Reproducing the Stylized Facts that Motivate Models of International Trade with Heterogeneous Firms using the World Bank Enterprise Surveys

Alejandro Riaño[†]

June 2nd 2023

Abstract

This article presents a teaching activity that utilizes the publicly-available firm-level surveys from the World Bank Enterprise Surveys (WBES). The activity consists of reproducing four key stylized facts that characterize firms' heterogeneity and their extent of global engagement, as established by Bernard et al. (2007) for manufacturing firms in the U.S., using data from one of the many countries available in WBES. This exercise connects the theory of trade with heterogeneous firms to the real-world empirical evidence that motivated the development of these models. It allows students to acquire skills handling and interpretating firm-level data and offers the opportunity to evaluate the extent to which the stylized facts established from U.S. data are also representative of countries of different size and stage of development.

Keywords: teaching international trade; heterogeneous firms; exports; imports; economic data; stylized facts

Subject classification codes: A22; F12.

[†] City, University of London, GEP and CESifo. Email: alejandro.riano@city.ac.uk. All remaining errors are my own.

1. Introduction

In an essay commemorating the centenary of Bertil Ohlin's birth entitled "What role for empirics in international trade?" Davis and Weinstein (2002) 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 and Jensen 1995; Roberts and Tybout 1996; Bernard and Jensen 1999) spurred a veritable revolution that put firm-level decisions front and centre 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) model – 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 the theory of trade with heterogeneous firms to the real-world empirical evidence that motivated the development of these models.

To fill this gap, I propose an activity that can be used to complement the teaching of theoretical models of international trade with heterogeneous firms. More specifically, I show how to use firm-level data from the Enterprise Surveys carried out

[1] Neither Feenstra and Taylor (2021), Gerber (2018) nor Salvatore (2019) discuss the Melitz model or heterogeneous firms.

by the World Bank (WBES), which are readily available in Stata format, to replicate key stylized facts that characterize the extent of global engagement of manufacturing firms, as documented by Bernard et al. (2007) for the U.S. Namely, the large differences in terms of size and productivity observed between individual firms operating within the same industry; the limited extent of global engagement (exporting, importing or being foreign-owned) by most manufacturing firms; the large heterogeneity in export intensity (the share of total sales accounted for by exports) among exporting firms; and the fact that globally-engaged firms tend to outperform non-globally-engaged firms across a wide range of performance measures such as employment, sales, and productivity, among others.

The empirical exercise proposed in this paper offers three key contributions to the teaching of international trade. First, it helps to bring the teaching of international trade at the undergraduate level closer to the way research at the cutting edge of the field is being presently carried out – that is, with a strong emphasis on data analysis and empirical work (Feenstra 2016). Second, it allows 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). Although 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 with confidentiality provisions. Third, because the WBES use a standardized questionnaire, it is straightforward to adapt the exercise I present in this paper to any country for which WBES are available, since the name of the variables used in the analysis are the same across all surveys. Furthermore, the availability of data across many countries allows instructors the possibility to provide a global perspective

to the topics discussed in the exercise (Lee 1992).[2] A key learning outcome for the exercise is to evaluate the extent to which the different stylized facts that Bernard et al. (2007) identify for U.S. firms also characterize the behavior of firms in countries of different size, comparative advantages and stage of development.

The paper is organized as follows: Section 2 describes the data available in WBES and discusses how to gain access to the raw data used in the exercise. Section 3 presents the empirical exercise that has been carried out as within a 50-minute tutorial session in a UK university using data from the 2017 survey for Colombia as an illustration. This section shows how to replicate each stylized fact using the WBES data and provides discussion points related to each set of statistics that students are asked to construct. Section 4 concludes.

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.[3] Since the vast majority of establishments surveyed report to be single-establishment ones, I will refer to them as "firms" hereafter. Since 2002 and up to May 2023, the project has interviewed close to 200,000 firms across 155 countries. Most countries covered are developing and transition ones, although surveys for a few EU countries have become available recently. The surveys cover a broad range of topics including firms' access to finance, corruption, infrastructure, and participation in international trade activities.

---

[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 only utilize U.S. data.

[3] An establishment is a physical location where business is carried out or industrial operations take place, which should have its own management and control over its workforce.

