Brazil - World Bank

Carlos Góes

December 2023

1 Variables Definitions

- $L_{r,i,f,t}$: formal employment in firm f coded in industry i (5-digit *CNAE95* sector code) and based in region r (meso- or micro-region) and year t (source: RAIS).
- $L_{r,i,t} \equiv \sum_f L_{r,i,f,t}$: formal employment in industry i (5-digit *CNAE95* sector code) based in region r (meso- or micro-region) and year t (source: RAIS).
- $L_{r,g,t}$: formal employment in region r (meso- or micro-region) and year t by gender $g \in \{male, female\}$. (source: RAIS).
- $L_{r,s,t}$: formal employment in region r (meso- or micro-region) and year t by schooling $s \in \{less-than-primary, primary, secondary, college-or-higher\}$. (source: RAIS).
- $L_{r,e,t}$: formal employment in region r (meso- or micro-region) and year t by environmental status $e \in \{green-sector/green-occupation, brown-sector/brown-occupation\}$. (source: RAIS + World Bank classification).
- $w_{r,t}$: nominal wages in region r (meso- or micro-region) and year t (sources: RAIS, PNAD).
- $w_{r,g,t}$: nominal wages in region r (meso- or micro-region) and year t by gender $g \in \{male, female\}$. (source: RAIS, PNAD).
- $w_{r,s,t}$: nominal wages in region r (meso- or micro-region) and year t by schooling $s \in \{less-than-primary, primary, secondary, college-or-higher\}$. (source: RAIS, PNAD).
- $L_{r,e,t}$: formal employment in region r (meso- or micro-region) and year t by environmental status $e \in \{green-sector/green-occupation, brown-sector/brown-occupation\}$. (source: RAIS + World Bank classification).

- $X_{r,i,t}$: exports of industry i (5-digit *CNAE95* sector code) based in region r (meso- or micro-region) and year t (source: MDIC?) – IF NOT AVAILABLE, WE CAN USE BARTIK STRATEGY AT AN AGGREGATE LEVEL (source: UNCOMTRADE).
- $Y_{d,t}$: Real GDP (or nominal USD GDP) for all trade partners of Brazil (source: WDI).
- Ask MDIC decide among Brazil team which other variables would be important on the left hand side.

2 Simple Explanation of the Methodology in a Static Setting — we will actually do the analysis in a dynamic setting but it is easier to understand the methodology in a static framing, so I am sending this version

Method In order to measure the impact of exports over local labor markets, we would ideally observe the share of exports produced in each region $r \in \mathcal{R}$. If we observe $X_{r,i,t}$ directly we will use it. However, these data are typically not available for most countries. Even for countries that do report export by region, these data typically do not account for location, but rather the location of the exporting firm – which could be an intermediary.

To circumvent such limitation, we use export growth in different industries $i \in \mathcal{I}$ interacted with differential exposure to industry-specific shocks across different local labor markets. Formally, we define local labor market exposure to exports growth as:

$$\Delta\widetilde{X}_{r,t+h} \equiv \sum_{i\in\mathcal{I}} \frac{L_{r,i,t}}{L_{r,t}} \cdot \Delta\widetilde{X}_{i,t+h} = \sum_{i\in\mathcal{I}} \frac{L_{r,i,t}}{L_{r,t}} \cdot \left(\widetilde{X}_{i,t+h} - \widetilde{X}_{i,t}\right)$$

where $\widetilde{X}_{i,t}$ denotes total exports of industry i at period t; $L_{r,i,t}$ denotes total employment in region r and industry i; and $L_{r,t} \equiv \sum_{i\in\mathcal{R}} L_{r,i,t}$ is total aggregate employment in region r. This kind of shift-share approach mirrors the approach of many papers that study the impact of trade shocks on local labor markets, including Autor, Dorn, and Hanson (2013) and Dix-Carneiro and Kovak (2015), for imports exposure, and Robertson et al. (2021) for exports exposure.

