The Journal of Economic Education

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/vece20

Reproducing the stylized facts that motivate models of international trade with heterogeneous firms

Alejandro Riaño

To cite this article: Alejandro Riaño (14 Dec 2024): Reproducing the stylized facts that motivate models of international trade with heterogeneous firms, The Journal of Economic Education, DOI: 10.1080/00220485.2024.2438048

To link to this article: https://doi.org/10.1080/00220485.2024.2438048

THE JOURNAL OF ECONOMIC EDUCATION
https://doi.org/10.1080/00220485.2024.2438048

Routledge
Taylor & Francis Group

RESEARCH IN ECONOMIC EDUCATION

Reproducing the stylized facts that motivate models of international trade with heterogeneous firms

Alejandro Riaño

Department of Economics, City St George's, University of London, London, United Kingdom

ABSTRACT

The author of this article shows how to use the publicly available firm-level surveys produced by the World Bank Enterprise Surveys (WBES) to reproduce the stylized facts that characterize firm heterogeneity and its relationship with global engagement. He describes how to incorporate this exercise into different teaching activities, such as small group classes, homework, and coursework assignments. The proposed activities allow students to develop skills handling and analyzing firm-level data, and offer a unique opportunity to evaluate the extent to which the stylized facts established from U.S. data are also representative of other countries. The activities connect the theory of international trade with heterogeneous firms to the real-world empirical evidence that motivated the development of these models.

KEYWORDS

Economic data; exports; heterogeneous firms; stylized facts; teaching international trade

JEL CODES

A22; F12

In an essay commemorating the centenary of Bertil Ohlin's birth entitled "What role for empirics in international trade?," Davis and Weinstein (2002, 363) wrote, "Our field shows little of the two-way interplay between theory and data that is the very life of many fields of economics, such as macro, labor, and others." Two decades later, the state of the international trade field could not be more different.

The increasing availability of micro-level surveys at the firm level since the mid-1990s (Bernard et al., 1995; Roberts and Tybout 1996; Bernard and Jensen 1999) spurred a veritable revolution that put firm-level decisions front and center in the effort to understand the causes and consequences of international trade (Melitz and Redding 2014). Nevertheless, and despite its popularity in the research sphere, the so-called "new-new" trade theory has been much slower to permeate the teaching of international trade at the undergraduate level. As Cook and Pantuosco (2022) demonstrate, only two of the most popular undergraduate textbooks in international trade, Krugman, Obstfeld, and Melitz (2023) and McLaren (2012), present the Melitz (2003) mode—the quintessential workhorse model of trade with heterogeneous firms.[1] Critically, neither of these textbooks offers exercises that use firm-level data to allow students to connect theory and empirical evidence.

In this article, I seek to fill the gap between the teaching of theoretical models of international trade with heterogeneous firms and the empirical evidence that motivated their development by leveraging the publicly available firm-level data provided by the World Bank Enterprise Surveys (WBES). More specifically, I show how to use WBES data, which are available for more than 150 countries since 2002, to reproduce key stylized facts that characterize the extent of global engagement of manufacturing firms, as documented by Bernard et al. (2007) for the United States, namely, the large differences in terms of size and productivity observed across firms operating in the same industry; the limited extent of global engagement—i.e., exporting, importing, or being foreign-owned—of most manufacturing firms; the substantial heterogeneity in export intensity (the share of total sales accounted for by exports) among

CONTACT Alejandro Riaño alejandro.riano@city.ac.uk Senior Lecturer and Research Fellow, Rhind Building, City St George's, University of London, London EC1V 0HB, United Kingdom.

exporting firms; and the fact that globally engaged firms, on average, outperform domestic firms across a wide range of performance measures such as employment, sales per worker, and capital and R&D-intensity.

I discuss in detail how the reproduction of stylized facts can be embedded into teaching in a variety of ways, including a small-group teaching activity, as well as homework and coursework assignments, and present different alternatives that instructors can utilize to customize the activity to best fit their needs. The main small group class activity requires students to be both familiar with Stata and having access to the software on their own laptops (or alternatively, would require the instructor to carry out the activity in a computer laboratory with access to the software) and having a basic knowledge of linear regression. Nevertheless, I discuss various alternatives that would allow instructors to use the activities even when the above conditions are not satisfied. For example, if students are not familiar with linear regression, they can carry out the performance comparison of globally and non-globally engaged firms using simple averages for both types of firms. The WBES Web site (https://www.enterprisesurveys.org/) also provides averages for a large number of questions in the surveys that can be further calculated for subgroups of firms according to their exporting and foreign-ownership status.

The activities I propose in this article offer three key contributions to the teaching of international trade. First, they help to bring the teaching of international trade at the undergraduate level closer to the way that research at the cutting edge of the field is conducted—that is, with a strong emphasis on data analysis and empirical work (Feenstra 2016). In so doing, they offer a natural complement to other teaching strategies like the interactive classroom simulation developed by Cook and Pantuosco (2022), which highlights the main contribution of incorporating firm heterogeneity into models of international trade—i.e., the fact that changes in trade barriers affect different firms differently—with large and productive firms expanding, and smaller, less-productive firms contracting and exiting the market.

Second, they allow students to familiarize themselves with firm-level data—how to manipulate and present it and draw conclusions from it—a set of skills that is highly valued by employers of economic graduates (Economics Network 2019). While firm-level surveys are generally not publicly available because of confidentiality requirements, the raw data from WBES is accessible to individuals for research and pedagogical purposes by simply creating a free account and agreeing to abide by its confidentiality provisions.

Third, it is straightforward to adapt the activities I present in this article using data from Colombia to any country for which WBES surveys are available because WBES uses a standardized questionnaire, which, in turn, means that the names of all the variables used in the statistical analysis are identical across all surveys. This notable feature enables instructors to provide a global perspective on the topics discussed in the exercises and allows students to evaluate the extent to which the different stylized facts Bernard et al. (2007) identify for U.S. firms also provide an accurate description of the behavior of firms in countries of different sizes, comparative advantages, and stages of development.[2]

The World Bank Enterprise Surveys data

The World Bank Enterprise Surveys (https://www.enterprisesurveys.org) project conducts establishment-level surveys that intend to be representative of a country's private sector. Since the vast majority of establishments surveyed report as operating only one establishment, I will refer to them as "firms" hereafter. Between 2002 and 2023, the project has interviewed close to 200,000 firms across 155 countries. Most countries covered are developing and transitional ones, although surveys for a few European Union countries have become available in recent years. The surveys cover a broad range of topics, including firms' access to finance, obstacles to conducting business, and participation in international trade activities, to name a few.