A crucial pedagogical advantage of the data available in WBES is that, for the most part, the survey questionnaire is standardized across countries and years.[4] This means that the name of Stata variables used to construct different indicators of global engagement or performance, do not change across countries or survey waves. Thus, the exercise proposed here can be readily applied to a dataset for a different country and year by simply downloading a different dataset.

Gaining access to the raw data

Instructors and students need to set up a free account to download the raw data.[5] After logging in, users can search for the specific survey they want to download by country and year. Enterprise surveys are usually conducted once every five years, which means that for most countries there are between 2 and 3 surveys, although for poorer and smaller countries often there is only one survey available.

The number of firms surveyed varies according to country size. For larger countries, a survey typically includes between 1,200 to 1,800 firms. I would recommend using a relatively larger country for the exercise to have more variability in terms of firms operating across different industries and greater availability of performance measures to analyze. Panel data is sometimes available for larger middle-income countries, but for the purposes of this exercise, only one cross-sectional survey is needed.

---

[4] All surveys have country-specific questions, but these are not used in the current exercise.

[5] First click in the Data tab, then Survey Datasets and lastly click on the Firm-level Datasets for Researchers link. In the right-hand side of the page you can find the link to create a new account: https://login.enterprisesurveys.org/content/sites/financeandprivatesector/en/signup.html, which requires an institutional email account and a one-paragraph summary of the research project.

The raw data obtained from WBES is available in Stata (.dta) format along with the questionnaire used in the survey. All variables in the Stata file are labelled and mapped directly to the survey questionnaire, where the name of the variable is written in red in the questionnaire's pdf file.

3. Replicating the stylized facts of firm-heterogeneity and trade

This activity asks students to replicate some of the key stylized facts that have motivated the research agenda on firm heterogeneity and international trade, as documented in Bernard et al. (2007) for U.S. manufacturing firms. The stylized facts that characterize firms' global engagement that can be analyzed using the WBES data are:

(1) There is substantial heterogeneity across firms 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 only operate domestically across a broad range of performance measures such as employment, skill- and capital-intensity, sales per worker and innovation.

For the purposes of this article, I illustrate the empirical exercise I carry out using the 2017 survey wave for Colombia, which is the one I have used in my own class. This dataset includes 993 firms, 569 of which operate in the manufacturing sector, and which constitute the sample that I use to carry out the empirical analysis. Appendix A describes the construction of each individual variable, and Appendix B provides the

Stata do-file to reproduce the statistics proposed in this activity, which can be readily used with any WBES dataset.

Set up

I have carried out this exercise during a 50-minute tutorial session (small group class with approximately 20 students) in a second-year undergraduate course in international trade and in a masters-level class on international business economics that I teach at City, University of London in the UK. These Students have taken classes in data analysis and introductory econometrics using Stata as the main statistical software. Most of the analysis, however, only requires students to calculate simple summary statistics, which could be easily done in Microsoft Excel.

In preparation for the exercise, I ask students to set up an account with WBES and download the raw data in advance and bring their laptop to class.[6] In the tutorial I ask them to work in groups of 3-4 students. I start the activity by opening both the Stata dataset and the questionnaire to show students the way in which the data is organized and remind students about basic commands used in Stata to tabulate and summarize the data. After letting students work on the data on their own (while circulating around the room answering clarifying questions), I ask groups to share their findings with the rest of the class and proceed to contrast the results with the findings of Bernard et al. (2007), which I have previously presented in the main lecture.