Given the shares $\frac{L_{r,i,t}}{L_{r,t}}$, we can potentially estimate the effect of exports over local labor markets by regressing some variable of interest $\Delta O_{r,i,t+h}$, which in our case will be primarily the change in the female-to-male employment ratio, on the shift-share regressand above, provided that the shifters $\Delta\widetilde{X}_{i,t+h}$ are as good as random (for a formal treatment, see Borusyak, Hull, and Jaravel 2020). If that were the case, we would be able to run the regression:

$$\Delta O_{r,t+h} = \alpha + \beta \Delta \widetilde{X}_{r,t+h} + \boldsymbol{Z'}_{r,t+h} \boldsymbol{\delta} + \varepsilon_{r,t+h} va$$

for which estimation of β is consistent if $E\left[\Delta \widetilde{X}_{i,t+h} \cdot \varepsilon_{r,t+h} \middle| \boldsymbol{Z}_{r,t+h},\ L_{r,i,t}\right] = 0$ for every i and r pair – i.e., if conditional on controls $\boldsymbol{Z}_{r,i,t+h}$ and on shares, changes in exports are uncorrelated with unobserved local labor markets shocks.

However, since exports depend partially on domestic human capital and technology use, which can be correlated with characteristics of local labor markets, the shifters $\Delta \widetilde{X}_{i,t+h}$ are likely not exogenous. For that reason, we instrument $\Delta \widetilde{X}_{r,i,t+h}$ with increases in foreign demand, proxied by changes in dollar GDP in foreign destinations.

For clarity, we need to introduce some notation. Let $\mathcal{S}$ be the set of countries in the world and let $s \in \mathcal{S}$ denote the source country —in this case, Brazil. Brazil exports to countries other than itself —or to destinations $d \in \mathcal{S} \setminus s$. We denote the exports of each industry i as the sum of its sales to every foreign destinations: $\widetilde{X}_{i,t} = \sum_{d \in \mathcal{S} \setminus s} \widetilde{X}_{d,i,t}$. Our foreign-demand instrument leverages the correlation between changes in exports to destination d and changes in dollar GDP, which is expected given the gravity structure typical of international trade. It is:

$$\Delta \bar{X}_{r,t+h} \equiv \sum_{i \in \mathcal{I}} \frac{L_{r,i,t}}{L_{r,t}} \cdot \sum_{d \in \mathcal{S} \setminus s} \frac{\widetilde{X}_{d,i,t}}{\widetilde{X}_{i,t}} \cdot \Delta Y_{d,t+h} \tag{1}$$

where $\frac{\widetilde{X}_{d,i,t}}{\widetilde{X}_{i,t}}$ denotes country d's share of industry i's exports; and $\Delta Y_{d,t+h}$ is the change in U.S. dollar GDP in country d. Note that this takes into account **every country that Brazil exports to** in **every industry**. Therefore, this instrumentation strategy is similar to those such as in Aghion et al. (2018), in which one takes considers demand to all foreign destinations as a proxy for exogenous variation in demand in a given industry.

Estimation now takes the form of two-stage least squares, with the first stage being:

$$\Delta \widetilde{X}_{r,t+h} = \omega + \gamma \Delta \bar{X}_{r,t+h} \ + \ \boldsymbol{Z'}_{r,t+h} \boldsymbol{\phi} + \bar{\varepsilon}_{r,t+h} \tag{2}$$

and the second stage is:

$$\Delta O_{r,t+h} = \alpha + \beta \Delta \hat{X}_{r,t+h} \ + \ \boldsymbol{Z'}_{r,t+h} \boldsymbol{\delta} + \varepsilon_{r,t+h}$$

where $\Delta \hat{X}_{r,t+h}$ are the predicted values of the first stage regression. Now estimation of β is consistent if $E\left[\Delta Y_{d,t+h} \cdot \varepsilon_{r,t+h} \middle| \boldsymbol{Z}_{r,t+h}, L_{r,i,t}, \widetilde{X}_{d,i,t}\right] = 0$ for every d and r pair – i.e. if changes in foreign demand are uncorrelated with unobserved factors that drive changes local labor markets in Brazil.