla E and Ron JF (2010). Food security, price volatility and trade. ICTSD Issue Paper No. 27. International Centre for Trade and Sustainable Development, Geneva.

Florax Raymond JCM, de Groot Henri LF and Reinout H (2002). The empirical economic growth literature. Tinbergen Institute Discussion Paper No. 2002-040/3.

Francois J and Hall HK (2002). Global simulation analysis of industry-level trade policy. Available at: *http://wits.worldbank.org/data/public/GSIMMethodology.pdf.*

Friedman J and Levinsohn JA (2002). Distributional impacts of Indonesia's financial crisis on household welfare: A rapid response methodology. *World Bank Economic Review*. 16 (3): 397–423.

Harrison A and Hanson G (1999). Who gains from trade reform? Some remaining puzzles. *Journal of Development Economics*. 59 (1): 125–154.

Järviluoma H, Moisala P and Vilkko A (2003). *Gender and Qualitative Methods*. Sage Research Methods, Thousand Oaks CA.

Levine R and Renelt D (1992). A sensitivity analysis of cross-country growth regressions. *American Economic Review*, 82 (4): 942–963.

Mankiw G, Phelps E and Romer P (1995). The growth of nations. *Brookings Papers on Economic Activity*, 1: 275–326. The Brookings Institution, Washington DC.

Metso M and Le Feuvre N (2006). Quantitative methods for analysing gender, ethnicity and migration. The University of York Working Paper.

Newman C (2002). Gender, time use, and change: Impacts of the cut flower industry in Ecuador. *World Bank Economic Review*, 16 (3): 375–396. Available at: *https://openknowledge.worldbank.org/bitstream/handle/10986/17212/773920JRN0200200CutoFlower0Industry.pdf?sequence=1.*

Nicita A (2009). The price effect of tariff liberalization: Measuring the impact on household welfare. *Journal of Development Economics*, 89 (1): 19–27.

Nicita A and Razzaz S (2003). Who benefits and how much? How gender affects welfare impacts of a booming textile industry. Policy Research Working Paper No. 3029. World Bank, Washington DC.

Porto G (2006). Using survey data to assess the distributional effects of trade policy. *Journal of International Economics*, 70 (1): 140–160.

Ramjoué M (2010). Can trade policies promote gender equality? Exploring the trade-growth-gender nexus. ARTNeT Publications. UNESCAP, Bangkok. Available at: *http://artnet.unescap.org/pub/alert7.pdf.*

Richards D and Gelleny R (2007). Women's status and economic globalization. *International Studies Quarterly*. 51: 855–876. Available at: *https://sites.google.com/site/drdavid68.*

Rodríguez F and Rodrik D (2001). Trade policy and economic growth: A skeptic's guide to the cross-national evidence. In: Bernanke B and Rogoff K, eds. *Macroeconomics Annual 2000*. MIT Press, Cambridge MA.

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

United Nations (2002). Classification by broad economic activity. Statistical Papers Series M No. 53, Revision 4. Available at: *http://unstats.un.org/unsd/publication/SeriesM/SeriesM_53rev4e.pdf.*

Warren C (1998). *Gender Issues in Field Research*. Sage University, Newbury Park CA.

Williams C (2007). Research methods. *Journal of Business and Economic Research*. 5 (3): 65–72.

Winters A, McCulloch N and McKay A (2004). Trade liberalization and poverty: The evidence so far. *Journal of Economic Literature*, 42 (1): 72–115.

Wood A (1995). How trade hurt unskilled workers. *Journal of Economic Perspectives*, 9 (3): 57–80.

MODULE 2
The microeconomic approach[30]

1 Introduction

This module analyses the gender effects of trade policies and shocks using household survey data. The approach follows the two-step methodology that characterizes recent trade and poverty literature. The first step studies how trade policies and trade shocks affect prices of goods and factor remuneration in the domestic economy. The second step uses household survey data to assess the welfare impact of those price changes.

Trade policies have diverse effects on individuals and households: some may benefit from trade liberalization or facilitation, others may suffer, and yet others may not be affected. The results depend mainly on two factors:

- The influence that trade policies and trade shocks have on domestic prices of goods and factors of production; and
- The degree of exposure that individuals have to the various goods and factors of production.

As we will see below, trade policies and shocks have three main effects on households:

- Consumption effect – the effect on the price of goods consumed by the households directly (traded goods) and indirectly (non-traded goods);
- Income effect – the effect on household income, including labour wages, earnings from the sales of agricultural products or any other goods, and other forms of income;
- Revenue effect – the effect on the generation and distribution of government revenues that may indirectly affect household welfare through government transfers and the provision of public goods.

This module will not consider the revenue effect but will rather focus on the first two effects. For methodological reasons, we will split the income effect into two components, the production effect and the labour income effect. We will explain the difference later on.

As we will show, once the welfare impact has been estimated for each household, we can aggregate it by the relevant dimension – geography, gender, or income level – to identify "winners" and "losers" from trade policy. In this module, we focus on the gender dimension.

Section 2 of this module provides a brief literature review of studies that have applied the microeconomic approach presented here as well as a list of types of microeconomic studies. Sections 3 and 4 provide the intuition behind the methodology, leaving the more technical presentation for the interested reader in the two annexes, A and B, at the end of this module. Section 5 offers the basics on non-parametric regression techniques, which are commonly applied to explore how trade-led price changes can influence household welfare. In the hands-on application in Section 6, we then explain step by step how to replicate in Stata the estimations from an UNCTAD (2011) study on gender and trade in Cape Verde. Section 7 presents concluding remarks.

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

- Describe the microeconomic approach relating changes in trade policy and/or shocks to changes in household welfare;
- Review and summarize the literature using the microeconomic approach to disentangle the effects of trade on gender;
- Understand the mechanisms linking changes in trade policy and/or shocks with changes in consumer and producer prices – i.e. the pass-through effect;
- Split the effects of trade on household welfare into three components: (a) the consumption effect, (b) the production effect, and (c) the labour income effect;
- Understand how the microeconomic approach can be used to analyse the relationship between trade and gender;
- Understand non-parametric estimation tools;
- Replicate, using Stata, the 2011 UNCTAD study on trade and gender in Cape Verde.

2 Review of the literature

The aim of this section is to discuss a series of papers that have used the microeconomic approach described in this module. The section will focus mostly on the trade and poverty literature with a gender dimension, for example, focusing on the impact of trade on female-headed households (see the hands-on application in Section 6). In this regard, it should be noted that this literature review is mostly concerned with providing a non-exhaustive collection of papers linking trade policy with welfare at the household and individual levels. At the end of the section, several studies on trade and gender are cited that have not necessarily employed the methodology presented in this module, thereby illustrating the existence of alternative microeconomic empirical approaches.

One of the most influential papers in the trade and poverty literature is Porto (2006), which

extends Deaton (1989a) to study the impact of international trade on household welfare. The work of Porto (2006) is innovative because it considers the impact of international trade liberalization both on household consumption and wages. In this respect, it represents a general equilibrium analysis. As such, it simultaneously explores more than one channel of interaction between trade and poverty – i.e. the consumption channel and the labour market channel. It is thus different from earlier partial equilibrium analyses that considered only one sector of the economy (for instance, labour markets) at a time (Goldberg and Pavcnik, 2007). Porto (2006) identifies two stages that link changes in trade policy with changes in household welfare: First, trade reforms cause direct changes in the prices of consumption goods; and second, these price changes affect households both as consumers (because households purchase consumption goods) and as income earners (because changes in the prices of goods affect the wages of workers living in the household).[31]

Porto (2006) applies his theoretical framework to study the distributional effects of the Southern Cone Common Market (MERCOSUR) on household welfare in Argentina. For this purpose, he employs household level data as well as data on the intra-zone tariffs applied to MERCOSUR members. The findings suggest that the MERCOSUR agreement had a pro-poor distributional effect and that trade liberalization was not responsible for the increase in poverty and income inequality experienced in Argentina throughout the 1990s.

Empirical studies employing Porto's framework have mushroomed over the years.[32] For example, Nicita (2009), Marchand (2012), and Borraz *et al.* (2012) analyse the cases of Mexico, India, and Paraguay and Uruguay, respectively. Their contribution is to adapt Porto's price transmission mechanisms from tariffs to prices by allowing trade costs (proxied by distance from borders), domestic producer prices, and exchange rates to influence prices of goods as well. In this scenario, there are more data requirements that cannot always be met, especially for developing countries. Despite analysing two very different regions of the world, Nicita (2009) and Marchand (2011) arrive at a similar conclusion: the distributional effects of trade liberalization on household welfare depend on the region where the individuals live, their consumption basket, and the factors of production owned (whether skilled or unskilled labour). Borraz *et al.* (2012) analyse the distributional implications of MERCOSUR on household welfare in Uruguay and Paraguay and also ac-

count for poverty and inequality effects separately. They find that preferential trade liberalization affected households in Paraguay and Uruguay differently, thereby suggesting that trade has different effects across countries as well. In sum, there is still no agreement on the welfare effects of trade liberalization at the microeconomic level, and there is scope for further research in this direction.

Despite its validity, the microeconomic approach introduced here represents only one of the existing ways to analyse the link between trade and gender. When it comes to exploring the gender implications of trade, one strand of literature focuses on intra-household dynamics. The basic assumption is that households are heterogeneous units comprised of men and women who differ in their control over resources and consumption preferences. By affecting prices and wages, trade may bring changes to the allocation of resources among members of the household and ultimately to the well-being of the household as a whole. This theoretical framework has been applied in a series of papers included in Bussolo and De Hoyos (2008). For example, by focusing on the case of Senegal, Bussolo *et al.* (2008) test the idea that in developing countries' rural households, education spending, mostly controlled by women, can be affected by a trade-related increase in the prices of export crops, from which men benefit the most. The findings support this hypothesis: men's income share increases relative to women's share, with negative though limited repercussions on the amount of education spending for children. De Hoyos (2008) focuses instead on the implications of trade for female wages. In particular, the study looks at better working opportunities for women, including lower gender wage gaps, offered by the *maquila* sector of Honduras that has developed and grown thanks to the liberalization policies implemented by the country since the early 2000s. The study also tries to explore the link between women's improved working conditions and poverty by simulating the level of poverty that Honduras would have reached if the *maquila* had not existed. The results show that in this scenario poverty levels would have been 1.5 percentage points higher.

In conclusion, what we want to say is that although this module focuses on a particular type of microeconomic approach, it is not the only approach. The aim of this brief literature review is to stimulate your interest and enrich your knowledge about the multiple approaches you can use to study the linkages between trade and gender at the microeconomic level.

3 From trade policy to prices

Going back to the two-step methodology introduced by the trade and poverty literature, this section explains the first step in detail. The key idea is that domestic and international prices are linked and that trade policy and trade shocks will therefore have an effect on domestic prices and factor remuneration. According to Brambilla and Porto (2009), standard models of international trade assume competitive markets (with homogeneous goods) and frictionless trade. In this scenario, markets are integrated and the law of one price holds:[33] domestic prices are equal to international prices converted into the local currency. That is, if a product costs $5 abroad and the exchange rate in your country is two units of the local currency for each unit of the foreign currency ($), then the product should cost 10 units in local currency. Of course this is a very simple model, as it does not consider transportation and distribution costs, or the fact that the price of the good may also be affected by trade policy instruments (such as tariffs). In this case, if p_i is the domestic price of an imported good, p_i^* is the international price, e_i is the exchange rate, tr_i is the sum of international transaction costs (i.e. transportation), and τ_i is the tariff rate applied to good i, then we may write:

$$p_i = p_i^* e_i \left(1 + tr_i \right) \left(1 + \tau_i \right) + \gamma_i \qquad (1)$$

where γ_i represents internal transportation, resale, marketing, and distribution costs. If good i is exported, then equation (1) becomes:

$$p_i = p_i^* e_i \left(1 - tr_i \right) \left(1 - r_i \right) - \gamma_i \qquad (2)$$

where r_i is the export tax if different from zero.

Let us assume for simplicity that internal costs γ_i are zero so we can focus on the response of domestic prices to changes in international prices, exchange rates, national trade policies, international trade policies, and transaction costs. Clearly, if these equations hold, then a proportional change in the exchange rate e_i, in the international price p_i^*, or in the tariff rate τ_i (or rather $(1+\tau_i)$) will be fully transmitted to domestic prices.[34] This can be formally derived from the log-linearized version of equation (1), where we have excluded internal costs γ_i :

$$\ln p_i = \ln p_i^* + \ln e_i + \ln(1 + tr_i) + \ln(1 + \tau_i) \qquad (3)$$

The derivative of $\ln p_i$ with respect to any determinant of the price (e.g. $\ln e_i$) is equal to one. This derivative corresponds to the elasticity of domestic prices with respect to the determinants of this elasticity. This implies that any relative change on the right-hand side of the expression would be fully transmitted to domestic prices. We call this perfect pass-through. However, there is strong evidence against this prediction, especially for exchange rates. Most studies consistently reject the law of one price for a variety of products and countries. There are many reasons why the law of one price may fail, some of which are presented by Feenstra (1989) and Nicita (2009). When the relative change in the domestic price of a good is lower than the relative change in international prices, tariffs, or exchange rates, we say that there is imperfect pass-through. For instance, if the international price increases by 10 per cent but the local price only increases by 5 per cent, keeping everything else (tariffs, transportation costs, and exchange rate) constant, the pass-through is 50 per cent. Sometimes prices take time to adjust following a shock because there are signed contracts, accumulated stocks, and other market frictions. For that reason, the pass-through is often lower in the short run than in the long run, and this should be taken into account when analysing the effects of trade policies.

Even though pricing equations (1) and (2) do not provide an accurate framework for measuring and estimating pass-through effects, they are useful to conceptually show different effects of price changes. For instance, it is often observed that governments react to exogenous changes in international prices by changing tariffs and export taxes (and sometimes also by altering the exchange rate via devaluations). In some instances, when the price of food imports increases, governments may reduce tariffs to alleviate the increase in domestic prices. In some exporting countries, the government's response to skyrocketing commodity prices has been to increase export taxes. Increasing export taxes reduces the incentives for national producers to export,[35] which increases the supply of the good in the domestic market and consequently decreases the domestic price of the good. Since food products represent consumption goods, such policies can be supported on distributional grounds. Export taxes are also a good source of public revenue, especially in the context of increasing international prices, which provides additional motivation for their implementation (UNCTAD, 2012).

There are several methodological approaches and models to study price changes assuming both perfect and imperfect pass-through. These studies are beyond the objective of this module, but we have included a reference to them in Annex A for the interested readers.

4 From prices to welfare impact

A useful way to study how trade affects household welfare is by noting that trade affects the prices faced by producers and consumers. In consequence, we can investigate the trade-welfare link by tracing how trade policy affects prices and, in turn, how prices affect welfare (Porto, 2006; Nicita, 2009). The previous section examined how trade affects prices. Here we focus on the second step of the two-step methodology discussed so far and look at how price changes translate into welfare effects.

The consumer and producer surplus measures covered in introductory microeconomics courses are useful to illustrate how price changes affect welfare. Consider all households consuming a good i whose initial price p_i changes. We can estimate the impact on consumer surplus by multiplying the amount of the price change by the quantity consumed by all households before (or after) the price shock. In the case of a price increase of good i, this would be an approximation of the loss in consumer surplus because each household has to pay a higher price for each unit it consumes. This loss would correspond to the shaded area in panel (a) of Figure 3. Conversely, if households produce and sell the good and its price increases, then all producing household will be better off as they will receive a higher price for each unit they sells in the market. In this case, the change in the producer surplus would be positive, and it can be approximated by the price differential multiplied by the production level before (or after) the shock that generated the price change (panel (b) in Figure 3).

Figure 3

Impact of a price increase ($p_0 \rightarrow p_1$) on consumption and production

(a) Consumer welfare loss from a price increase
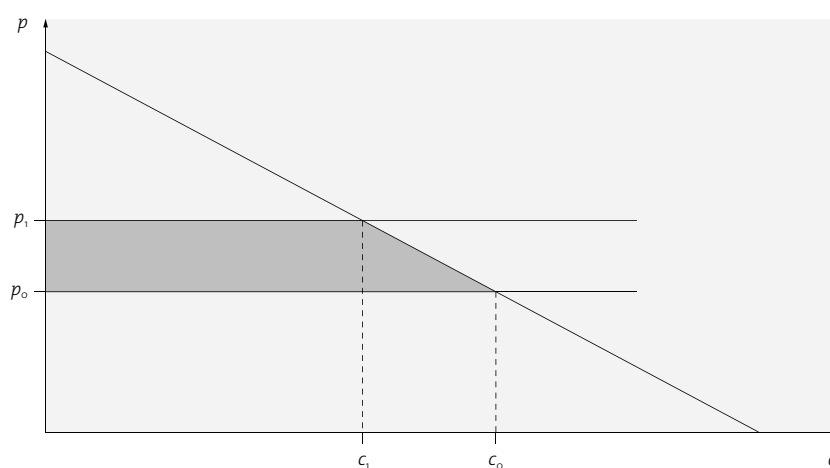
(b) Producer welfare gain from a price increase
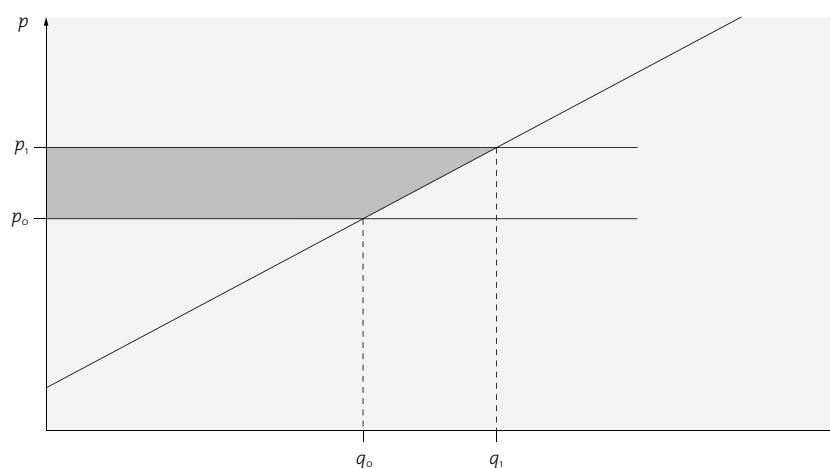
Source: UNCTAD (2012).
Note: p_0 is the initial price which increases to p_1 after an exogenous price shock.

To evaluate the overall impact of the price change on household welfare, we need to consider the changes in both the consumer and producer surpluses. If the household is a net producer of the good (i.e. its production exceeds consumption), the loss in the consumer surplus is lower than the gain in the producer surplus and the welfare of the household will increase. On the other hand, if the price of the good increases and the household is a net consumer (i.e. its consumption exceeds production), the welfare of the household will decrease.[36]

If a household is a net consumer of the good (consumption > production), the loss of welfare after the price increase can be approximated by:

$$\Delta W = -\Delta p \left(c_0 - q_0 \right) \tag{4}$$
or
$$\Delta W = -\Delta p \left(c_1 - q_1 \right) \tag{5}$$

where Δp is the amount of the price change, c_0 and q_0 are the quantities consumed and produced before the price shock (respectively), and c_1 and q_1 are the quantities consumed and produced after the price shock. Panel (a) in Figure 4 shows the loss for the net consumer.

In the case of a net producer household (production > consumption), the impact of a price change on welfare will be positive and can be approximated by:

$$\Delta W = \Delta p \left(q_0 - c_0 \right) \tag{6}$$
or
$$\Delta W = \Delta p \left(q_1 - c_1 \right) \tag{7}$$

Panel (b) in Figure 4 shows the gain for the net producer.

Figure 4

Loss and gain from price increase of good *i*

(a) Welfare loss from a price increase for a net consumer ($c_0 > q_0$ or $c_1 > q_1$)
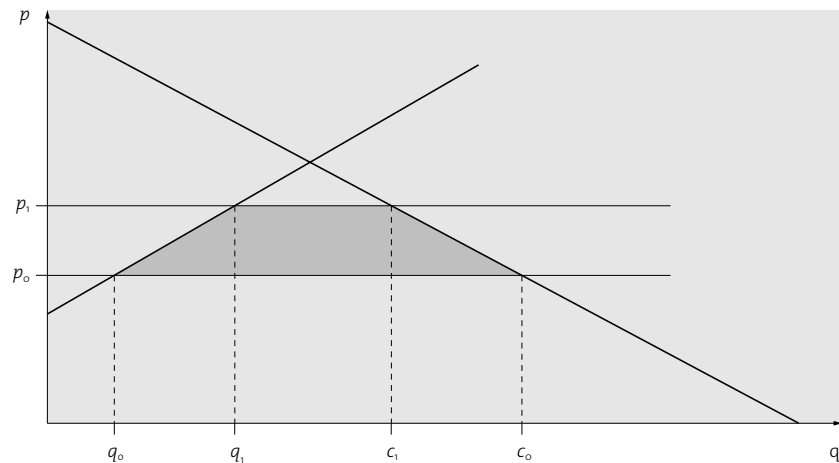
(b) Welfare gain from a price increase for a net producer ($q_0 > c_0$ or $q_1 > c_1$)
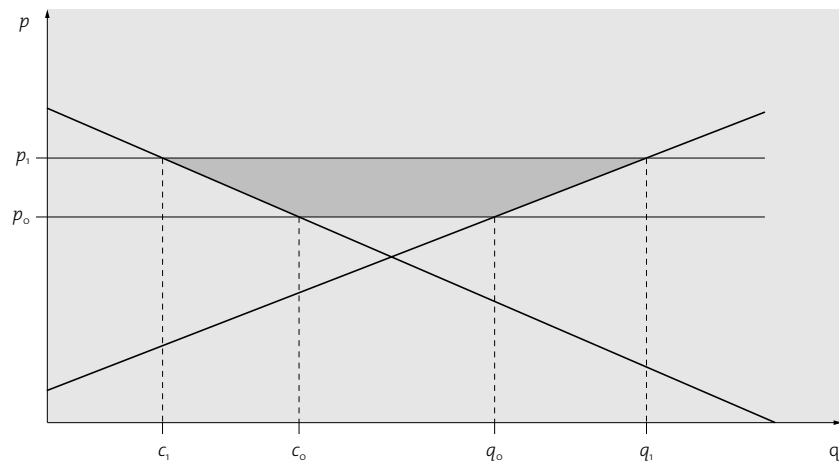
Source: UNCTAD (2012).

Unfortunately, in most surveys the quantities households consume and produce are not observed and therefore it is not possible to use this type of intuitive approximation in practical applications. As discussed below, this can be overcome by using consumption expenditure shares and production income shares, information that is often available in household surveys. Aside from this difficulty, this simple approximation does not take into account possible labour market changes following the price change. It also does not consider the general equilibrium effects that can affect the price of non-traded goods as a response to changes in the price of traded goods. We will look at these issues now.

Figure 5 provides us with the basic intuition of the transmission of a traded good price increase to prices of non-traded goods and the labour market equilibrium. In panel (a), we have a representation of the labour demand curve (which decreases as the wage level rises) and the labour supply curve (which increases as the wage level rises; their intersection determines the equilibrium wage paid to the workers. Assuming an exogenous price increase of a traded good ($p_o \rightarrow p_1$), companies selling that good would like to sell more of it. Consequently, at each level of wage, the labour demand from firms will increase, shifting the demand curve upwards ($D_o \rightarrow D_1$). As a result, there will be a new equilibrium in this market, with higher wages ($w_o \rightarrow w_1$) and more labour employed ($L_o \rightarrow L_1$) in the economy.

This change in a particular market will have spillover effects on other markets. For instance, this could be the case because the traded goods are an input in the production of non-traded goods. It could also be the case that following an increase in the wages paid to workers in traded sectors, the firms in the non-traded goods sectors would have to increase the wages paid to their employees in order to keep them.[37] In panel (b) of Figure 5, we observe the effect of this wage increase on non-traded goods. The wage increase implies a higher production cost of non-traded goods, causing the supply curve for the non-traded goods to move to the left ($S_o \rightarrow S_1$). Finally, we observe that the equilibrium price of the non-traded goods increases ($p^k\uparrow$) and the quantity sold in the market decreases ($q^k\downarrow$).

Figure 5

Effect of a price increase on labour market equilibrium and prices of non-traded goods

(a) Effect on labour market equilibrium

(b) Effect on prices of non-traded goods
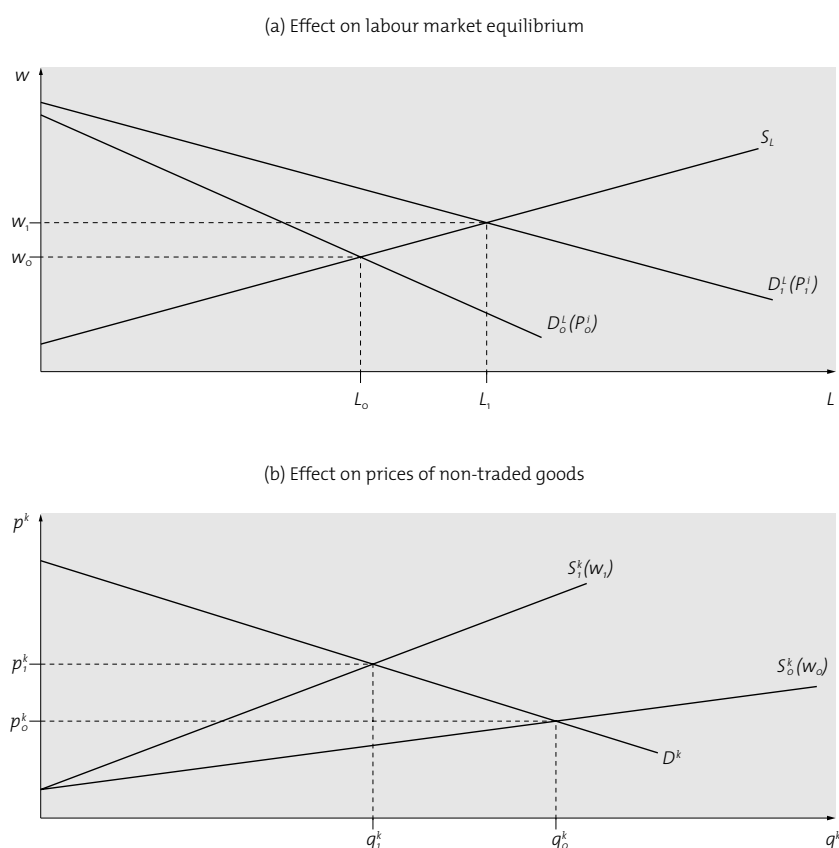
Source: UNCTAD (2012).

The welfare effect will now depend not only on how trade affects the price of the goods the household consumes and produces, but also on the effect of trade on labour income. As mentioned previously, we rarely observe the quantities of a good a household buys or sells in the household surveys. However, we often have information on budget shares and income shares. We know what percentage of total household expenditures a household spends on a good i. We also often know from the household surveys about the source of income of the household, that is, what percentage of the total income of the household comes from selling good i or from selling its labour in exchange for wages. Therefore, we can approximate the welfare impact of a trade policy or shock by estimating three effects:

$$\left(\begin{array}{c}\text{Consumption effect}\\\text{for household } h\end{array}\right) = -\left(\begin{array}{c}\text{Share of good } i \text{ in total}\\\text{expenses of household } h\end{array}\right) \cdot \left(\begin{array}{c}\text{Change in the price}\\\text{of good } i\end{array}\right)$$

$$\left(\begin{array}{c}\text{Production effect}\\\text{for household } h\end{array}\right) = -\left(\begin{array}{c}\text{Share of income household } h\\\text{gets from selling good } i\end{array}\right) \cdot \left(\begin{array}{c}\text{Change in the price}\\\text{of good } i\end{array}\right)$$

$$\left(\begin{array}{c}\text{Labour income effect}\\\text{for household } h\end{array}\right) = -\left(\begin{array}{c}\text{Share of income household } h\\\text{gets from selling its labour}\end{array}\right) \cdot \left(\begin{array}{c}\text{Change in the wage}\\\text{perceived by household } h\end{array}\right)$$

Finally, the total welfare effect for the household is the combination of the three effects:

Welfare effect = Consumption effect
　　　　　 + Production effect
　　　　　 + Labour income effect

Annex B presents a detailed derivation of this result.

5 Methodological approach: Non-parametric regressions

Most of the papers in the trade and gender literature that use the microeconomic approach follow the standard two-step approach of the trade and poverty literature presented above and in Annex B. However, the 2011 UNCTAD study that we will review below skips the first step. Rather than trying to answer how trade liberalization would increase or reduce prices and household income, the paper just assumes those changes and tries to estimate the welfare impact at the household level and determine if that impact is different, depending on the gender of the household head and the number and share of women in the household.

The second step of the approach is based on non-parametric analysis.[38] Non-parametric methods let the data show the "shape" of the relationship between the y and the x variables without any parameters that would, for instance, appear in linear regression analysis. The UNCTAD study uses non-parametric methods to estimate densities (Figures 2–5 in the paper) and regressions (Figures 6–15 in the paper).[39] We provide the in-

tuition behind both the non-parametric densities and regressions and briefly explain how to estimate them using Stata.

5.1 Basic idea of density estimations

Suppose you have a large number of observations on a variable x and you would like to "draw a picture" of the density function of x. The simplest method is to divide the range of x into a small number of intervals and count the number of times x is observed in each interval – basically, to use a histogram. You need to choose the number of "bins" (number of columns) into which you will split the data. If you choose too few, you will not capture the shape of the distribution very well. On the other hand, choosing too many will make the distribution too erratic due to the small number of points in each bin. The larger the sample, the more scope for using smaller bins (Cameron and Trivedi, 2005).

There are two problems with histograms. First, for a given number of bins, moving their exact location (boundary points) can change the figure. Second, from the technical point of view, the density function produced is a step function and the derivative either equals zero or is not defined (at the cut-off point for two bins).[40] This represents a problem if we are trying to maximize a likelihood function[41] that is characterized by a step function of the distribution. The first problem with histograms – i.e. the arbitrariness in the location of the bin cut-off points – can be avoided by having a "moving" bin that is defined for every possible value of x. The second problem of discontinuities in the estimated ("empirical") density function can also be avoided by using kernel

estimation (Glewwe, 2013). For continuous data taking many values, kernel density plots are preferable to histograms, as they result in a smooth curve that directly connects the midpoints of a series of histograms rather than forming the histogram step function.

Figure 6 shows graphically the difference between a histogram and its non-parametric estimation using kernel densities for numbers generated randomly from a given distribution. You can easily see that the non-parametric estimation is smoother.

Histogram and kernel density function
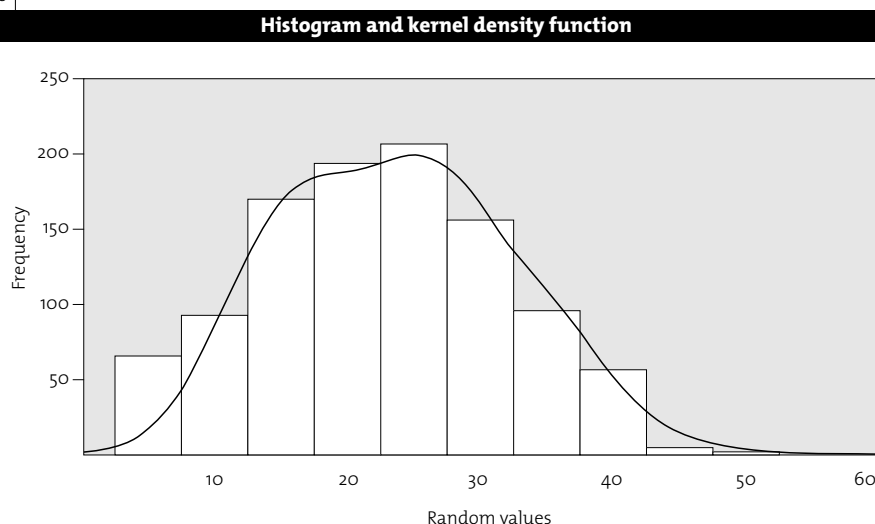
Source: UNCTAD Secretariat.

In Stata, the command for kernel densities is `kdensity`. It has different options:

- `bwidth(#)` specifies the half-width of kernel, i.e. the width of the density window around each point;
- `generate` *(newvar_x newvar_d)* stores the estimation points in *newvar_x* and the density estimate in *newvar_d*;
- `n(#)` indicates the number of points at which the density estimate should be evaluated; if not specified, the default is *min(N,50)* with *N* the number of observations;
- `at(`*var_x*`)` estimates a density using the values specified by *var_x*;
- `nograph` suppresses the graph.

We will use some of these options in our Stata estimation below.