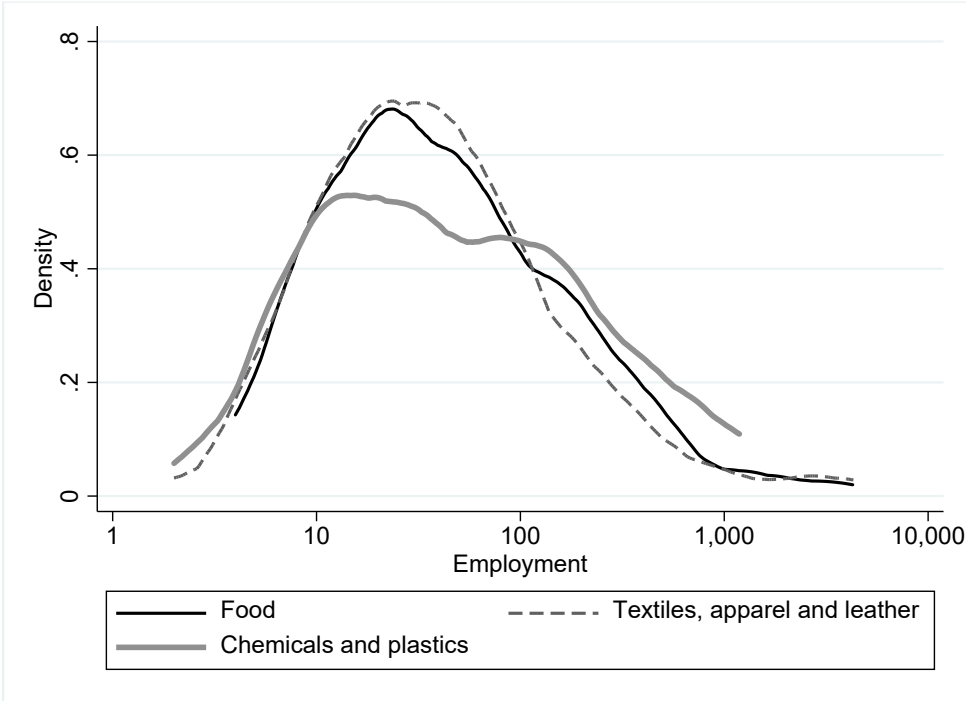
[6] Students at City, University of London have virtual access to Stata in their own devices through an online portal called Appsanywhere.

Assessing the extent of firm heterogeneity

Summarizing total permanent employment (variable l.1 in WBES surveys) 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. Consistent with the international evidence, there are a few 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 within industries.

Figure 1 is crucial to convey the key innovation of models of trade with heterogeneous firms – firms differ quite substantially within industries. Crucially, as the next sections of the exercise will show, the extent to which firms engage with the rest of the world also differs substantially across firms.

Figure 1: Size Distribution of Firms – Selected Industries (Colombia, 2017)



How prevalent is global engagement?

The second stylized fact that students replicate pertains to the extent of firms' global engagement – i.e. whether they export and import and if they are foreign-owned. To this end, students must first construct indicator variables for firms' export-, import- and foreign-owned-status.

Students are asked to tabulate each indicator, first across all manufacturing firms and then for each manufacturing industry. The key message from computing these summary statistics, which is also illustrated by Figure 2 below, is the notable degree of heterogeneity in terms of the margins 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; in contrast, importing is much more prevalent among Colombian firms – two thirds of them import any materials or inputs, while in the U.S. the corresponding share is only 14%. Foreign ownership is the least common dimension of global engagement, accounting for only 7% of manufacturing firms in Colombia – a figure consistent with the stylized facts reported by Àntras and Yeaple (2014).

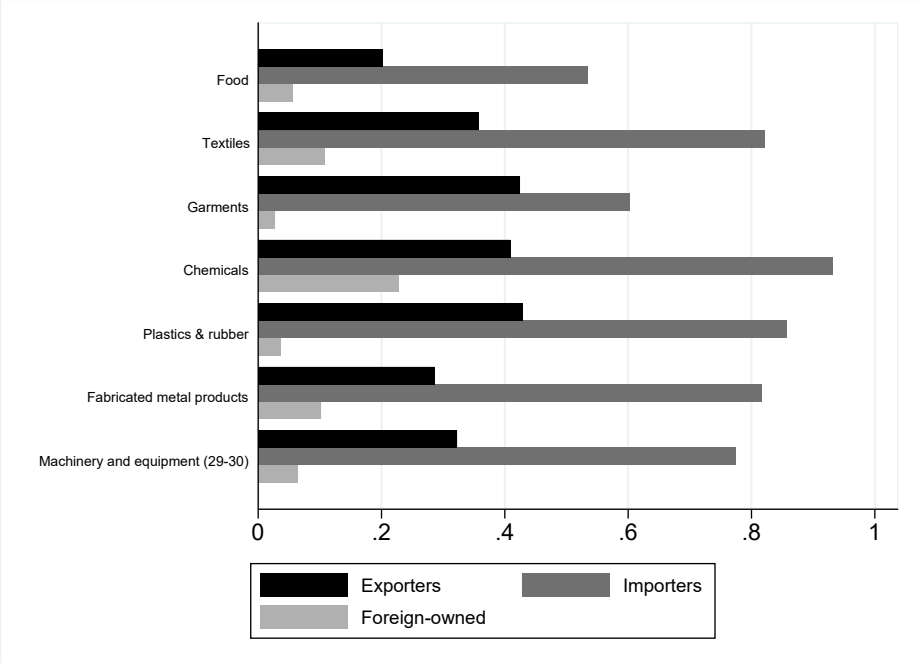
The large differences in the extent of firms' global engagement across industries offers an opportunity to revisit topics that are likely to have been covered earlier in an international trade course. For example, before starting the exercise, I ask students what they think are the main industries in which the country under analysis has comparative advantage and then compare their answers with the WBES data.