Gaining access to the raw data

Both instructors and students need to set up a free account to download the raw data (detailed instructions for creating an account with WBES and downloading the raw data and questionnaires are provided

in appendix A). It is critical that instructors remind students of the need to create an account and download the data by themselves. In so doing, students agree to abide by the confidentiality agreement regarding the use of the data, which explicitly prohibits users from directly or indirectly, in any way, revealing, reporting, publishing, disclosing, transferring, or otherwise using any firm-specific information except aggregate data that does not identify individual firms.[3]

After logging in to the "Firm-level Datasets for Researchers" section of WBES, students can search for the specific survey they want to download by country and year. Enterprise surveys are usually conducted once every five years; this means that, for most countries, there are two to three survey waves available, while smaller and less-developed countries have only one survey. The raw data obtained from WBES are available in Stata (.dta) format, along with the questionnaire used in the survey (in pdf format). All variables in the Stata file are labeled and mapped directly to the survey questionnaire, where the variable's name is written in red in the questionnaire.

The number of firms surveyed varies according to country size. For larger countries such as Argentina, China, India, Türkiye, or Vietnam, a survey typically includes between 1,200 and 1,800 firms.[4] I recommend using relatively larger countries for the teaching activities in order to have more variability in terms of firms operating across different industries and greater availability of performance measures to analyze. Panel data are sometimes available for larger middle-income countries, but the exercises proposed in this article require only cross-sectional surveys.

Reproducing the stylized facts that characterize firm heterogeneity and global engagement

This section uses data from the 2017 survey wave for Colombia (the country I have used in the tutorial activity described in the next section) to illustrate the key statistical regularities that can be computed using the WBES data. This dataset includes 993 firms, 569 of which operate in the manufacturing sector and constitute the sample that I use to carry out the empirical analysis below.[5]

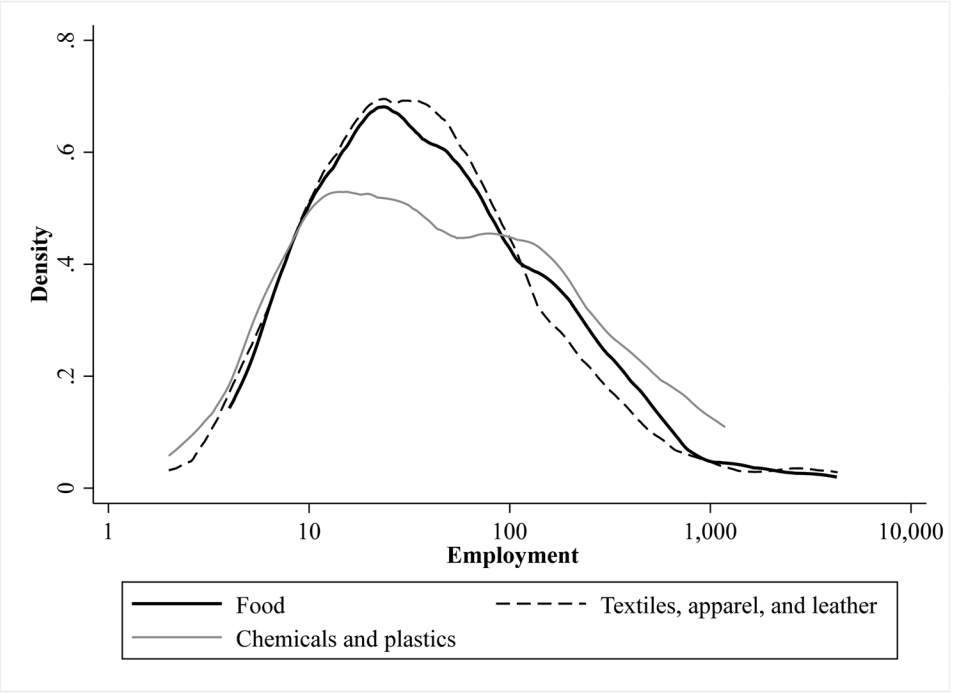
The stylized facts that characterize firms' global engagement that I propose to analyze using the WBES data are:

1. There is substantial heterogeneity across firms in terms of size (employment) within the same industry;
2. In most industries, globally engaged firms—i.e., firms that export, import, or are foreign-owned—are a minority;
3. There is large heterogeneity in terms of export intensity—the share of total sales accounted for by exports among exporting firms—across firms in the same country;
4. Globally engaged firms are "better" than firms that operate only domestically across a broad range of performance measures such as employment, skill- and capital-intensity, sales per worker, and innovation.

How large are differences in firm size within the same industry?

Summarizing total permanent employment reveals the large degree of heterogeneity observed across manufacturing firms. For instance, the median manufacturing firm in the Colombian survey has 32 permanent employees, while approximately one-quarter of the surveyed firms employ fewer than 15 workers. At the same time, and consistent with the international evidence, there are a handful of large firms employing more than 1,000 workers (Axtell 2001; Cabral and Mata 2003). Figure 1 shows that the same pattern emerges when we examine the size distribution of firms within industries.

Figure 1 is crucial to conveying the key innovation of models of trade with heterogeneous firms—namely, that firms differ quite substantially within industries, in contrast to other workhorse models of international trade such as the Ricardian, Hecksher-Ohlin and even the monopolistically competitive model of Krugman (1980). The following sections show that the extent of global engagement also differs substantially across firms.



Notes: The figure reports the share of globally engaged firms (exporters, importers, and foreign-owned firms) in manufacturing industries with at least 20 firms.

Figure 1. Size distribution of firms—Selected industries (Colombia, 2017).

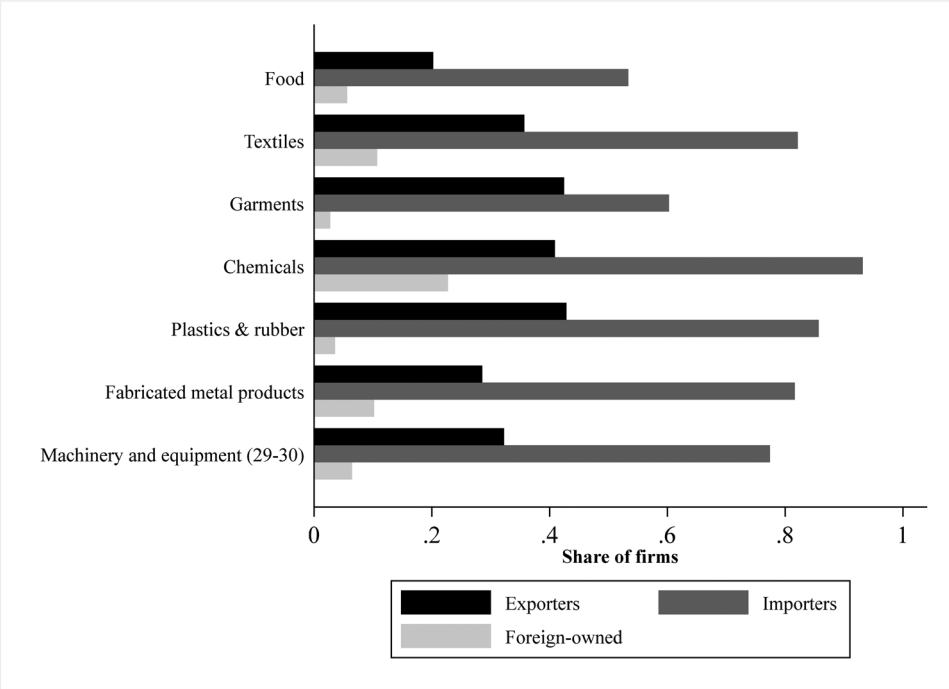
How prevalent is global engagement?