5.2 Intuition behind non-parametric regressions

Suppose you want to regress a variable *y* on a single explanatory variable *x*, without using any functional form on the relationship (e.g. without imposing the functional form of the relationship). This is not the same as looking at a density estimate of a bivariate distribution; a plot of a bivariate distribution has not yet defined one variable as the dependent variable and the other as the independent variable. With regressions, we are ultimately interested in the expectation of *y* conditional on *x*. Assuming away problems

of endogeneity of *x*, measurement error, etc., the relationship that we are interested in estimating is the expectation of *y* conditional on *x*: E[*y*|*x*]. The relation of interest is:

$$y_i = m(x_i) + \varepsilon_i \qquad i = 1, 2, \dots N \qquad (8)$$

where $\varepsilon_i \sim i.i.d\,(0, \sigma_\varepsilon^2)$. The functional form of *m(.)* is unknown and we will not approximate it using some parametric functional form. Kernel regressions are weighted average estimators that use kernel functions as weights. Technically, a kernel regression estimator is a local constant regression because it sets *m(x)* equal to a constant in the very small "neighbourhood" of *x* (Glewwe, 2013).

Figure 7 is an example of a graph generated by the non-parametric regression of the share of income that households in Ghana earn from selling cocoa (vertical axis) on the logarithm of the household per capita consumption (horizontal axis). The non-parametric regression will be useful to assess the effect of a price change in cocoa. As cocoa is a cash crop that in most cases is not consumed by Ghanaian households, the consumption effect is nil. The figure shows that the share of income coming from cocoa in Ghana increases with the level of income. It also shows the non-parametric regressions conditional on the location of the household (rural or urban areas) – we can see that rural households have a larger share of their income coming from cocoa than urban households.

Figure 7

Non-parametric regression of share of income from cocoa and per capita expenditure in Ghana

National ——
Urban - - -
Rural ········
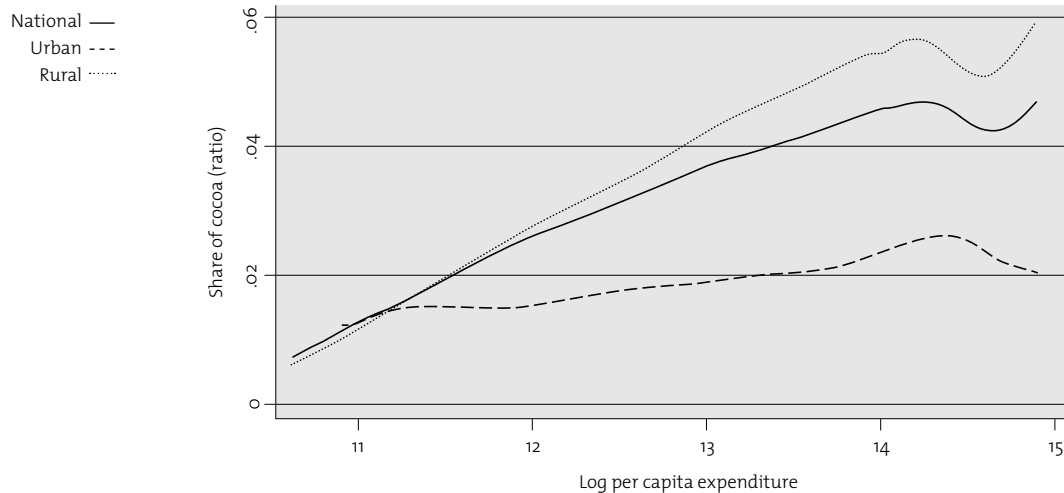
Source: UNCTAD Secretariat, based on the Ghana LSMS 5.

Several methods can be applied for non-parametric regressions. In Stata, two commands are used: `lpoly` and `lowess`. In the paper which we will review below, the author uses `lpoly` that performs a kernel-weighted local polynomial regression of *y* on *x* and displays a graph of the smoothed values. See the command `help` in Stata (`h lpoly -> Manual: [R] lpoly`) for a more extensive and complete explanation about smooth regression and local weighted regressions.

6 Hands-on application: "Who is benefiting from trade liberalization in Cape Verde? A gender perspective" (UNCTAD, 2011)[42]

6.1 Context and overview

The objective of this study developed by UNCTAD's Trade, Gender and Development Section is to explore whether there are any differential effects of trade policy on men and women and, in particular, to analyse whether there is a gender bias in the potential gains from trade in Cape Verde. A small and open country, Cape Verde is largely dependent on imports, remittances, and tourism. The export sector is small and limited to primary and low technology-intensive goods. In general, the country has a very large deficit in its balance of trade. A portion of this deficit is financed by tourism and travel receipts (19.5 per cent of GDP in 2008), and remittances from emigrants (8.5 per cent of GDP in 2008). Based on its analysis and results, the study also provides some policy recommendations.

The preparation of the study involved the following methodological steps:

- Preparation of a country profile that included identifying vulnerable groups and key economic sectors, with special emphasis on women;
- Description of the trade sectors, which identified major imports and exports, partners, trade agreements, and markets;
- Assessment of some of the effects of trade liberalization on household welfare (with a focus on gender issues), including an assessment of whether the effects depend on the location of the household (if located on an island or not), and an analysis of the effects from food prices, remittances, and tourism;
- Drafting of a set of policy recommendations.

This section focuses on the third step, i.e. the assessment of some of the effects of trade liberalization on household welfare using the microeconomic approach. In particular, we look at the effects of trade on some gender outcomes. The study follows the two-step methodology used in the trade and poverty literature described above. However, it does not cover the first step in detail, i.e. it does not look at the impact of trade liberalization on changes in food prices, tourism receipts, or remittances, but just assumes these changes. Given these assumptions, it then tries to estimate their welfare impact at the household level in Cape Verde, following the second step of the analysis. Table 4 provides a summary of the country case study of Cape Verde.

Table 4

Overview of UNCTAD (2011)

Objective

The objective of the study is to analyse some of the effects of trade liberalization on household welfare in Cape Verde, with a focus on gender issues. It looks at the effects stemming from:

- Changes in food prices: Cape Verde is highly dependent on food imports (80 per cent of the food it consumes is imported). In particular, the study simulates (a) an increase in international food prices; and (b) trade policy changes, including the phasing out of the Everything But Arms initiative (duty-free access to the European Union) on 1 January 2012 and subsequent implementation (although negotiations are still ongoing) of the Economic Partnership Agreement (duty-free access to the European Union but on the basis of reciprocity).
- Increase in income from remittances from the Cape Verdean diaspora.
- Expansion of the tourism subsectors (hotels and restaurants, trade and transport), given the potential of the tourism sector for economic empowerment of women.

Methodology

- Two-step approach: (a) assumed induced price and income changes; (b) used price and income changes and household data to study the trade impact on household welfare.
- Non-parametric estimation techniques to describe the effects of changes in food prices, remittances and tourism, conditional on the level of income, the geographical area where the household is located, and the gender of the head of the household.

Equation estimated

In non-parametric methods, we do not estimate equations, as there are no parameters to estimate. Results are displayed in terms of graphs and/or summary tables.

Dependent and independent variables

The dependent variables are:

- The share of consumption spent on food (*share_food*)
- The share of income coming from remittances (*part_rem*)
- The gains coming from the simulated increase in remittances (*gain*)
- The gains coming from the simulated increases (three scenarios) in expenditure related to the tourism sector (*gain1, gain2,* and *gain3*)

The independent variables are:

- The log per capita expenditure of the household (*logpce*)
- The area, rural or urban, where the household is located (*area*)
- The location (which island) of the household (*ilha*)
- The gender of the household head (*female_head*)
- The proportion and number of women in a household (*female* and *sfemale*)

Results

- The price effects will be more strongly felt in rural areas than in urban areas but the gender differentiated effects from food price changes would only be marginal.
- In general, a drop in food prices will have a pro-poor effect, while higher food prices will have an anti-poor bias.
- The simulated increases in remittances and income from tourism seem to favour female-headed households. Female-headed households located in rural areas benefit in particular from increases in remittances, while those located in urban areas gain more from increases in income related to tourism.
- While the reduction of prices has a pro-poor impact, the increase in remittances and income from tourism would benefit mostly middle- and high-income families.

Source: UNCTAD (2011).

6.2 Data sources

The data for the analysis are already saved in Stata format (*graphs.dta*). There are two main sources of data:

- Cape Verde's 2002 Household Expenditure and Income Survey (*Inquérito Às Despesas e Receitas Familiares* – IDRF);
- Cape Verde's 2007 Questionnaire on Basic Welfare Indicators (*Questionário Unificado de Indicadores Básicos de Bem Estar*).

Detailed information regarding the methodology and questionnaires used in these two surveys can be found on the website of the Cape Verdean National Institute of Statistics.[43] When we go through the methodology step by step,

we will also discuss the content of the different variables.

The IDRF does not contain data that allow for identifying net producers and net consumers of food. However, considering that Cape Verde imports 80 per cent of the food it consumes, it seems reasonable to assume that most households consume imported food.

6.3 Step-by-step explanation of how to do the estimations in Stata

We will now show you how to estimate the kernel densities and non-parametric regressions of the Cape Verde study in Stata. To do that, we will describe each of the steps and commands in the do-file that is provided with this teach-

ing material "doFiguresCapeVerde.do".[44] As explained in the previous module, although it is possible to use Stata interactively, in this and the following modules we will work with do-files. The advantage of writing a do-file is that you do not have to type the same commands again and again. You can run your programme as many times as you wish and correct typos or wrong commands.[45]

Before going through the steps in the Stata do-file, we strongly recommend that you read the UNCTAD (2011) study.

Step 1: Open the database and explore the variables

The first step is to ask Stata to clean any result and data that may be in the memory and set an appropriate memory space level to work with. The commands are as follows:

```
clear matrix
clear
set mem 100m
```

We then tell Stata in which folder in the computer we will be working. You need to upload the data file (*dta* file) to the computer first, and it is also where Stata will save the outputs by default. The command will appear as follows:

```
cd "c:\...\...\...\"
```

Note that you need to specify the disk ("c" here), the folder, subfolders, etc. In sum, you need to specify the folder path. Note that we use "" to indicate the whole path.

We now need to tell Stata what database we will use in the analysis:

```
use "Data\graphs.dta", clear
```

Often we use different data files, and it is convenient to work with macros that allow us to call them separately.[46] The command to define a local macro is `local`. We can tell Stata to create one and then to use it. The database we will use for the analysis (graphs.dta) is located in the subfolder *Data* that is inside the folder we specified above with the command `cd`. The command to create a local macro is the following:

```
local base_in_1 ="Data\graphs.dta"
```

We then tell Stata to use the dataset:

```
use "`base_in_1'", clear
```

Now you have the data uploaded, and you can see the name of all file variables in the variables window. What does each variable contain? We can use the command `describe` to take a first look at the data. As you will see from the output of the command, most of the labels are in English but a few of them are in Portuguese, so we explain them below:

- *Actividad* is the economic sector where the individual works;
- *Ilha* is the island where the household is located;
- *des_indi* is the per capita expenditure of the household;
- *area* takes two values: 1 corresponds to urban households, 2 corresponds to rural households;
- *pens_est* corresponds to the amount of foreign pensions received;
- *rem_emig* corresponds to the amount of remittances received.

Step 2: Describe the income distribution using kernel densities

We will use a kernel density function to describe the distribution of the log of per capita expenditure for all households, all urban households, all rural households (one graph combining the three), by island, and by the gender of the household head. This corresponds to Figures 2–5 in the UNCTAD (2011) study.

Note that the data do not contain the logarithm of per capita expenditure but the per capita expenditure (*des_indi*). We can create the logarithm of per capita expenditure using the command `gen`:

```
gen logpce = log(des_indi)
```

There are some variables and labels we will use repeatedly in the graphs, so it would be convenient to define them. In particular, we will often split the household data by area (rural vs. urban households). Let us make a value label called `area1` to label the values of the variable *area*. This is a two-step process in which you first define the label and then assign the label to the variable. The command `label define` creates the value label called `area1` that associates 1 with an urban household and 2 with a rural household.

```
label define area1 1 "Urban" 2 "Rural"
```

The command `label values` associates the variable *area* with the label `area1`.

```
label values area area1
```

If we use the command `describe`, we can see that the variable *area* has a value label called `area1` assigned to it.

We are now ready to produce our kernel densities of log of per capita expenditure for the whole population and its different subsamples. We will use the command `kdensity` to produce the kernel densities and the command `twoway` to produce a graph. We could first produce the densities, save the outputs, and then build the graphs using those outputs, or combine everything in one single programme sentence. We will use the latter option as it is faster.

The Stata command to reproduce Figure 2 in the UNCTAD (2011) study is:

```
twoway (kdensity logpce[w=pondera],
legend(lab(1 "National")))
(kdensity logpce [w=pondera] if
area== 1, lpattern(dash)legend
(lab(2 "Urban"))) (kdensity
logpce [w=pondera] if area== 2,
lpattern(dash_dot) legend(lab
(3 "Rural"))), ytitle(Density)
xtitle(Log per capita expenditure)
```

Note that we estimate kernel densities for all households (`kdensity logpce [w=pondera]`), and then for those in urban areas (`kdensity logpce [w=pondera] if area== 1`) and those in rural areas (`kdensity logpce [w=pondera] if area== 2`). When producing the density estimations, we ask Stata to give each household a different weight (`w=pondera`) because of the representativeness of each household in the population.

We use the command `twoway` to plot the three densities on a single graph. For each density, we specify its `legend` (national, urban and rural) and the type of line pattern (`lpattern`) (`solid` (which is the default), `dash`, and `dash_dot`). The syntax of the command `twoway` requires each set of line variables and its associated parameters to be included in a pair of parentheses. We can also specify the title of the y and x axes using `ytitle` and `xtitle`, respectively.

The output of the command is shown in Figure 8 (Figure 2 in the study).

Figure 8

Distribution of income for national, urban and rural households
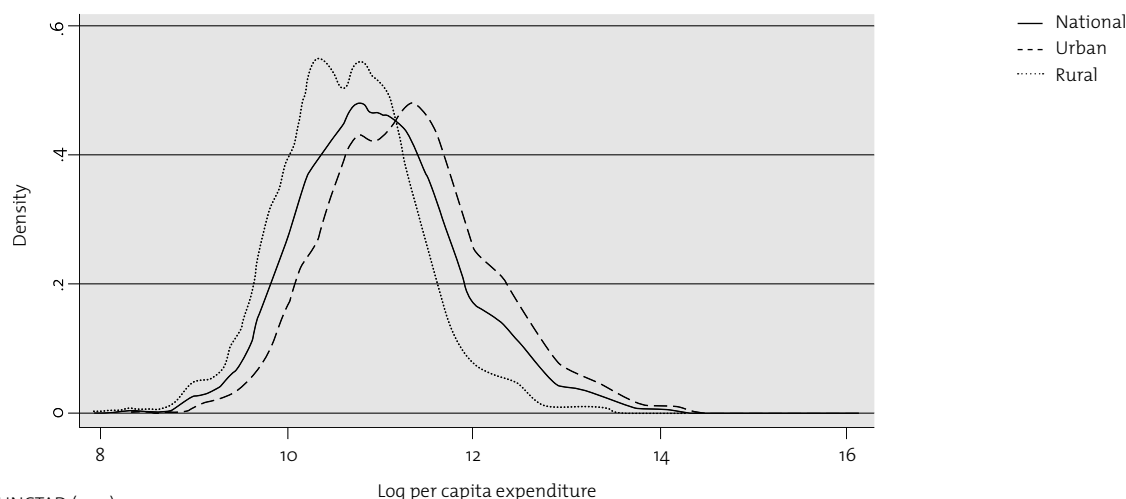
Source: UNCTAD (2011).

Figure 8 shows the distribution of the log of per capita expenditure in Cape Verde for all households, urban households, and rural households. The distribution is somehow similar to a normal distribution, with the urban density shifted to the right as urban households tend to be wealthier than rural ones.

We can split the population not only by area (urban and rural) but also by gender of the head of the household and obtain Figure 9 (Figure 3 in the study). The Stata command is similar to the previous one:

```
twoway (kdensity logpce [w=pondera]
if female_head==1 , legend(lab(1
"Female- headed"))) (kdensity logpce
[w=pondera] if female_head== 0,
lpattern(dash) legend(lab(2
"Male-headed"))), by(area)
ytitle(Density) xtitle(Log per
capita expenditure)
```

Distribution of income for female- and male-headed households in urban and rural areas

Female-headed ——
Male-headed - - -

(a) Urban   (b) Rural

Log per capita expenditure
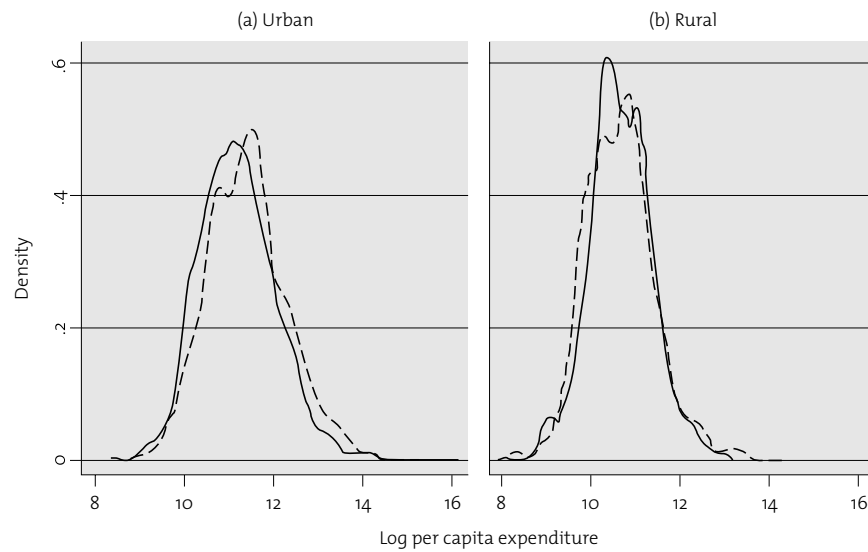
Source: UNCTAD (2011).

Figure 9 compares the income distribution among urban and rural female- and male- headed households. There is a left shift of the female-headed income distribution relative to the male-headed income distribution, in particular in urban areas. Therefore, the data show that, on average, female-headed households tend to be poorer than their male counterparts.

We can also produce a similar figure, shown here as Figure 10 (Figure 4 in the study), by island and area using:

```
graph twoway (kdensity logpce
[w=pondera] if area == 1 , legend(lab
(1 "Urban"))) (kdensity logpce
[w=pondera]if area==2,lpattern(dash)
legend(lab(2 "Rural"))),
ytitle("Density",size(small))
xtitle("Log per capita
expenditure",size(small)) by(ilha,
cols(3)) legend(size(small))
```

By specifying `cols(3)` in the `by` option, we are telling Stata to display the panels in three columns.

Distribution of income for urban and rural areas for each island

Urban ——
Rural - - -

S. Antão   S. Vicente   S. Nicolau

Sal   B. Vista   Maio

Santiago   Fogo   Brava

Log per capita expenditure
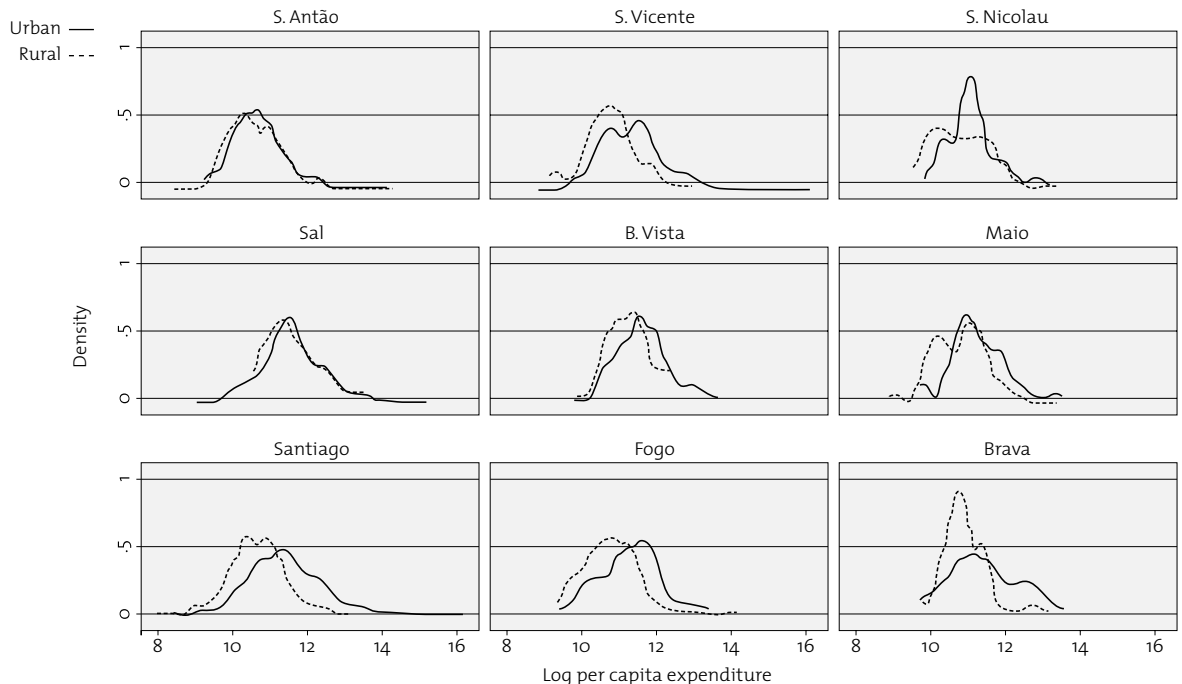
Source: UNCTAD (2011).

Finally, Figure 11 (Figure 5 in the study) shows the distribution of income for each household by the gender of the head of the household. The syntax of the Stata command is similar:

```
graph twoway (kdensity logpce
[w=pondera] if female_head==1 ,
legend(lab(1 "Female-headed")))
(kdensity logpce [w=pondera] if
female_head== 0,lpattern(dash)
legend(lab(2 "Male-headed"))),
ytitle("Density",size(vsmall))
```

```
xtitle("Log per capita
expenditure",size(vsmall)) legend
(size(vsmall)) by(ilha, cols(3)
rescale)ylabel(,labsize(vsmall))
xlabel(,labsize(vsmall))
```

In this case, we add `rescale` to the `by` option to allow the x and y axes to be different for each group. Also, the `ylabel` and `xlabel` options are added to the command `graph twoway` to indicate the size of the values on the y axis and x axis.

Figure 11

Distribution of income by gender of the household head for each island
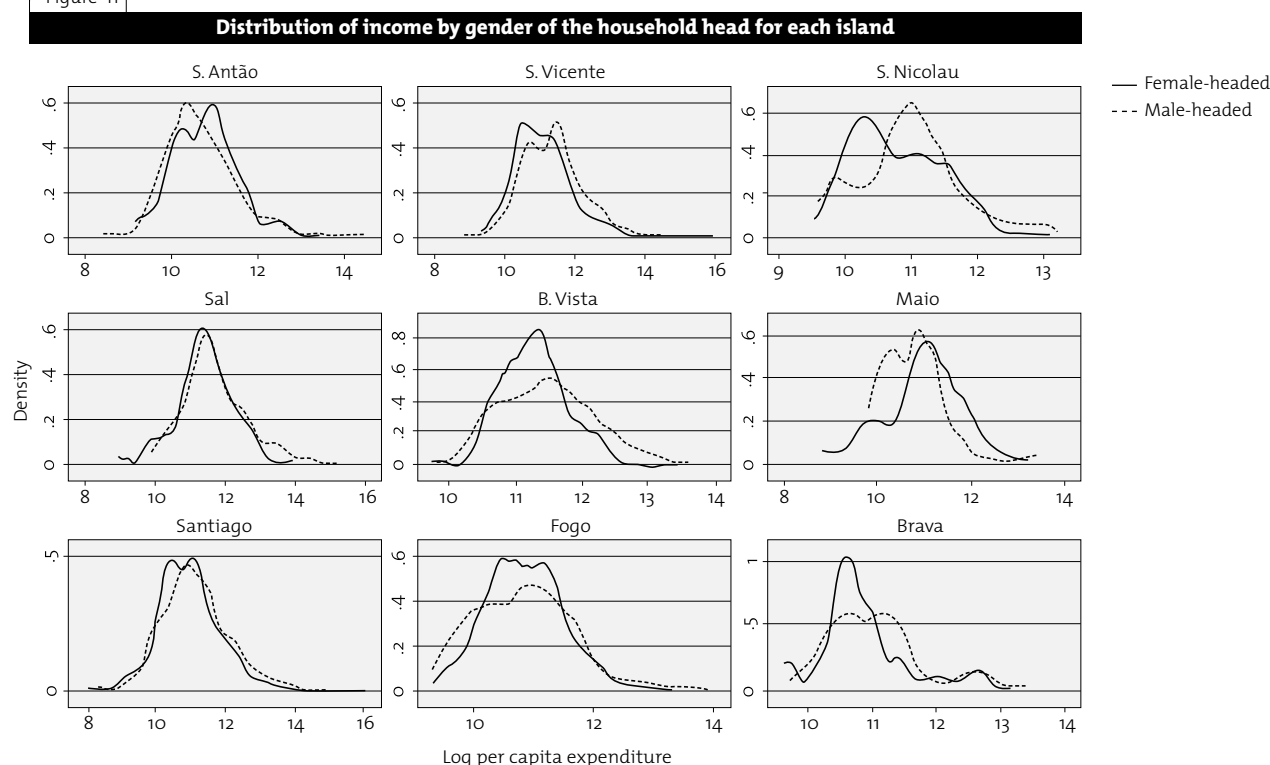
Source: UNCTAD (2011).

The figures by island and gender of the household head show that there are fewer disparities in the distribution of income for male- and female-headed households in Sal and Santiago. São Vicente and São Nicolau show a higher mode for men. In contrast, the mode seems higher for women in São Antão. Boa Vista, Brava, and Fogo show a greater dispersion for the log of per capita expenditure of female-headed households than of male-headed households.

Step 3: Create graphs of the non-parametric regression of log of per capita income, food, and remittances shares

In this step, the study analyses the relationship between two variables through figures derived from non-parametric regressions. The aim is to relate the level of livelihood to the consumption of food and the income from remittances. This allows us to see how changes in food prices or remittances affect different "types" of households.

Non-parametric regressions fit a local relationship between two variables, *y* and *x*. By "local" we mean that separate fitted relationships are obtained at different values of the explanatory variable, *x*. Two commands can be used to do this in Stata: `lpoly` and `lowess`.[47] There are several other methods for running this type of analysis, but we will not discuss them in this material. Here, we will exclusively use `lpoly`, which performs a kernel-weighted local polynomial regression of *y* on *x* and displays a graph of the smoothed values with (optional) confidence bands.[48]

Figure 12 (Figure 6 in the study) shows the non-parametric regressions of the share of food (how

much a household in Cape Verde spends on food as a proportion of its total expenses) on the level of livelihood of the family (proxied by the log of per capita expenditure of the household). The figure has three different panels, one for the whole sample, one for households in urban areas, and one for households in rural areas. In each panel, we are interested in knowing, for each level of income, how much a family with a female household head spends on food compared to how much a male-headed household spends.

The command to create the figure is once again `twoway` and therefore we do not need to explain its syntax. This time, however, each line will not be a density function but a non-parametric regression. We then replace the command `kdensity` with `lpoly`. We regress *share_food* on *logpce*, recognizing once again that each observation in the sample has a different weight in the population by including the option [`w=pondera`], and we perform the regressions for female-headed households (if `female_head==1`) and male-headed households (if `female_head==0`). We also want to eliminate possible outliers, so we remove the first and last observation in the sample (`logpce <16 & logpce>= 9`).

The Stata command for all households is:

```
twoway (lpoly share_food logpce
[w=pondera] if female_head==1 &
logpce <16 & logpce>= 9 ,legend(lab
(1 "Female-headed"))) (lpoly
share_food logpce [w=pondera] if
female_head== 0 & logpce <16 &
logpce>= 9,lpattern(dash) legend
(lab(2 "Male-headed"))),
ytitle(Share of food (ratio))
xtitle(Log per capita expenditure)
```

The Stata command for urban households only is:

```
twoway (lpoly share_food logpce
[w=pondera] if female_head==1 &
area ==1 & logpce <16 & logpce>= 9,
legend(lab(1 "Female-headed")))
(lpoly share_food logpce [w=pondera]
if female_head== 0 & area==1 &
logpce <16 & logpce>= 9,
lpattern(dash) legend(lab(2
"Male-headed"))), ytitle(Share
of food (ratio)) xtitle(Log per
capita expenditure)
```

The Stata command for rural households only is:

```
twoway (lpoly share_food logpce
[w=pondera] if female_head==1
& area == 2 & logpce <14 & logpce>=
9 , legend(lab(1 "Female-headed")))
(lpoly share_food logpce [w=pondera]
if female_head== 0 & area ==2 &
logpce <14 & logpce>= 9,
lpattern(dash) legend(lab(2
"Male-headed"))), ytitle(Share
of food (ratio)) xtitle(Log per
capita expenditure)
```

The outputs of these three commands are the non-parametric regressions panels in Figure 12. These regressions estimate the average food share at different levels of expenditure. The food budget share at the left tail of the income distribution is approximately 45 per cent (in both urban and rural areas). In rural areas, there are some differences between female-headed and male-headed households in the share spent on food. As expected and in accordance with Engel's law, the share spent on food declines with the increase in the level of household well-being. It follows that lower food prices will have a pro-poor bias, while higher food prices will have an anti-poor bias when considering the consumption effect on welfare.

Figure 12

Share of food and level of livelihood

(a) National

Female-headed ——
Male-headed - - -

Log per capita expenditure

Figure 12

Share of food and level of livelihood

(b) Urban

—— Female-headed
--- Male-headed

(c) Rural
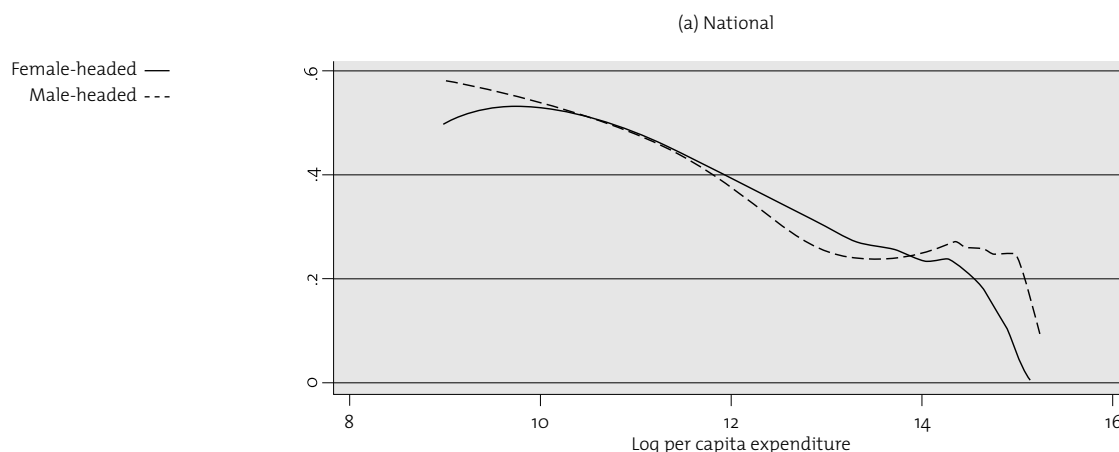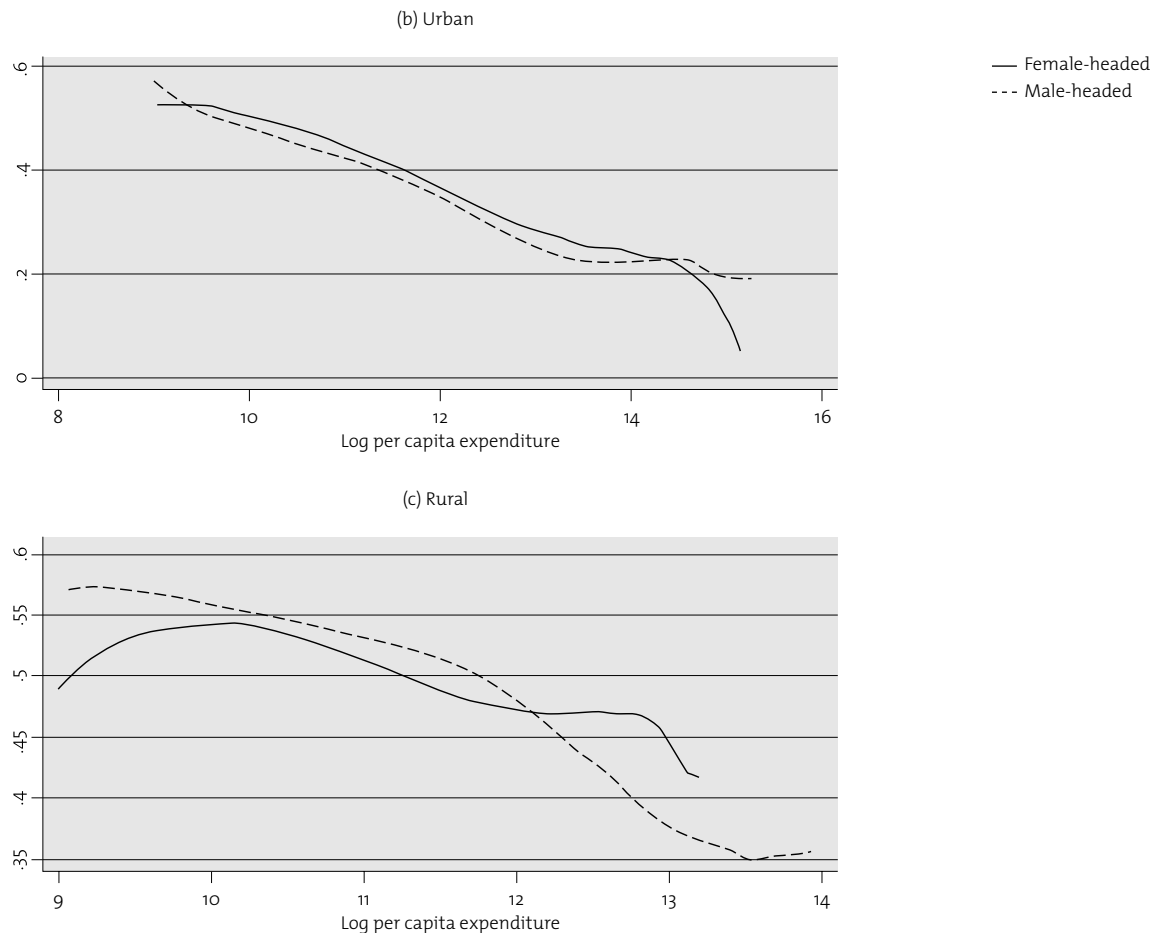
Source: UNCTAD (2011).