Are there any surprises in terms of the industries that exhibit a higher prevalence of exporters? When I have carried out this activity in my own class, students expect Colombia to have a comparative advantage in labor-intensive sectors like food or textiles – on the basis of comparative advantage based on countries' relative

endowments as in the Hecksher-Ohlin model – and thus find surprising that the share of exporters is higher in capital-intensive industries like chemicals and machinery and equipment. I then ask them if there other factors that could explain these data patterns. In the case of Colombia, I use data of exports by industry and destination country publicly available in the Observatory of Economic Complexity website level (https://oec.world/) to show students that capital-intensive exports are often shipped to neighboring countries, which are less capital abundant than Colombia.

Comparing the prevalence of importers and exporters within a given industry can also 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?

Figure 2: Share of Globally Engaged Firms by Industry (Colombia, 2017)



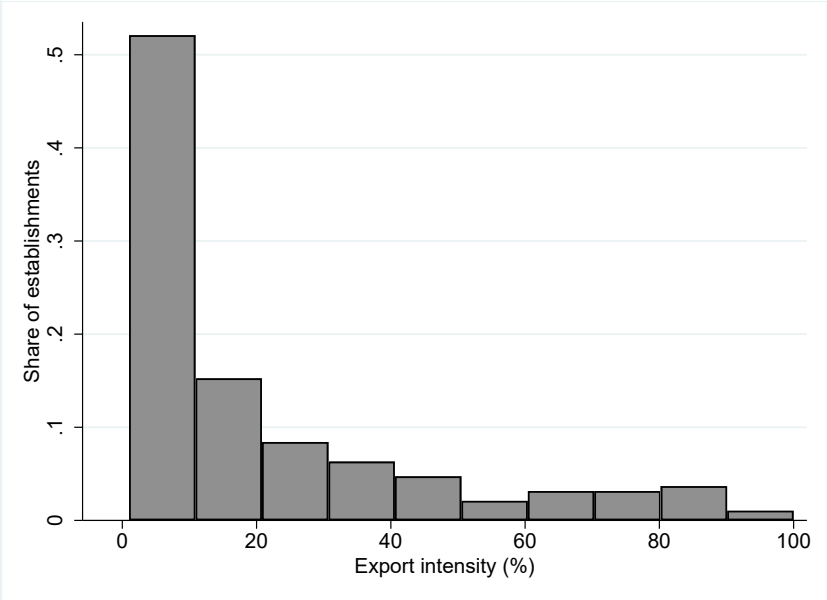
The figure reports the share of globally engaged firms in manufacturing industries with at least 20 firms.

How important are export sales for exporting firms?

Bernard et al. (2007) show that in the U.S. the majority of exporting firms sell most of their output domestically – or in the 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% or less of their total sales, and only a small minority of firms export most of their output. Defever and Riaño (2022), however, using WBES data for 72 countries, show that this stylized fact is in fact not that robust across the world. Export intensity distributions vary tremendously across countries and often display bimodality – large shares of both low- and high-intensity exporters coexisting alongside each other within a country.