The second stylized fact pertains to the extent of firms' global engagement—i.e., whether they export, import, or are foreign-owned. To this end, students must use the information provided in the WBES questionnaire to construct indicator variables for firms' export-, import- and foreign-owned-status.

Figure 2 displays the share of firms engaged across each margin of global engagement defined above for selected manufacturing industries. The key message provided by this figure is the high level of firm heterogeneity in terms of global engagement across industries. On average, one-third of manufacturing firms in Colombia export some of their output—a similar figure to the one reported by Bernard et al. (2007) for U.S. firms and across the 72 countries included in the sample used by Defever and Riaño (2022), in which 40 percent, on average, of manufacturing firms export. Importing, in contrast, is much more prevalent among Colombian firms, with two-thirds of them importing any materials or inputs, while the corresponding share in the United States is only 14 percent. Foreign ownership is the least common dimension of global engagement, accounting for only 7 percent of manufacturing firms in Colombia—a figure consistent with the stylized facts reported by Antràs and Yeaple (2014).

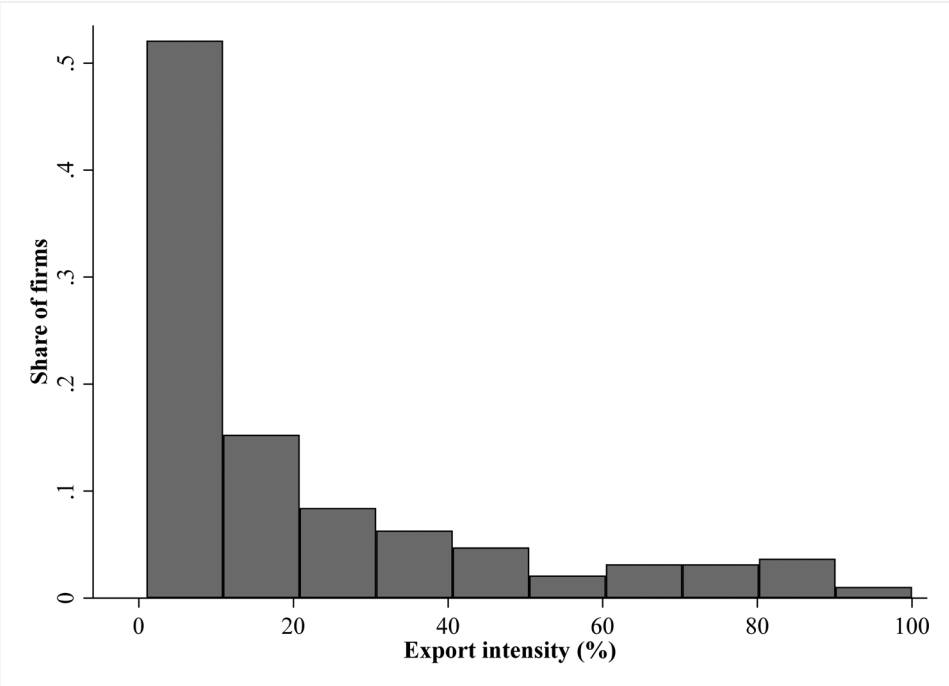
How important are export sales for exporting firms?

Bernard et al. (2007) show that in the United States. the majority of exporting firms sell most of their output domestically—or, in other words, most exporters exhibit low export intensity. Figure 3 presents a histogram of export intensity for Colombian exporters, which shows a similar pattern—more than half of manufacturing exporters export 10 percent or less of their total sales, and only a small minority of firms export most of their output. Defever and Riaño (2022), however, show that this stylized fact is not very robust across the world. Export-intensity distributions vary tremendously across countries and are often bimodal, with large shares of both low- and high-intensity exporters coexisting alongside each other within a country (see Figure 8 in their paper).



Notes: Export intensity is defined as the share of an exporter's total sales that are exported. Because this measure is calculated conditional on exporting, it takes values on the interval (0.100 percent].

Figure 2. Share of globally engaged firms—Selected industries (Colombia, 2017).



Notes: Export intensity is defined as the share of an exporter's total sales that are exported. Because this measure is calculated conditional on exporting, it takes values on the interval (0,100%].

Figure 3. Export intensity of manufacturing firms (Colombia, 2017).

Performance premia associated with global engagement

A cornerstone result of trade models with heterogeneous firms is that globally engaged firms are "better"—i.e., larger, more productive, more capital- and skill-intensive, and conduct more investment in innovation than firms that operate only domestically. This stylized fact is established by estimating bivariate regressions of the type:

$$\ln y_i = \alpha + \beta GLOB_i + \varepsilon_i, \tag{1}$$

where $\ln y_i$ denotes the natural logarithm of a given performance outcome for firm i and $GLOB_i$ is an indicator variable taking the value 1 if firm i is globally engaged (exporting, importing, or being foreign-owned) and 0 otherwise. Table 1 reports the performance premium in percentage terms, i.e., $\exp\left(\breve{\beta}\right)-1$, for each combination of performance outcome and margin of global engagement, with $\breve{\beta}$ denoting the estimated coefficient associated with the variable *GLOB* in regression (1).

The results of estimating the regressions described above reveal that Colombian manufacturing exporters are 203 percent larger in terms of employment than non-exporters, while importers are 51.3 percent larger along the same dimension than non-importing firms, and so forth.[6] The results presented in table 1 reveal that, similarly to what Bernard et al. (2007) report for U.S. firms, globally engaged firms in Colombia are significantly larger (both in terms of employment and sales), more productive, and carry out more R&D spending than their non-globally engaged counterparts. These results are consistent with the findings in Defever and Riaño (2022), showing that non-exporting firms in their sample are, on average, 50 percent smaller in terms of employment, have 50 percent lower sales overall, and 20 percent lower sales per worker than exporting firms, in accordance to the predictions of the Melitz (2003) model. In contrast, globally engaged firms in Colombia are not significantly more capital- or skill-intensive. Another important insight that can be gleaned from table 1 is that the size and productivity premia of being foreign-owned are substantially larger than those of being an exporter, consistent with the predictions in Helpman, Melitz, and Yeaple (2004).