As the study notes, "Looking at the difference between female- and male-headed households does not necessarily capture all the impacts on women. [...] Women living in male-headed households can also benefit from lower food prices. This is explored by examining the relationship between the food share on the one hand and the total number of females in the household or alternatively the share of females in the household on the other." UNCTAD (2011: 40).

To this end, we need to create the variables capturing the number of females in a household and the proportion of women in a household. To do that, we create an auxiliary variable *aux* that takes the value 1 if the member of the household is a woman and 0 otherwise. We then create a new variable *female* with the command egen to count how many women live in that particular household (*houseid*). The syntax of the command is egen *<new variable>= <function>(<expression(s)>* or *<variable(s)>)* [, by *(<variables>)*]. The functions actually determine what the command egen will do. There are many functions, all described in the manual.[49] We also need to create a variable (*sfemale*)

indicating the share of women in a household. The Stata commands are as follow:

```
gen aux = 1 if male == 0
egen female = sum(aux), by(houseid)
drop aux
gen sfemale = female / hsize
```

The variable *hsize* stands for the number of household members. As before, non-parametric regressions are estimated and reported in Figure 13 (Figure 7 in the study) according to the area where the household is located. The Stata commands are similar to the ones used before. Specifically, for plotting the relationship between the share of food and the number of females living in a household we use the following command:

```
twoway (lpoly share_food female
[w=pondera] if head == 1 &
female<11, msymbol(none) legend(lab
(1 "National")))(lpoly share_food
female [w=pondera] if head ==
1 & area == 2 & female<11,
lpattern(dash) msymbol(none)
legend(lab(2 "Rural"))) (lpoly
```

```
share_food female [w=pondera] if
head == 1 & area == 1 & female<11,
lpattern(dash_dot) msymbol(none)
legend(lab(3 "Urban"))), ytitle(Share
of food (ratio)) xtitle(Number of
females)
```

For plotting the relationship between the share of food and the share of females living in a household, we use the following command:

```
twoway (lpoly share_food sfemale
[w=pondera] if head == 1 ,
msymbol(none) legend(lab(1
```

```
"National")))(lpoly share_food
sfemale [w=pondera] if head == 1 & area
== 2, lpattern(dash) msymbol(none)
legend(lab(2 "Rural"))) (lpoly
share_food sfemale [w=pondera] if
head == 1 & area == 1, lpattern
(dash_dot) msymbol(none) legend(lab
(3 "Urban"))), ytitle(Share of food
(ratio)) xtitle(Share of females
(ratio))
```

The option msymbol indicates the marker symbol that should be used in the figure.

Figure 13

Share of food and women in the household

(a) Share of food and number of women in the household

(b) Share of food and share of women in the household
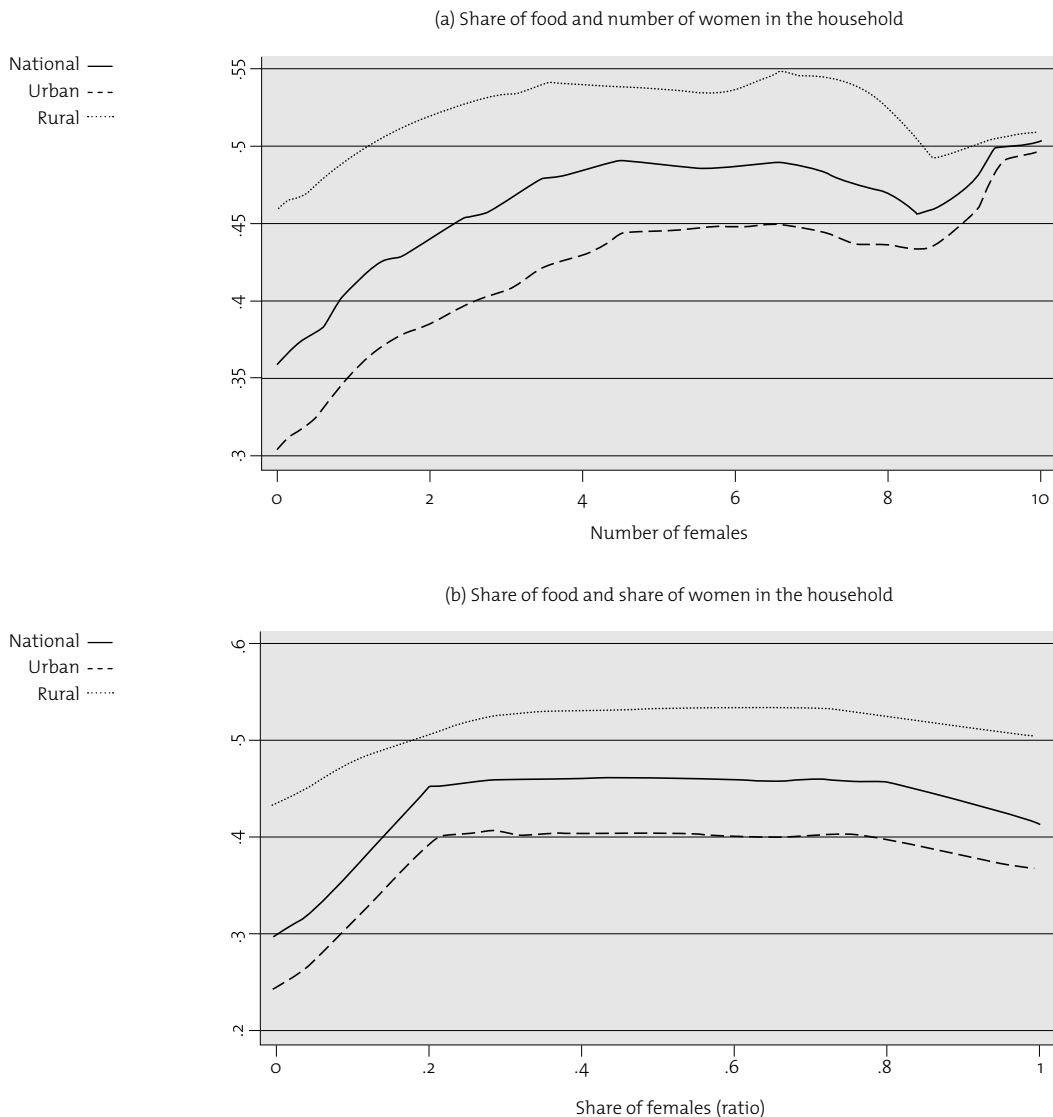
Source: UNCTAD (2011).

What we can see from the results in Figure 13 is that, in principle, households with more women tend to allocate a slightly higher share of their expenditure to purchase food, especially in rural areas, and thus these households will enjoy higher gains from lower food prices.[50] Note, however, that there are few differences in food shares for different gender structures (share of females) in Cape Verdean households.

The study then follows the same methodology to estimate the relationship between the share of income the household gets from remittances and the level of expenditure, again by area and by gender of the household head. The analysis is also carried out for the number and share of females in the household. The structure of the Stata command is very similar to the one used before. In particular, at the national level, the command to plot the figure is the following:

```
twoway (lpoly part_rem logpce
[w=pondera] if head == 1 &
female_head == 1, msymbol(none)
legend(lab(1 "Female-headed")))
(lpoly part_rem logpce [w=pondera]
if head == 1 & female_head == 0,
lpattern(dash) msymbol(none)
legend(lab(2 "Male-headed"))) ,
ytitle(Share of remittances (ratio))
xtitle(Log per capita expenditure)
```

Panel (a) in Figure 14 (Figure 9 in the study) presents the non-parametric regressions at the national level and shows that the share of remittances in total income is always higher for female-headed than for male-headed households, except for very poor households. For females, the share increases sharply with the level of livelihood until reaching values higher than 15 per cent of total income, then decreasing to less than 5 per cent of total income.

To plot the figure for households living in urban areas only, the command is the following:

```
twoway (lpoly part_rem logpce
[w=pondera] if head == 1 &
female_head == 1 & area == 1,
```

```
msymbol(none) legend(lab(1
"Female-headed"))) (lpoly part_rem
logpce [w=pondera] if head == 1 &
female_head == 0 & area == 1,
lpattern(dash) msymbol(none) legend
(lab(2 "Male-headed"))), ytitle(Share
of remittances (ratio)) xtitle(Log
per capita expenditure)
```

Panel (b) shows that in urban areas the share of remittances is higher for female-headed households at the left (poorest) tail of the distribution and in the middle, but the shares seem to converge at the richest tail.

To plot the figure for households living in rural areas only, the command is the following:

```
twoway (lpoly part_rem logpce [w=
pondera] if head == 1 & female_head
== 1 & area == 2, msymbol(none)
legend(lab(1 "Female-headed")))
(lpoly part_rem logpce[w=pondera] if
head == 1 & female_head == 0 & area
== 2, lpattern(dash) msymbol(none)
legend(lab(2 "Male-headed"))),
ytitle(Share of remittances (ratio))
xtitle(Log per capita expenditure)
```

As shown in panel (c), in rural areas the share of remittances in total income is low for the poorest households but increases sharply as the level of income increases. This analysis reveals that remittances are an important source of income, and more so for female-headed households than for male-headed households, reaching more than 30 per cent of the income of the richest rural households.

Figure 14

Share of remittances and level of expenditure

(a) National

Share of remittances and level of expenditure

(b) Urban

Female-headed ——
Male-headed - - -

Log per capita expenditure

(c) Rural

Female-headed ——
Male-headed - - -

Log per capita expenditure
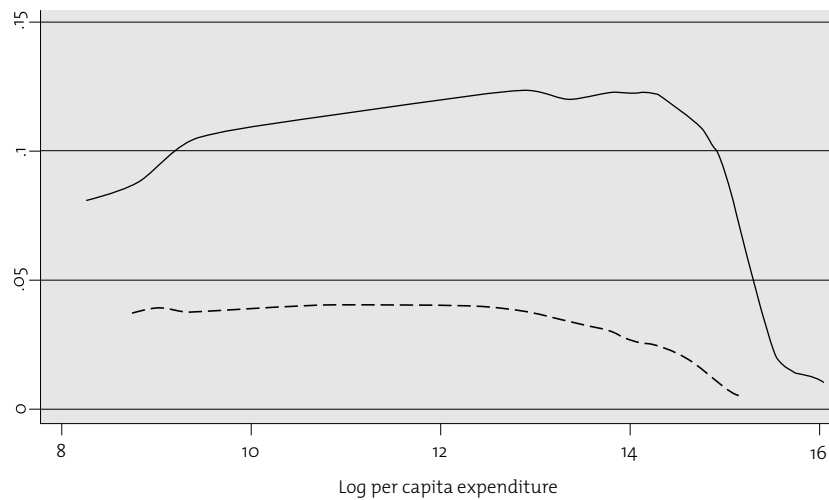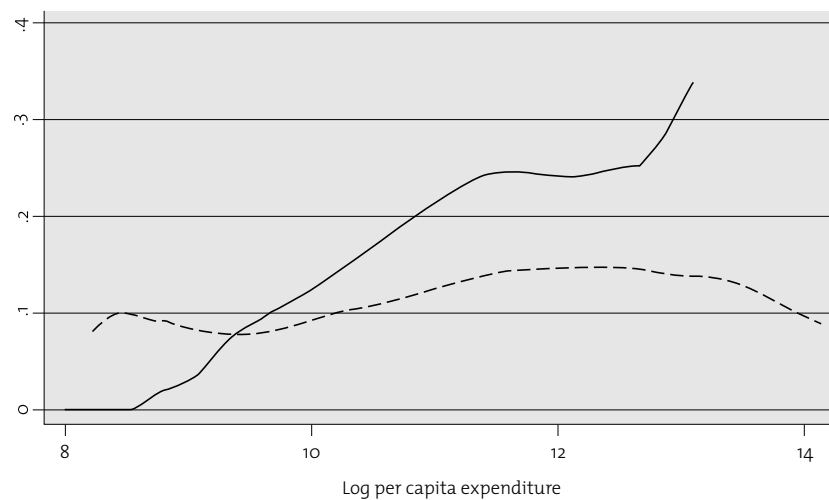
Source: UNCTAD (2011).

Finally, the study plots the relationship between the share of remittances in total income and the number of females living in the household, and the share of females in the total number of household members according to the area where the household is located (Figure 10 in the study, not displayed here). The code lines are, respectively:

```
twoway (lpoly part_rem female
[w=pondera] if head == 1 &
female<11,msymbol(none)legend(lab(1
"National")))(lpoly part_rem female
[w=pondera] if head == 1 & area == 1
& female<11, lpattern(dash) msymbol
(none) legend(lab(2 "Urban")))
(lpoly part_rem female [w=pondera]
if head == 1 & area == 2 & female<11,
lpattern(dash_dot) msymbol(none)
legend(lab(3 "Rural"))) , ytitle
(Share of remittances (ratio))
xtitle(Number of females)
```

```
twoway (lpoly part_rem sfemale
[w=pondera] if head == 1 , msymbol
(none) legend(lab(1 "National")))
(lpoly part_rem sfemale [w=pondera]
if head == 1 & area == 1, lpattern(dash)
msymbol(none) legend(lab(2 "Urban")
))(lpoly part_rem sfemale [w=pondera]
if head == 1 & area == 2, lpattern
(dash_dot) msymbol(none) legend(lab
(3 "Rural"))), ytitle(Share of
remittances (ratio)) xtitle(Share
of females (ratio))
```

Step 4: Plot welfare gains in simulated scenarios

Our last task is to study the welfare effects of different scenarios. The first simulation is an increase of 20 per cent in remittances. We want to see how this would affect household per capita income for different types of households. We will need to create a few new variables. First, we cre-

ate the log of per capita income using the same command `gen` we employed to create the logarithm of per capita expenditure (*logpce*):

```
gen lipcf = log(ipcf)
```

Second, we generate a variable that would be the new household per capita income simulated, that is, the initial income plus an increase of 20 per cent of the part of the income that comes from remittances:

```
gen ipcf_sim = ipcf + 0.2 *
part_rem * ipcf
```

Third, we create the logarithm of the simulated income in the following way:

```
gen lipcf_sim = log(ipcf_sim)
```

Finally, we generate a new variable *gain* as the difference between the simulated and the original income, as follows:

```
gen gain = lipcf_sim - lipcf
```

We are now ready to create Figure 15 (Figure 12 in the study) using the command `twoway` and the same syntax as before. We are also eliminating potential outliers by restricting the analysis to households with a logarithm of per capita expenditure below 15. The code is the following:

```
graph twoway (lpoly gain logpce
[w=pondera] if female_head == 1 &
logpce <15 , legend(lab(1
"Female-headed"))) (lpoly gain logpce
[w=pondera] if female_head == 0 &
logpce <15 , lpattern(dash) legend
(lab(2 "Male-headed"))), ytitle("Gain
(ratio)",size(vsmall)) xtitle("Log
per capita expenditure",size(small))
legend(size(small)) by(area,
cols(2)) ylabel(,labsize(small))
xlabel(,labsize(small))
```

As shown in Figure 15, the effect would be significant, in particular for female-headed households both in urban and rural areas. While the effect is more or less constant in urban areas, it increases according to the level of income in rural areas.

Figure 15

Welfare gains from increase in remittances by gender of the household head and area
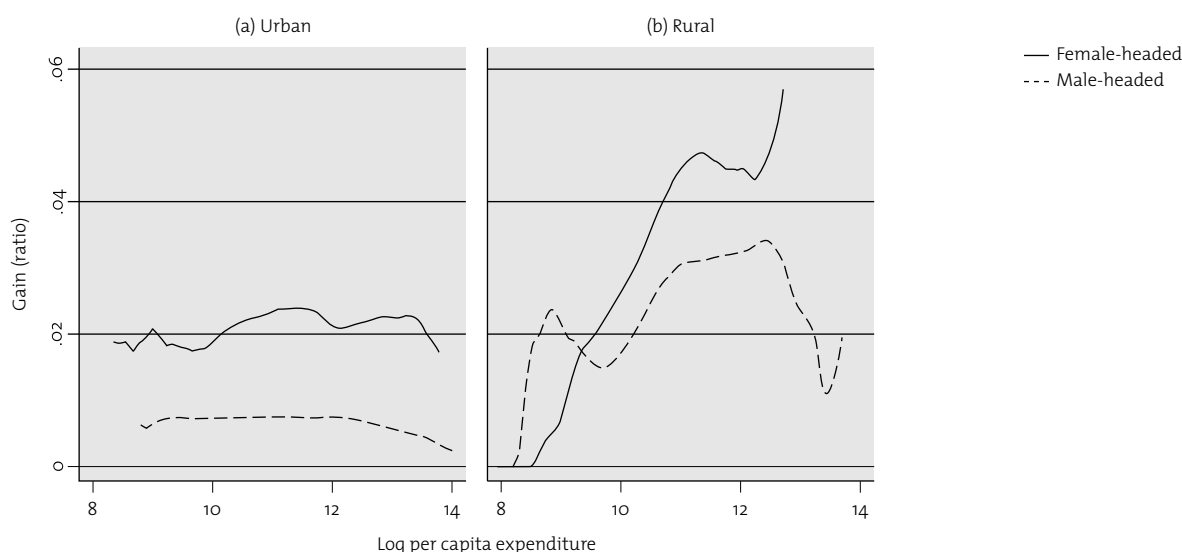
Source: UNCTAD (2011).

We now move to the analysis of the gains from an increase in tourism. This simulation is more complex than the previous one because tourism generates income not only for those who work directly in this sector but also for those who provide services such as transport and communications, or who work in the retail sector. Therefore, we should consider simulations with both direct and indirect effects of tourism on workers in Cape Verde. We will follow four steps to do these estimations.

Step 1: We generate a new dataset that contains the original dataset twice.[51] This is done by attaching the dataset to the original dataset again. We generate a variable called *case* and give it the value 1 for all households in the original dataset, and in the repeated dataset we assign the values 2 for households in rural areas and 3 for households in urban areas to the variable *case*.

```
use "`base_in_1'", clear
gen case = 1
append using "`base_in_1'"
replace case = 2 if case ==.
& area == 2
replace case = 3 if case ==.
& area == 1
```

Step 2: We define labels also for graphical reasons as follows:

```
label define case1 1 "National" 2
"Rural" 3 "Urban"
label values case case1
label define female 0 "Male-headed" 1
"Female-headed"
label values female_head female
```

Step 3: We define the variables of interest using the command gen as follows:

```
gen pce = des_indi
gen lpce = log(pce)
```

We generate variables recording the activity of the household head for tourism, retail trade (commerce), and transport and communications. The variable *actividad* in the data records the economic sector, with 7 being tourism, 6 retail trade (trade), and 9 transport and communications (transport).[52]

```
* Tourism

gen aux_head_tourism = 0
replace aux_head_tourism = 1 if
actividad == 7 & head ==1
egen head_tourism =
sum(aux_head_tourism), by(houseid)
replace head_tourism = 1 if
head_tourism == 2

* Trade

gen aux_head_trade = 0
replace aux_head_trade = 1 if
actividad == 6 & head ==1
egen head_trade =
sum(aux_head_trade), by(houseid)
replace head_trade = 1 if
head_trade == 2

* Transport

gen aux_head_transport = 0
replace aux_head_transport = 1
if actividad == 9 & head ==1
egen head_transport =
sum(aux_head_transport), by(houseid)
```

```
replace head_transport = 1 if
head_transport == 2
```

head_tourism, head_trade and *head_transport* are thus dummy variables that take the value 1 if the household head works in tourism, trade, or transport and communications, or 0 otherwise.

Step 4: As you can see in the study, the idea is to simulate the welfare effects of an increase in income from tourism and related activities (retail trade, and transport and communications). In particular, the country case study proposes different scenarios (cases).

Case 1: Tourism 30 per cent (called Tourism in the labels of the figures)

30 per cent increase in per capita expenditure (*pce*) of households with the head employed in tourism. The code to build the variable corresponding to the gain obtained by households in this first case (*gain1*) is the following:

```
gen pce_sim1 = pce
replace pce_sim1 = pce * 1.3 if
head_tourism == 1
gen lpce_sim1 = log(pce_sim1)
gen gain1= lpce_sim1- lpce
```

Case 2: Tourism 30 per cent and trade 10 per cent

10 per cent increase in *pce* of households with the head employed in retail trade. The code to build the variable corresponding to the gain obtained by households in this second case (*gain2*) is the following:

```
gen pce_sim2 = pce_sim1
replace pce_sim2 = pce_sim2 * 1.10
if head_trade == 1
gen lpce_sim2 = log(pce_sim2)
gen gain2= lpce_sim2- lpce
```

Case 3: Tourism 30 per cent, trade 10 per cent, and transport 10 per cent

10 per cent increase in *pce* of households with the head employed in transport and communications. The code to build the variable corresponding to the gain enjoyed by households in this third (*gain3*) case is the following:

```
gen pce_sim3 = pce_sim2
replace pce_sim3 = pce_sim3 * 1.10
if head_transport == 1
gen lpce_sim3 = log(pce_sim3)
gen gain3= lpce_sim3- lpce
```

We are now ready to generate Figures 16 and 17 (Figures 13 and 14 in the study). For Figure 16 (Figure 13 in the study), we calculate the gains for each of the three cases described above. We are also interested in finding out whether there are any differences in the distribution of gains between rural and urban households. The Stata command to perform these simulations is similar to those presented before:

```
graph twoway (lpoly gain1 lpce
[w=pondera] if lpce<15, legend(lab(1
"Tourism"))) (lpoly gain2 lpce
[w=pondera]if lpce<15,lpattern(dash)
```

```
legend(lab(2 "Tourism and Trade")))
(lpoly gain3 lpce [w=pondera] if
lpce<15 , lpattern(dash_dot) legend
(lab(3 "Tourism, Trade and
Transport"))) , ytitle("Gain
(ratio)",size(vsmall)) xtitle("Log
per capita expenditure",size(vsmall))
legend(size(vsmall)) by(case,
cols(3)) ylabel(,labsize(vsmall))
xlabel(,labsize(vsmall))
```

To eliminate any outliers, we only consider households with a logarithm of per capita expenditure smaller than 15.

Figure 16

Welfare gains from tourism by area
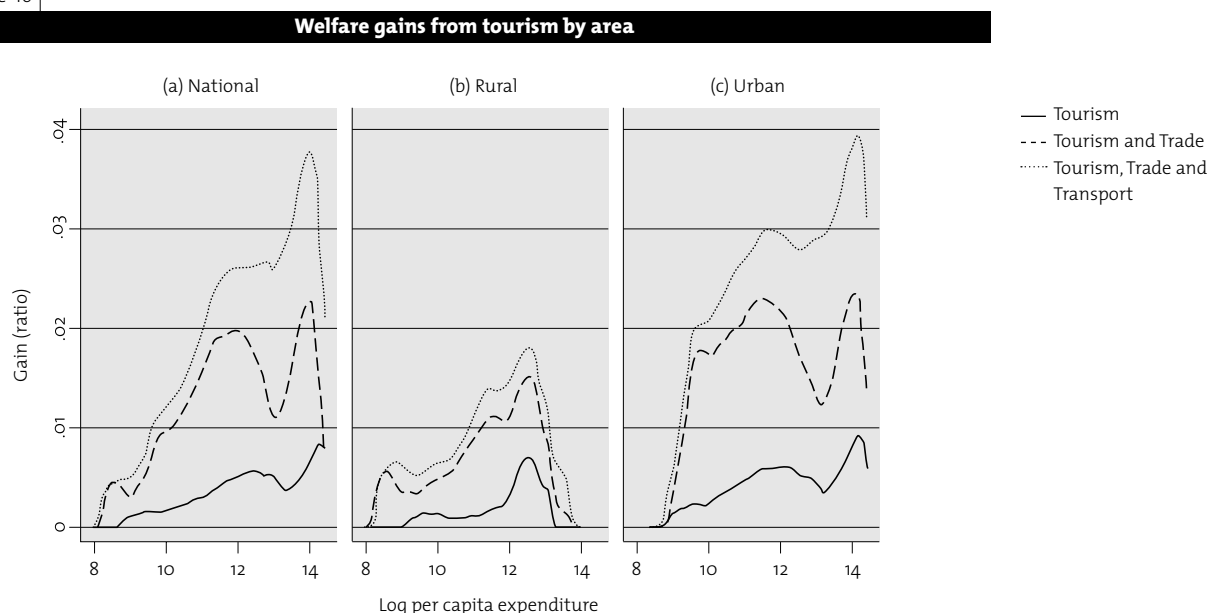
Source: UNCTAD (2011).

The results in Figure 16 show that the gains mostly increase with the level of livelihood and are larger in urban than in rural areas. The panels also show that the effects can be important when we take into account both the direct and indirect effects of tourism.

Figure 17 (Figure 14 in the study) explores whether the expected gains are different depending on the gender of the household head (we find that they are larger for female-headed households). For that reason, we run the non-parametric regression for the whole sample (case==1) by gender of the head of the household (by female_head). The Stata command is as follows:

```
graph twoway (lpoly gain1 lpce
[w=pondera] if case == 1 & lpce<15,
legend(lab(1 "Tourism"))) (lpoly
gain2 lpce [w=pondera] if case == 1
& lpce<15, lpattern(dash) legend
(lab(2 "Tourism and Trade")))(lpoly
gain3 lpce [w=pondera] if case == 1
& lpce<15, lpattern(dash_dot)
legend(lab(3 "Tourism, Trade and
Transport"))) , ytitle("Gain
(ratio)",size(vsmall)) xtitle("Log
per capita expenditure",size(small))
legend(size(small)) by(female_head,
cols(2)) ylabel(,labsize(small))
xlabel(,labsize(small))
```

Figure 17

Welfare gains from tourism by gender of the household head

Tourism ——
Tourism and Trade - - -
Tourism, Trade and ·······
Transport

(a) Male-headed (b) Female-headed

Log per capita expenditure
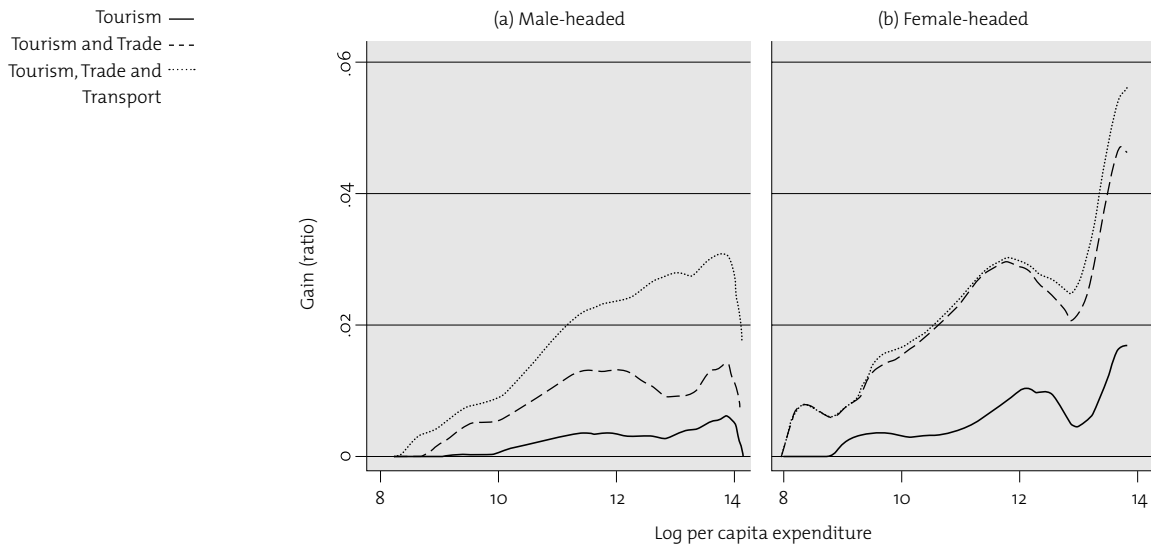
Source: UNCTAD (2011).

Figures 15 (a) and 15 (b) in the study (not displayed here) run the same non-parametric regressions of the gains by gender of the household head and for urban and rural areas, as follows:

```
graph twoway (lpoly gain1 lpce
[w=pondera] if case == 1 & lpce<15 &
area == 1, legend(lab(1 "Tourism")))
(lpoly gain2 lpce [w=pondera] if
case == 1 & lpce<15 & area == 1,
lpattern(dash) legend(lab(2 "Tourism
and Trade")))(lpoly gain3 lpce
[w=pondera] if case == 1 & lpce<15
& area == 1, lpattern(dash_dot)
legend(lab(3 "Tourism, Trade and
Transport"))) , ytitle("Gain
(ratio)",size(small)) xtitle("Log
per capita expenditure",size(small))
legend(size(small)) by(female_head,
cols(2)) ylabel(,labsize(small))
xlabel(,labsize(small))
```

```
graph twoway (lpoly gain1 lpce
[w=pondera] if case == 1 & lpce<15 &
area == 2, legend(lab(1 "Tourism")))
(lpoly gain2 lpce [w=pondera] if
case == 1 & lpce<15 & area == 2,
lpattern(dash) legend(lab(2 "Tourism
and Trade")))(lpoly gain3 lpce
[w=pondera] if case == 1 & lpce<15
& area == 2, lpattern(dash_dot)
legend(lab(3 "Tourism, Trade and
Transport"))) , ytitle("Gain
(ratio)",size(small)) xtitle("Log
per capita expenditure",size(small))
legend(size(small)) by(female_head,
cols(2)) ylabel(,labsize(small))
xlabel(,labsize(small))
```

6.4 Discussion of findings and limitations of the analysis

By focusing on the second step of the two-step methodology (see Section 3 of this module), this study uses the microeconomic approach to analyse the potential welfare impact of further trade liberalization in Cape Verde. The study looks both at the household consumption impact of trade-induced changes in the prices of goods and the employment/income changes from the potential increase in remittances and tourism. The simulations find that the price effects would be more strongly felt in rural than in urban areas, but that the differences between male- and female-headed households from food price changes would only be marginal. On the other hand, the simulated increase in remittances and income from tourism seems to favour female-headed households. Female-headed households located in rural areas would benefit in particular from increases in remittances, while those located in urban areas would gain more from increased income from tourism. While the reduction of prices has a pro-poor impact, the increase in remittances and tourism would benefit mostly middle- and high-income families.

One limitation of this analysis is that it does not estimate the effects that trade liberalization would have on food prices, remittances, and tourism. The analysis assumes those changes and then studies the welfare impact, leaving open the question of the amount of gains there are for Cape Verde, which is an important question for policymakers. Another limitation of the analysis is that, due to the lack of data, it is limited to a comparison between female- and male-headed

households, thus overlooking any potential in-tra-household reallocation effect. Male-headed households often have females living in the household, and it is therefore important to see how a trade shock affects the distribution of tasks within the household (e.g. whether domestic household activities are distributed fairly between males and females) and resources among various competing categories of expenditure (e.g. how much the household spends on food and on education).

