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

This online appendix describes the construction of variables used to carry out the statistical analysis proposed in the paper. It also presents the do-file used to carry out the analysis.

1 Variables that can be constructed with the raw data

Each survey provides information on a firm's main sector of operation, i.e. manufacturing, retail, and other services (a0), and within manufacturing, 3-digit ISIC Rev. 4 industry (20 industries) (a4b). Below I discuss how to construct variables related to firms' global engagement and performance indicators used in this paper to exemplify the type of empirical analysis that can be carried out with WBES surveys. As noted in Section **??**, WBES surveys cover a broad range of topics, and therefore, the following list of variables is not exhaustive.

1.1 Global engagement variables

- *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). The export indicator variable 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. Students can explore whether the number of exporting firms changes significantly when restricting the definition to only include direct exporters and discuss potential reasons why some firms rely on intermediaries to sell abroad.

- *Export intensity.* Question d.3 is also used to create the export intensity variable for exporters. It is important to remember to create export intensity only for firms that have strictly positive export sales, or, in other words, for those firms for which the export status indicator defined above is equal to 1.

- *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%.

1.2 Performance variables

- *Sales.* Question d.2. asks the value of the establishment's total annual sales across all products and services it sold 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—i.e. those involved in the firm's management, administration and sales (l3b).

- *Capital stock.* Question n.6a provides information on the net book value (i.e. value of assets after depreciation) for machinery, vehicles, and equipment in local currency units.

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

2 Stata do-file

This appendix provides the Stata do-file used to carry out the analysis performed in the paper.

```
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 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
```