Teaching activities using the World Bank Enterprise Surveys data

In this section, I describe in detail how I have implemented the teaching activity in the classroom and discuss the pedagogical choices I have made in the process, primarily in the context of a tutorial session (small group class with 20 to 30 students), while at the same time considering potential alternatives that the referees have kindly suggested. I also discuss how the activities carried out in the tutorial can be further developed into a homework assignment or coursework (term-project) assessment.

Tutorial (small group class)

I have carried out a teaching activity in which students reproduce the stylized facts presented by Bernard et al. (2007) in a 50-minute tutorial, both in an undergraduate course in international trade and in a

Table 1. Performance premia of different indicators of global engagement (Colombia, 2017).

	Premium		
Performance measure	Exporter	Importer	Foreign-owned
Employment	2.034***	0.513***	3.055***
Sales	4.635***	0.927***	8.777***
Sales per worker	0.881***	0.195***	1.342***
Capital per worker	0.423	0.195	0.710
Skill intensity	−0.067	0.053	0.153
R&D expenditure/sales	4.104***	1.131***	4.278***

Notes: The table reports $\exp(\breve{\beta})-1$, in which $\breve{\beta}$ denotes the estimated coefficient in a bivariate regression of the type represented in equation (1), where the dependent variable is the performance measure in a given row and the independent variable is the respective indicator of global engagement in the column. All regressions have been estimated using robust standard errors.
*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

master's-level class on international business economics that I teach at City, St George's University of London in the UK. These students have previously taken classes in data analysis and introductory econometrics using Stata as the main statistical software.

Appendix A presents the instructions that I provide students in the week the tutorial takes place, i.e., when I introduce the topic of firm heterogeneity in international trade. I have presented the Bernard et al. (2007) paper in the lecture before the tutorial in which students will use the WBES data. An alternative approach that could be implemented in a flipped-classroom framework would instead ask students to use the WBES data to "discover" rather than reproduce the stylized facts presented in the previous section.

The benchmark teaching activity consists of three parts: (i) preliminary tasks to be carried out before the tutorial; (ii) tasks to be conducted in the classroom during the tutorial; and (iii) debriefing and discussion. I now discuss these three stages in more detail.

Preliminary tasks to be carried out before the tutorial

The most important thing for students to do prior to the tutorial is to create an account with WBES, which allows them to download the raw data and corresponding survey questionnaire. To create an account, students need to use their university email account and provide a one-paragraph description of how they will use the data. This summary can be based on the description of the tutorial activity available in appendix A.

A key decision that the instructor needs to make in preparing for the tutorial is whether to ask students to carry out the statistical analysis using data from a single country or assign a different country to each group. In the context of a 50-minute tutorial, I have asked students to use data from a single country and then compare their findings with those reported by Bernard et al. (2007). While working with a single country involves lower coordination costs for the instructor and permits incorporating a discussion about the country's trade policy to offer more context to the results, analyzing different countries has the potential to enrich the stage in which each group shares their results with the rest of the class by offering a truly global perspective to the topic (Lee 1992).

In the latter case, the instructor needs to choose the countries to analyze with the objective of achieving a diverse sample along a range of characteristics such as country size, stage of development, and location, as well as in terms of the statistics that students are going to compute: for instance, including countries with low shares of exporters like Senegal and Zambia with countries in which most firms export, such as Tunisia or Slovenia; or countries where the majority of exporters sell most of their output abroad, like Bangladesh or Mauritius, and countries in which the distribution of export intensity is markedly bimodal like China or Uruguay.

Once students have had their account approved by WBES, they can download the data and questionnaire for the country they will analyze in the tutorial. At this stage, I ask students to open the data in their computers and use the Stata command "describe" to familiarize themselves with the way in which items in the questionnaire are mapped into the dataset; I also ask students to read pages 105–12 of Bernard et al. (2007) or review the slides that I have presented in the lecture in preparation for the tutorial.

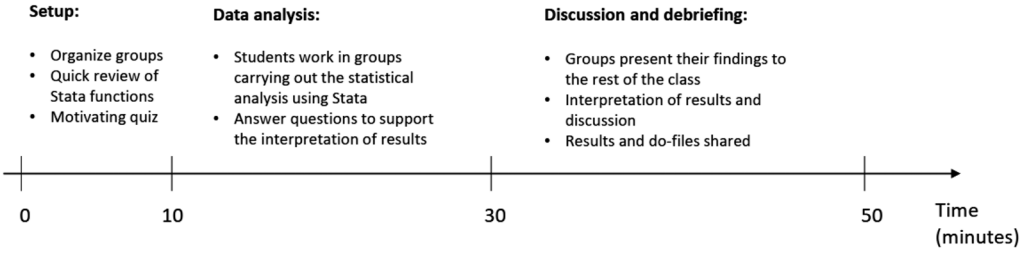


Figure 4. Timeline of tutorial activity.

Tasks to be completed in the classroom during the tutorial

The tutorial itself is further subdivided into three parts: (i) setup; (ii) data analysis; and (iii) discussion and debriefing. Figure 4 summarizes the activities taking place and the approximate amount of time devoted to each stage.

Once students have organized themselves into their respective groups (of 3 to 4 members), I do two things before they open the data and begin their analysis. First, I run a poll (using *polleverywhere* or a similar program) in which I ask students what percentage of firms they think are exporters; what manufacturing industries they think have a higher number of firms and exporters; what is the average export intensity among exporters and the average difference in terms of employment between exporting and non-exporting firms. I use the results of this poll in the discussion stage to contrast these responses with the figures that groups produce in their analysis. Secondly, I open the dataset in Stata and show students how the variables relate to the survey questionnaire and briefly review how to use basic Stata commands such as "tabulate," "summarize," and "regress," how to create variables, and use conditional statements. After doing this, I present a cheat-sheet slide that I leave on during the Data Analysis stage with these instructions that students can refer to when they work with the data.

In the tutorial, I give students 20–30 minutes to conduct the statistical analysis, writing a Stata do-file that they will use to carry out the calculations set out in Appendix A. While students work on this in their respective groups, I circulate around the room, answering clarifying questions. It is important to note that in the tutorial, I focus on exporting as the only margin of global engagement to analyze (i.e., I do not ask them to calculate the number of firms engaged in importing or that are foreign-owned nor to estimate the performance premia associated with these two modes of global engagement). I do so both because of time considerations and because the Melitz (2003) model, which is the next topic in my course's syllabus, emphasizes firms' exporting decisions; analyzing the other two modes of global engagement can be naturally incorporated into a homework or coursework assignment, as I discuss in more detail below.