7 Conclusions

This module has introduced the microeconomic approach (household and market channel) to study the welfare effect of trade policy at the household level and identify different gender outcomes. The approach follows the trade and poverty literature, where we first estimate the effects of trade on domestic prices and remuneration and then use those changes to estimate the welfare impact at the household level. For this purpose, this module uses information collected by household surveys. As these surveys often contain information about the gender of the household head and other members of the family, we can use this information to depict potential gender-differentiated effects. In the application we reviewed, the emphasis is on the effects of trade on earnings and expenditure, but a similar methodological approach could be used, for instance, to estimate health and education outcomes of trade policy.

The methodology reviewed in this module uses a specific definition of trade that refers to trade policy or reform and trade facilitation, as explained in Box 1 of Module 1 of this volume. The methodology can be used to analyse welfare effects of bilateral, regional, or multilateral trade agreements. Trade policy and reform can be measured, for example, by changes in tariffs. We reviewed the main sources of trade data in Module 1 of this volume. Additionally, we can analyse

more general trade costs such as those linked to trade procedures, transportation, availability of infrastructure, and access to credit or inputs. When data are available, this methodology also makes it possible to study price transmission issues, including those related to market structure and competition that may have an effect on the degree of pass-through of international prices to domestic prices. The analysis can be applied both *ex ante* and *ex post*. In *ex-post* analysis, we compare various outcomes (wages, incomes, employment, expenditures, etc.) before and after episodes of trade liberalization or reform. We can sometimes distinguish those effects by the gender of the individual or the head of the household, but that often depends on the content of the dataset. In *ex-ante* analysis, we work with a two-step methodology, examining first the likely transmission effects from trade to prices, and then the effects from prices to households and/or firms. As described in Module 1, there are various open-source tools you can employ to conduct an *ex-ante* analysis. Again, how much we can say about differential gender effects may depend on the available data.

The method used in this module is based on non-parametric econometric techniques. The advantage of non-parametric analysis compared to the more traditional regression analysis is that you do not have to make any assumption regarding the relationship between the independent variable and the dependent variable. In other words, you let the data choose the best shape of the functional form. Additionally, in welfare analysis, the non-parametric approach allows you to look at the distributional impact of trade policy along the entire income distribution of households. Trade policy usually has different indirect effects on rich and poor households, since they have different consumption baskets and sources of income. For example, the case study of Cape Verde shows that, overall, the share of household income spent on food declines with the increase in the level of household well-being.

ANNEX

Annex A   Modelling price changes[53]

This annex presents three ways of modelling *ex-ante* price changes: (a) econometric estimation within a model (homogeneous goods); (b) simulation models (heterogeneous goods across countries); and (c) CGE models.

A1   Econometric estimation within a model

The first methodology to study price changes in an open economy setting was the model by Hoekman *et al.* (2005). These authors combined a simple structural model with parameter estimation. The model is used to estimate the welfare effects of the Doha Round of multilateral trade negotiations: it has a multi-country and multi-product setting and allows for estimations of changes in prices of more than 5,000 products. Based on estimates of demand and supply elasticities, the authors use their structural model to solve for equilibrium prices of agricultural products following a change in tariff rates.

The import demand and export supply functions are given by $m_c (p_c, Z_c^m)$ and $x_c (p^*, Z_c^x)$, where m_c and x_c are the import demand and supply vectors, respectively, for country c across all goods; p_c is the domestic price vector of imported goods in country c; p^* is the vector of world prices; and Z_c^m and Z_c^x represent matrices of exogenous variables that determine imports and exports, respectively, in country c. Each good g is homogeneous across all countries, but within each country, it is an imperfect substitute for all other traded goods.

World markets for each good clear so that $\sum_c m_c (p_c, Z_c^m) - \sum_c x_c (p^*, Z_c^x) = 0$ and the solution with respect to p* yields equilibrium world prices. To see how this works, assume that all world and domestics markets are perfectly competitive so that $p_c = p^* . \tau_c$, where τ_c is a vector of all goods of the form $\tau_{c,g} = (1 + t_{c,g})$, in which $t_{c,g}$ is the level of protection in country c for good g. We obtain world prices $\hat{p}^*$ by solving:

$$\hat{p}^* = \left[\sum_c E_c^x - \sum_c E_c^m \right]^{-1} \sum_c E_c^m \hat{\tau}_c \qquad (A1)$$

where E_c^x and E_c^m are square matrices in which elements are equal to the elasticity of import supply and demand (respectively) in country c, multiplied by the share in world trade of each country's imports of good g. We can use equation (A1) to estimate the price effects of any change in trade protection in the local economy or abroad.

Balat *et al.* (2007) use this methodology to estimate the effects of the Doha Round on Zambia. They follow the two-step methodology of the trade and poverty literature. In the first step, they look at price changes resulting from the implementation of the Doha Round agreement for the main crops produced and consumed in Zambia. They find that cotton prices will increase by 3.5 per cent, hybrid maize prices will increase by almost 4 per cent, and tobacco prices will increase by 1.3 per cent. They find modest price reductions in vegetables and groundnuts. In the second step, given these price changes, the authors explore the net effect of trade liberalization at the household level, using Zambia's 1998 Living Conditions Monitoring Survey.

The average budget share spent on food consumption is high, with poor households dedicating more than 70 per cent and non-poor households almost 60 per cent of their budget to food (see Table 5). The estimated price increases from the Doha Round agreement would therefore certainly have a negative impact on consumers. However, the overall effect needs to be carefully assessed, as most households are both consumers and producers of agricultural goods. Therefore, it is also important to consider the production effect, which in the case of Zambia will operate through the 10 per cent (5 per cent) of income derived from sales of poor (non-poor) households (see shares of food and non-food crop sales in Table 6).[54]

Finally, the authors calculate the net welfare effect of price increases (Table 7). The consumption effect is negative for every decile, particularly for the poorest households, given that a large share of their expenditure is spent on food. The income effect is positive, but not strong enough to overcome the negative impact on consumption (except for households in the sixth decile). It appears therefore, that the Doha agreement would have on average a negative welfare effect in Zambia.

Table 5

Average household budget shares in Zambia, 1998 (per cent)			
	Total	**Poor**	**Non-poor**
Food	67.5	71.8	59.3
Clothing	5.6	4.8	7.1
Alcohol & tobacco	3.6	2.9	4.9
Personal goods	7.1	6.8	7.6
Housing	4.5	4.2	5.0
Education	2.5	2.6	2.3
Health	1.4	1.3	1.6
Transport	4.2	3.2	5.9
Remittances	1.3	0.7	2.4
Other	2.4	1.7	3.9

Source: Balat *et al.* (2007).

Table 6

Sources of household income in Zambia, 1998 (per cent)			
	Total	**Poor**	**Non-poor**
Own production	28.3	33.3	19.1
Sales of food crops	6.3	7.6	3.8
Sales of non-food crops	2.5	3.0	1.3
Livestock & poultry	5.5	6.8	2.9
Wages	20.8	14.4	32.9
Non-farm	22.3	20.9	24.9
Remittances	4.9	5.0	4.8
Others sources	9.5	9.0	10.3

Source: Balat *et al.* (2007).

Table 7

Average household welfare effects in Zambia, 1998 (per cent)			
Income decile	**Consumption**	**Income**	**Total**
1	-1.44	0.84	-0.60
2	-1.26	0.78	-0.48
3	-1.22	0.37	-0.85
4	-1.14	0.75	-0.39
5	-1.04	0.72	-0.32
6	-0.93	0.96	0.03
7	-0.87	0.67	-0.20
8	-0.77	0.23	-0.54
9	-0.66	0.30	-0.36
10	-0.49	0.17	-0.32
Total	-0.98	0.58	-0.40

Source: Balat *et al.* (2007).

A2 Global Simulation Model

An additional modelling framework that can be used to study price changes is the GSIM developed by Francois and Hall (2009). The authors build a partial equilibrium framework of global trade policy changes at the industry level. This model can be used for the analysis of global, regional, and unilateral trade policy changes. One of the differences from the previous model is that here the authors assume imperfect substitutability across goods.

The WITS platform (mentioned above) has a computational package that can be used to estimate price changes, and combines modelling with data analysis. The simulation package requires the in-put of bilateral trade volumes for key partners in the agreement and the rest of the world, together with key export supply elasticities, import demand elasticities, and elasticities of substitution.

Porto (2007) uses the GSIM methodology to study the effects of the Central America-Dominican Republic-United States Free Trade Agreement (CAFTA) in Guatemala. The paper uses the two-stage approach. First, the model is used to estimate the effect of CAFTA on prices of several products. For instance, the author finds that following the agreement, yellow maize prices would decrease by 19 per cent, while prices of white maize would decrease by 38 per cent and prices of chicken by 18.6 per cent. In the second stage, the author uses the estimated price changes

and household data for Guatemala[55] to study the welfare changes. Food is the most important expenditure for all Guatemalan households, and the average food budget share is particularly high for poor and indigenous households (almost half of the budget, as shown in Table 8). Here, price decreases of food staples would be positive for consumers, but again, the net effect would also depend on the income effect, as some households are producers of agricultural goods.

Table 8

	Average household budget shares in Guatemala – total and indigenous population, 2000 (per cent)					
	Total			**Indigenous**		
Household budget item	**Total**	**Poor**	**Non-poor**	**Total**	**Poor**	**Non-poor**
Food	35.3	45.4	26.7	42.4	47.0	32.8
Clothing	12.7	13.4	12.1	14.7	15.2	13.8
Alcohol & tobacco	0.04	0.06	0.02	0.01	0.01	0.01
Personal goods	21.2	19.4	22.8	19.7	18.8	21.6
Housing	3.3	1.9	4.6	2.1	1.8	2.7
Education	3.8	2.9	4.6	2.7	2.3	3.4
Health	5.5	4.0	6.8	4.1	3.4	5.6
Transport	10.5	6.0	14.3	7.9	5.4	13.1
Remittances	0.28	0.07	0.45	0.13	0.03	0.34
Other	7.3	6.9	7.7	6.3	6.1	6.8

Source: Porto (2007).

Table 9 shows that own production is an important source of income for indigenous people (31 per cent of income) and the effect of price decreases will operate through sales of food crops by households. This negative production effect should in principle be a source of concern. However, as seen in Table 10, the net effect would be positive on average for all households, because the positive consumption effect would be larger than the negative production effect across all levels of income. In conclusion, according to this analysis, the CAFTA agreement is expected to have a positive effect on the welfare of Guatemalan households.

Table 9

	Average household income shares in Guatemala – total and indigenous population, 2000 (per cent)					
	Total			**Indigenous**		
Income source	**Total**	**Poor**	**Non-poor**	**Total**	**Poor**	**Non-poor**
Own production	18.1	30.1	7.9	30.8	36.6	18.6
Sales of food crops	7.1	10.3	4.3	10.3	11.3	8.0
Sales of non-food crops	0.1	0.1	0.1	0.2	0.1	0.3
Livestock and poultry	0.9	0.9	0.9	0.9	0.8	1.0
Wages	45.2	39.0	50.5	34.6	32.7	38.5
Income non-farm	18.2	13.9	21.9	16.6	13.8	22.5
Remittances	5.1	3.2	6.7	3.9	3.0	5.7
Other sources	5.2	2.4	7.6	2.9	1.7	5.3

Source: Porto (2007).

Table 10

	Welfare effects of CAFTA in Guatemala, 2000 (per cent)						
Income decile	**Consumption effects**				**Income effects**		**Net total effects**
	Vegetables	**Cereals**	**Fruits**	**Total**	**Maize**	**Total**	
1	1.17	2.92	1.14	5.23	-3.52	-3.41	1.82
2	0.58	4.38	1.09	6.05	-4.38	-4.24	1.81
3	0.43	4.19	1.23	5.85	-4.17	-4.1	1.75
4	0.42	4.35	1.1	5.87	-4.62	-4.56	1.31
5	0.37	3.82	0.9	5.09	-3.62	-3.49	1.6
6	0.33	4.03	1.05	5.41	-4.06	-3.92	1.49
7	0.33	4.07	0.87	5.27	-4.24	-4.08	1.19
8	0.27	3.47	0.95	4.69	-3.39	-3.33	1.36
9	0.28	2.33	0.7	3.31	-2.66	-2.57	0.74
10	0.07	3.66	0.19	3.92	-3.27	-3.16	0.76
Total	0.425	3.722	0.922	5.069	-3.793	-3.686	1.383

Source: Porto (2007).

A3 Computable general equilibrium model

Finally, another modelling alternative is the CGE model. CGE models are built to represent a given economy (region, country, group of countries) and assume optimizing behaviour by agents (firms, consumers). They not only apply market-clearing conditions as in the previous models – i.e. partial equilibrium models – but also deal with government and household budget constraints, labour market decisions, profit maximization, and other features. Data are used to infer ("calibrate" in technical terms) the parameters of the model in order to obtain an accurate representation of the economy under study. Concerning resulting price changes, a key feature of CGE models is that these results embody not only the direct price effects of the trade policy change, but also "second-round" indirect effects on the prices of non-traded goods and on factor returns, including effects operating through the government's budget constraint. The solution of the model and its comparative statics provide predictions of the change in variables, such as prices, output, and economic welfare resulting from a change in a tariff, for instance. These price changes can then be used with household survey data to analyse the welfare impact of trade shocks.

An example of this type of methodology is Chen and Ravallion (2004). This paper evaluates the welfare effects of price changes in goods and factors following the accession of China to the WTO, using the GTAP model.[56] Table 11 shows the price changes reported by the model over the period 2001–2007. With those price changes, the authors estimate the net income effect and the net welfare effect on households in rural and urban areas. As columns 5 and 6 of Table 11 show, the welfare effects are mostly negative for the rural population and mostly positive for the urban population. This suggests that trade policy has a diverse impact according to the household type and region, associated with differences in consumer behaviour and income sources.

Table 11

Predicted price changes from the GTAP model and welfare effects on rural and urban populations in China

			Rural		Urban	
	Wholesale prices	Consumer prices	Net revenue	Mean welfare change	Net revenue	Mean welfare change
	Per cent change	Per cent change	Renminbi	Renminbi	Renminbi	Renminbi
Expenditures						
Rice	-1.4	0.7	73.66	-1.39	-109.33	-0.75
Wheat	-1.5	0.7	40.86	-0.92	0	0
Feed grains	-3.7	2.1	117.04	-4.90	0	0
Vegetable and fruits	-2.6	-0.6	123.41	-4.02	-378.69	2.24
Oilseed	-5.7	-5.9	37.05	-2.10	-1.04	0.06
Sugar	-2.8	-3.5	13.74	-0.34	-174.06	6.01
Plant-based fibers	1.6	4.1	36.84	0.56	0	0
Livestock & meat	-1.5	0.7	194.62	-5.21	-500.65	-3.40
Dairy	-2.4	-0.5	2.50	-0.09	0	0
Other food	-3.1	-2.7	-81.60	2.04	-343.13	9.32
Beverages and tobacco	-5.6	-7.7	-72.98	5.62	-197.2	15.09
Extractive industries	-0.4	1.7	17.99	-0.86	-173.03	-2.92
Textiles	-0.2	-1.5	-11.08	0.17	-53.5	0.82
Apparel	2.6	0.8	-64.13	-0.51	-394.3	-2.98
Light manufacturing	-0.6	0.5	-16.15	-0.08	-82.96	-0.43
Petrochemical industry	-1.1	0.8	-325.39	-2.60	-398.23	-3.19
Metals	-0.6	1.3	-15.30	-0.20	-24.02	-0.31
Autos	-3.8	-4	-52.27	2.09	-37.76	1.52
Electronics	-1.2	-1.4	-24.27	0.34	-162.69	2.20
Other manufactures	-0.8	0.8	-264.61	-2.12	-431.16	-3.46
Trade and transport	-0.4	1.7	-18.70	-0.32	-110.53	-1.85
Construction	-0.4	1.7	0	0	-31.11	-0.52
Communication	-0.4	1.7	-16.72	-0.28	-152.04	-2.54
Commercial services	-1.1	0.9	-61.37	-0.55	-533.33	-4.72
Other services	-0.7	1.3	-414.45	-5.39	-680.99	-8.76

Table 11						
Predicted price changes from the GTAP model and welfare effects on rural and urban populations in China						
			Rural		**Urban**	
	Wholesale prices	**Consumer prices**	**Net revenue**	**Mean welfare change**	**Net revenue**	**Mean welfare change**
	Per cent change	Per cent change	Renminbi	Renminbi	Renminbi	Renminbi
Income sources						
Farm unskilled labour	-0.3	-0.3	313.58	-0.85		
Non-farm unskilled labour	1	1.0	287.19	2.96	1227.51	12.64
Skilled labour	0.4	0.4	360.87	1.55	3391.11	14.58
Land	-4.7	-4.7	17.08	-0.80		
Capital	0.6	0.6	21.14	0.13	126.01	0.80

Source: Chen and Ravallion (2004).

The first panel of Table 12 summarizes the aggregate welfare impact of China joining the WTO over the time frames covering 1995–2001 and 2001–2007. In the short run (1995–2001), this trade reform has a positive impact on all Chinese households (a gain of 55.49 renminbi per capita). However, in the long run (2001–2007), the negative effect on rural households (–18.07 renminbi per capita) makes the aggregate effect at the national level slightly negative (–1.54 renminbi per capita during this time period). The second panel shows the changes in inequality measured by the Gini index with respect to the baseline. Inequality increases in all areas, yet mainly in rural areas (from 33.90 to 34.06 per cent). A similar conclusion can be drawn for the poverty indices for which all estimates show an increase of overall poverty in China.

§ Percentage of mean income.

§§ Official poverty line is from China National Bureau of Statistics; $1/day and $2/day poverty lines are from Chen and Ravallion (2001).

Table 12			
Summary statistics on aggregate welfare impacts			
Item	**Rural**	**Urban**	**National**
1. Mean gains (renminbi per capita)			
1995–2001	34.47	94.94	55.49 (1.54)**§**
2001–2007	–18.07	29.45	–1.54 (-0.04)**§**
2. Inequality effects (Gini index, per cent)			
Baseline, 1999	33.95	29.72	39.31
Simulated: Less gains 1995–2001	33.90	29.68	39.27
Simulated: Plus gains 2001–2007	34.06	29.65	39.35
3. Poverty effects (headcount index, per cent)§§			
Official poverty line			
Baseline (1999)	4.38	0.08	2.92
Simulated: Less gains 1995–2001	4.56	0.08	3.04
Simulated: Plus gains 2001–2007	4.57	0.07	3.04
$1/day (1993 purchasing power parity)			
Baseline (1999)	10.51	0.29	7.04
Simulated: Less gains 1995–2001	10.88	0.28	7.28
Simulated: Plus gains 2001–2007	10.81	0.28	7.23
$3/day (1993 purchasing power parity)			
Baseline (1999)	45.18	4.07	31.20
Simulated: Less gains 1995–2001	46.10	4.27	31.88
Simulated: Plus gains 2001–2007	45.83	3.97	31.6

Source: Chen and Ravallion (2004).

Annex B Formal derivation of the first-order welfare effect[57]

This annex aims to formally derive the first-order welfare effect that we presented intuitively in Section 3 of this module. The framework builds on standard agricultural household models, as in Singh *et al.* (1986), which we will modify to take into account that most urban households in middle-income countries are wage earners and do not produce agricultural goods. The unit of analysis is the household, denoted by *h*. To measure welfare changes, we begin by adopting the indirect utility function approach, as in Deaton (1997). We will later derive the same result using the expenditure function (as in Dixit and Norman, 1980) where we will incorporate the effects of labour income.

The indirect utility function of household *h* depends on a vector of prices *p* and on household income y^h:

$$V^h(p, y^h) = V^h(p, x_o^h + \sum_j \pi_j^h(p_j))$$ (B1)

where the vector *p* comprises consumer prices for all goods. In this equation, household income comprises profits from the production of goods *j*, $\pi_j^h(p_j)$, and exogenous income, x_o^h. We purposefully leave labour income, transfers, and other sources of income (e.g. capital income) out for the moment.

Let us now consider the impact of changes in the price of commodity *i* (with $i \in j$). The short-term effects on the household can be derived by differentiating the indirect utility function. This delivers:

$$\frac{\partial V^h}{\partial p_i} = \frac{\partial V^h}{\partial p_i} + \frac{\partial V^h}{\partial y^h}\frac{\partial \pi_i^h}{\partial p_i}$$ (B2)

Next, recall first that Roy's identity indicates that the consumption of good *i*, c_i^h is given by:

$$c_i^h = -\frac{\partial V^h / \partial p_i}{\partial V^h / \partial y^h}$$ (B3)

Second, recall also that Hotelling's lemma establishes that household production q_i^h is given by:

$$q_i^h = \frac{\partial \pi_i^h}{\partial p_i}$$ (B4)

Dividing and multiplying the first term on the right-hand side of (B2) by $(\partial V^h)/(\partial y^h)$ results in:

$$\frac{\partial V^h}{\partial p_i} = \frac{\partial V^h / \partial p_i}{\partial V^h / \partial y^h}\frac{\partial V^h}{\partial y^h} + \frac{\partial V^h}{\partial y^h}\frac{\partial \pi_i^h}{\partial p_i}$$ (B5)

Substituting (B3) and (B4) in (B5) yields:

$$\frac{\partial V^h}{\partial p_i} = \frac{\partial V^h}{\partial y^h}(q_i^h - c_i^h)$$ (B6)

Leaving aside, for the moment, the factor $(\partial V^h)/(\partial y^h)$, this equation shows that the welfare impact of a price change depends on the difference between the production and the consumption level of good *i* by the household. Before discussing the implications of this result, note that, in empirical work, we seldom observe consumed and produced quantities. Instead, we observe total expenditure on various goods and services, and total income from various production activities. In order to be able to apply the framework to the data, we need some manipulation of equation (B6). In short, multiply and divide by p_i and by total household income y^h to get:

$$\frac{\partial V^h}{\partial p_i}\frac{p_i}{y^h} = \frac{\partial V^h}{\partial y^h}(q_i^h - c_i^h)\frac{p_i}{y^h}$$ (B6.1)

Furthermore, multiply each side of equation (B6.1) by y^h, as follows:

$$\frac{\partial V^h}{\partial p_i}\frac{p_i}{y^h}y^h = \frac{\partial V^h}{\partial y^h}(q_i^h - c_i^h)\frac{p_i}{y^h}y^h$$ (B6.2)

to get, after applying the properties of logarithms:

$$\frac{\partial V^h}{\partial ln p_i} = \frac{\partial V^h}{\partial ln y^h}(\phi_i^h - s_i^h)$$ (B7)

The left-hand side is the object we are trying to measure. On the right-hand side, $(\partial V^h)/(\partial ln y^h)$ is the marginal utility of money to household *h*; ϕ_i^h is the share of household income derived from the production of goods *i* equal to $q_i^h(p_i/y^h)$; and s_i^h is the budget share spent on good *i*, equal to $c_i^h(p_i/y^h)$. In Deaton (1989b, 1997), the quantity $\phi_i^h - s_i^h$ is the net benefit ratio, which is what we care about with regard to policy implications. In fact, $\phi_i^h - s_i^h$ is the money equivalent of the losses or gains for different individuals. The benefit ratios are easily retrieved from the household surveys. Note that $(\partial V^h)/(\partial ln y^h)$ is the private marginal utility of income, which is not the focus of our analysis. Instead, we care about the social marginal utility of money, which informs us about the amount of resources the social planner[58] needs to transfer to household *h*.

We can now turn to the interpretation of this equation. Households are affected both on the consumption and income sides. On the consumption side, consumers are worse off if prices go up and better off if prices go down. In a first-order

approximation, these effects can be measured with budget shares, s_i. On the income side, there is also a direct impact on profits if the household produces goods i, which depends on the share of income derived from these goods, $\emptyset_i$. In rural economies, this source of income can account for a large portion of total income. In more urbanized economies with more developed labour markets (as in many countries in Latin America or Asia), the role of the direct production of (agricultural) goods tends to be much less important and may be treated as zero. When we do not consider labour income, the total effect of a price change will depend on whether the household is a net consumer or a net producer of the goods under study.

In a small open economy that faces exogenous commodity prices (determined in international markets), wages will respond to changes in those prices mainly because the demand for labour depends on prices (labour supply can be affected by prices as well, but we defer this discussion for the moment). Changes in relative product prices cause some sectors to expand and others to contract. If sectors use factors of production in different proportions, then the relative demand for factors (including skilled labour, unskilled labour, and capital) will change. Even with a fixed labour supply, wages will adjust. If the labour supply reacts as well, an additional channel emerges. In practice, the link between wages and prices depends on the way product prices affect factor demand and supply, and the way changes in factor demand and supply transmit to wages. It is possible to imagine situations where wages would not react to a change in a given price, or situations where wages would increase or decrease. The prices of non-traded goods can also be affected. In the simplest mechanisms, a change in the price of traded goods affects factor prices, as discussed above, and this, in turn, affects the cost of production of non-traded goods. As a result, the prices of these goods may change as well. How these prices (including wages) respond to trade policy is an empirical question.

It is relatively simple to amend the theoretical framework to account for these responses. We begin with wage adjustments. To illustrate them, we work with the expenditure function approach as in Dixit and Norman (1980). As before, the unit of analysis is the household, denoted by h. In equilibrium, household expenditure (including savings) has to be financed with household income (including transfers).

$$e^h(p, u^h, x^h) = \sum_m w^m + \sum_i \pi_i^h(p, \varphi) + T^h + x_o^h \quad \text{(B8)}$$

The expenditure function $e(.)$ of household h on the left-hand side of equation (B8) is defined as the minimum expenditure needed to achieve a given level of household utility u^h. It depends on a vector of prices of consumption goods, p, on the level of utility u^h the household wishes to achieve, and on other household characteristics, x^h (such as household composition).

Income comprises the sum of the wages of all working members m, (w^m), and the sum of the profits π_i made in different economic activities i. Profits include, for instance, the net income from agricultural production or farm enterprises. They depend on prices, technical change, and key household characteristics (summarized by φ). Note that profits are defined as sales net of purchases of inputs so that some of the effects caused by protection on inputs or intermediate goods can be captured by π_i. In equation (B8), T^h measures transfers (public or private), savings, and other unmeasured factor returns. Finally, we add exogenous income x_o^h for technical reasons.

It is evident from equation (B8) that household welfare depends on equilibrium variables such as prices and wages (that affect household choices) and also on household endowments. For instance, household consumption depends on the prices of consumer goods and household income depends on the labour endowment (skilled, unskilled), the wage rate, and the prices of key outputs. It follows that changes in commodity prices affect welfare directly via consumption and production decisions, and that these effects are heterogeneous insofar as they depend on household choices and endowments.[59]

The first-order impact of changes in the price of good i can be derived by differentiating equation (B8) (while keeping utility constant and adjusting x^h) and following a similar procedure to the one above in the case of the indirect utility function. Specifically, the terms of equation (B8) are re-arranged as follows:

$$x_o^h = e^h(.) - \sum_m w^m - \sum_i \pi_i^h(p, \varphi) - T^h \quad \text{(B8.1)}$$

Assuming T^h constant or equal to zero, the differentiation for an exogenous change in the prices of consumption goods i (p_i) yields:

$$dx_o^h = \frac{\partial e^h(.)}{\partial p_i} \partial p_i - \sum_m \frac{\partial w^m}{\partial p_i} \partial p_i - \frac{\partial \pi_i^h(.)}{\partial p_i} \partial p_i \quad \text{(B8.2)}$$

Dividing all terms by $e^h(.)$ and manipulating the right-hand side of equation (B8.2), it follows that:

$$\frac{dx_o^h}{e^h} = \frac{\partial e^h(.)}{\partial p_i}\ \frac{dp_i}{p_i}\ \frac{p_i}{e^h(.)} - \sum_m \frac{\partial w^m}{\partial p_i}\ \frac{dp_i}{p_i}\ \frac{p_i}{w^j}\ \frac{w^m}{e^h(.)}$$

$$- \frac{\partial \pi_i^h(.)}{\partial p_i}\ \frac{dp_i}{p_i}\ \frac{p_i}{e^h(.)} \qquad\qquad (B8.3)$$

The right-hand side of equation (B8.3) is composed of three terms.

First, the consumption effect for good *i*, represented by:

$$\left[\frac{\partial e^h(.)}{p_i}\ \frac{p_i}{e^h(.)} \right] \frac{dp_i}{p_i} = s_i^h\ \frac{dp_i}{p_i} \qquad\qquad (B8.4)$$

where s_i^h represents the budget share of household *h* spent on good *i* and derives from Shepard's lemma, according to which $\partial e^h(.) / \partial p_i$ is equal to the demand of household h for good *i* (x_i^h).

Second, the labour income effect, equal to:

$$\left[\sum_m \frac{\partial w^j}{\partial p_i}\ \frac{p_i}{w^j}\ \frac{w^m}{e^h(.)} \right] \frac{dp_i}{p_i} = \left[\sum_m \varepsilon\,w_i^m\,\theta^m \right] \frac{dp_i}{p_i} \ (B8.5)$$

where $\varepsilon\,w_i^m$ is the elasticity of the wage earned by household member *m* with respect to p_i and θ^m is the share of wage income of the household member *m* in total household expenditure $e^h(.)$.

Finally, the third term on the right-hand side of equation (B8.3) is equal to:

$$\left[\frac{\partial \pi_i^h(.)}{\partial p_i}\ \frac{p_i}{e^h(.)} \right] \frac{\partial p_i}{p_i} = \phi_i^h\ \frac{\partial p_i}{p_i} \qquad\qquad (B8.6)$$

with ϕ_i^h representing the share of household income from the production of good *i*, derived from Hotelling's lemma (see equation (B4) above).

Substituting s_i^h, ϕ_i^h and the labour income effect in equation (B8.3), it follows that:

$$\frac{dx_o^h}{e^h} = \left[s_i^h - \phi_i^h - \sum_m \theta^m\,\varepsilon\,w_i^h \right] \frac{dp_i}{p_i} \qquad (B8.7)$$

where dp_i / p_i can also be written as $dlnp_i^h$ for the properties of logarithms. Multiplying each term of equation (B8.4) by -1[60] yields:

$$- \frac{dx_o^h}{e^h} = \left[-s_i^h + \phi_i^h - \sum_m \theta^m\,\varepsilon\,w_i^m \right] dlnp_i^h \qquad (B8.8)$$

which finally results in:

$$- \frac{dx_o^h}{e^h} = (\phi_i^h - s_i^h)\,dlnp_i^h + \sum_m \theta^m\,\varepsilon\,w_i^m\,dlnp_i^h$$

$$= cv^h \qquad\qquad\qquad (B9)$$

where $cv^h = -dx_o^h / e^h$ is a measure of the compensating variation (as a share of initial expenditure) at the household level associated with a change in the ith price. The compensating variation is the revenue that the social planner needs to transfer to households to compensate them for the price change. If a household loses from a price increase, the compensating transfer of income from the planner is $-dx_o/e$ and the compensating variation cv is negative (i.e. a deficit for the planner). Instead, if the household benefits from a price increase, the compensating variation is positive because it actually represents a transfer from the household to the planner (so that dx_o/e is negative).

Equation (B9) summarizes the first-order effects of a price change. The first term on the right-hand side re-establishes the net consumer/net producer result, as described in Section 3 of this module. Additionally, price changes affect wages. This channel is described by the second term on the right-hand side of equation (B9). The mechanisms are in principle simple. When there is a price change, demand for different types of labour (and also labour supply) can change, thus affecting equilibrium wages. In equation (B9), these responses are captured by the elasticities $\varepsilon\,w_i^m$, which vary from one household member to another provided different members are endowed with different skills (unskilled, semi-skilled, or skilled labour) – also known as skill wage premiums – or if they work in different sectors – also known as industry wage premiums. These effects on labour income depend on the share of income contributed by the wages of different members, θ^m. Clearly, if countries differ in technologies, endowments, or labour regulations, the responses of equilibrium wages to prices can be heterogeneous across different economies.

In the presence of wage adjustments, the standard net consumer/net producer proposition needs to be modified. Consider the case where a household consumes a product but does not produce it at all, yet its members earn an income from selling labour. Omitting wages, this household is a net consumer and could thus be hurt by a price increase. But if wages respond positively to prices, the final welfare effect may not necessarily entail a loss.

REFERENCES

Balat J, Brambilla I and Porto G (2007). Zambia. In: Hoekman B and Olarreaga M, eds. *Global Trade and Poor Nations: The Poverty Impacts and Policy Implications of Liberalization*. Brookings Press, Washington DC.

Borraz F, Rossi M and Ferres D (2012). Distributive effects of regional trade agreements on the 'small trading partners': Mercosur and the case of Uruguay and Paraguay. *The Journal of Development Studies*, 48 (12): 1828–1843.