Using a country with a high prevalence of high-intensity exporters such as Bangladesh, Ireland or Turkey to give just a few examples, can be useful to discuss the reasons for the existence of these exporters, which do not arise in the simplest two-country version of the Melitz model when the fixed cost of exporting is higher than the fixed cost of operating domestically. Are these firms producing goods for which there is little to no domestic demand such as wool sweaters in Bangladesh (Díaz de Astarloa et al. 2013) ? Or are they producing highly-specialized goods that are exported to the next stage of production within a global value chain (Àntras 2020)? Perhaps the country under consideration encourages firms to locate in special economic zones, where they enjoy a wide array of fiscal incentives, but are often required to export most of their output (Defever and Riaño 2017; Defever et al. 2019).

Figure 3: Export Intensity Distribution (Colombia, 2017)



Export intensity is defined as the share of a firm's total sales accounted for by exports conditional on exporting and thus lies on the interval (0,100%].

Performance premia of global engagement

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

$$\ln Y_i = \alpha + \beta GLOB_i + \varepsilon_i, \quad (1)$$

Where $\ln Y_i$ denotes the natural logarithm of a given performance outcome for firm i, and $GLOB_i$ is a dummy variable taking the value 1 if firm i is globally engaged (exporting, importing or being foreign owned) and 0 otherwise.[7] Table 1 reports the

[7] Bernard et al. (2007) also report regressions like equation (1) that include industry fixed effects and controlling for firms' employment. These regressions can be replicated too, but the addition of new controls might confuse students without providing further economic insights.

performance premium in percentage terms, i.e. $\exp(\hat{\beta}) - 1$, for each combination of performance outcome and margin of global engagement, where $\hat{\beta}$ is the estimated coefficient associated with the variable $GLOB$ in regression (1). Thus, Colombian manufacturing exporters are 203% larger in terms of employment than non-exporters, while importers are 51.3% larger along the same dimension than non-importing firms, and so on. If some students are not familiar with regression analysis, the same analysis can be carried out by simply calculating the difference in the mean of performance measures for globally-engaged and non-globally engaged firms.

Table 1: Performance Premia of Different Indicators of Global Engagement (Colombia, 2017)

Performance measure	Performance premia		
	Exporter	Importer	Foreign-owned
log employment	2.034[a]	0.513[a]	3.055[a]
log sales	4.635[a]	0.927[a]	8.777[a]
log sales per worker	0.881[a]	0.195[c]	1.342[a]
log capital intensity	0.423	0.195	0.710
log skill intensity	-0.067	0.053	0.153
log R&D expenditure	4.104[a]	1.131[a]	4.278[a]

The table reports $\exp(\hat{\beta}) - 1$, in which $\hat{\beta}$ denotes the estimated coefficient in a bivariate regression of the type represented in equation (1) in which 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. [a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level.

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. In contrast, globally-engaged firms in Colombia are not significantly more capital nor skill-intensive than non-globally engaged firms. Another important insight that can be gleaned from the results in Table 1 is that the size and productivity premia of being foreign-owned are substantially larger than those from being an exporter. This result is consistent with the

theoretical model of horizontal FDI of Helpman et al. (2004) and can be used to motivate the proximity-concentration hypothesis as a motivation for firms to reach foreign consumers by means of foreign affiliates rather than exporting.

An interesting question to ask students after having computed these performance premia is whether these can be given a causal interpretation. On one hand, the workhorse models of firm heterogeneity like Melitz (2003) or Helpman et al. (2004) assume that global engagement is fully explained by firms' selection into these activities as a consequence of their higher 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 et al. 1998; van Biesebroeck 2005).

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 et al. 2017).[8] The podcast, which only lasts 18 minutes, not only illustrates vividly the difficulties in moving from correlation to causation, but also helps to show concrete mechanisms through which exposure to international markets can improve firms' performance and the policy instruments governments can use to achieve these goals.

Conclusion

In this article I described an instructional empirical activity that allows students to work with otherwise hard to access firm-level data from a wide range of countries with the

[8] See https://tradetalkspodcast.com/podcast/62-randomized-trade/.

objective of reproducing key stylized facts that characterize the large differences observed across firms and the extent of their global engagement.

Students who have participated in the activity 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 to achieve the learning objective of applying theoretical tools of international trade theory to analyse 'real world' situations related to the international movement of goods and services.