One of the key learning outcomes that I seek to achieve through this activity is for students to develop the ability to use raw data to construct empirical counterparts to the variables in theoretical models. For instance, the surveys do not explicitly ask firms whether they are exporters or not; students need to figure out that they can construct an export indicator using the question in the survey that asks the percentage of a firm's sales that were: (i) national; (ii) indirect exports; or (iii) direct exports. Similarly, in order to construct the export-intensity variable, students need to notice that they have to use conditional statements to ensure that this variable is defined only for exporters (in other words, they need to make sure that the export-intensity variable does not take the value of 0 for firms that sell only domestically).

Another important skill that this activity enables students to cultivate—and which arguably often takes up to 80 percent of data analysts' time (Békés and Kézdi (2021)—is the importance of "cleaning" data. That is, learning what type of variables are included in a dataset, how to identify and address problems with observations and variables, and creating a reproducible workflow to organize the data. For instance, students should note that when a firm does not answer a question, this variable is coded in the dataset as taking the value −9, and therefore should be aware that these observations need to be excluded when calculating statistics. Alternatively, the instructor could provide students with the do-file and ask them to explain what different commands are doing.[7] A third approach could instead present the figures and tables in the previous section and focus the discussion on the interpretation of these results.

Debrief and discussion

After students have finished analyzing the data, the last part of the tutorial entails presenting the results they have obtained, discussing them with the rest of class. To operationalize this, when analyzing data from a single country, I ask that a different group present the calculations and results associated with each question. In order to facilitate the discussion, I present the responses from the initial poll and contrast those with the actual results as well as with the results presented by Bernard et al. (2007).

When revisiting the poll results about the industries with the largest number of firms and exporters, I ask students to justify their answers. For instance, when working with the Colombian data, students often think the country has a comparative advantage in labor-intensive sectors such as food or textiles—based on its relative endowments, as suggested by the Hecksher-Ohlin model, and thus tend to find it surprising that the share of exporters is higher in capital-intensive industries like chemicals and machinery

and equipment. I then ask them if there are other factors that could explain these data patterns. In the case of Colombia, I use data on exports by industry and destination country publicly available in the Observatory of Economic Complexity Web site (https://oec.world/) to show students that capital-intensive exports are often shipped to neighboring countries such as Ecuador and Venezuela, which are less abundant in capital than Colombia.[8]

If the tutorial studies a country with a high prevalence of high-intensity exporters, such as Bangladesh, Ireland, or the Philippines, the instructor can ask students about the reasons why the distribution of export intensity of these countries looks so different than that in the United States. Are these firms producing goods for which there is little to no domestic demand, such as wool sweaters in Bangladesh (Diaz de Astarloa et al. 2013)? Or are they producing highly specialized goods that are exported to the next production stage within a global value chain (Antràs 2020)? These results can also be used to introduce the concept of special economic zones—geographically bounded areas in which customs, tax, and investment regulations are more liberal than in the rest of the country. These zones are ubiquitous across the world, are one of the most important tools of industrial policy in developing countries (Rodrik 2004), and often require firms to export most of their output in order to benefit from the fiscal benefits they provide (Defever and Riaño 2017; Defever et al. 2019).

After students have presented the results related to the performance premia of exporters, I then turn the discussion to whether we can give these estimates a causal interpretation. On one hand, the workhorse model of firm heterogeneity assumes that exporting is fully explained by firms' selection into this activity as a consequence of their higher intrinsic productivity; alternatively, exposure to international best practices and access to higher-quality inputs allows firms to become more productive once they are globally engaged—the so-called "learning-by-exporting" hypothesis (Clerides, Lach, and Tybout 1998; van Biesebroeck 2005).

I like to finish the discussion with a reflection on the implications of the results from the performance premia regressions to economic policy. Should governments subsidize or support firms to become exporters? Or does this mean that governments would end up spending resources helping the firms that are already the best performers? To anchor this discussion, I ask students to listen to the *Trade Talks* podcast episode in which the hosts, Chad Bown and Soumaya Keynes, interview Amit Khandelwal about the randomized controlled experiment that he and his coauthors ran with rug producers in Egypt (Atkin, Khandelwal, and Osman 2017) after concluding the tutorial.[9] The podcast, which lasts only 18 minutes, not only vividly illustrates the difficulties involved in establishing a causal link between exporting and firm outcomes but also highlights concrete mechanisms through which exposure to international markets can improve firms' performance and the policy instruments that governments can use to achieve these goals.

Students who have participated in the tutorial were highly enthusiastic about it, noting in student evaluations of teaching that the use of real-world data helps to enliven the theoretical models and facilitate their comprehension of the economic mechanisms at play. The activity is a critical tool for achieving the learning objective of applying the theoretical tools of international trade theory to analyze "real-world" situations related to the international movement of goods and services.

Alternative ways to use the reproduction of stylized facts in teaching

Homework assignment

A natural extension to the activities proposed for the tutorial, as described in the tutorial, is to carry out the statistical analysis for importing and foreign ownership. Comparing the prevalence of importers and exporters within a given industry can be used to motivate a discussion about the importance of inter- and intra-industry trade—are the industries in which exporters are more prevalent also the ones in which a higher percentage of firms import inputs as would be the case in models with monopolistic competition, or are these two activities negatively correlated with each other as the Ricardian or Hecksher-Ohlin models would predict?

The estimation of performance premia can also be extended to estimate the regressions that Bernard et al. (2007) report in columns (2) and (3) of Table 3 of their paper. That is:

$$\ln y_{ij} = \alpha + \beta GLOB_{ij} + f_j + \varepsilon_{ij} \tag{2}$$

and

$$\ln y_{ij} = \alpha + \beta GLOB_{ij} + f_j + \gamma \ln EMP_{ij} + \varepsilon_{ij}, \quad (3)$$

where subindices *i* and *j* reference firms and industries, respectively. Thus, f_j denotes industry-level fixed effects and $\ln EMP_{ij}$ is the natural logarithm of employment of firm *i* operating in industry *j*. Regression (2) adds industry fixed effects to regression (1), while regression (3) includes both industry fixed effects and the logarithm of firm's employment. The key insight to be obtained from these is that now the performance comparison between globally engaged and domestic firms is carried out within the same industry in the case of regression (2), which in turn means that the difference in, say, employment, is not due to a majority of globally engaged firms operating in sectors characterized by high economies of scale, like machinery and equipment, and domestic firms being more prevalent in sectors favoring low-scale operation, such as textiles or apparel. The intuition behind regression (3) is similar, but now the regression compares globally engaged and domestic firms of similar size and within the same industry.