Brambilla I and Porto G (2009). Commodity prices: Impacts and adjustments. Universidad Nacional de la Plata, Argentina.

Bussolo M and De Hoyos RE, eds. (2008). *Gender Aspects of the Trade and Poverty Nexus - A Macro-Micro Approach*. Palgrave Macmillan and World Bank, Washington DC.

Bussolo M, De Hoyos RE and Wodon Q (2008). Higher prices of export crops, intra-household inequality, and human capital accumulation in Senegal. In: Bussolo M and de Hoyos RE, eds. *Gender Aspects of the Trade and Poverty Nexus - A Macro-Micro Approach*. Palgrave Macmillan and World Bank, Washington DC.

Cameron C and Trivedi P (2005). *Microeconometrics: Methods & Applications*. Cambridge University Press, Cambridge UK.

Chen S and Ravallion M (2001). How did the world's poor fare in the 1990s? *Review of Income and Wealth*, 47 (3): 283–300.

Chen S and Ravallion M (2004). Welfare impacts of China's accession to the World Trade Organization. *World Bank Economic Review*, 18 (1): 29–57.

Cherkaoui M, Khellaf A and Nihou A (2011). The price effect of tariff liberalization in Morocco: Measuring the impact on household welfare. Economic Research Forum Working Paper No. 637.

Deaton A (1989a). Rice prices and income distribution in Thailand. *The Economic Journal*, 99 (395): 1–37.

Deaton A (1989b). Household survey data and pricing policies in developing countries. *World Bank Economic Review,* 3 (2): 183–210.

Deaton A (1997). *The Analysis of Household Surveys. A Microeconometric Approach to Development Policy*. John Hopkins University Press for the World Bank, Baltimore.

De Hoyos RE (2008). Can maquila booms reduce poverty? Evidence from Honduras. In: Bussolo M and de Hoyos, RE, eds. *Gender Aspects of the Trade and Poverty Nexus - A Macro-Micro Approach*. Palgrave Macmillan and the World Bank, Washington DC.

Dixit A and Norman V (1980). *Theory of International Trade: A Dual, General Equilibrium Approach*. Cambridge University Press, Cambridge UK.

Feenstra R (1989). Symmetric pass-through of tariffs and exchange rates under imperfect competition: An empirical test. *Journal of International Economics*, 27 (1, 2): 25–45.

Francois J and Hall K (2009). Global simulation analysis of industry-level trade policy: The GSIM model. IIDE Discussion Paper No. 20090803. Institute for International and Development Economics, Rotterdam.

Glewwe P (2013). Econometric analysis II, lecture notes. University of Minnesota.

Goldberg PK and Pavcnik N (2007). Distributional effects of globalization in developing countries. *Journal of Economic Literature*, 54 (March): 39–82.

Hoekman B, Nicita A and Olarreaga M (2005). Can Doha deliver the "development" round? World Bank, Washington DC.

Marchand BU (2012). Tariff pass-through and the distributional effects of trade liberalization. *Journal of Development Economics*, 99 (2): 265–281.

Nicita A (2009). The price effect of trade liberalization: Measuring the impacts on household welfare. *Journal of Development Economics*, 89 (1): 19–27.

Porto G (2006). Using survey data to assess the distributional effects of trade policy. *Journal of International Economics*, 70 (1): 140–160.

Porto G (2007). Globalization and poverty in Latin America: Some channels and some evidence. *The World Economy*, 30 (9): 1430–1456.

Porto G, Depetris Chauvin N and Olarreaga M (2011). *Supply Chains in Export Agriculture, Competition, and Poverty in Sub-Saharan Africa*. Centre for Economic Policy Research, London.

Singh I, Squire L and Strauss J, eds. (1986). *Agricultural Household Models: Extensions, Applications and Policy*. Johns Hopkins Press for the World Bank, Baltimore.

Sourabh Bikas P (2009). Distributional effects of tariff reforms in India. Working Paper. University of British Columbia, Vancouver.

UNCTAD (2011). Who is benefiting from trade liberalization in Cape Verde? A gender perspective. United Nations publication. UNCTAD/OSG/2011/2. New York and Geneva. Available at: *http://unctad.org/en/PublicationsLibrary/osg2011d2_en.pdf*.

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

MODULE 3
The macroeconomic approach[61]

1 Introduction

This module uses country-level data to focus on the interconnection between trade openness and gender outcomes in terms of improving women's economic, political, and social status. Because we use this level of aggregation in the data, we decided to title this module "macroeconomic approach", but this does not necessarily imply that we will only use variables associated with macroeconomic theory such as inflation, unemployment, and growth.

The first macroeconomists in this area of research looked at the relationship between trade openness and macroeconomic outcomes by exploring the effects of trade openness on economic growth. It was only at a later stage that macroeconomists started to analyse the repercussions of trade on inequality, with some authors focusing on poverty among specific groups, such as women, and on gender inequality. It is well documented that opening an economy to international trade often produces significant changes that go beyond the changes associated with growth rates. For example, international trade often triggers structural transformation of the local economy, prompting shifts in employment with consequent feminization or de-feminization of the workforce (Tejani and Milberg, 2010). Moreover, access to new markets for exporting firms from low- and middle-income countries may generate higher income for workers in the exporting sectors. Local firms can also access higher-quality inputs and better technologies, which can help them close the productivity gap observed in most developing countries.

But international trade can also increase unemployment, poverty, and income inequality in the short and medium term due to stiff international competition, as well as create a socially, economically, and politically unsustainable situation in which the potential benefits of trade sometimes do not materialize. One reason for this could be rigidities in labour markets. For the gains from trade to occur, resources need to be reallocated from less productive activities to more productive ones, which may not happen in the presence of imperfect labour markets. This suggests that the relationship between international trade and labour market outcomes is complex and that there are important complementarities between trade and labour market policies. All the more, even when we observe aggregate benefits from increased trade, some groups, including women, may lose as a result. The overall effect of trade openness in a developing country may depend on complementary policies, institutions,

and infrastructure – which is why public policies are important and policymakers need to consider the gender effects of trade.

Besides its direct employment effect, international trade connects countries in a way that such matters as standards, laws, cultural norms, and gender roles in a given country may have spillover effects in other countries, particularly in those linked by commercial and financial flows. Of particular interest to us is the effect of globalization, proxied by a measure of trade openness, on women's economic, political, and social status. We will see below how econometric techniques can be used to assess this question.

Consequently, we need to ask ourselves whether gains from trade actually happen,[62] and then for the purpose of this teaching material, whether they reduce or increase existing gender inequalities.[63] It is also important to evaluate whether international trade empowers women,[64] and what its effects are beyond income-generating opportunities. While some dimensions of the effect of trade on gender inequality (e.g. employment and wages, mostly linked to export expansion) are better documented than others (e.g. intra-household resources and time allocation), there remains great scope for more research on developing countries.

Section 2 of this module summarizes the macroeconomic literature on trade and gender. Section 3 provides the intuition behind the methodology used to link international trade to several gender outcomes using aggregate data. In particular, we review a collection of panel data techniques. For the hands-on application in Sections 4 and 5, we present two papers estimating the effects of globalization, captured by a variety of measures of trade openness, on women's status. In this sense, we depart from typical economistic analyses that study the relationship between trade and labour market outcomes and instead present papers that take a broader view of globalization and women's rights. The selected applications also allow us to see how similar data can be used with different estimation techniques. Section 6 draws some conclusions.

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

- Use the macroeconomic approach to analyse the link between trade and gender;
- Review and summarize the literature employing the macroeconomic approach to investigate the linkages between trade and gender;
- Understand how the macroeconomic ap-

proach differs from the microeconomic approach presented in the previous module;

- Have a basic understanding of econometric models, such as panel data (including dynamic panel data) with fixed and random effects;

- Compare the fixed-effects model with the random-effects model and identify which is more appropriate to use according to the research question of interest;

- Replicate, using Stata, the results of the paper by Richards and Gelleny (2007) titled "Women's Status and Economic Globalization";

- Replicate, using Stata, the results of the paper by Neumayer and de Soysa (2011) titled "Globalization and the Empowerment of Women: An Analysis of Spatial Dependence via Trade and Foreign Direct Investment".

2 Review of the literature

This section reviews a brief collection of macroeconomic studies on trade and gender. Its aim is to familiarize readers with a few well-known papers in the literature on trade and gender, rather than serve as an exhaustive literature review. The papers cited below may also prove useful when you are carrying out your own research.

The topic of trade and gender inequality is fairly recent. The policy prescription of trade liberalization was promoted in the 1970s and 1980s as a means to address (a) the efficiency distortions generated by the import substitution industrialization strategy, and (b) the disappointing economic performance of the inward-oriented Latin American countries in the 1960s and 1970s, which contrasted sharply with the success of the outward-oriented East Asian "Tigers". Accordingly, the primary focus of economists in the empirical literature has been to establish the links between trade and growth and understand whether the latter could contribute to poverty alleviation and income equality, especially in developing countries, on the grounds that trade has greater potential as an engine of growth in countries with widespread poverty than in other countries. Economists have studied the empirical research on the relationship between trade liberalization and gender only at a later stage, and there is a need for more evidence, in particular for developing countries.

A number of papers have assessed cross-country evidence using a macroeconomic approach. The early work of Adrian Wood (1991) explores the changes in the gender composition of manufacturing employment for a set of developed and developing countries and investigates the extent to which these changes were caused by trade.

His results suggest that the expansion of trade between developing and developed countries coincided with an increase in the intensity of female employment in the former but, contrary to prior evidence (Schumacher, 1984; Baldwin, 1984), did not result in a reduction in the demand for female workers in the latter. Wood (1991) provides several explanations for this asymmetry, including the possibility that it is easier for female workers in developed countries to relocate from one manufacturing sector to the other, (such as from textiles manufacturing to the manufacturing of food and beverages) since males are assumed to have more sector-specific skills. Kucera and Milberg (1999) update the work of Wood and focus on the employment effects of trade in terms of gender in the manufacturing sector of ten OECD countries. They find that in most cases trade with developing countries has adversely affected female manufacturing employment. Both the papers by Wood (1991) and Kucera and Milberg (1999) are part of a long-standing academic debate on the macroeconomic effects of trade on gender inequality. Another interesting study in this field is that of Bussmann (2009), who looks at 134 countries and discovers yet again that trade openness increases female labour force participation in developing countries, whereas the share of working women in OECD countries declines.

Also using cross-sectional data, Baliamoune-Lutz (2006) finds evidence that globalization and growth seem to have no effect on gender equality (measured as the difference between women's and men's illiteracy rates) in non-sub-Saharan African developing countries, but exacerbate gender inequality in sub-Saharan African countries. Wamboye and Seguino (2014) focus on 14 sub-Saharan African countries and claim that the employment effects of trade in terms of gender depend on the structure of a country's economy rather than on its level of economic development. They find that trade liberalization plays a different role in women's relative employment according to each country's endowment of physical infrastructure (electrification, clean water, transport and communication infrastructure), which influences women's care burdens and thus their labour supply. In a cross-country study on the effects of trade and FDI on the gender wage gap, Oostendorp (2009) finds evidence that increased trade and FDI contribute to narrowing the gender wage gap in developed countries but not in developing countries.

Other papers have explored indirect channels of interaction between trade and gender. For example, Black and Brainerd (2002) tested Becker's (1959) hypothesis[65] according to which there is

a negative correlation between employer discrimination and the degree of competition in the product market. Using data from United States manufacturing industries, the authors show that higher competition as a result of trade reduces the ability to discriminate against women in concentrated industries, and thus trade openness contributes to shrinking the gender wage gap. Berik *et al.* (2004) apply this framework to study the effect of trade on wage discrimination in Taiwan Province of China and the Republic of Korea. Contrary to Becker's theory, the authors find that trade is linked to higher gender wage discrimination in more concentrated industries where women seem to be more affected by the cost-cutting strategy of their employers.

As explained in Volume 1 of this teaching material, the relationship between trade and gender is bi-directional. For this reason, among the macroeconomic literature on trade and gender you will also find studies that examine how gender affects trade. In particular, there is a wide range of studies that explore the linkages between gender inequalities and export performance. The basic idea of this strand of literature is that firms in labour-intensive, export-led sectors rely on cheap female labour to thrive in international markets. In this context, gender norms and stereotypes also play an important role in the clustering of the female workforce in labour-intensive manufacturing (see Module 3 of Volume 1). For instance, Seguino (2000) studies a group of semi-industrialized, export-oriented countries and shows that gender inequality reflected in lower wages for women contributed to higher growth through its positive effect on exports. This paper opened a lively discussion with Schober and Winter-Ebmer (2011) who find that gender inequality is bad for economic growth and criticize Seguino (2000) for promoting gender inequality as a growth-enhancing strategy. Seguino (2011) replies by raising some empirical concerns about the approach of the paper by Schober and Winter-Ebmer. She concludes: "A finding that gender wage inequality is a stimulus to growth is not a vote or indeed justification for inequality. Rather, it is an evidence-based approach for assessing how things stand and what we need to do at the policy level to promote equity-led growth." (Seguino, 2011: 1487).

Busse and Spielmann (2006) adopt a broader definition of gender inequality, including wages, labour market access, and education inequality. They argue that gender bias does not influence the amount of export flows but rather the type of products exported. In their view, gender inequality may favour the export of labour-intensive products but would disincentivize countries to switch to higher-value products and thus limit their growth potential.

These are just a few macroeconomic studies in the area of trade and gender in addition to the two studies reviewed in Sections 4 and 5 of this module. The papers reviewed here differ in their scope and theoretical approach, and use different definitions of gender inequality and trade. For example, they define gender inequality in terms of women's status rather than measuring inequality in terms of labour market outcomes (e.g. gender wage gap). For a more detailed literature review you may read Çağatay (2001) and Fontana (2008). In conclusion, the aim of this literature review was to demonstrate how extensive the macroeconomic literature on trade and gender is, but also to make you realize that the debate on the interlinkages between trade and gender is still open. We encourage you to contribute to it through your own research.

3 Methodological approach: Panel data models

Panel data are repeated measures of a variable (*i*) over time (*t*). This variable may be related to individuals, households, firms, countries, etc. over a period of time. The main characteristic of a panel dataset is the two-dimensionality of the data. Most papers adopting the macroeconomic approach in the trade and gender literature use panel data where the main variable is defined in the country and the year dimension.

A micro-panel dataset is a panel for which the time dimension *T* is largely less important than the individual dimension *N*. A macro-panel dataset is a panel for which the time dimension *T* is similar to the individual dimension *N*. In panel data with *N* units and *T* periods, we could independently estimate *N* time series models or *T* cross-section models. However, there are several advantages of using panel data rather than independent regressions. One of them is that panel data allow us to control for unobserved heterogeneous characteristics. Panel data also allow us to aggregate information in some way, and the more information we have to run the estimation of a set of parameters, the more efficient such estimation is. However, panel data also have some disadvantages. Sometimes it is not possible to aggregate cross-sectional and temporal data. Panels, especially for micro data, are often expensive and difficult to assemble. There may be selection problems in panels, as some individuals may disappear, decide not to answer some specific questions, or be selected for the panel because

they have particular characteristics of interest for the purpose of specific research. In a nutshell, selection problems usually occur when individuals are not selected randomly to be included in the sample.

The basic model in a panel data framework is:

$$y_{it} = x'_{it}\beta + u_{it} \qquad (1)$$

where y is the dependent variable of interest (e.g. the gender wage gap); x is a matrix of independent variables or covariates (e.g. GDP per capita, a measure of trade openness, etc.); β are the coefficients that identify the statistical relationship between y and x; and

$$u_{it} = \mu_i + \delta_t + \varepsilon_{it} \qquad (2)$$

is the error term. The latter includes three components that represent the three sources of unobserved variability: μ_i represents unobserved variability across individuals (some individual-specific characteristics that we are not able to capture); δ_t represents unobserved variability across periods of time (in a particular period, variables may be affected by something we cannot observe); and ε_{it} is pure unobservable variability specific to the individual and the time observation.

Assume that δ_t is zero and that ε_{it} satisfies all the classic assumptions.[66] In the case that all individual-specific components μ_i are zero, we have $u_{it} = \varepsilon_{it}$, the model becomes $y_{it} = x'_{it}\beta + \varepsilon_{it}$ and the panel data structure does not add any useful information for the estimation of the parameters.

3.1 Fixed-effects model

In a nutshell, a fixed-effects model is an econometric specification of panel data model that allows us to "net out" from the estimation results the effects of unobserved time-invariant and individual-specific characteristics (μ_i) that are probably correlated with the independent variables x_{it}. If we do not account for these fixed effects and they are related to the independent variables, we create an "omitted variable bias". Assume that the model now is:

$$y_{it} = x'_{it}\beta + \mu_i + \varepsilon_{it} \qquad (3)$$

Here we can see the advantage of using panel data, as parameters μ_i cannot be estimated with a cross-section but can be with a panel. If μ_i is correlated with x_{it}, the independent variables may be endogenous with respect to μ_i but not to ε_{it}. This panel model can be seen as a linear model where each individual (or firm, country, etc.) has its own y-intercept. We can estimate this model using $N–1$ dummy variables per individual in the sample. Having a panel allows us to control for individual omitted variables that do not vary over time.

3.2 Random-effects model

A random-effects model is a statistical model that is used when we assume that μ_i – the omitted or time-invariant component of the error term – are uncorrelated with x_{it} – the independent variables. Take the same model as before:

$$y_{it} = x'_{it}\beta + \mu_i + \varepsilon_{it} \qquad (3)$$

but now instead of considering μ_i as constant for each individual, we assume that it is a random variable. In other words, the random-effects assumption refers to the uncorrelated individual-specific effects with the independent variables (the time-invariant characteristic μ_i is purely random and uncorrelated with x_{it}). Contrary to random effects, the fixed-effects assumption is that the individual-specific effects are indeed correlated with the independent variables. We can estimate a panel data model with either of the two models. Which one will be better (more efficient)? It depends on which one of the two assumptions is true in the data. The discussion of the tests[67] to assess which model fits the data best is beyond the scope of this teaching material but it is worthwhile noting that in practice most papers use fixed-effects models (see Box 2).

Box 2

Fixed effects vs. random effects

In econometrics, fixed-effects models control for, or "partial out" the effects of omitted (unobserved) time-invariant variables that are correlated with one or more observed variables included in the model. This is true regardless of whether or not the time-invariant variables are explicitly measurable. By time-invariant variables, we mean variables whose value does not change over time; by time-invariant effects, we mean that the variables have the same effect on the dependent variable over time.

In a random-effects model, the omitted variables are assumed to be uncorrelated with (or, more strongly, statistically independent of) all the observed variables included in the model. This assumption will often not be valid, but a random-effects model may be still desirable under some circumstances.

<div style="border:1px solid">

Box 2

<div style="background:gray">**Fixed effects vs. random effects**</div>

There are three criteria you may wish to follow for choosing whether random or fixed effects best fit the model you need to estimate:

(a) The nature of the variables that have been omitted from the model

- If you think that you are not omitting any additional explanatory variables, or if you can justify that the omitted variables are uncorrelated with the explanatory variables included in the model, then a random-effects model is probably what you should use. It will produce unbiased estimates of the coefficients, use all the data available, and produce the smallest standard errors.
- However, if there are omitted variables, and these variables are correlated with the variables included in the model, then a fixed-effects model is the most appropriate for controlling for omitted variable bias. The idea is that whatever effects the omitted variables have on the dependent variable at a given time, they will also have the same effect at a later time; hence, their effects will be constant or "fixed". Note that in order to apply the fixed-effects model, the omitted variables must satisfy the condition of being time-invariant values with time-invariant effects.

(b) The amount of variability within the independent and dependent variables

- If subjects – i.e. the key dependent variable and independent variable – change little or not at all over time, a fixed-effects model may not work very well. There needs to be within-dimension variability in the variables. If there is little variability, then the standard errors produced by the fixed-effects model may be too large, and it would be better to employ a random-effects model.
- A random-effects model will often produce smaller standard errors. However, there is a trade-off: the coefficients produced by applying random-effects are more likely to be biased if the assumption that there are no omitted variables is violated (which is most often the case).

(c) Control vs. effect of time-invariant variables

- A fixed-effects model does not estimate the effects of variables whose values do not change over time. Instead, it controls for them or "partials them out".
- A random-effects model estimates the effects of time-invariant variables, but the estimates may be biased if you incorrectly assume that there are no omitted variables in the model.

Source: Allison (2009).

</div>

3.3 Dynamic panel data models

We get a dynamic panel data model when we add to the independent variables of a panel data model the lagged value of its dependent variable. In other words, dynamic panel data models introduce the temporal dependency of the dependent variable into the equation. In equations, the latter can be represented as:

$$y_{it} = \delta y_{i,t-1} + x'_{it}\beta + \mu_i + \varepsilon_{it} \qquad (4)$$

where $y_{i,t-1}$ is the lagged value of y_{it}. In this model, we have two sources of persistence: $y_{i,t-1}$ and μ_i. If we use an estimator like ordinary least squares or fixed effects, the estimations will be biased and inconsistent because by construction, both y_{it} and $y_{i,t-1}$ depend on μ_i, and $y_{i,t-1}$ is correlated with $\mu_i + \varepsilon_{it}$. We therefore need a different estimator that gets around this problem. Arellano and Bond (1991) provide an appropriate solution. Assuming for simplicity that $\beta = 0$, our model is now:

$$y_{it} = \delta y_{i,t-1} + \mu_i + \varepsilon_{it} \qquad (5)$$

If we subtract $y_{i,t-1}$ from both sides we get:

$$\Delta y_{it} = \delta \Delta y_{i,t-1} + \Delta \varepsilon_{it} \qquad (6)$$

In this equation μ_i, which was creating problems, has disappeared. The first period for which we have this relationship is $t = 3$, where we have:

$$\Delta y_{i3} = \delta \Delta y_{i,2} + \Delta \varepsilon_{i3} \qquad (7)$$

In this case, $y_{i,1}$ is a valid instrument as it is correlated with $\Delta y_{i,2} = y_{i,2} - y_{i,1}$, but is not correlated with $\Delta \varepsilon_{i3}$. In $t = 4$, the relationship is:

$$\Delta y_{i4} = \delta \Delta y_{i,3} + \Delta \varepsilon_{i4} \qquad (8)$$

In this case, $y_{i,2}$ and $y_{i,1}$ are valid instruments. Using the same logic, the valid instruments for T are $y_{i,1}$, $y_{i,2}, \dots, y_{i,T-2}$. Arellano and Bond provide an estimator, the generalized method of moments (GMM), which optimally combines all these instruments.

In the following two sections, we will see how variants of these panel data models are used in the trade and gender literature.

4 Hands-on application I: "Women's status and economic globalization" (Richards and Gelleny, 2007)[68]

4.1 Context and overview

Richards and Gelleny (2007) study the relationship between economic globalization and the status of women by looking at arguments advanced both by proponents and sceptics of globalization. According to the authors, addressing this issue has value for two main reasons. The first is because economic globalization is state-induced ("globalization from above"): those typically most adversely affected by globalization have no voice in its implementation. Therefore, it is important to provide a clear understanding of its nature. Second, the authors want to bridge a theoretical disconnect that exists in the literature between papers applying macro-level analytical methods and those looking at the effects of globalization on particular vulnerable groups that often do not use theoretically oriented methodologies at the country-year or macro levels of analysis. The authors use a panel of 130 countries between 1982 and 2003 and check robustness using alternative measures of economic globalization and women's status. Table 13 provides an overview of their paper.

Table 13

Overview of Richards and Gelleny (2007)
Objective
The objective of the paper is to analyse the relationship between economic globalization and the status of women.
Methodology

• The authors use five different measures of women's status as the dependent variable.
 • The choice of estimation techniques depends on the nature of the dependent variable:
  - For the United Nations indicators of women's status, the authors use a generalized estimation equation (GEE) technique with robust standard errors;
  - For the Cingranelli-Richards (CIRI)[§] indicators (ordinal variables) of women's status, the authors use the ordered logit estimation technique.

Sample description
• 130 countries   • Period of analysis: 1982–2003   • Panel data
Equation estimated[§§]
$y_{it} = \alpha + \beta_1 Open_{it} + \beta_2 FDI_{it} + \beta_3 Portfolio_{it} + \beta_4 SAP_{it} + \beta_5 Development_{it} + \beta_6 Regime_{it} + \varepsilon_{it}$ Note: • In the CIRI indicators equations, the authors also control for the lagged dependent variable (y_{it}) to account for serial correlation;   • All variables are observed for country i and year t.
Dependent and independent variables
The dependent variable (y) are five different measures of women status: • Gender-Related Development Index (GDI) from the United Nations *Human Development Report*   • Gender Empowerment Measure (GEM) from the United Nations *Human Development Report*   • Women's economic status from the CIRI database   • Women's political status from the CIRI database   • Women's social status from the CIRI database   Six independent variables (x) are used to measure globalization: • Trade openness: total value of a country's imports and exports of goods and services as a percentage of GDP (*Open*)   • Foreign direct investment: net inflow of investment as a percentage of GDP (*FDI*)   • Portfolio investment: net amount of transactions in equity securities and debt securities, expressed as a percentage of a country's GDP (*Portfolio*)   • Structural Adjustment Programmes: a dummy variable to account for the effects of the IMF and World Bank Structural Adjustment Programmes (*SAP*)   • Development: log of per capita GDP as a proxy of the level of development of the country (*Development*)   • Democracy: ordinal regime-type indicator from the Polity IV[§§§] to measure the level of democracy in the country (*Democracy*)
Results

• Women's status in a given country is associated with that country's involvement in the global economy.
 • On average, economic globalization improves women's status: 67 per cent of the statistically significant globalization coefficients indicated an association with improved women's status.
 • However, the relationship between economic globalization and women's status differs by type of globalization, type of status, and era:
  - Trade openness has a generally positive influence on women's status;
  - Portfolio investment is associated with lower scores on the CIRI women's economic and social rights indicators and the UN's GEM measure;
  - FDI does not show a statistically significant effect on women's status;
  - There is weak empirical support for the proposition that SAP implementation affects women's status.

Source: Richards and Gelleny (2007).

[§] Cingranelli and Richards (2014) is the latest version of the CIRI Human Rights Dataset.

[§§] The authors do not specify the equation estimated in the paper. The equation displayed here is the result of UNCTAD's own elaboration.

[§§§] Polity IV is a dataset providing information on the regime authority of 167 countries over the period 1800–2013. The Polity score ranges from –10 (most authoritarian) to +10 (most democratic). More information on the Polity IV project is available at: *http://www.systemicpeace. org/polityproject.html*

4.2 Data sources

The dependent variables in the analysis are indicators that measure women's status. The authors use five different indicators to control for the sensitivity of any finding to a particular indicator.

The first two indicators of women's status are related to human rights and are drawn from the United Nations' annual *Human Development Report*. This report includes two gender-specific indices covering the political, economic, and social dimensions of development: (a) the Gender-Related Development Index, and (b) the Gender Empowerment Measure. The GDI is a composite index that measures longevity (measured by life expectancy at birth), knowledge (measured by a combination of the adult literacy rate and the combined primary, secondary, and tertiary gross enrolment ratio), and standard of living (measured by GDP per capita in United States dollars in purchasing power parity terms). These features are combined in a manner so as to penalize gender inequality. The GEM is a composite index measuring gender inequality in three dimensions of empowerment: economic participation and decision-making power (measured as female shares of professional/technical positions and female shares of positions as legislators, senior officials, and managers); political participation and decision-making (measured as the female share of parliamentary seats); and power over economic resources (measured as women's estimated earned income as compared to that of men). Both indexes, the GDI and the GEM, range from 0 to 1.0,[69] with a higher score being more desirable than a lower score.[70]

The authors also use three government indicators for women's economic, political, and social rights from the CIRI Human Rights Dataset (Cingranelli and Richards, 2005). The CIRI economic rights indicator includes women's rights to equal pay for equal work, free choice of employment, gainful employment without the need to obtain spousal consent, equality in hiring and promotion practices, job security, non-discrimination by employers, freedom from sexual harassment in the workplace, and the right to work at night, to work in dangerous occupations, and to work in the military and police force. The CIRI political rights indicator includes women's rights to vote and/or run for political office, hold elected and appointed government positions, join political parties and petition government parties. The CIRI social rights indicator includes women's rights to equal inheritance; enter into marriage on a basis of equality with men; travel abroad; obtain a passport; confer citizenship to children or

a husband; initiate a divorce; own, acquire, manage, and retain property brought into marriage; participate in social, cultural, and community activities; obtain an education; freely choose a residence/domicile; to be free from female genital mutilation of children and adults without their consent; and to be free from forced sterilization. All three indicators are ordinal and range from 0 to 3, with a score of 3 representing the highest level of government respect for women's rights, both in law and in practice.

The most important explanatory variables in Richards and Gelleny (2007) are those related to globalization. The authors use four indicators to account for a country's level of economic globalization: FDI (net flows as a percentage of GDP), portfolio investment (net amount of transactions in equity securities and debt securities, expressed as a percentage of a country's GDP), trade openness (total value of a country's imports and exports of goods and services as a percentage of GDP), and structural adjustment policy implementation (dichotomous measure to account for the IMF and World Bank Structural Adjustment Programmes). The first three measures are from the World Bank's 2005 World Development Indicators dataset. The last measure is from Abouharb and Cingranelli (2006).

Richards and Gelleny (2007) also control for the level of democracy in the country (using the Polity IV dataset) and its level of development (proxied by log of per capita GDP from the World Bank's WDI, 2005). The expectations are that, on the one hand, democratic regimes are more likely to implement and enforce laws promoting women's rights, and, on the other, countries with a higher level of economic development are more likely to provide all citizens with higher levels of income, thereby potentially giving women better prospects of social, political, and economic empowerment.

4.3 Empirical methodology

Richards and Gelleny (2007) use two different empirical methodologies according to the nature of the dependent variable.

Both the GEM and GDI data are interval-level, pooled cross-section, time-series data (with significantly more cross-section observations than temporal units). To deal with this type of data, the authors use the GEE estimation technique with robust standard errors.[71] This approach extends generalized linear models to a regression setting with correlated observations within subjects, i.e. repeated observations on individuals over time

and/or clustered observations. In these cases, the use of ordinary models to analyse data with correlated observations tends to produce incorrect standard errors and *p*-values for regression coefficients. Models that ignore clustering tend to underestimate standard errors of regression coefficients for covariates. However, with time-varying covariates, standard models may tend to overestimate standard errors. To solve this, we use GEE. This estimation technique can be used with a variety of models (linear, logistic, Poisson, etc.) and uses robust estimation of standard errors to allow for clustering. The robust standard errors are derived using the observed variability in the data rather than the variability predicted by an underlying probability model (which produces model standard errors).

The methodology applied in the estimations of the equations with the GDI and GEM indicators as dependent variables cannot be applied when the dependent variables come from the CIRI database, since the three CIRI indicators are ordinal (taking values 0, 1, 2, and 3). The authors therefore use the ordered logit estimation technique to estimate these models.[72] A logit regression is a non-linear regression model that forces the output (predicted values) to be either 0 or 1 (estimating the probability of an outcome). When a dependent variable has more than two categories and the values of each category have a meaningful sequential order, where a value is indeed "higher" than the previous one, then you can use ordinal logit. This type of model is an extension of the logistic regression model that is applied to dichotomous dependent variables, allowing for more than two (ordered) response categories. The only difference is that the ordered logit model estimates the probability of each outcome as a subtraction of a cumulative probability. The authors incorporate a lagged dependent variable in this model to account for serial correlation, and also adjust standard errors to account for country-specific clustering. To account for the fact that the level of globalization was very different at the beginning and end of their sample period, the authors divide their sample into pre-globalization and globalization eras.

4.4 Step-by-step explanation of how to do the estimations in Stata

Together with this module, you are provided with the do-files[73] that Richards and Gelleny created to run the estimations included in the paper. The authors created a do-file that only contains the commands for applying the GEE and the ordered logit model. Therefore, you need to open the statistical software first, set a memory level that is reasonable for the analysis (for instance, 100m), as we have explained in Module 1 of this volume, and then open the dataset (*DR_RG_ISQ_07a.dta*) manually. You can then open the do-file "STATA Richards".

The structure of the do-file has eight command lines, two for each table. The first two lines estimate the GEE model, the results of which are shown in Table 1 in the Richards and Gelleny paper. One line corresponds to the GEM variable (*gem*) and the other to the GDI variable (*gdi*). Lines 3–8 estimate the ordered logit model for the other three variables: women's economic rights (*wecon*) (see Table 2 in the paper), women's political rights (*wopol*) (see Table 3 in the paper), and women's social rights (*wosoc*) (see Table 4 in the paper). For each of the three variables, the data are divided into two sub-samples – pre- and post-1992 – because 1992 is the cut-off the authors choose to indicate the pre-globalization and globalization eras. Note that to do this, the authors created a dummy variable *globz*, taking a value of zero for data corresponding to years before 1992 and a value of one after 1992.