To conclude, it is important to reemphasize the richness of WBES in terms of the information they contain. These data can be effectively used in a similar manner to what this article suggests across a broad range of courses in development economics, industrial organization, labor and public economics.

References

Àntras, P. 2020. Conceptual Aspects of Global Value Chains. *World Bank Economic Review* 34 (3): 551-574.

Àntras, P. and S. R. Yeaple. 2014. Multinational Firms and the Structure of International Trade. In *Handbook of International Economics*, edited by G. Gopinath, E. Helpman, and K. Rogoff, Vol. 4. Elsevier, 55–130.

Atkin, D., A. K. Khandelwal and A. Osman. 2017. Exporting and Firm Performance: Evidence from a Randomized Experiment. *Quarterly Journal of Economics* 132 (2): 551-615.

Axtell, R. L. 2001. Zipf Distribution of U.S. Firm Sizes. *Science*, 293 (5536): 1818–1820.

Bernard, A. B. and J. B. Jensen. 1995. Exporters, Jobs, and Wages in U.S. Manufacturing: 1976–1987. *Brookings Papers on Economic Activity* 26: 67–119.

Bernard, A. B. and J. B. Jensen. 1999. Exceptional Exporter Performance: Cause, Effect, or Both? *Journal of International Economics* 47 (1): 1–25.

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.

Cabral, L. M. B. and J. Mata. 2003. On the Evolution of the Firm Size Distribution: Facts and Theory. *American Economic Review* 93 (4): 1075-1090.

Clerides, S. K., S. Lach and J. R. Tybout. 1998. Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico, and Morocco. *Quarterly Journal of Economics* 113 (3): 903-947.

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.

Davies, D. R. and D. E. Weinstein. 2002. What role for empirics in international trade? In *Bertil Ohlin: A Centennial Celebration, 1899-1999*, edited by Findlay, R., L. Jonung and M. Lundahl. Cambridge: 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): 532–568.

Defever, F. and A. Riaño. 2017. Subsidies with Export Share Requirements in China. *Journal of Development Economics* 126 (C): 33–51.

Defever, F. and A. Riaño. 2022. The Twin Peaks of the Export Intensity Distribution. *Journal of the European Economic Association* 22 (3): 1347–1394.

Díaz 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. Working paper, International Growth Centre.

Economics Network. 2019. Survey of Employers of Economics Graduates Results Report. Available at: https://www.economicsnetwork.ac.uk/projects/surveys/employers2019

Feenstra, R. C. 2016. *Advanced International Trade*. 2nd ed. Princeton University Press.

Feenstra, R. C., and A. Taylor. 2021. *International trade*. 5th ed. Macmillan international higher education.

Gerber, J. 2018. *International economics*. 7th ed. Boston: 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.

Krugman, P., M. Obstfeld, and M. Melitz. 2023. *International trade theory & policy, global edition*. 12th ed. New York: Pearson.

Lee, D. Y. 1992. Internationalizing the Principles of Economics Course: A Survey of Textbooks. *Journal of Economic Education* 23 (1): 79-88.

McLaren, J. 2012. *International trade*. 1st ed. Hoboken, NJ: John Wiley & Sons.

Melitz, M. J. 2003. The Impact of Trade on Intra-Industry Reallocations and Aggregate Productivity. *Econometrica* 71 (6): 1695–1725.

Roberts, M. J. and J. R. Tybout. 1996. *Industrial Evolution in Developing Countries: Micro Patterns of Turnover, Productivity and Market Structure*. Oxford University Press.

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): 373-391.

World Bank Enterprise Surveys, http://www.enterprisesurveys.org.

Appendix A: Variables that can be constructed with the raw data

Each survey provides information on a firm's main sector of operation, manufacturing, retail, and other services (a0), and within manufacturing, 3-digit ISIC Rev. 4 industry (20 industries) (a4b). Below I discuss how to construct the different variables used in the exercise.

Export status. Question d.3 asks the percentage of a firm's sales that were (i) national sales (d3a); (ii) indirect exports (sold domestically to a third party that exports products) (d3b) and direct exports (d3c). An export indicator takes the value 1 if the percentage of a firm's sales accounted for either direct or indirect exports is strictly positive or zero otherwise.