Beyond the process of constructing the relevant variables, some interesting questions for students to analyze are: (i) Do performance premia remain significant as we conduct the comparison within narrower groups of firms? (ii) If the latter, what other factors explain the performance differences between globally engaged and domestic firms? (iii) What happens to the magnitude of the performance premia estimates? Alternatively, instructors could combine the activities described in the tutorial and homework sections into a scaffolded term portfolio project in which each group of students analyzes a different country and submits each component of the tutorial and homework assignments in a staggered way.

Coursework assignment

One of the main objectives of the WBES surveys is to document the obstacles that private businesses face when operating in different countries. I have used this information in a coursework (take-home) assignment in a master's-level course in international business economics, in which groups of 3 to 4 students write a business intelligence report (with a 1,600-word limit) for a country I assign them from WBES.

The assignment has two parts. In the first one, students provide a brief description of the country's overall trade performance and recent key developments in trade policy (e.g., whether the country has recently lowered import tariffs; measures that the country has taken to integrate local firms in global value chains; signature of free trade agreements, etc.). The second part asks students to use WBES data to describe the main obstacles that inhibit firms' growth and to investigate to what extent these barriers differ across exporters and domestic firms. Appendix B provides the description of this assignment, the criteria used to evaluate the assessment, and a group-audit form that gives students the opportunity to assess their contribution toward the group's submission.

Students do not need to use the raw WBES data I have discussed so far in the article to carry out the analysis in the second part of the assignment. Clicking on the "Economy Snapshots" in the WBES Web site, students can scroll down and see the survey's questions related to obstacles to growth organized in different tabs, e.g., regulations and taxes, corruption, management practices, and so on. Clicking on a given tab reveals the specific questions related to the topic (e.g., percent of firms identifying tax administration as a major constraint; percent of firms expected to give gifts in meetings with tax officials), which can in turn be broken down by different characteristics such as sector, size, location, and, crucially for the purposes of the assignment, exporter type, and whether the firm is foreign-owned or not. Any figures are reported for the country in question (e.g., Azerbaijan) benchmarked against the broad geographic group to which the country belongs (Europe; Central Asia) and all countries. This feature of the WBES Web site can significantly reduce the barriers to using this resource for teaching when students do not have access to or are not familiar with Stata to work with the surveys' raw data.

Conclusion

In this article, I have described how to use the publicly available, firm-level data provided by the World Bank Enterprise Surveys (WBES) to reproduce stylized facts that characterize firm heterogeneity and its

relationship with different dimensions of global engagement. To do so, I provide detailed instructions about how to access the data and how to calculate the different statistics reported by Bernard et al. (2007) in their seminal paper. I also discuss how this exercise can be embedded into several teaching activities, including a small-group teaching activity, homework and coursework assignments, and present different alternatives that instructors can customize.

Reproducing the stylized facts that characterize firm heterogeneity and its relationship to exporting and other forms of global engagement makes the theoretical models of international trade more stimulating to students by strengthening their connection to the real world. At the same time, working with raw firm-level data helps students to develop data analysis skills and improve their understanding of how to connect the theoretical constructs in economic models to observable data. To conclude, I emphasize the richness of the data that WBES makes available for researchers, instructors, and students. These data can be effectively used in a manner similar to what I suggest in this article across a broad range of courses in development economics, industrial organization, labor, and public economics.

Notes

1. Neither Feenstra and Taylor (2021), Gerber (2018), nor Salvatore (2019) discuss the Melitz model or heterogeneous firms.
2. It is important to remark that even in textbooks that emphasize the importance of data analysis, such as Feenstra and Taylor (2021), most of the empirical exercises offered utilize only U.S. data.
3. The full Data Access Protocol can be found here: https://login.enterprisesurveys.org/content/sites/financeandprivatesector/en/terms.html.
4. Defever and Riaño (2022) use data drawn from WBES across 72 countries for the period 2002–16. Across this period, the average survey wave includes 442 firms, with substantial heterogeneity in terms of survey size—ranging from 28 firms in Latvia in 2002 to 7,165 firms in India in 2014. These data, which contain several of the variables used in this article are available here: https://academic.oup.com/jeea/article/20/3/1347/6521445/#supplementary-data.
5. The Stata do-file used to produce the results presented in this section is available at: https://alejandroriano.weebly.com/uploads/6/0/7/9/6079661/firmheterogeneitystylizedfacts.do. Due to the standardization of WBES, this Stata code can be readily used with any other survey because the names of the variables remain constant across datasets.
6. If students are not familiar with linear regression, the same analysis can be carried out by simply calculating the difference in the means of performance measures for globally engaged and non-globally engaged firms.
7. I make the do-file available to students after the tutorial.
8. Another advantage of focusing the tutorial on one country is that the instructor can more easily embed country-specific knowledge into the discussion, which grounds the conversation and relates it more to "real-world" issues.
9. The podcast episode can be found here: https://tradetalkspodcast.com/podcast/62-randomized-trade/.

Acknowledgments

The author thanks JEE Associate Editor William Bosshardt, Guglielmo Volpe, and two anonymous referees for their helpful comments and suggestions, which have greatly improved the quality of the paper.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Alejandro Riaño http://orcid.org/0000-0002-8824-4055

References

Antràs, P. 2020. Conceptual aspects of global value chains. *The World Bank Economic Review* 34 (3): 551–74. doi: 10.1093/wber/lhaa006.

Antràs, P., and S. R. Yeaple. 2014. Multinational firms and the structure of international trade. In *Handbook of International Economics*, ed. G. Gopinath, E. Helpman, and K. Rogoff, vol. 4, 55–130. Amsterdam: Elsevier, North-Holland.