The Stata command to implement the GEE technique is `xtgee` (please refer to Stata `help xtgee` for a detailed explanation of the command). This command estimates longitudinal models and allows you to specify the within-group (within-subject) correlation structure. Using the command `xtgee` is equivalent to using the command `xtreg, pa` (the command for the estimation of the parameters of a lineal panel data model using population averages).

Step 1: Run the GEE model for the GEM variable

For GEM as the dependent variable, the complete Stata command to estimate column 1 in Table 1 of the paper is:

```
xtgee gem tradeopenness_unlogged
fdi_unlogged portgdp
sap_implementation gdppercap_logged
democracy, i (country) t (time)
robust
```

The command `xtgee` is followed by the dependent variable (*gem*), the four measures of globalization (*tradeopenness_unlogged, fdi_unlogged, portgdp*, and *sap_implementation*), the measure of economic development (*gdppercap_logged*), and the level of democracy (*democracy*). The dimensions `i` (*country*) and `t` (*time*) define the unit of analysis and the time dimension of the panel data;[74] the option `robust` is included to obtain cluster-robust standard errors.

Step 2: Run the GEE model for the GDI variable

For the GDI variable (column 2 in Table 1), the command is the same, just replacing *gem* for *gdi*:

```
xtgee gdi tradeopenness_unlogged
fdi_unlogged portgdp
sap_implementation gdppercap_logged
democracy, i (country) t (time)
robust
```

Step 3: Run the ordered logit model for the women's economic rights variable

The Stata command to implement the ordered logit technique is `ologit` (see Stata `help ologit` for a detailed explanation of the command). This command fits ordered logit models of an ordinal dependent variable. The actual values taken on by the dependent variable are irrelevant, except that larger values are assumed to correspond to "higher" outcomes. The sign of the regressions parameters can be interpreted as determining whether the (latent) dependent variable increases with the regressor. If β_i is positive, then an increase in x_{it} reduces the probability of being in the lower category, and increases the probability of being in a higher category of y_i.

For women's economic rights (*wecon*) as the dependent variable, the complete Stata command to estimate column 1 in Table 2 of the paper is:

```
ologit wecon wecon_lag1
tradeopenness_unlogged fdi_unlogged
portgdp sap_implementation
gdppercap_logged democracy if
globz==0 & year>1981, cluster(country)
robust
```

The command `ologit` is followed by the dependent variable (*wecon*). As regressors, we have the time lag of the dependent variable (*wecon_lag1*), the four measures of globalization (*tradeopenness_unlogged, fdi_unlogged, portgdp* and *sap_implementation*), the measure of economic development (*gdppercap_logged*), and the level of democracy (*democracy*). We ask Stata to consider only the observations before 1992 (`globz==0`) and after 1981 using the `if` qualifier. The option `robust` is included to obtain cluster-robust standard errors, and we indicate that clustering is done at the country level.

We have the same command syntax (column 2 in Table 2) for the period of globalization (`globz==1`):

```
ologit wecon wecon_lag1
tradeopenness_unlogged fdi_unlogged
portgdp sap_implementation
gdppercap_logged democracy if
globz==1 & year>1981, cluster(polity)
robust
```

Step 4: Run the ordered logit model for the women's political rights variable

The do-file then repeats the same estimation procedure replacing the women's economic rights (*wecon*) variable with the political (*wopol*) and social (*wosoc*) rights measures.

For Table 3, column 1 in the paper:

```
ologit wopol wopol_lag1
tradeopenness_unlogged fdi_unlogged
portgdp sap_implementation
gdppercap_logged democracy if
globz==0 & year>1981, cluster(polity)
robust
```

For Table 3, column 2 in the paper:

```
ologit wopol wopol_lag1
tradeopenness_unlogged fdi_unlogged
portgdp sap_implementation
gdppercap_logged democracy if
globz==1 & year>1981, cluster(polity)
robust
```

Step 5: Run the ordered logit model for the women's social rights variable

For Table 4, column 1 in the paper:

```
ologit wosoc wosoc_lag1
tradeopenness_unlogged fdi_unlogged
portgdp sap_implementation
gdppercap_logged democracy if
globz==0 & year>1981, cluster(polity)
robust
```

For Table 4, column 2 in the paper:

```
ologit wosoc wosoc_lag1
tradeopenness_unlogged fdi_unlogged
portgdp sap_implementation
gdppercap_logged democracy if
globz==1 & year>1981, cluster(polity)
robust
```

4.5 Discussion of findings and limitations of the analysis

The authors find that in most specifications, there is a relationship between the level of globalization of that country and the status of women. However, the relationship depends on the type of globalization we are considering, the type of status (economic, political, or social), and the period under consideration (before or after 1992). In most cases, globalization is associated with improvements in women's status (a positive statistically significant coefficient), but there are negative coefficients for some variables (in particular the ones associated with the portfolio variable), and in some cases the coefficient is not statistically significantly different from zero.

These results show one of the difficulties of trying to establish relationships between aggregate macroeconomic variables using cross-section and panel regressions. The estimations of Richards and Gelleny show that, depending on the variables and time period we select for the analysis, the result could be different. There is also the possibility of omitted variables that may bias some of the results, which shows the difficulties the analyst faces when deciding on the set of relevant control variables that will be needed to account for the many dimensions in which the countries included in the sample differ. Another drawback with this type of analysis is the difficulties encountered when trying to establish causality between the variables. For instance, women's status and the level of globalization could be caused by a common variable that was omitted in the analysis (e.g. the level of institutional development of the country).

Note that the authors assume somehow that what matters for women's status is the level of globalization and not how this globalization happens (for instance, they do not ask if it matters with what countries an economy is trading). It could be argued that countries that trade with partners where women enjoy better status may benefit from spillover and demonstration effects. Something similar could be argued about FDI flows. In what follows, we will present a second paper that deals with the issues that were not addressed by Richards and Gelleny.

5 Hands-on application II: "Globalization and the empowerment of women: An analysis of spatial dependence via trade and foreign direct investment" (Neumayer and de Soysa, 2011)[75]

5.1 Context and overview

Similar to Richards and Gelleny (2007), Neumayer and de Soysa (2011) also look at the effect of general openness to trade and FDI on women's rights. However, their paper aims to identify a specific channel through which trade may affect the status of women, and it systematically addresses the question of whether trade and investment linkages can enable the empowerment of women. What matters in their analysis is not only if you trade but with whom you trade. However, the paper does not analyse the effects of certain policies often associated with globalization, such as capital account liberalization, trade liberalization, investment incentives, etc. The authors also do not analyse other important aspects of globalization, such as migration and the illegal trafficking of people.

The authors depart from previous studies in two important ways. First, they employ broader measures of women's rights that include both economic and social rights (such as marriage and divorce rights, the right of movement, the right to property, the right to participate in social activities, the right to education, the right to inherit, etc.) as a better gauge of women's empowerment than simple measures of the wage gap and employment ratios. Second, they examine whether it matters with whom one trades and from whom one receives FDI, whereas existing studies have only examined general openness to trade and FDI. For example, if a country mainly trades with and receives FDI from countries that violate rights, we would not expect domestic rights to be enhanced.

The paper is similar to the one by Richards and Gelleny both in terms of the questions being addressed and some of the data used. However, it is interesting to discuss these issues here to show how the same questions can be addressed from a different angle and using a different methodology. Table 14 provides an overview of the Neumayer and de Soysa paper.

Table 14

Overview of Neumayer and de Soysa (2011)

Objective

The objective of the paper is to analyse whether the foreign country with which a country trades and from which it receives FDI matters for women's economic and social rights in the home country. The main question is the following: If a country trades and receives FDI from countries where women's status is high, would that lead to a higher status for women in the home economy?

Methodology

- The authors apply the "ordered logit" estimation method.
- However, ordered logit models do not allow to estimate country-fixed effects; alternatively, the authors use regional dummies.
- When the lagged dependent variable is added as an explanatory variable, Arellano and Bover's system-GMM estimator is employed.

Sample description

- 152 countries
- Period of analysis: 1981–2007
- Panel data

Equation estimated

$$y_{it} = \alpha_i + \beta_1 y_{it-1} + \beta_2 GDPpc_{it} + \beta_3 democracy_{it} + \beta_4 \frac{trade_{it}}{GDP_{it}} + \beta_5 \frac{FDI_{it}}{GDP_{it}} + \beta_6 \sum_k w_{ikt-1}^{trade_{it}} y_{kt-1} + \beta_7 \sum_k w_{ikt-1}^{FDI_{it}} y_{kt-1} + \delta_t + u_{it}$$

- Variables are observed for country i, year t, and partner country k (for trade or FDI).
- w are the weights estimated as the share foreign countries (k) have in trade and FDI stock of partner country i under observation.

Dependent and independent variables

Two dependent variables (y), both from Cingranelli and Richards' (2009) Human Rights Database:
- Measure of women's economic rights
- Measure of women's social rights

Two main explanatory variables (the spatial lag variables):
- Women's rights in foreign countries weighted by the amount of trade of each country with its trading partners ($w^{trade} y_{kt-1}$)
- Women's rights in foreign countries weighted by the amount of FDI received by each country from foreign countries ($w^{FDI} y_{kt-1}$)

Other control variables:
- *GDPpc*: log of per capita income in constant 2000 $ at market exchange rates (from the World Bank)
- *democracy*: Polity II variable as a measure of democracy (from the Polity IV dataset)
- trade/GDP – trade openness measured as the ratio of the sum of exports and imports to GDP (from the World Bank)
- FDI/GDP – trade openness to FDI measured as the value of the total stock of inward FDI relative to GDP (from UNCTAD)
- δ – time-fixed effects

Results

The paper finds that who you trade with matters for the status of women:
- Spillover effects on women's economic and social rights:
  - The trade-weighted spatial lag effect is positive and significant in most specifications;
  - One exception: the sample including only low-income countries.
- FDI links seems to matter less for women's rights:
  - Spillover effects on women's economic and social rights are limited and weak;
  - The FDI-weighted spatial lag effect is positive and significant only for economic rights in middle-income countries.
In general terms, trade openness seems to be conducive to stronger women's economic rights, whereas general FDI openness seems to not matter much.

Source: Neumayer and de Soysa (2011).

5.2 Data sources

The measures of women's economic and social rights, which represent the dependent variables in the analysis, are taken from the CIRI Human Rights Database. These data are also utilized by specialized agencies monitoring the progress of women in the economic and social spheres of their lives (UNIFEM, 2008).

Table 15 lists women's economic and social rights covered in the database that are used for the estimations in Neumayer and de Soysa (2011). An older version of this dataset (Cingranelli and Richards, 2005) was used in the Richards and Gelleny (2007) paper.

Table 15

Economic and social rights from Cingranelli and Richards (2009)
Economic rights
• Equal pay for equal work
• Free choice of profession or employment without the need to obtain a husband's or male relative's consent
• The right to gainful employment without the need to obtain a husband's or male relative's consent
• Equality in hiring and promotion practices
• Job security (maternity leave, unemployment benefits, no arbitrary firing or layoffs, etc.)
• Non-discrimination by employers
• The right to be free from sexual harassment in the workplace
• The right to work at night
• The right to work in occupations classified as dangerous
• The right to work in the military and the police force
Social rights
• The right to equal inheritance
• The right to enter into marriage on a basis of equality with men
• The right to travel abroad
• The right to obtain a passport
• The right to confer citizenship to children or a husband
• The right to initiate a divorce
• The right to own, acquire, manage, and retain property brought into marriage
• The right to participate in social, cultural, and community activities
• The right to an education
• The freedom to choose a residence/domicile
• Freedom from female genital mutilation of children and adults without their consent
• Freedom from forced sterilization

Source: Neumayer and de Soysa (2011).

The main explanatory variables are the spatial lagged variables. They capture the dependent variable (i.e. women's rights) in foreign countries, weighted by some link function connecting each country to its trading partners and the source countries of FDI, w^{trade} and w^{FDI}, respectively.[76]

Other control variables included in the analysis are trade openness, measured as the ratio of the sum of exports and imports to GDP, taken from the World Bank's WDI (2009) (*trade/GDP*); openness to FDI, measured as the value of the total stock of inward FDI relative to GDP, taken from UNCTAD (2009) (*FDI/GDP*); the natural logarithm of per capita income in constant 2000 $ at market exchange rates, taken also from the World Bank's WDI (*GDPpc*); and a measure of democracy from the Polity IV dataset (*democracy*).[77]

5.3 Empirical methodology

The spatial patterns[78] in women's rights are often not caused by spatial dependence[79] but by observable as well as unobservable phenomena – such as cultures and customs, preferences and perceptions, constitutions and institutions, etc. – that are typically spatially clustered. These unobservable variables might lead to spatial patterns in the dependent variable, even in the absence of spatial dependence. A popular method for mitigating the problem created by spatial clustering is the inclusion of country-fixed effects. Such models take out all of the "between" variation in the data and are estimated based on the "within" variation of the data in each observational unit

only (each of the countries in the study). This reduces bias because any spatial clustering or unobserved spatial heterogeneity in the levels of women's rights is fully captured by the fixed effects.

However, the authors cannot apply country-fixed effects here because of the nature of their dependent variable (women's economic and social rights taken from Cingranelli and Richards, 2009). Women's economic and social rights are measured as ordered categorical variables, which take on values 0, 1, 2, or 3. Thus, we need to use an ordered logit or probit model, and this type of econometric technique does not allow for using country-fixed effects. As a compromise, the authors include regional rather than country-fixed effects in ordered logit estimations, using dummy regional variables. They later consider a model that adds a lagged dependent variable and use Arellano and Bover's (1995) system-GMM estimator to perform the analysis. This estimator is preferable to a standard fixed-effects estimator because it can treat both the lagged dependent variable and the spatial lagged variables as endogenous.

Another problem of spatial analysis in cross-sectional time series analysis is that of common shocks and common trends, such as a general increase in awareness of women's rights over time. The authors control for this by including year-fixed effects representing separate intercepts for each year of the period under study, as well as the temporally lagged dependent variable.

5.4 Step-by-step explanation of how to do the estimations in Stata

The file "Article for World Development (womens rights).do" contains the do-file to reproduce the tables in Neumayer and de Soysa (2011).[80] The structure of the file is slightly more complicated than the previous one. We will therefore split the task into several steps.

Step 1: Declare type of data

Sometimes Stata does not recognize that the data have a time series dimension, so we need to indicate that using the command `tsset`. We can tell Stata we have a time series using `tsset` plus the name of the variable measuring time. If we have a panel, we use the same command followed by the variable recording the individual dimension (in this case the country) and the time dimension (e.g. year).

In this particular case, the command is:

```
tsset countryid year
```

Step 2: Fix the estimation sample and construct basic descriptive statistics

We will create a table with sample statistics of the main variables (Table 2 in the Neumayer and de Soysa paper) and their correlation matrix (Table 3 in the paper). When we run and present multiple analyses for the same paper, we often want to keep the same sample across all our models and estimations. If we do not indicate this to Stata, different models can have different sample sizes because different variables have different patterns of missing data (e.g. unbalanced panels). We can avoid this by using the command `e(sample)` that creates a variable that records the estimation sample, i.e. the sample used in the most recent statistical command. For this reason, the authors first run the estimation of the ordered logit model they will later use in order to fix the sample for the analysis. They use the command `quietly` because they do not want to display the results of the estimation. Since the dependent variable is categorical, as in the Richards and Gelleny paper reviewed above, the estimation command is `ologit`. We use the command `xi` to tell Stata to create dummy variables for variables preceded by an i. prefix. For example, by adding *i.year* to the list of dependent variables, Stata will create year dummy variables. We regress the variable measuring women's economic status (*wecon*) on the log of per capita GDP (*lngdpconstpc*), the level of democracy (*polity2*), trade openness (*trade*), FDI levels (*fdiinstocktogdp*), the lagged values (using command `l.`) of the geographically weighted

measures of women's economic status using both FDI stocks (*l.weconfdiinstockslrowst*) and bilateral trade flows (*l.wecontradeslrowst*) as weights, and, finally, the year dummies (*i.year*) and the region dummies. Note that the authors could have listed the regional dummies one by one (*reg_eap reg_eca reg_lac reg_mena reg_na reg_sa reg_ssa reg_we*),[81] but they preferred to use the suffix *, which is a shortcut that tells Stata to use all the variables that have the same root (in this case "*reg_*").

```
xi: quietly ologit wecon
lngdpconstpc polity2 trade
fdiinstocktogdp
l.weconfdiinstockslrowst
l.wecontradeslrowst i.year reg_*,
robust cluster(country)
```

We can summarize the variables of interest to obtain the number of observations, mean, standard deviation, and minimum and maximum values of each variable using the command `summarize`, here abbreviated as `su`. Note that we are asking Stata to use the observations that were included in the previous regression and defined as the sample we will work with throughout the entire paper.

```
xi: su wecon lngdpconstpc polity2
trade fdiinstocktogdp
l.weconfdiinstockslrowst
l.wecontradeslrowst if e(sample)
```

Similarly we ask Stata to produce the correlation matrix between the relevant variables using the Stata command `corr`. Note that we have also included here women's social status and the trade-weighted spatial lag and FDI-weighted spatial lag variables, using the social status of women abroad.

```
xi: corr wecon lngdpconstpc polity2
trade fdiinstocktogdp
l.weconfdiinstockslrowst
l.wecontradeslrowst wosoc
l.wosocfdiinstockslrowst
l.wosoctradeslrowst if e(sample)
```

We repeat the first quiet regression using women's social status (*wosoc*) instead of economic status (*wecon*) as the dependent variable, and we summarize the variables related to women's social status to include them in Table 2 of the paper.

```
xi: quietly ologit wosoc
lngdpconstpc polity2 trade
fdiinstocktogdp
l.wosocfdiinstockslrowst
l.wosoctradeslrowst i.year reg_*,
robust cluster(country)
```

```
xi: su wosoc l.wosocfdiinstockslrowst
l.wosoctradeslrowst if e(sample)
```

Step 3: Estimate the results for women's economic rights

Table 4 in the paper shows estimates for all countries (columns 1–3) and developing countries only (columns 4–6). The estimator used is ordered logit in models 1–2 and 4–5 and system-GMM in models 3 and 6. Models 1–2 and 4–5 contain regional dummy variables, while models 3 and 6 contain country-fixed effects. Year-specific fixed effects are always included.

As before, we use the command `ologit` to estimate the ordered logit model. We ask Stata to create and include year dummy variables using `xi` and `i.year`. We regress women's economic status (*wecon*) on the given independent variables and ask Stata to estimate robust standard errors using the country as the cluster by adding the option `robust cluster(country)`.[82]

```
xi: ologit wecon lngdpconstpc
polity2 trade fdiinstocktogdp
l.weconfdiinstockslrowst
l.wecontradeslrowst i.year reg_*,
robust cluster(country)
```

We repeat the procedure including a lag of the dependent variable (*l.wecon*). Results in column 2 of Table 4 in the paper are found by running the following command:

```
xi: ologit wecon l.wecon
lngdpconstpc polity2 trade
fdiinstocktogdp
l.weconfdiinstockslrowst
l.wecontradeslrowst i.year reg_*,
robust cluster(country)
```

However, when we include a lag of the dependent variable, our panel becomes dynamic. By construction, the unobserved country-level effects are correlated with the lagged dependent variable, making standard estimations like the one above inconsistent. Arellano and Bond (1991) and Arellano and Bover (1995) derived a consistent GMM estimator for the parameters of this type of model. The Stata command for this procedure is `xtabond2`.[83] There are other closer command versions `xtabond` and `xtdpd`, but the explanation of the differences between them is beyond the scope of this material. Note that `xtabond2` is not an official Stata command, but a free contribution to the research community (see Roodman, 2009). To install it, type `ssc install xtabond2, replace` in Stata. If you do not want to install `xtabond2`, you can use the command `xtabond` and get similar results. See Stata help for `xtabond`. An interesting feature of the `xtabond2` is that it allows you to determine the variables you would like to include in the GMM estimation as instrumental variables (IV).

```
xi: xtabond2 wecon l.wecon
lngdpconstpc polity2 trade
fdiinstocktogdp
l.weconfdiinstockslrowst
l.wecontradeslrowst i.year,
robust iv(lngdpconstpc polity2 trade
fdiinstocktogdp i.year)
gmm(l.wecon l.weconfdiinstockslrowst
l.wecontradeslrowst, lag (2 8))
```

The results presented in Table 4 of the paper are those generated by the command `xtabond2`.

We can repeat the same procedure but considering a sub-sample of developing countries only. We will use the command `preserve` to tell Stata that we will modify the data but that we want the programme to keep (preserve) the original dataset to eventually recover it (with the command `restore`):

```
preserve
```

We only keep those countries that are not high-income OECD countries (`inc_highoecd==0`) using the command `keep`:

```
keep if inc_highoecd==0
```

We run the same estimation procedure we applied for columns 1–3 to get results in columns 4–6.

```
xi: ologit wecon lngdpconstpc
polity2 trade fdiinstocktogdp
l.weconfdiinstockslrowst
l.wecontradeslrowst i.year reg_*,
robust cluster(country)
```

```
xi: ologit wecon l.wecon
lngdpconstpc polity2 trade
fdiinstocktogdp
l.weconfdiinstockslrowst
l.wecontradeslrowst i.year reg_*,
robust cluster(country)
```

```
xi: xtabond2 wecon l.wecon
lngdpconstpc polity2 trade
fdiinstocktogdp
l.weconfdiinstockslrowst
l.wecontradeslrowst i.year, robust
iv(lngdpconstpc polity2 trade
fdiinstocktogdpi.year) gmm(l.wecon
l.weconfdiinstockslrowst
l.wecontradeslrowst, lag (2 8))
```

We recover the original dataset (both developing and OECD countries) by typing:

```
restore
```

Step 4: Estimate results for women's social

To generate Table 5 in the paper, we repeat Step 3 replacing the economic rights variables (*wecon*) by the social rights variable (*wosoc*):

```
** All countries

xi: ologit wosoc lngdpconstpc
polity2 trade fdiinstocktogdp
l.wosocfdiinstockslrowst
l.wosoctradeslrowst i.year reg_*,
robust cluster(country)

xi: ologit wosoc l.wosoc
lngdpconstpc polity2 trade
fdiinstocktogdp
l.wosocfdiinstockslrowst
l.wosoctradeslrowst i.year reg_*,
robust cluster(country)

xi: xtabond2 wosoc l.wosoc
lngdpconstpc polity2 trade
fdiinstocktogdp
l.wosocfdiinstockslrowst
l.wosoctradeslrowst i.year,
robust iv(lngdpconstpc polity2 trade
fdiinstocktogdp i.year) gmm(l.wosoc
l.wosocfdiinstockslrowst
l.wosoctradeslrowst, lag (2 8))

** Developing countries only

preserve

keep if inc_highoecd==0

xi: ologit wosoc lngdpconstpc
polity2 trade fdiinstocktogdp
l.wosocfdiinstockslrowst
l.wosoctradeslrowst i.year reg_*,
robust cluster(country)

xi: ologit wosoc l.wosoc
lngdpconstpc polity2 trade
fdiinstocktogdp
l.wosocfdiinstockslrowst
l.wosoctradeslrowst i.year reg_*,
robust cluster(country)

xi: xtabond2 wosoc l.wosoc
lngdpconstpc polity2 trade
fdiinstocktogdp
l.wosocfdiinstockslrowst
```

```
l.wosoctradeslrowst i.year, robust
iv(lngdpconstpc polity2 trade
fdiinstocktogdp i.year) gmm(l.wosoc
l.wosocfdiinstockslrowst
l.wosoctradeslrowst, lag (2 8))

restore
```

Step 5: Estimate the results for women's economic and social rights in middle- and low-income countries

We repeat the steps above for the sub-sample of low- and middle-income countries (Table 6 in the paper) using `keep if inc_low==1` and `keep if inc_middle==1`. Do not forget to use the commands `preserve` and `restore` to avoid modifying your original dataset. Note that all estimations in this step include the lagged dependent variable (*l.wecon*).

```
** Economic rights
* Low-income countries only

preserve

keep if inc_low==1

xi: ologit wecon l.wecon
lngdpconstpc polity2 trade
fdiinstocktogdp
l.weconfdiinstockslrowst
l.wecontradeslrowst  i.year  reg_*,
robust cluster(country)

xi: xtabond2 wecon l.wecon
lngdpconstpc polity2 trade
fdiinstocktogdp
l.weconfdiinstockslrowst
l.wecontradeslrowst i.year, robust
iv(lngdpconstpc polity2 trade
fdiinstocktogdp i.year) gmm(l.wecon
l.weconfdiinstockslrowst
l.wecontradeslrowst, lag (2 8))

restore

* Middle-income countries only

preserve

keep if inc_middle==1

xi: ologit wecon l.wecon
lngdpconstpc polity2 trade
fdiinstocktogdp
l.weconfdiinstockslrowst
l.wecontradeslrowst i.year reg_*,
robust cluster(country)
```

```
xi: xtabond2 wecon l.wecon
lngdpconstpc polity2 trade
fdiinstocktogdp
l.weconfdiinstockslrowst
l.wecontradeslrowst i.year, robust
iv(lngdpconstpc polity2 trade
fdiinstocktogdpi.year) gmm(l.wecon
l.weconfdiinstockslrowst
l.wecontradeslrowst, lag (2 8))

restore

** Social rights
* Low-income countries only

preserve

keep if inc_low==1

xi: ologit wosoc l.wosoc
lngdpconstpc polity2 trade
fdiinstocktogdp
l.wosocfdiinstockslrowst
l.wosoctradeslrowst i.year reg_*,
robust cluster(country)

xi: xtabond2 wosoc l.wosoc
lngdpconstpc polity2 trade
fdiinstocktogdp
l.wosocfdiinstockslrowst
l.wosoctradeslrowst i.year, robust
iv(lngdpconstpc polity2 trade
fdiinstocktogdp i.year) gmm(l.wosoc
l.wosocfdiinstockslrowst
l.wosoctradeslrowst, lag (2 8))

restore

* Middle-income countries only

preserve

keep if inc_middle==1

xi: ologit wosoc l.wosoc
lngdpconstpc polity2 trade
fdiinstocktogdp
l.wosocfdiinstockslrowst
l.wosoctradeslrowst i.year reg_*,
robust cluster(country)

xi: xtabond2 wosoc l.wosoc
lngdpconstpc polity2 trade
fdiinstocktogdp
l.wosocfdiinstockslrowst
l.wosoctradeslrowst i.year, robust
iv(lngdpconstpc polity2 trade
fdiinstocktogdp i.year) gmm(l.wosoc
l.wosocfdiinstockslrowst
```

```
l.wosoctradeslrowst, lag (2 8))

restore
```

5.5 Discussion of findings and limitations of the analysis

The authors studied whether stronger women's rights abroad translate into stronger economic and social rights in the home country via international trade and FDI linkages. On the one hand, the authors find evidence of spillover effects working via trade links for both women's economic and social rights: $w^{trade} y_{kt-1}$ is statistically significant across most of the estimated models. This result holds for all the sub-samples except the one restricted to low-income countries. On the other hand, the authors only find weak and limited evidence of spillover effects via FDI links for women's economic or social rights: $w^{FDI} y_{kt-1}$ is statistically significant only in a few cases. They also find that general trade openness ($trade_{it}/GDP_{it}$) improves women's economic and social rights in the home country whereas general FDI openness (FDI_{it}/GDP_{it}) contributes to the improvement of social rights, but not in developing countries.

Despite providing insightful evidence on the beneficial impact of trade openness on women's rights, the paper has a few limitations. First, it cannot provide any conclusion on whether the improvement in women's economic and social rights in absolute terms translates into greater gender equality in rights because the analysis does not consider any measure of men's economic and social rights. Second, it does not provide any evidence supporting the assumption that improved rights for women leads to improved material outcomes for them.

From a technical point of view, most of the caveats we discussed for Richards and Gelleny (2007) – including the possibility of establishing causality, the problems of defining the variables, and of controlling for the many dimensions of heterogeneity, etc. – apply as well to Neumayer and de Soysa (2011).

6 Conclusions

This module discussed the macroeconomic approach that employs country-level data to study the relationship between trade and gender. Since the increased level of globalization starting in the 1990s and the liberalization policies carried out by many developing countries in the following decade, economists' primary concern has been to study the relationship between trade, growth,

and poverty and income inequality. Subsequently, economists also started to be interested in establishing the links between trade liberalization and gender inequality. The basic idea is that trade has gender consequences because it brings structural transformations that have different repercussions on men and women depending on the role they play in the economy as a whole.

Although most studies have focused on the effects of trade on labour market outcomes (i.e. women's share of employment in the manufacturing sector, the gender wage gap, etc.), some studies have analysed if and how trade, and more broadly globalization, can contribute to the empowerment of women. For example, the papers reviewed in this module represent a cross-country assessment on how trade openness in practice and trade orientation can improve women's status in political, economic, and social terms. In terms of trade openness in practice, the papers analysed here seek to capture the level of liberalization of countries by measuring the amount of trade flows; in terms of trade orientation, the aim is to capture the spillover effect on women's rights from one country to another by focusing on a country's main trading partners rather than the volume of trade. The trade analysis can be extended to include other measures of trade openness, such as changes in tariffs that, however, do not necessarily imply adjustments in the volume of exports and imports. As regards the status of women, the most widely used sources of data are the United Nations *Human Development Report*,[84] the World Bank's WDI, and the CIRI dataset (Cingranelli and Richards, 2014). There are also more recent sources of information on gender-related outcomes collected at the macroeconomic level, such as the *Global Gender Gap Report* launched by the World Economic Forum[85] in 2006 or the Demographic and Health Surveys Program.[86] Each trade- and gender-related measure has its advantages and shortcomings and there is no universal rule as to which measure you should choose for your own research. One point you should be aware of, however, is that the most appropriate measure of trade and gender inequality to employ is highly dependent on the scope of your study.

For the purpose of this module, the method reviewed is based on panel data econometric techniques that are useful when dealing with country-level data. The main characteristic of panel data is that they contain observations for a number of variables over a specified timeframe, thereby allowing us to assess the effect of a potential explanatory or independent variable on the dependent variable of interest over time (e.g. the impact of trade liberalization on the female labour force participation rate). This module described and compared two estimation methods by means of which we can estimate panel data models: the fixed-effects and random-effects models. Additionally, the module introduced dynamic panel data models that are used when we want to include among our explanatory variables (at least) the first-time lagged values of the dependent variable, thereby allowing us to control whether the dependent variable at the time of analysis is influenced by its past value. When following these methods, the macroeconomic approach becomes an *ex-post* analysis whereby we can investigate the effect of trade and gender after changes in a country's trade patterns have taken place.

REFERENCES

Abouharb MR and Cingranelli D (2006). The human rights effects of World Bank structural adjustment lending, 1981–2000. *International Studies Quarterly*, 50 (2): 233–262.

Allison PD (2009). Fixed effects regression models. Quantitative Applications in the Social Sciences Series Vol. 160. Sage Publications, Thousand Oaks CA.

Arellano M and Bond S (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58 (2): 277–297.

Arellano M and Bover O (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68 (1): 29–51.

Baldwin RE (1984). Trade policies in developed countries. In: Jones RW and Kenen PB, eds. *Handbook of International Economics* 1. North Holland, Amsterdam.

Baliamoune-Lutz M (2007). Globalisation and gender inequality: Is Africa different? *Journal of African Economies*, 16 (2): 301–348.

Becker GS (1959). *The Economics of Discrimination*. The University of Chicago Press, Chicago.

Berik G., van der Meulen Rodgers Y and Zveglich J (2004). International trade and gender wage discrimination: Evidence from East Asia. *Review of Development Economics*, 8 (2): 237–254.

Black SE and Brainerd E (2002). Importing equality? The impact of globalization on gender discrimination. NBER Working Paper No. 9110. National Bureau of Economic Research, Cambridge MA.

Busse M and Spielmann C (2006). Gender inequality and trade. *Review of International Economics*, 14 (3): 362–379.

Bussmann M (2009). The effect of trade openness on women's welfare and work life. *World Development*, 37 (6): 1027–1038.

Çağatay N (2001). *Trade, Gender and Poverty*. UNDP, New York.

Cameron C and Trivedi P (2005). *Microeconometrics: Methods & Applications*. Cambridge University Press, Cambridge.

Cingranelli DL and Richards DL (2005, 2009, 2014). The Cingranelli-Richards (CIRI) Human Rights Dataset. Available at: *http://www.humanrightsdata.com/p/data-documentation.html.*

Fontana M (2008). The gender effects of trade liberalization in developing countries: A review of the literature. In: Bussolo M and De Hoyos R, eds. *Gender Aspects of the Trade and Poverty Nexus: A Micro-Macro Approach*. Palgrave Macmillan, New York.