Importer status. Question d.12 asks the percentage of a firm's material inputs or supplies that were (i) of domestic origin (d12a) and (ii) of foreign origin (d12b). An importer indicator takes the value 1 if the percentage of material inputs or supplies of foreign origin is strictly positive or zero otherwise.

Foreign ownership. Question b.2 asks the percentage of the firm owned by (i) private domestic individuals, companies or organizations (b2a); (ii) private *foreign* individuals, companies or organizations (b2b); government or state (b2c); and other (b2d) which can be used to construct an indicator of foreign ownership based on whether the variable b2b is greater than 10% or 50%.

Sales. Question d.2. asks the value of the establishment's total annual sales across all products and services it sells during last fiscal year.

Employment. Question l.1. asks for the number of permanent, full-time individuals employed in the firm in the last fiscal year. Total permanent employment is further decomposed in production (l3a) and non-production, e.g. employees – e.g. managers, administration, sales (l3b).

Capital. Question n.6a provides information on the net book value (i.e., value of assets after depreciation) for machinery, vehicles, and equipment.

Expenditure in R&D. Question h9 asks how much did an firm spent on R&D, either in-house or externally in local currency units.

Appendix B: Stata do-file

```
use Colombia-2017-full-data.dta, replace

* Keeping only manufacturing firms
keep if a0==1

* Determining how many manufacturing firms are there in the
data
     tab a0

*Firm heterogeneity: examples

gen log10emp = log10(l1) twoway (kdensity log10emp if
a4b==15) (kdensity log10emp if a4b==17 | a4b==18 | a4b==19)
(kdensity log10emp if a4b==24 | a4b==25)


*Indicators of global engagement: exporting, importing,
being foreign-owned

* Create a dummy variable called 'export' = 1 if an
establishment exports some of its output (directly or
indirectly) and 0 otherwise

gen pctexp = d3b + d3c
gen export=.
replace export=0
replace export=1 if pctexp>0

* Create a dummy variable called 'import' = 1 if an
establishment imports some of its inputs and 0 otherwise

gen import=.
replace import=1 if d12b>0 & d12b~=.
replace import=0 if d12b==0 & d12b~=-9

* Create a dummy variable called 'foreign' = 1 if the % of
establishment owned by foreigners is at least 10% and 0
otherwise
```

```
gen b2 = b2a + b2b + b2c + b2d
gen foreign=.
replace foreign=1 if b2b>=10
replace foreign=0 if b2b<10 & b2~=.

tab export
tab import
tab foreign

* Calculate observations per industry
egen obsind = count(id), by(a4b)

* plotting incidence of exporting, importing and being
foreign-owned in sectors with more than 20 establishments:

graph hbar (mean) export import foreign if obsind>=20,
over(a4b)

* Plotting export intensity distribution:
hist pctexp if pctexp>0, bin(10) frac xtitle(export
intensity)


* Constructing performance indicators:
* log employment
gen logemp = log(l1)

* log skill-intensity (share of permanent non-production
workers in total permanent workers)
gen skillint = l3b/(l3a + l3b)
replace skillint=. if skillint<0 | skillint>1
gen logskillint = log(skillint)

* log sales
gen logsales = log(d2)

* log sales per worker
gen salespw = d2/l1
replace salespw=. if salespw<0
gen logsales_pw = log(d2/l1)

* log capital per worker
gen kpw = n6a/l1
gen logk_pw = log(kpw)

* log female employment
gen femempint = (l5a + l5b)/l1
replace femempint=. if femempint<0 | femempint>1
gen logfemempint = log(femempint)

* log R&D expenditure/sales
```

```
gen rd = h9
replace rd=. if rd<0
gen logrd = log(rd)

* export regressions
reg logemp export,r
reg logsales export, r
reg logsales_pw export,r
reg logk_pw export,r
reg logskillint export,r
reg logrd export,r

* import regressions
reg logemp import,r
reg logsales import, r
reg logsales_pw import,r
reg logk_pw import,r
reg logskillint import,r
reg logrd import,r

* foreign-owned regressions
reg logemp foreign,r
reg logsales foreign, r
reg logsales_pw foreign,r
reg logk_pw foreign,r
reg logskillint foreign,r
reg logrd foreign,r
```