Atkin, D., A. K. Khandelwal, and A. Osman. 2017. Exporting and firm performance: Evidence from a randomized experiment. *The Quarterly Journal of Economics* 132 (2): 551–615. doi: 10.1093/qje/qjx002.
Axtell, R. L. 2001. Zipf distribution of U.S. firm sizes. *Science* 293 (5536): 1818–20. doi: 10.1126/science.1062081.
Békés, G., and G. Kézdi. 2021. *Data analysis for business, economics, and policy*. Cambridge, UK: Cambridge University Press.
Bernard, A. B., and J. B. Jensen. 1999. Exceptional exporter performance: Cause, effect, or both? *Journal of International Economics* 47 (1): 1–25. doi: 10.1016/S0022-1996(98)00027-0.
Bernard, A. B., J. B. Jensen, and R. Z. Lawrence. 1995. Exporters, jobs, and wages in U.S. manufacturing: 1976–1987. *Brookings Papers on Economic Activity* 1995 (1): 67–119. doi: 10.2307/2534772.
Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott. 2007. Firms in international trade. *Journal of Economic Perspectives* 21 (3): 105–30. doi: 10.1257/jep.21.3.105.
Cabral, L., and J. Mata. 2003. On the evolution of the firm size distribution: Facts and theory. *American Economic Review* 93 (4): 1075–90. doi: 10.1257/000282803769206205.
Clerides, S., S. Lach, and J. R. Tybout. 1998. Is learning by exporting important? Micro-Dynamic evidence from Colombia, Mexico, and Morocco. *The Quarterly Journal of Economics* 113 (3): 903–47. doi: 10.1162/003355398555784.
Cook, N. P. S., and A. Pantuosco. 2022. International trade with heterogeneous firms: An interactive classroom simulation. *Journal of Economic Education* 53 (1): 31–42. doi: 10.1080/00220485.2021.2004275.
Davis, D. R., and D. E. Weinstein. 2002. What role for empirics in international trade?. In *Bertil Ohlin: A centennial celebration, 1899-1999*, ed. R. Findlay, L. Jonung, and M. Lundahl, 363–88. Cambridge, MA: MIT Press.
Defever, F., J.-D. Reyes, A. Riaño, and M. E. Sánchez-Martín. 2019. Special economic zones and WTO compliance: Evidence from the Dominican Republic. *Economica* 86 (343): 532–68. doi: 10.1111/ecca.12276.
Defever, F., and A. Riaño. 2017. Subsidies with export share requirements in China. *Journal of Development Economics* 126 (2): 33–51. doi: 10.1016/j.jdeveco.2016.12.003.
———. 2022. The twin peaks of the export intensity distribution. *Journal of the European Economic Association* 20 (3): 1347–94. doi: 10.1093/jeea/jvac006.
Diaz de Astarloa, B., J. Eaton, K. Krishna, B. Y. Aw-Roberts, A. Rodríguez-Clare, and J. R. Tybout. 2013. Born-to-export firms: Understanding export growth in Bangladesh. International Growth Centre (IGC) Working Paper. London: London School of Economics.
Economics Network. 2019. Survey of employers of economics graduates. 2019 results report. https://www.economicsnetwork.ac.uk/sites/default/files/Ashley/Full%20Report_web.pdf (accessed April 3, 2024).
Feenstra, R. C. 2016. *Advanced international trade*. 2nd ed. Princeton, NJ: Princeton University Press.
Feenstra, R. C., and A. Taylor. 2021. *International trade*. 5th ed. New York: MacMillan Learning.
Gerber, J. 2018. *International economics*. 7th ed. London: Pearson Education.
Helpman, E., M. J. Melitz, and S. R. Yeaple. 2004. Export versus FDI with heterogeneous firms. *American Economic Review* 94 (1): 300–316. doi: 10.1257/000282804322970814.
Krugman, P. 1980. Scale economies, product differentiation, and the pattern of trade. *American Economic Review* 70 (5): 950–59.
Krugman, P. R., M. Obstfeld, and M. J. Melitz. 2023. *International trade: Theory & policy*. 12th ed. London: Pearson.
Lee, D. Y. 1992. Internationalizing the principles of economics course: A survey of textbooks. *Journal of Economic Education* 23 (1): 79–88. doi: 10.1080/00220485.1992.10844742.
McLaren, J. 2012. *International trade*. Hoboken, NJ: John Wiley & Sons.
Melitz, M. J. 2003. The impact of trade on intra-industry reallocations and aggregate productivity. *Econometrica* 71 (6): 1695–725. doi: 10.1111/1468-0262.00467.
Melitz, M. J., and S. J. Redding. 2014. Heterogeneous firms and trade. In *Handbook of international economics*, ed. G. Gopinath, E. Helpman, and K. Rogoff, vol. 4, 1–54. Amsterdam: Elsevier, North-Holland.
Roberts, M. J., and J. R. Tybout. 1996. *Industrial evolution in developing countries: Micro patterns of turnover, productivity and market structure*. Oxford, UK: Oxford University Press.
Rodrik, D. 2004. Industrial policy for the twenty-first century. CEPR Discussion Paper 4767. Paris and London: Center for Economic Policy Research.
Salvatore, D. 2019. *International economics*. 13th ed. Hoboken, NJ: John Wiley & Sons.
van Biesebroeck, J. 2005. Exporting raises productivity in sub-Saharan African manufacturing firms. *Journal of International Economics* 67 (2): 373–91. doi: 10.1016/j.jinteco.2004.12.002.

Appendix A

Tutorial instructions

The objective of this tutorial is for you to reproduce some of the key stylized facts that characterize firms' export behavior, as documented by Bernard et al. (2007) for the United States and as reviewed in the lecture slides for this week. To do so, you will use a survey of Colombian manufacturing firms conducted in 2017 by the World Bank Enterprise Surveys (WBES). You need to form groups of 3 to 4 students to carry out the in-class activities in the Tutorial.

Preliminary tasks to be completed before the tutorial

Create an account with the World Bank Enterprise Surveys and download the 2017 Stata dataset and questionnaire for Colombia. To do so, follow the steps below:

- Go to http://www.enterprisesurveys.org
- Click on the "Data" tab on the top of the page
- Click on the "Survey Datasets" tab
- Click on the "Firm-level Datasets for researchers" tab
- Sign up for a new account on the right-hand side of the page. This last step requires you to use your university email account and provide a one-paragraph summary of the research project, for which you can use the description of the tutorial provided above. It is critical that you create an account and download the data by yourself, and that you abide by the confidentiality agreement regarding the use of the data (see https://login.enterprisesurveys.org/content/sites/financeandprivatesector/en/terms.html)
- Once your account has been approved, you can then search for Colombia in the "Data by Economy" search bar
- Download the files (The former is the Stata file and the latter zip folder has the survey's questionnaire):
  - Colombia-2017-full data.dta
  - Colombia2017_Enterprise_Survey_Documentation.zip
- Open the Colombia-2017-full data.dta in Stata. Browse the data and use the command "*describe*" to see the variables included in the dataset, their definitions, and how each variable in the Stata file maps to questions in the questionnaire. Note that the variable names in red in the questionnaire pdf file are the same variable names in the Stata file.
- How many observations are included in the dataset? Tabulate the variable "a0" (questionnaire) to determine the number of firms in the survey that operate in the manufacturing sector.
- Please review this week's lecture slides on firms' decision to export and read pages 105–112 of Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott. 2007. Firms in International Trade, *Journal of Economic Perspectives* 21 (3): 105–130.