Gaddis I and Pieters J (2012). Trade liberalization and female labour force participation: evidence from Brazil. IZA Discussion Paper No. 6809. Institute for the Study of Labour, Bonn.

Kucera D and Milberg W (1999): Gender segregation and gender bias in manufacturing trade expansion: Revisiting the "Wood asymmetry". *World Development,* 28 (7): 1191–1210.

Marshall MG, Jaggers K and Gurr TR (2006). Polity IV data sets. University of Maryland, Center for International Development and Conflict Management.

Neumayer E and de Soysa I (2011). Globalization and women's empowerment: An analysis of spatial dependence via trade and foreign direct investment. *World Development*, 39 (7): 1065–1075. Available at: *http://www.lse.ac.uk/geographyandenvironment/whoswho/profiles/neumayer/pdf/globalization_women_empowerment.pdf.*

Oostendorp RH (2009). Globalization and the gender wage gap. *World Bank Economic Review*, 23 (1): 141–161.

Richards D and Gelleny R (2007). Women's status and economic globalization. *International Studies Quarterly*, 51 (4): 855–876. Available at: *https://sites.google.com/site/drdavid68.*

Roodman D (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, 9 (1): 86–136.

Schober T and Winter-Ebmer R (2011). Gender wage inequality and economic growth: Is there really a puzzle? *World Development*, 39 (8): 1476–1484.

Schumacher D (1984) North-South trade and shifts in employment. *International Labour Review*, 123 (3): 333–348.

Seguino S (2000). Gender inequality and economic growth: A cross-country analysis. *World Development*, 28 (7): 1211–1230.

Seguino S (2011). Gender inequality and economic growth: A reply to Schober and Winter-Ebmer. *World Development*, 39 (8): 1485–1487.

Tejani S and Milberg W (2010). Global defeminization? Industrial upgrading, occupational segmentation and manufacturing employment in middle-income countries. Schwartz Center for Economic Policy Analysis and Department of Economics Working Paper 2010-1. New School for Social Research, New York.

Tseloni A, Tsoukis C and Emmanouilides C (2011). Globalization, development, and gender inequality across the world: A multivariate multilevel approach. *Quantitative and Qualitative Analysis in Social Sciences*, 5 (1): 1–25.

UNCTAD (2009). Foreign direct investment database. UNCTAD, Geneva. Available at: *http://unctadstat.unctad.org/wds/ReportFolders/reportFolders.aspx?sRF_ActivePath=P,5,27&sRF_Expanded=,P,5,27.*

UNDP (2014). *Human Development Report 2014. Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience.* United Nations Development Programme, New York. Available at: *http://hdr.undp.org/en/2014-report/download.*

UNIFEM (2008). *Who Answers to Women: Gender and Accountability. In Progress of the World's Women 2008/2009.* United Nations Development Fund for Women, New York.

Veerbeck M (2012). *A Guide to Modern Econometrics*. Fourth edition. Wiley, New York.

Wamboye E and Seguino S (2014). Gender Effects of Trade Openness in Sub-Saharan Africa. *Feminist Economist* [online].

WEF (2013). *2013 Global Gender Gap Report*. World Economic Forum, Geneva. Available at: *http://www.weforum.org/issues/global-gender-gap.*

Wood A (1991). North-South trade and female labour in manufacturing: An asymmetry. *Journal of Development Studies*, 27 (2): 817–838.

Wooldridge JM (2009). *Introductory Econometrics: A Modern Approach*. Forth edition. South-Western, Cengage Learning, Mason OH.

World Bank (2009). World Development Indicators. World Bank, Washington DC. Available at: *http://data.worldbank.org/data-catalog/world-development-indicators.*
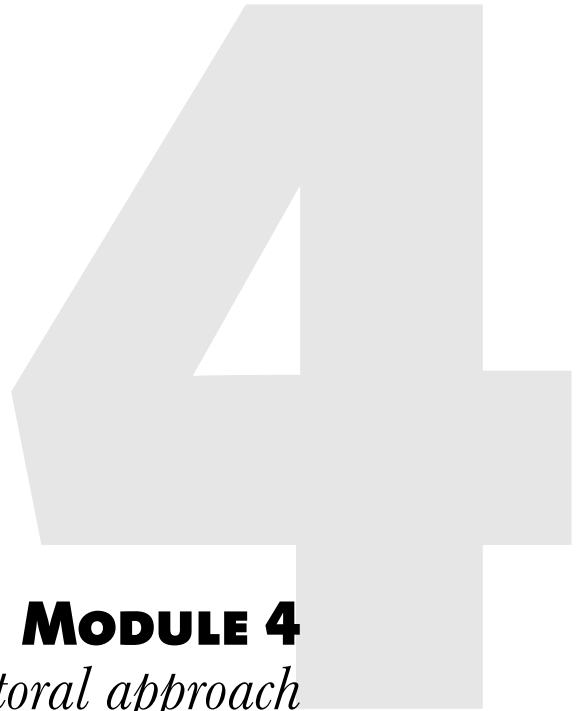
MODULE 4
The sectoral approach

4

MODULE

1 Introduction

The aim of this module is to introduce the methodologies that can be used to analyse the effects of trade on gender at the sectoral level. In particular, we will look at how the existing empirical literature investigates the impact of trade shocks and trade policies on women engaged in specific sectors and industries of the economy. As mentioned in Module 1 of this volume, while there is value to qualitative studies on trade and gender (e.g. Shayo, 2012), the focus here is on a quantitative analysis.

Trade shocks and trade policies will affect different sectors in different ways. Some sectors will contract, as they will not be able to compete with imports, and others will expand as a result of trade-led specialization. This implies that resources, and in particular labour, will have to move from one sector to another. However, it is important to note that this process is not automatic, but is rather influenced by labour market frictions,[87] which vary across countries and sectors and shape countries' patterns of specialization.[88] Sector- and country-specific labour market frictions also imply that unemployment rates vary by sector and country. As mentioned in Module 1 of Volume 1 of this teaching material, women also tend to be segregated into fewer economic sectors – what we have called "horizontal gender segregation". Therefore, it is important to also evaluate the gender effects of trade at the sectoral level. With this type of study, we can estimate, for instance, export premiums or mobility costs by gender, and explore whether trade affects women more than men. We can also study indirect effects on important issues such as domestic labour-sharing among members of the household, investment in children's education, and even gender selection at birth.

There has been a recent effort in economic literature to develop trade models that explore how the correlation between comparative advantage and labour market frictions at the sectoral level can explain the heterogeneous impact of trade on unemployment and gender inequality. Our objective here is not to review these new theoretical developments, but to look at some of the empirical studies that document the evidence these models try to explain.

Section 2 of this module is a summary of studies on trade and gender using the sectoral approach. In particular, we briefly introduce the literature on global value chains, which represents an additional method of exploring the relationship between trade and gender at the sectoral level

beyond the study in the hands-on application, examined later in this module. Section 3 provides the intuition behind some of the methodologies applied in the empirical sectoral studies. For the hands-on application in Section 4, we have selected a paper that uses a quasi-experimental approach to study how labour income opportunities for women in a non-traditional exporting sector may affect the time women and men spend on housework. Some final remarks are offered in Section 5.

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

- Apply the sectoral approach to the research on the relationship between trade and gender;
- Review and summarize the literature exploiting the sectoral approach to study the relationship between trade and gender, including the literature on global value chains;
- Identify the differences between the sectoral approach and the microeconomic and macroeconomic approaches presented in the previous modules;
- Understand and describe the difference between truncated and censored data;
- Understand the econometric techniques for the treatment of truncated and censored data, such as the Tobit and Heckman sample selection models;
- Replicate, using Stata, the results of the paper by Newman (2002) titled "Gender, Time Use, and Change: The Impact of the Cut Flower Industry in Ecuador".

2 Review of the literature

This section discusses a few studies that have used the sectoral approach to analyse the interlinkages between trade and gender inequality. In particular, the papers cited here look at the agricultural sector – where women represent the majority of casual and seasonal workers – and the garment sector – where women are mostly employed as subcontractors and home-based workers. This section is a non-exhaustive survey of the literature but it examines a few interesting papers that you may wish to add to your reading list.

We start by presenting the literature on global value chains, which includes a wide selection of detailed studies on trade and gender inequality at the sectoral level. The basic idea of this strand of literature is that developing countries should take advantage of global value chains to improve their economic performance and thus achieve better labour conditions or social improvements.

Participation in global value chains may on the one hand incentivize firms to produce higher quality/value-added goods, and thus foster the use of a more skilled and formalized labour force; on the other hand, firms may react to increasing global competitive pressures by cutting labour costs and increasing the flexibility of the work force. In the latter case, women, among other weaker groups of the population, would be the most affected because they represent a large portion of irregular and informal workers, and because gender constraints often limit their ability to access work opportunities in global production (Barrientos and Kabeer, 2004). Overall, the aim is to understand how both firms and workers can stand to benefit from increased participation of developing countries in global value chains (University of Manchester, 2010).

Barrientos (2013) offers a gender perspective to the analysis of global value chains. The author focuses on global value chains in the cocoa sector of Ghana and India and finds that the participation of female cocoa farmers and workers in global production contributes to their empowerment and to more sustainable and quality production. According to the study, women in both countries have been long relegated to household activities and forms of casual work due to gender social norms and practices. However, they play a relevant role in the production of quality cocoa because their work is mostly concentrated in activities (early plant care, fermentation, and dying) that are considered crucial to ensure good yields, thereby attracting foreign chocolate companies. Another paper on trade and gender inequality in the context of global value chains is Tallontire *et al.* (2005), who focus on the African horticultural sector, where women represent the bulk of insecure seasonal and casual workers. The paper explores whether ethical trading practices (codes of conduct covering employment conditions, and environmental and social standards) applied by European buyers in the horticultural value chain may reach women workers, among others, and whether those practices contribute to the improvement of their working conditions. The horticultural sector is also the focus of Maertens and Swinnen (2009), who present the various mechanisms through which women are directly affected by the emergence of modern supply chains. They also review existing empirical evidence and present new quantitative evidence for the high-value horticulture supply chains in Senegal. Their findings suggest that the growth of modern horticulture supply chains has been associated with direct beneficial effects on rural women, and that it has reduced gender inequalities in rural areas.

With regard to the textiles and garment sector, the integration of export-oriented firms in global value chains has been found to worsen the working conditions of women. Dedeoğlu (2010) studies the case of Turkey, where garment exporters have increasingly relied on informal labour through subcontracted and home-based female workers as a strategy to reduce production costs. On a similar note, Rossi (2001) finds that in the garment sector of Morocco, women, who generally have less bargaining power, are mostly engaged to perform unskilled activities, including packing and loading. In conclusion, firms in the low-value segments of global value chains, which are mostly concentrated in developing countries, make use of informal and low-paid jobs, for which female workers are usually preferred, to compete in foreign markets.

We have already mentioned a couple of other types of sectoral studies in Section 2.3 of Module 1 of this volume. Here we include a few more, including Porto *et al.* (2011), who look at how the internal structure of agricultural export markets and the level of competition affect poverty and welfare in rural areas of Africa. They conduct twelve case studies and find that in nine of the twelve simulations the increased competition has a larger positive income effect on male-headed households than on female ones. This is not surprising, given that women normally face entry barriers to participation in cash crop production (Vargas-Hill and Vigneri, 2011). Ackah and Aryeetey (2012) instead focus on the cocoa sector. They assess whether cocoa production, and therefore the income generated by it, is controlled by males and whether this in turn causes gender inequalities to be reinforced by the promotion of cash cropping in rural areas. They find that this is not the case, at least not for cocoa in Ghana.

Of the three methodological approaches described in this volume, the sectoral approach is perhaps the most heterogeneous from a technical point of view. We have seen that sectoral studies on trade and gender can be carried out with a gender analysis of global value chains through simulations or with more traditional econometric techniques. The paper reviewed in Section 4 employs yet another technique based on quasi-experimental data. If your research question on trade and gender requires the adoption of a sectoral approach, you should also think about the most appropriate technique to use.

3 Methodological approach

While the microeconomic approach explained in Module 2 in this volume was taken from the trade and poverty literature, and the macroeconomic approach described in Module 3 was mainly based on the growth, trade openness, and women's empowerment literature, empirical analyses on the relationship between trade and gender at the sectoral level can be carried out in many different ways.

Therefore, our focus in this section will be on providing the intuition of a series of econometric methods that are often used in case studies, but without exhausting the toolkit available to the researcher. These methods relate to the statistical treatment of truncated and censored data as described below.

3.1 Truncation and censoring

We define a variable as truncated or censored when we cannot observe all the possible values that this variable takes. Specifically, a variable *Y* is censored when we only know the true value of *Y* for a restricted range of observations. Values of *Y* are in a certain range and reported as a single value or there is significant clustering around a particular value, for instance zero. An example of a censored variable is when we consider data on consumers and prices paid for a good: if a consumer's willingness to pay for a particular good is negative, we will have observations with consumers' information but not the "real" price for that good, as price observations are censored at zero.

On the other hand, a variable is truncated when we only observe values of *X* in case *Y* is not censored. In this case, we do not have a full sample for {*Y*,*X*} as we exclude observations based on the characteristics of *Y*. The truncation is a result of sampling only part of the distribution of the outcome variable. A variable can be truncated because of the survey design (you only have a sample of women who work) or because of a particular incidence (if you are studying the wage offered to married women, you only have wage information for those women who actually are in a job). Truncated data differ from censored data in that for the latter we still observe values for *X* when *Y* is censored.

3.2 Tobit model[89]

The Tobit model is a censoring model applied to a linear model with normal residuals. The textbook example of a Tobit model is that of the labour supply of married women. In fact, labour supply

is a two-stage decision process. In the first stage, a woman has to decide whether or not to work (is the salary higher than her reservation wage?).[90] This is a probit model[91] because her participation decision is a binary outcome variable (0 = no work, 1 = work). If she decides to work, the second decision is about the number of hours she is prepared to work. This can be considered as a linear regression model where several factors can explain how many hours (a continuous variable) she decides to work. Both decisions are related: factors that make a married woman more likely to participate in the labour market tend to make her work more hours. Typically, as regressors we would include variables such as education level, non-labour income, spouse's income, number of children, general economic conditions. However, to estimate the labour supply, we require information on wage offers and we do not observe this information for those women who are not working. Moreover, a wage offer is also likely to be related to unobservable characteristics that also affect the decision to work. If we use only the observed wages and labour supply decisions to estimate the regression, the estimated coefficient will be inconsistent under ordinary least squares because of the selection bias.[92] The Tobit model proposes an alternative to this, as it is a combination of a linear regression model for the estimation of the variables that influence the number of hours supplied and a probit model for the estimation of the likelihood for an individual to participate in the labour market. This model, which uses a full sample (i.e. both women who work and those who do not), can be estimated using maximum likelihood. Under normality, it provides consistent estimators.

3.3 Heckman sample selection model

The Heckman (1979) sample selection model is a type of Tobit model.[93] In the context of our example, the equation that estimates the labour supply of women (or the wage paid to those who work) is:

$$w_i = \beta X_i + u_i \qquad (1)$$

where w_i represents women's (hourly) wages, X_i is a set of individual characteristics, and u_i is the error term. However, there is a second equation that determines if a person is willing to work or not. This second equation is called the sample selection equation and it takes the following form:

$$h_i = \gamma_i (w_i - w_i^*) = \pi Z_i + \varepsilon_i \qquad (2)$$

Here, the reservation wage w_i^* is the minimum wage at which individual *i* is willing to work and

h_i is a dichotomous variable that takes values 0 or 1 depending on whether or not the individual decides to work. On the one hand, if the wage offered w_i is below the individual's reservation wage w_i^*, the individual chooses not to work and the indicator variable for work h_i takes the value zero. On the other hand, if the wage that is offered exceeds the individual's reservation wage, the individual decides to work and the indicator function h_i takes value 1. This decision process can be explained by a set of variables Z_i influencing women's decision to work or not; ε_i is the error term.

The Heckman model assumes:

- $(\varepsilon, u) \sim N(0, 0, \sigma_\varepsilon^2, \sigma_u^2, \rho_{\varepsilon u})$: both error terms ($\varepsilon$ and u) are normally distributed, with mean equal to 0 and variances σ_ε^2 and σ_u^2, respectively. The error terms are also correlated and $\rho_{\varepsilon u}$ indicates the correlation coefficient;
- (ε, u) is independent of X and Z: the error terms are independent of both sets of explanatory variables.

The key problem that we try to address with the Heckman sample selection model is that by regressing wages on characteristics (equation (1)) we are only looking at women in employment. Therefore, we are not observing the population as a whole but only that of women who are working – i.e. we are only observing the wages for working women. In this sense, the dependent variable in equation (1) above, w_i, is truncated. Consequently, the results will tend to be biased (sample selection bias). This problem comes from the fact that the error term ε is restricted to be above a certain value ($\varepsilon_i > -Z_i\pi$): those individuals who do not satisfy this value are excluded from the regression. Heckman showed that this can be approached as an omitted variable problem where $E[u_i | \varepsilon_i > -Z_i\pi]$ is the "omitted variable". An estimate of the omitted variable would solve this problem and at the same time eliminate the sample selection bias. This involves following two interconnected steps:

- In the first step, the researcher develops a model, based on economic theory, to estimate the probability for an individual to work (probit model) – the so-called "selection equation";
- In the second step, the researcher corrects for self-selection by incorporating a transformation of these predicted individual probabilities as an additional explanatory variable in the wage equation (1).

For this procedure to work, however, we need at least one variable with a significant regression coefficient in the selection equation (the first-step

one) that is not correlated with the error term of the second-step equation of interest (an "instrument" or so-called "exclusion restriction"). The other key assumption in this estimation method is that the errors in the two equations are jointly normal. If this is not the case, the estimator is generally inconsistent. Semi-parametric and other robust alternatives can be used in such cases.

3.4 Censored least absolute deviation model[94]

The previous two models require that specific assumptions on the residuals must hold. Unlike those standard estimators, the censored least absolute deviation (CLAD) estimator is robust to heteroskedasticity and is consistent and asymptotically normal for a wide class of error distributions. In the paper analysed in the hands-on application below, as the residuals fail the normality test, the author uses the CLAD method to estimate the censored model.

The CLAD estimation method was proposed by Powell (1984). In its linear model version, the method of least absolute deviations produces regression coefficient estimates by minimizing the sum of absolute residuals:

$$\sum_{i=1}^{n} |y_i - \max(0, x_i'\beta)| \qquad (3)$$

Chay and Powell (2001) explain the intuition of the methodology. CLAD is similar to a typical regression model but in the context of censored data. When you run a standard linear regression model, you are looking at sample mean relationships between variables y and x. In CLAD, you are looking at the sample median (instead of the mean) relationship between the variables in a context where the dependent variable is censored.

Let us consider a latent variable model where the variable y is only observed when it is positive:

$$y^* = x\beta + \varepsilon \qquad (4)$$

$$y = max(0, y^*) \qquad (5)$$

Assume that conditional on x the error term ε has a median of zero: $med(\varepsilon|x)=0$. This assumption implies that the median of y conditional on x is $x\beta$ if $x\beta > 0$ and zero if $x\beta < 0$. Then, because not all true values of the dependent variable y^* are observed, we cannot directly use the least absolute deviation method[95] to estimate the unknown coefficients. The CLAD estimator applies extra censoring (e.g. applies censoring from above if the outcome is already censored from below) to a least absolute deviations (median regression) estimator. Unlike least squares

regression, least absolute deviations regression does not have an analytical solving method. Therefore, an iterative approach is required. For this reason, and as we will see in the hands-on application in this module, the CLAD command needs to be run more times before arriving at the final estimation output.

One of the solving methods is the iterative linear programming algorithm proposed by Buchinsky (1991). This is the procedure used by the command `clad` in Stata. The method consists of estimating successive quantile regressions,[96] dropping in each estimation the observations for which the predicted value of the dependent variable is less than the censoring value (in our example, zero).[97] The procedure is stopped when no negative predicted values are obtained in two consecutives estimations.

4 Hands-on application: "Gender, time use, and change: Impacts of the cut flower industry in Ecuador" (Newman, 2002)[98]

4.1 Context and overview

Newman (2002) studies women's household time allocation in a context where economic reforms have led to the growth of a non-traditional agricultural export sector, specifically the Ecuadorian cut flower industry. The paper uses quasi-experimental data from Ecuador to understand the impact of an increase in women's employment opportunities in the cut flower industry on household paid and unpaid labour allocation. For this purpose, the paper compares data from the cut flower industry area (with high demand for female labour) with data from areas that are culturally and ecologically similar to the previous one, but that have been less influenced by the boom of the non-traditional agricultural exporting sector.

The paper addresses two main questions concerning the changes in household time allocation as a result of the expansion of the cut flower industry. The first is whether women who work in the flower industry are working more hours as a result of combining market labour with unpaid household labour. The second is whether some responsibilities for unpaid household labour have been shifted from female to male members of the household.

The analysis in the paper shows that with the increase in labour market opportunities for women, women's total time spent in the labour market remains the same, while men's time in unpaid labour increases. Table 16 provides an overview of Newman's paper.

Table 16

Overview of Newman (2002)
Objective

The objectives of the paper are to:
• Examine the effects of women's employment in an export industry on the allocation of paid and unpaid labour within the household.
• Separate the bargaining effects of wages from the substitution effect of wages on time use:
  - Substitution effect: following a wage increase, the individual substitutes housework for paid work;
  - Bargaining effect: higher wages increase the person's ability to survive independently of the household and this reduces his/her domestic work supply.

Methodology

Three models are considered in the paper:
(a) Heckman model
  - Two-step estimation model, initially chosen to correct for sample selection;
  - Requires the identification of at least one exclusion restriction in the first-stage equation model;
  - The results of the estimation greatly depend on the choice of the exclusion restriction; to avoid this problem, the author implements the Tobit model.
(b) Tobit model
  - Single-equation model that takes into account the censoring of values at zero;
  - Allows for all the variables to be tested as possible determinants of shares of participation;
  - Requires the errors to be normally distributed, but the test for normality fails, so the author decides to use the censored least absolute deviation as an alternative methodology.
(c) CLAD model
  - Does not require any assumptions about the distribution of the error term;
  - The disadvantage of using this estimator compared to the Tobit is in precision (larger standard errors).

Sample description
• 562 households were surveyed, resulting in 2,567 individual observations. • Quasi-experimental survey in two distinct regions of northern Ecuador: Cotocachi (control group – no export flower industry) and Cayambe (treatment group – export flower industry). • The survey was modeled after the World Bank's LSMS.

Overview of Newman (2002)
Equation estimated: Tobit vs. CLAD

$$y_i^* = \beta_0 + \beta'_x\, x_i + \beta_w\, w_i + \beta'_h\, h_i + \varepsilon_i$$

$y_i = 0$ if $y_i^* \leq 0$

$y_i = y_i^*$ if $y_i^* > 0$

The model is estimated first with Tobit and then with CLAD. Data are collected at the individual level i and the sample is split in two sub-samples: men and women.

Dependent and independent variables

- The dependent variables y are the share of time spent on unpaid and paid work of men and women above 10 years of age.
- The independent variables are:
  - Household characteristics (h): regional location, number of children, ratio of females to males in the household, houshold's assets, and urban location;
  - Individuals' characteristics (x): age, education, age difference between husband and wife, educational difference between husband and wife, migrant dummy, and marital status;
  - Hourly wage (w): own and husband's/wife's.

Note: In the CLAD model, x_i is a vector of household and individual characteristics, including hourly wages.

Results

Women's employment in the non-traditional exporting industry (cut flower industry) affects the allocation of paid and unpaid labour within the household:

- Married men in Cayambe spent twice as much time on housework as did men in Cotocachi.
- This result is related to women's increased participation in the labour market and their increased bargaining power.
- Regardless of the growth of the cut flower industry, the paper finds that women worked more than men when both paid work and housework were included.

Source: Newman (2002).

4.2 Data sources

For the purpose of the analysis, a survey was designed and data collected in two regions of northern Ecuador – Cotocachi and Cayambe. These two regions are about 200 kilometres apart and are similar in their cultural and ecological characteristics. A total of 562 households were surveyed, resulting in 2,567 individual observations, covering all household members above the age of 10. The survey was modeled after the World Bank's LSMS, which we reviewed in Module 1. The survey includes detailed modules on expenditures, economic activity, health, education, fertility, and credit and savings, as well as a detailed accounting of time use. Two types of time-use data were collected to capture different time allocation habits and, accordingly, two measures were calculated. The first one is an accurate indicator for activities carried out in the last 24 hours. The problem with this measure is that it may miss unusual or irregular activities. Therefore, a second indicator was designed to capture general time dedicated to housework, rest, recreation, and work over the prior week. These data are often less precise and more subject to recall error, but they have the advantage of being less burdensome for the interviewee. The 24-hour recall data were collected only for the male and female heads of the household. The weekly data were collected for all household members interviewed.

4.3 Empirical methodology

The paper uses as a framework a model in which household decisions are derived from the maximization of the weighted sum of individuals' utilities. The survey was specially designed to calculate indicators that may influence the relative decision-making power of the individuals (Browning and Chiappori, 1998) and created utility weights based on them. In this model, wages affect the labour and domestic work supply functions directly as well as indirectly through the distribution function (through, for instance, a bargaining effect). As wage opportunities change among household members, the amount of labour supplied by each of them can be affected beyond the traditional substitution and income effects of a wage change. The bargaining effect of a higher relative wage would have an impact on domestic work supply as well as on unpaid work.

To capture this effect, the author used a quasi-experimental approach. Two groups were selected – the treatment group (households in Cayambe) and the control group (households in Cotocachi)[99] – and the survey was detailed enough to capture many differences between these two groups. However, since the experiment was applied in a real economy, there are likely to be some unobservable differences that may affect the result. When carrying out this type of analysis, the main concern is the endogeneity problem. In this context, a problem would arise if the location of

flower production was correlated with the qualities of the workforce that might also influence time allocation decisions of working individuals. However, this seems not to be the case here as the flower producers interviewed for the study reported that the characteristics of the workforce are irrelevant to their choice of location, which is strictly guided by the unique combination of ecological characteristics in Cayambe, as well as its proximity to a regional airport. If this assumption is true, then we have a relatively clean quasi-experimental setting to study the question of interest.

The paper uses different methodologies. To compare the different time allocations between the treatment (Cayambe) and control (Cotocachi) areas, the author uses average time spent on different economic activities, household tasks, and wages received by region, gender, marital status, and labour market participation (Tables 2–5 in the paper). She then compares them to check whether there are significant differences among those averages (easily performed with the command `ttest` in Stata). The paper then asks the question of which characteristics are the determinants of time allocation differences between individuals' activity choices. The paper controls for exogenous factors that might influence individuals' decisions by including variables such as human capital, age, marital status, location, and other household and regional characteristics.

The objective of the paper is to separate the wage effect from the bargaining effect. If men in the treatment group work more at home, is it because of wage differences or because women in that region have gained bargaining power through increased access to paid work opportunities in the export-oriented industry? To capture this effect, the model includes a Cayambe dummy that, under the assumption that the two samples are comparable, captures the additional effect of the presence of the flower industry. The author also controls for own individuals' wages and spouses' wages to help separate the substitution effect from the bargaining effect. Indeed, while the Cayambe dummy is more likely to capture the bargaining effect as a result of changing social norms, the wages are more likely to capture the substitution effect because theoretically they have the most direct impact on utility.

From a technical point of view, three models are used to test the paper's hypothesis. The Heckman model was originally chosen to correct for sample-selection, but a Tobit model was estimated instead. As a matter of fact, the Tobit model is preferred to the Heckman model because it allows for all the variables to be tested as possible determinants of the share of time spent by men and women on unpaid and paid work, whereas in the Heckman model, some variables need to be excluded for identification purposes without a clear theoretical guideline for making this choice. The Tobit model, however, requires errors to be normally distributed and, in this paper, the normality test fails for the sample analysed. Therefore, the paper estimates a CLAD model because it has the advantage of not requiring any assumptions about the distribution of the errors. However, the CLAD model has the disadvantage of being less precise (larger standard deviation) than the Tobit model.

4.4 Step-by-step explanation of how to do the estimations in Stata

The data file we will use for this application is called *perscv.dta*. We will use a separate do-file for each table in the paper, but you may also choose to write a single do-file that contains the commands to get all the tables and regressions included in the paper. In our case, each do-file is named after the table it generates.[100]

Step 1: Generate Table 1 in the paper – Demographics of Cayambe and Cotocachi areas

To perform the first step, we use the do-file "Table 1 Final.do". First of all, we tell Stata to clear anything that was initially stored in the memory (`clear`). We then tell the output to scroll down without requiring user assistance (`set more 1`), and we define the line size in the Stata do-file to have 132 characters (`set linesize 132`).

We then define in which directory we will work (`cd`) and what data we will use for the analysis (`use perscv`).

To keep an output generated by the do-file, we will create a log file that captures all of the output from the time you open it until you close it, regardless of how much output is produced. You can later open this file with a text editor and edit the content to include it in your paper. You start a log file with the command:

```
log using name_of_file.log
```

and close it using the command:

```
log close
```

We want to create a table that compares variables across two populations, one being the treatment group (`treat==1`) and the other the

control group (`treat==0`). One of the characteristics of Cayambe is that it has a larger number of migrants. For this reason, the author wants to see if Cayambe's subpopulation has different characteristics from that of Cotocachi, and also show a separate column for the subpopulation who are migrants (`migrant==1`) in the treatment group. To be able to tabulate the information using the command `by`, we need to first sort the data (`sort treat`). Then the tabulation is performed using:

```
by treat: tab name_variable
```

where the *name_variable* is the age group (*agegroup*), the education level (*edlevel*), marital status (*mar*), whether the household head is a female (*fhh*), if they have children less than 15 years old (*child*), if they have children less than 6 years old (*childlt6*), and the relationship to the head of the household (*rel2*). The commands always have the same structure:

```
by treat: tab name_variable
tab name_variable if treat==1 &
migrant==1
```

The first line generates the columns "Cayambe All" and "Cotocachi". The second/third line generates the column "Cayambe Migrants".

Step 2: Generate Table 2 in the paper – Use of time, by gender and marital status

Next, we use the do-file "Table 2 Final.do".[101] The first command lines are as before. Now we have a new command `svyset` that declares a survey design for the dataset and the option `strata(dominio)` that indicates that *dominio* is the name of the variable identifying the strata:

```
svyset, strata(dominio)
```

Once Stata knows about the design of the survey via the command `svyset`, you can use the `svy:` prefix using syntax that is quite similar to the non-survey versions of the commands. For example, if you use the command `svy: regress`, it is like using a regular command `regress`, but it uses the information you have provided about the survey design and does the computations taking this information into consideration. In this application, we use `svy: mean`

```
svy: mean farm pdwork comwork hwork24
totwk recre percare totime if
sex==1 , over(treat)
```

We use the command `svy: mean` with the option `over (treat)` to get the means for each variable (*farm, pdwork, comwork, hwork24, totwk, recre, percare*, and *totime*) for each subgroup in the variable *treat* (here our two distinct populations in Ecuador). The variables are the ones in Table 2 of the paper and represent different activities performed by the individual. The analysis is first done for all men (`sex==1`), and then only for married men (`if sex==1 & mar<=2`) and non-married men (`if sex==1 & mar>2`). Replacing sex by `sex==2`, we obtain the same tables as before, but this time for women.

We want to know if the estimated means are different across the two populations. For this purpose, we use the command `lincom` (in this case for all men, time spent in farm work):

```
lincom [farm]1 – [farm]0
count if farm~=. & sex==1 & treat==1
count if farm~=. & sex==1 & treat==0
```

`lincom` computes point estimates, standard errors, *t* or *z* statistics, *p*-values, and confidence intervals for linear combinations of coefficients (here the difference in the mean time spent by all men in farm activities in the two populations) after any estimation command (in this case, `lincom` will consider the calculated coefficients with the command `svy: mean`).

The command `count` provides the number of observations in the category indicated by the expression in `if`.

A similar procedure is then implemented to see if there are differences in the mean across gender groups and not only across populations. For instance, to display the means for men and women living in Cayambe you need to run the following command:

```
svy: mean farm pdwork comwork
hwork24 totwk recre percare totime
if treat==1 , over(sex)
lincom [farm]Male – [farm]Female
```

The bottom of Table 2 in the paper shows the ratio of men's to women's time in each activity. These numbers can be obtained from the means displayed with the commands above.

Step 3: Generate Table 3 in the paper – Time spent performing household tasks, by gender, marital status, and labour market participation

Here we turn to the do-file "Table 3 Final.do". We want to estimate if there is a significant difference (`ttest`) in the time spent on household tasks (*hwork24*) conditional on the spouse work-

ing (`spwks==1`) or not (`spwks==0`) across the two populations (`by(treat)`).