Activities to be carried out in class

Use the dataset Colombia-2017-full data.dta to compute the following statistics:

1. Using information from Section L of the questionnaire, summarize the number of full-time employees in manufacturing firms. What is the average size of manufacturing firms? What are the 25th, 75th, and 90th percentiles of the size distribution of firms?
2. The number of firms operating in each manufacturing industry, as in column 1 of table 2 in Bernard et al. (2007).
3. Use information from Section D of the questionnaire to create a dummy variable called "export" that equals 1 if a manufacturing firm exports some of its output and 0 otherwise. Tabulate the number of exporters in each manufacturing industry, as in column 2 of Table 2 in Bernard et al. (2007). Which industries have the highest percentage of exporters?
4. Calculate the mean export intensity (i.e., the percentage of exporter firms' sales accounted for by exports) across all firms and by manufacturing industries, as in column 3 of Table 2 in Bernard et al. (2007). Use a histogram to display the distribution of export intensity across all exporting firms.
5. Create the following variables:

   - Natural logarithm of total permanent employment
   - Natural logarithm of skill intensity, i.e., natural logarithm of the ratio of the number of nonproduction workers to total employment
   - Natural logarithm of total sales per worker
   - Natural logarithm of capital stock per worker, i.e., natural logarithm of net book value (value of assets after depreciation for machinery, vehicles, and equipment) divided by total employment
   - Natural logarithm of expenditure in research and development intensity, i.e., natural logarithm of R&D expenditure divided by total sales
   - Natural logarithm

   And for each of these variables estimate the following regressions by ordinary least squares:

$$\ln y_i = \alpha + \beta \mathrm{EXP}_i + \varepsilon_i,$$

   where y_i denotes the outcome variable of interest (employment, skill intensity, etc.), EXP_i is the export dummy you constructed above, and ε_i is the error term. Interpret the estimated coefficients $\breve{\alpha}$ and $\breve{\beta}$.

Debrief and discussion

The following questions intend to help you discuss and analyze your results with your fellow group members. In the last 20 minutes of the tutorial, each group will report their results to the rest of the class.

- In light of the results your group obtained in question (1), do you think the assumption that all firms in the same industry are identical fits the data well? What factors can explain the differences in firm size that you observe within a given industry?
- Compare the results that your group obtained in questions (2)–(4) with Table 2 in Bernard et al. (2007). Does Colombia have a comparative advantage in the same industries as the United States? If not, what factors could potentially explain these differences? Which models that you have seen in the course (Ricardian, Hecksher-Ohlin, monopolistic competition) can be useful to explain the difference in the number of exporters across industries in Colombia?
- Do most exporters operate at a similar export intensity, or are there significant differences in export intensity across Colombian exporters? What factors could explain why some firms do not sell domestically and instead export all their output?
- Can we use the estimates of performance premia regressions carried out in question (6) to argue that exporting improves firms' performance or vice versa?

Appendix B

Coursework instructions

The objective of this coursework is for you to write a business intelligence report (1,600 words maximum) for a country that I will assign you. This is a group assessment because, in the real world, this type of analysis is carried out by teams of economists—that the sharing of ideas, thoughts, views, and opinions makes the final output richer and more creative than if a single individual wrote it. My expectation is that you will work together on the report by actively collaborating, sharing the work, and consulting each other on a regular basis. Any concern about the group's ability to work together should be communicated to me as soon as possible.

The report should be structured in two sections:

1. Introduction: (not more than 500 words): brief description of the country's overall trade performance, main exports, and trade partners, and a succinct summary of key recent developments in the country's trade policy (e.g., whether the country has recently lowered import tariffs; measures that the country has taken to integrate local firms into global value chains; signature of free trade agreements, etc.). Which topics are more relevant depends on the country that you are analyzing. Keep in mind that the space devoted to this section is limited, so you need to choose carefully what policy developments merit discussion.
2. Business environment: This section should discuss the main obstacles to growth experienced by private firms in the country under analysis, using data obtained from the World Bank's Enterprise Surveys (www.enterprisesurveys.org). Discuss how these obstacles affect exporting and non-exporting firms in different ways. It is crucial that the figures you report to support your analysis are benchmarked against the same figures for the region to which the country belongs, e.g., if the country you are analyzing is Greece, its figures should be compared with those for the Europe & Central Asia region. The Enterprise Surveys Web site provides this information automatically.

The United States International Trade Administration produces country commercial guides (https://www.trade.gov/country-commercial-guides) that, while covering a broader range of topics, provide very similar information to what your report should contain.

The World Bank and the World Trade Organization provide useful information on a country's trade policy and business environment:

- World Integrated Trade Solutions (https://wits.worldbank.org/)
- Trade Policy Review gateway (https://www.wto.org/english/tratop_e/tpr_e/tpr_e.htm)

You should feel free to use information from other sources in writing your report.

Assessment criteria

Your report will be assessed against the following criteria:

- **Analysis and application (40 percent):** Is there evidence that the report provides an insightful analysis of the country's trade policy and business environment? Where appropriate, is there evidence of references to the theories and empirical evidence discussed in the module?
- **Research effort (40 percent):** Does the report show evidence that you have collectively familiarized yourselves

with the facts for the country you are analyzing, correctly identified salient patterns in the data, and offered thoughtful and plausible explanations for them? Is there evidence of independent research into the sources of information and insight relevant to writing the report? Does the report avoid mistakes—e.g., factually untrue statements and logical errors?

- **Communication and structure (20 percent):** Is the report well presented, clearly written and accessible? Is it structured in a logical way that makes it easy to follow its findings? Is there fluency between the different arguments made in the report and consistency in the quality of graphs, tables, writing, and style? Does it use proper referencing, diagrams labeled and integrated, correct grammar and spelling, spacing and visual appeal, etc.?

Teamwork auditing form

Marking teamwork can be difficult. Awarding the same mark to each student is not a problem if everyone contributes equally to a team project but, if for any reason, some students contribute more than others, it can have the effect of penalizing students who do more work and in extreme situations can reward a student who did nothing at all. In addition to your coursework, you have to submit a self-assessment account that describes both the tasks each student was allocated before starting the project and their actual contribution at the end of it. This self-assessment method of differential marking is designed to make assessment fairer by awarding you a mark that reflects the standard of the project and the contribution of each student.

How does this work? It's important that you decide how you will allocate tasks among yourselves before you start to work so that each member of the team is clear about what they are expected to do. At the end of the project, you should meet again, try to reach an agreement you are all happy with and complete and sign off pn the contribution self-assessment form. If, for any reason, one or more of you is unhappy with the agreement, you should put a mark in the disagreement box next to your name and submit the form before the deadline.

You can calculate the total number of marks awarded among your group by multiplying the mark you are awarded by the number of students in your group. For example, if 3 students worked on a project that was awarded a mark of 65, you would have $65 \times 3 = 195$ marks to share among you. If the team agrees that they all contributed equally, then you all get the same mark: $195 / 3 = 65$. Alternately, suppose that students agreed that one student (A) had contributed more than the other two (B and C), so they agreed to allocate 36 percent of the mark to student A, and 32 percent to students B and C each. As a result, student A's mark was $0.36 \times 195 = 70.2$, and students B and C's marks are each $0.32 \times 195 = 62.4$ instead. Please note that small differences in the allocation of agreed contributions can have a substantial impact on individual marks.