The do-file first looks at households where the head is a man (`sex==1`) who is married or in a free union (`mar<=2`). The file first looks at men who work and men who do not work (not used in the table because of too few observations) as well as men who work in the flower sector in the treatment region compared to men who work in any other sector in the control region. For each of them, the do-file also shows the results based on whether the spouse works (`spwks==1`) or not (`spwks==0`).

The command ttest is then used to compare the means between the two regions.

```
*Men who work

ttest hwork24 if sex==1 & mar<=2 &
rel==1 & wklast==1, by(treat)

ttest hwork24 if sex==1 & mar<=2
& rel==1 & wklast==1 & spwks==1,
by(treat)

ttest hwork24 if sex==1 & mar<=2
& rel==1 & wklast==1 & spwks==0,
by(treat)

*Men who don't work -- not used in
table because too few observations
available

ttest hwork24 if sex==1 & mar<=2 &
rel==1 & wklast==0, by(treat)

ttest hwork24 if sex==1 & mar<=2
& rel==1 & wklast==0 & spwks==1,
by(treat)

ttest hwork24 if sex==1 & mar<=2
& rel==1 & wklast==0 & spwks==0,
by(treat)

*Men who work in flowers in treat
compared to men who work in other
sectors in control

drop if wkflowrs==0 & treat==1

ttest hwork24 if sex==1 & mar<=2 &
rel==1 & wklast==1, by(treat)

ttest hwork24 if sex==1 & mar<=2
& rel==1 & wklast==1 & spwks==1,
by(treat)
```

```
ttest hwork24 if sex==1 & mar<=2
& rel==1 & wklast==1 & spwks==0,
by(treat)
```

The procedure is repeated for women to identify whether the same differences apply in the time spent on household chores across regions. Specifically, we look at households where the head is a woman as well as where the head is a man. The commands and syntax are similar to the previous ones and for this reason we do not specify them here.

Step 4: Generate Table 4 in the paper – Wages by gender, marital status, and work type

We use the do-file "Table 4 Final.do". The command we use here is summarize (`sum`). It provides mean, standard deviation, minimum, maximum, and number of observations for a variable. We can ask for more detailed information using summarize, detail (`sum, d`), which provides us with additional statistics, including skewness, kurtosis, the four smallest and four largest values, and various percentiles.

The variable we are interested in describing is wage (*jperhr*) and we want to display it by region (*treat*), marital status (*mar*), gender (*sex*), and for those who work in the flower industry versus those who work in all other sectors (*wkflowrs*).

For example, if we want the summary statistics for wages in Cayambe for women who work in the flower sector and are married, the command line is:

```
sum jperhr if treat==1 & sex==2 &
wkflowrs==1 & mar<=2, d
```

Step 5: Generate Table 5 in the paper – Average hours per week spent performing main activities, by gender and marital status

We now turn to the do-file "Table 5 Final.do", which creates the table with the number of hours spent on different activities by gender and marital status for the two regions, using the time variable recorded on a weekly basis instead of the 24-hour measure of activities. The variables of interest here are paid work (*hrsw*), housework (*hrsh*), recreation (*hrsr*), and sleeping time (*hrsl*).

The commands used here (`svy: mean` and `lincom`) are the same as those in step 2 "Table 2 Final.do".

Step 6: Generate Table 6 in the paper – CLAD and Tobit estimates of men's share of time performing housework and paid work (dependent variable: individual's share of housework)

We use the do-file "Table 6 Final.do", where we estimate the Tobit and CLAD models of the determinants of time spent on household activities for men only.

We first run the Tobit estimation for the share of time that the individuals spend on housework. To generate a Tobit model in Stata, we use the command `tobit`, followed by the outcome variable and the predictors. We also can specify the lower limit and/or upper limit of the outcome variable. While the lower limit is specified in parentheses after `ll`, the upper limit is specified in parentheses after `ul`. A Tobit model can be used to predict an outcome that is censored from above, from below, or both.

```
tobit hwsh2w age age2 educ married
widdiv sucre1 hhsize numchil ratiofm
assets urban treat migrant if sex==1
& age>=10, ll(0)
```

Here *hwsh2w* is the share of unpaid labour and is explained by a list of individual characteristics, household characteristics, the individual's own wage, and the dummy variable for the Cayambe location (*treat*). The analysis is performed only for males who are at least 10 years old. The share cannot be lower than zero, and therefore the lower limit `ll(0)` is specified.

A result not shown in the paper is the test for normality for the residuals of the Tobit model. As mentioned before, the Tobit model requires residuals to be normally distributed. The do-file performs this test by running a Tobit regression, predicting the estimated values of the dependent variable and generating the residuals (`gen res=hwsh2w-yhat`). We can then use the command `sktest` to perform a test for normality based on skewness and another one based on kurtosis, and then combine the two tests into an overall test statistic. The residuals fail the normality test, suggesting that using a Tobit model may not be the best approach in this case.

The author then runs CLAD estimations for housework (*hwsh2w*) and paid work (*pdsh2w*) for all men and married household heads. This generates columns 2–5 in Table 6 of the paper. One possibility to run the estimation is to use the command `clad`. This is not an original Stata command and therefore you need to install

it typing `net install sg153.pkg`. This will provide you with the programme for estimating Powell's (1984) CLAD model and obtaining bootstrap estimates of its sampling variance. The CLAD estimator is a generalization of the least absolute deviations estimator, which is implemented in Stata in the command `qreg`. This programme sidesteps the issue of programming analytical standard errors and provides instead bootstrapped estimates of the sampling variance. See Newton *et al.* (2000) for details on how to write the command line.

At the time of writing the paper, the command `clad` did not exist; the author therefore had to write an alternative routine in which she reproduced the algorithm explained in Chay and Powell (2001). We have left that routine for the interested reader in the do-file. Fortunately, today we have the command `clad` that simplifies the implementation of the procedure.[102] The command line is the following:

```
clad hwsh2w age age2 educ married
sucre1 hhsize numchil ratiofm assets
urban treat migrant
```

The same procedure is then repeated to generate the other columns in Table 6 of the paper.

Step 7: Generate Table 7 in the paper – CLAD and Tobit estimates of women's share of time performing housework and paid work (dependent variable: individual's share of housework)

Finally, we use the do-file "Table 7 Final.do". This estimates the Tobit and CLAD models of the determinants of time spent on household activities for women only.[103]

The commands and syntax used here are the same as those used in the previous table, the only difference being that we focus on females (`sex==2`) instead of males (`sex==1`).

4.5 Discussion of findings and limitations of the analysis

The main findings of the paper analysed in this module show that increased participation of women in the labour market has a bearing on household labour allocation. Married men in the treatment group spend double the time on housework compared to men in the control group, and this is clearly related to women's increased participation in the labour force because of the increase in employment opportunities in the non-traditional exporting sector. Women in the

treatment group, especially married women, do less housework than those in the control group. The author also controls for other determinants of time use, including household characteristics, such as the ratio of female to male members, and social characteristics, such as marital status and the attitude towards women in society. Overall, the gender impact of the growth of the cut flower industry is perceived as positive based on the author's idea that the increase in the participation of women in the labour market itself leads to cultural changes. One of the issues the study does not consider is the possibility that some of the housework is transferred to the children within the household, which might leave them with little time to go to school or to play.

From the technical point of view, most of the problems with the quasi-experimental approach arise from self-selection and sample selection, as well as the comparability issues between the treatment and control group. Quasi-experimental estimates of effects are subject to contamination by confounding variables. The lack of random assignment in the quasi-experimental design method may allow studies to be more feasible, but this also poses many challenges for the researcher in terms of internal validity. Newman (2002) makes all possible efforts to find two comparable populations and exploits the fact that the location choice for flower firms seems to be unrelated to the characteristics of the labour force and only depends on agronomic conditions and available export infrastructure.

5 Conclusions

This module used the sectoral approach to examine the effects of trade on gender. In other words, we looked at the impact of trade on women engaged in particular sectors and industries of the market as well as the non-market economy instead of looking at the individual-level (microeconomic) and aggregate-level (macroeconomic) implications of trade on gender-related outcomes. When data are available, the sectoral approach allows us to look at the shifts in a country's production and export and import structure patterns and if these have translated into changes in women's economic and/or social status. For instance, the economic reforms implemented by Ecuador in the 1990s contributed to the development of non-traditional agricultural exporting sectors, most notably the cut flower industry. Being an industry that is traditionally female-intensive, there have been relevant gender repercussions, particularly changes

in the intra-household allocation of time and tasks, which are the object of the paper reviewed in this module.

When deciding to adopt a sectoral approach to the study of trade and gender, it is thus important to conduct an *ex-ante* assessment of the concentration of women workers at the sectoral level to understand the gender patterns of employment as well as the relevance of horizontal gender segregation for the economy at hand. The impact of trade on gender at the sectoral level will depend both on the kind of structural transformations resulting from trade and on where women workers are most concentrated. For instance, if trade causes an expansion of a female-labour-intensive sector, women wage workers employed in that sector could potentially gain by, for instance, experiencing an increase in their wages; conversely, if trade causes a contraction of a female-intensive sector, women employed in that sector could be adversely affected by being displaced. Furthermore, women workers could be concentrated in non-tradable sectors, most notably non-tradable services such as health and education, and thus be unexposed to trade-related changes in the structural composition of the economy. In this case, women in the non-tradable sector could still be indirectly affected when, for example, trade results in an increasing specialization of the country towards the production of non-tradable services. One of the causes of changes in the structural composition of the economy could be a trade shock. For example, in the paper by Newman (2002), the shock is caused by economic reforms, including trade reforms, that have led to a boom in the growth of non-traditional agricultural exports, specifically cut flowers, and thus to the growth of the cut-flower industry that is traditionally a female-labour-intensive sector.

As regards the method of analysis, this module presented a series of censored and truncated regression models that are often useful to estimate empirical specifications of women's participation in the labour market. In the hands-on application, we looked at the difference in the allocation of housework responsibilities between men and women, comparing one area where there are work opportunities in the exporting sector with another where those opportunities do not exist. The technique adopted in the paper analysed is the CLAD, which offers many statistical advantages compared to the most commonly used Heckman and Tobit models. However, as for the choice of data, the most appropriate econometric tool you should use depends on the purpose of your investigation. As already

mentioned, the CLAD is one of many models that have been applied in sectoral studies. Other authors have also looked at the effect of trade on gender outcomes at the sectoral level using different methodologies. For instance, we have seen how the analysis of global value chains has been implemented to explore the relationship between trade and gender. Moreover, Nicita and Razzaz (2003) explore whether an increase in textile and apparel exports in Madagascar benefits the poor and examine its effect across gender groups by first using a propensity score. Depetris Chauvin and Porto (2011) use a methodology akin to the microeconomic approach. All these studies provide an *ex-post* assessment of the impact of trade on gender, but you can also decide to use the sectoral approach to carry out an *ex-ante* analysis. In this case, you can employ the simulation tools described in Module 1 of this volume.

REFERENCES

Ackah C and Aryeetey E (2012). Cash cropping, gender, and household welfare: Evidence from Ghana. In: Ackah C and Aryeetey E, eds. *Globalization, Trade and Poverty in Ghana*. Sub-Sahara Publishers for the International Development Research Centre, Accra.

Barrientos S (2013). Gender production networks: Sustaining cocoa-chocolate sourcing from Ghana and India. Working Paper No. 186. University of Manchester, Brooks World Poverty Institute. Available at: *http://www.bwpi.manchester.ac.uk/medialibrary/publications/working_papers/bwpi-wp-18613.pdf*.

Barrientos S and Kabeer N (2004) Enhancing female employment in global production: Policy implications. *Global Social Policy*, 4 (2): 153–169.

Browning M and Chiappori PA (1998). Efficient intra-household allocations: A general characterization and empirical test. *Econometrica*, 66 (6): 1241–1278.

Buchinsky M (1991). Methodological issues in quantile regressions. Chapter 1 of the Theory and Practice of Quantile Regressions. PhD Dissertation. Harvard University.

Cameron C and Trivedi P (2005). *Microeconometrics: Methods & Applications*. Cambridge University Press, Cambridge.

Chay K and Powell J (2001). Semiparametric censored regression models. *Journal of Economic Perspectives*, 15 (4): 29–42.

Dedeoğlu S (2010). Visible hands – invisible women: Garment production in Turkey. *Feminist Economics*, 16 (4): 1–32.

Depetris Chauvin N and Porto G (2011). Market competition in export cash crops and farm income in Africa. Available at: *http://acetforafrica.org/wp-content/uploads/2011/09/paper-depetris-porto-acet-web.pdf*.

Heckman J (1979). Sample selection bias as a specification error. *Econometrica*, 47 (1): 153–161.

Maertens M and Swinnen J (2009). Are African high-value horticulture supply chains bearers of gender inequality? Paper presented at the FAO-IFAD-ILO Workshop on Gaps, Trends and Current Research in Gender Dimensions of Agricultural and Rural Employment: Differentiated Pathways Out of Poverty. Rome, 31 March – 2 April.

Newman C (2002). Gender, time use, and change: Impacts of the cut flower industry in Ecuador. *World Bank Economic Review*, 16 (3): 375–396. Available at: *https://openknowledge.worldbank.org/bitstream/handle/10986/17212/773920JRN0200200CutoFlower0Industry.pdf?sequence=1*.

Newton HJ, Cox N, Garrett J, Pagano M, Royston JP and Weesie J (2000). Censored least absolute deviations estimator: CLAD. *Stata Technical Bulletin*, 58 (November): 13–15. Available at: *http://www.stata-press.com/journals/stbcontents/stb58.pdf*.

Nicita A and Razzaz S (2003). Who benefits and how much? How gender affects welfare impacts of a booming textile industry. Policy Research Working Paper No. 3029. World Bank, Washington DC.

Porto G, Depetris-Chauvin N and Olarreaga M (2011). *Supply Chains in Export Agriculture, Competition, and Poverty in Sub-Saharan Africa*. Centre for Economic Policy Research, London

Powell JL (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25 (3): 303–325.

Rossi A (2011). *Economic and Social Upgrading in Global Production Networks: The Case of the Garment Industry in Morocco*. University of Sussex, Institute of Development Studies, Brighton.

Shayo R (2012). Assessment of gender impacts of SADC trade protocol in selected countries. African Trade Policy Centre Work in Progress No. 88. Available at: *http://www.uneca.org/sites/default/files/publications/atpcworkinprogress88.pdf*.

Tallontire A, Dolan C, Smith S and Barrientons S (2005). Reaching the marginalized? Gender value chains and ethical trade in African horticulture. *Development in Practice*, 15 (3–4): 559–571.

University of Manchester (2010). Capturing the gains. *Development@manchester: Cross-Disciplinary Perspectives on Development Research*, 6 (October). Available at: *http://www.sed.manchester.ac.uk/idpm/research/publications/dev@man/devman-research-note-201006.pdf*.

Vargas-Hill R and Vigneri M (2011). Mainstreaming gender sensitivity in cash crop market supply chains. ESA Working Paper No. 11-08. United Nations Food and Agriculture Organization, Rome.

Wooldridge JM (2013). *Introductory Econometrics*. A Modern Approach. Fifth edition. South-Western, Cengage Learning, Mason OH.

ENDNOTES

1 Ethnographic studies, which find their origin in anthropology and have subsequently been adopted by economists to perform qualitative research, focus on a group of individuals who share the same culture. In practice, the studies consist of observing a group over a prolonged period of time to identify daily behaviours, as well as the norms, beliefs, and social structure that shape these behaviours. Ethnographic studies allow for understanding, among other issues, the factors that trigger changes in a group's culture over time (Williams, 2007).

2 This subsection and those that follow are largely based on the teaching material for the online course on trade and poverty developed by the UNCTAD Virtual Institute (UNCTAD, 2012).

3 Labour markets are defined as imperfect when they are not perfectly competitive. This is due to the scarcity or even complete lack of information for the actors involved. For example, workers might be unaware of better employment conditions elsewhere, which makes it possible for their current employer to exercise monopsony power and keep wages low.

4 While general equilibrium models study the effects that changes in demand/supply in one market bring to the markets that are connected to it, partial equilibrium models focus on one market in isolation and thus do not allow for capturing the interconnections between different sectors of the economy.

5 See Section 2 of Module 2 for further details on the difference between perfect and imperfect pass-through.

6 See the paper by Richards and Gelleny (2007) presented in Module 3.

7 See Module 1 of Volume 1 for an explanation of the gender wage gap.

8 See Module 1 of Volume 1 for a deeper discussion of the different measures of trade openness.

9 Country- and time-fixed effects are usually included in panel data regressions because they allow for netting out any country- and time-invariant, unobserved effects from the coefficients of interest. See Module 3 for further details on the fixed-effects model.

10 See Ramjoué (2010) for an analysis of the trade-growth-gender nexus.

11 Module 2 of Volume 1 describes in more detail the theoretical expectations of standard economic theory regarding the employment and wage effects of trade on women as workers. The heterodox theory comes to different conclusions.

12 In statistics, we say that a variable is endogenous when it is correlated with the error term. This can happen because of measurement errors, simultaneity, and omitted variables.

13 When two variables influence each other, we say that there is a simultaneity problem that, if not accounted for, may lead to biased results. The instrumental-variable (IV) approach is the most widely used methodology to address this problem.

14 In the context of Newman (2002), endogeneity can arise if the location of the export flower industry is correlated with unobservables in the labour market participation equation for women. If this is the case, then the estimated coefficients will be biased and inconsistent (we will discuss this in detail in Module 4).

15 Trade data are available for the following product groups: agricultural products, fuels and mining products, and manufactures.

16 The HS is an internationally standardized and recognized system to classify traded products. It was introduced in 1988 and is currently maintained by the World Customs Organization (WCO – formerly the Customs Co-operation Council), an intergovernmental organization based in Brussels. Since 1988, the HS has been revised four times. The latest version is known as HS 2012. The previous revisions were undertaken in 1996, 2002, and 2007. The HS comprises approximately 5,300 article/product descriptions.

17 The SITC is a system of classification of traded products not only on the basis of their physical properties, but also according to the stage of processing, their economic use, and their technological properties. It was developed by the United Nations with the aim to facilitate economic analysis. The SITC maintains a correspondence with the HS and is less detailed than the HS.

18 The BEC, introduced by the UN in 1970, is a three-digit classification system that groups tradable goods according to their main end use. The seven main categories are: food and beverages; industrial supplies not elsewhere specified; fuels and lubricants; capital goods (except transport equipment) and parts and accessories thereof; transport equipment and parts and accessories thereof; consumer goods not elsewhere specified; and goods not elsewhere specified (United Nations, 2002).

19 The ITC is a joint agency of the WTO and UNCTAD. Its aim is to help businesses in developing countries become more competitive in global markets, thus speeding up economic development and contributing to the achievement of the MDGs.

20 GTAP is a network of researchers and policymakers conducting quantitative analysis on international policy issues.

21 The WITS manual can be downloaded at: *http://wits.world bank.org/data/public/WITSGlobalTariffCutsandTradeSimu latorUserManual.pdf.*

22 The Swiss formula is a non-linear formula in which tariff cuts are proportionally higher for tariffs that are initially higher. For instance, a country that has an initial tariff of 40 per cent on a product will have to undertake proportionally higher cuts than a country that has an initial tariff of 20 per cent on the same product.

23 This subsection is largely based on the teaching material for the online course on trade and poverty developed by the UNCTAD Virtual Institute (UNCTAD, 2012).

24 A labour force survey is a standard household-based survey of work-related statistics. The International Labour Organization (ILO) lists the countries and territories that make their labour force surveys available online. See *http://www. ilo.org/dyn/lfsurvey/lfsurvey.home.*

25 *See http://www.ihsn.org/home.*

26 Not all the economies recorded in the database are independent countries.

27 The PISA is a worldwide study by the OECD of the scholastic performance of 15-year-old students in mathematics, science, and reading. It was first performed in 2000 and then repeated every three years. The study is conducted with a view to improving education policies and outcomes. The test is implemented in a large number of developed and developing countries.

28 The ILO defines "working poverty" as the proportion of employed persons in a household whose members are living below the $1.25/day poverty line.

29 See Subsection 4.2.2 below for further details on do-files.

30 Sections 2 and 3 and Annex A of this module are based on the teaching material for the online course on trade and poverty developed by the UNCTAD Virtual Institute (UNCTAD, 2012).

31 This assumption derives from the Stolper-Samuleson theorem, according to which there is a direct and positive relationship between changes in the price of a good and changes in the return to the factor intensively used in its production (and a direct and negative relationship with the return to the other factor). For example, an increase in the price of clothing – a relatively low-skilled labour-intensive sector – causes an increase in the wages of low-skilled workers and a decline in the wages of high-skilled workers.

32 Preliminary works of this kind include Sourabh Bikas (2009) and Cherkaoui *et al.* (2011), who focus on the case of India and Morocco, respectively.

33 The law of one price is an economic concept that states that a good must sell for the same price in all locations. Under incomplete pass-through, the law of one price does not hold.

34 Furthermore, if the law of one price holds and the country is a small producer and a small consumer, domestic demand or supply shocks do not affect domestic prices, given that supply in international markets is infinitely elastic at a fixed price (the country can buy or sell as much of the good at the international given price as it wishes).

35 This can also be seen as a reduction of the price received by exporters for each unit sold in international markets.

36 We are here considering only first-order effects, i.e. the direct impact of consumption and production price changes on household welfare. A complete assessment of the welfare implications of price changes would include the analysis of second-order effects, i.e. of the changes in the household's consumption bundle deriving from the substitution of more expensive goods with cheaper goods. In this case, we would also have to calculate elasticities of substitution, or, in other words, we would need to have more information about the degree of substitutability between goods.

37 Assuming free movement of labour across sectors, workers move from the sectors with the lowest wages to sectors with the highest wages.

38 Empirical methods (regressions and densities) are defined as non-parametric when the resulting estimators do not have any fixed functional form (structure) and depend on all the data points to return an estimate.

39 For further details on non-parametric densities and regression, see Section 9.2 of Cameron and Trivedi (2005).

40 In this case, we say that the function presents a discontinuity.

41 A likelihood function is a function of the parameters that characterize a statistical model. The maximization of the likelihood function is done to find the estimated value of the parameters for a given probability function which will make it more likely to observe the data used in the maximization.

42 The UNCTAD study is available for download at: *http://unc tad.org/en/PublicationsLibrary/osg2011d2_en.pdf.*

43 See *http://www.ine.cv* (in Portuguese only).

44 All application files are contained in the accompanying CD, and are available for download at: *http://vi.unctad.org/tag.*

45 In general, Stata will stop running at a line that it cannot execute. When a Stata do-file stops running, you need to find the error in your do-file, correct it, and run the do-file again.

46 Macros are an alternative and more efficient way to recall a list of variables in Stata. Imagine macros as boxes where you store a set of variables. When you need to work on those variables, you can just recall the macro instead of recalling each of the variables.

47 `lowess` is the Stata command to carry out a locally weighted regression. The procedure is a combination of classical regression models and non-parametric ones, as it uses linear and non-linear regression techniques in a very small neighborhood but does not require the same functional form to be used globally.

48 See the manual for `lpoly` (enter `help lpoly` in the command window) for a more extensive and complete explanation of smooth regressions and local weighted regressions.

49 Enter `help egen` in the command window.

50 Households with a larger number of females usually have a larger number of household members. In developing countries, and especially in rural areas, poverty is concentrated in large households, which have a higher share of expenditure on food products.

51 This is done only for graphical reasons, as we want to apply the command `by` for all, urban, and rural households simultaneously.

52 The lines that begin with an asterisk (*) are treated as user comments and Stata does not take any action. They can however be useful for the user.

53 This annex is based on Module 4 of the teaching material for the online course on trade and poverty developed by the UNCTAD Virtual Institute (UNCTAD, 2012).

54 Moreover, one-third of poor households' income and one-fifth of non-poor households' income come from own production, which does not pass through the market.

55 Data taken from *Encuesta Nacional de Condiciones de Vida* (2000).

56 The GTAP model is a CGE model commonly used as a tool for measuring trade effects. See Module 1 of this volume for further details on GTAP.

57 This annex is based on Modules 2 and 3 of the teaching material for the online course on trade and poverty developed by the UNCTAD Virtual Institute (UNCTAD, 2012) and on Depetris and Ramos (2012).

58 In textbooks, a social planner is a decision-maker above all other parties. In practice, the role of the social planner is played by the government of a country.

59 In addition, there are short-term effects when households do not adjust, medium-term effects when households make partial adjustments, and long-term effects when growth, investments, and long-term choices have been made. However, we do not discuss these effects here.

60 The compensation variation is expressed in negative terms, see further on.

61 The authors would like to thank professors David Richards and Eric Neumayer for sharing their data and do-files.

62 See Gaddis and Pieters (2012), who investigate the impact of trade liberalization on labour force participation of women in Brazil.

63 See Baliamoune-Lutz (2007), who explores the effects of increased trade openness and growth on gender inequality in Africa, and Tseloni *et al.* (2011), who study the link between globalization and socio-economic development and gender inequality for a panel of 68 countries.

64 The concept of women's empowerment, and its difference from gender equality, is described in Module 1 of Volume 1 of this teaching material.

65 For more details on Becker's theory of discrimination applied to international economics, see Section 2.2 of Module 2 of this teaching material's Volume 1.

66 In brief, these include the so-called Gauss-Markov assumptions, which are the following: (a) the expected value of the error term is $E[\varepsilon_i] = 0$ for $i = 1, ..., N$; (b) all explanatory variables are uncorrelated with the error term – i.e. $[x_i,..., x_N]$ and $[\varepsilon_i, ...,\varepsilon_N]$ are independent; (c) all the error terms have the same variance, which is referred to as homoskedasticity – i.e. $V[\varepsilon_i] = \sigma^2$ for $i = 1, ..., N$; and (d) there is zero correlation between error terms, which excludes any form of autocorrelation – i.e. $corr[\varepsilon_i, \varepsilon_j] = 0$ for $i, j = 1, ..., N$ and $i \neq j$. For a complete presentation of the classic assumptions underlying regression analysis, see Wooldridge (2009).

67 To determine whether the fixed- or random-effects model best fits your panel data, you can use the Hausman test. It tests the assumption about the independence of μ_i with respect to x_{it}. It basically compares the results from the fixed-effects and random-effects models and tells you which one is the most suitable. The Hausman test can be implemented in Stata with the `hausman` command.

68 The paper by Richards and Gelleny is available for download on Professor Richards' personal webpage at: *https://sites.google.com/site/drdavid68*.

69 This is true in general, but note that for the specific sample used in the paper and contained in the Stata .dta file provided, the GDI ranges between 0.15 and 0.96 while the GEM ranges between 0.10 and 0.93.

70 In 2010, the GDI and GEM were replaced by the Gender Inequality Index (GII) to remedy the shortcomings of the previous indicators. More information on the GII is available at: *http://hdr.undp.org/en/faq-category/faq-category-2-0*.

71 For more details on GEE, see Section 23.2.6 of Cameron and Trivedi (2005).

72 For the interested reader, ordered logit models are explained in Section 15.9.1 of Cameron and Trivedi (2005).

73 All application files are contained in the accompanying CD, and are available for download at: *http://vi.unctad.org/tag*.

74 Basically, this option replaces the command `tsset`, usually run before the estimation to define the unit of analysis and the time dimension of the panel data. See Section 4.4 below and the Stata help on command `tsset` for further details.

75 The paper by Neumayer and de Soysa is available for download at: *http://www.lse.ac.uk/geographyandenvironment/whoswho/profiles/neumayer/pdf/globalization_women_empowerment.pdf*.

76 These weights are standardized according to a particular methodology. More specifically, the weighting matrices are row-standardized. Row standardization is often used in spatial econometric models to yield standardized spatial weights.

77 See Marshall *et al.* (2006) for more information on the Polity IV dataset.

78 Spatial patterns refer to the geographical distribution of a feature (here women's right) in the world.

79 Spatial dependence is a characteristic of the distribution of geographic data, meaning that the realization of a given estimated effect of any independent variable on the outcome variable in any regression model depends on the location variable (e.g. country, region, city level).

80 All application files are contained in the accompanying CD, and are available for download at: *http://vi.unctad.org/tag*.

81 The regions are East Asia and Pacific (eap), Eastern Europe and Central Asia (eca), Latin America and the Caribbean (lac), Middle East and North Africa (mena), North America (na), South Asia (sa), sub-Saharan Africa (ssa), and Western Europe (we).

82 If you use Stata 12.1 to run your estimations, the results you obtain for the enhanced ordered logit model – i.e. with the lag of the dependent variable – might differ from those presented in the paper. This is because the authors used an earlier version of Stata.

83 We strongly recommend that your read Roodman (2009), in particular Section 3, for a detailed explanation of the system and different GMM estimators.

84 The latest *Human Development Report* (UNDP, 2014) can be downloaded at: *http://hdr.undp.org/en/2014-report/download*.

85 The *Global Gender Gap Report* assigns a score to each of the 136 countries analysed measuring inequality between men and women based on economic, political, education, and health-based criteria. The score ranges between 0 and 1, with 1 indicating complete gender equality. The 2013 Report (WEF, 2013) can be downloaded at: *http://www.weforum.org/issues/global-gender-gap*.

86 The DHS Program is funded by the United States Agency for International Development and implemented by ICF International. It collects various data measuring women's status and empowerment. For more information on the DHS dataset see *http://dhsprogram.com/Topics/Womens-Status-And-Empowerment.cfm*. The DHS dataset is downloadable at: *http://dhsprogram.com/Data/*.

87 Frictions can basically be considered as anything that interferes with the smooth and instantaneous exchange of goods and services and, in the context of the labour market, the movement of productive factors, capital, and labour across sectors.

88 Anything that might create a difference in the productivity of two countries will influence their patterns of specialization. Labour market frictions are not an exception, as legislation about the length of the work week, minimum wages, skill endowments, market participation rates, etc., all have an effect on a country's production patterns.

89 For the reader interested in gaining a deeper understanding of binary outcomes, the Tobit model and selection models, see Chapters 14 and 16 of Cameron and Trivedi (2005).

90 In labour economics, the reservation wage is the minimum wage rate at which an individual will accept employment.

91 A probit model is a type of regression where the dependent variable can only take two values, e.g. "works" (variable takes value 1) and "does not work" (variable takes value 0). The purpose of the model is to estimate the probability that an event, condition, or characteristic will occur.

92 Selection bias occurs when the statistical population is not the outcome of random sampling and there are other external factors influencing it. If not duly addressed, selection bias may lead to distorted results.

93 The Heckman model is specifically used to correct the estimations from sample selection bias, or, in other words, to correct the bias deriving from using a sample that is dependent on the variable of interest (e.g. studying the determinants of wages for the whole population based only on the sample of individuals who report their wages).

94 Semi-parametric selection models (including CLAD) are treated in Section 16.9 of Cameron and Trivedi (2005).

95 This method minimizes the sum of absolute errors. The errors are the vertical "residuals" between points generated by the function and corresponding points in the data.

96 Quantile regression is a type of regression analysis that seeks to estimate either the conditional median or other quantiles of the dependent variable. In Stata, the command to estimate quantile regressions is `qreg`.

97 If the censoring part of the distribution is instead the right tail, the procedure is similar but we drop the observations for which the predicted values are greater than the censoring value.

98 The paper by Newman (2002) is available for download at: *https://openknowledge.worldbank.org/bitstream/handle/ 10986/17212/773920JRN020020oCutoFlower0Industry. pdf?sequence=1.*

99 Quasi-experimental studies try to replicate the randomized experimental studies in non-random settings by generating control and treatment groups as if they were randomly generated. The treatment group is identical to the control group except that it is the one exposed to the treatment (in the paper presented here, households living in the region of Cayombe). The control group (households living in the region of Cotocachi) is used as a benchmark to assess the impact of the treatment on the treated group.

100 All application files are contained in the accompanying CD, and are available for download at: *http://vi.unctad.org/tag.*

101 Note that in the do-file, there are some commands that are preceded by an asterisk (*). That means that those commands were not executed, as the author did not need them for the final paper.

102 As already mentioned, you may not immediately obtain the results shown in the paper when running the command `clad`. This may happen for two reasons: First, `clad` is a relatively new command that the author did not use at the time she wrote the paper. She generated her own algorithm, which is still included in the do-file contained in the accompanying CD. Second, both the CLAD and the algorithm need to be run a couple of times before they converge to the results presented in the paper, since they are based on an iterative approach.

103 Please note that there is a typo in the heading of the first column of Table 7 in Newman (2002): it should read "Tobit All women" instead of "CLAD All women".

UNCTAD
VIRTUAL
INSTITUTE

UNITED NATIONS
UNCTAD

vi@unctad.org
http://vi.unctad.org