

Worker Heterogeneity, Wage Inequality, and International Trade: Theory and Evidence from Brazil*

Rodrigo Adão

Massachusetts Institute of Technology

Job Market Paper

November 18, 2015

[latest version]

Abstract

This paper proposes a new approach to quantify the distributional effects of international trade. The starting point of my analysis is a Roy-like model where workers are heterogeneous in terms of their comparative and absolute advantage. In this environment, I show that the schedules of comparative and absolute advantage (i) determine changes in the average and the variance of the log-wage distribution, and (ii) are nonparametrically identified from the cross-regional variation in the sectoral responses of employment and wages to observable sector-level demand shifters. I then use these theoretical results to quantify the distributional consequences of the recent movements in world commodity prices in Brazil. I find that shocks to world commodity prices account for 5–10% of the fall in Brazilian wage inequality between 1991 and 2010.

*Author contact information: rradao@mit.edu. I am extremely grateful to Daron Acemoglu, Arnaud Costinot and Dave Donaldson for invaluable guidance and support. I also thank Pol Antràs, David Atkin, David Autor, Arthur Bragança, Ariel Burstein, Nicolas Caramp, Rafael Dix-Carneiro, Sebastian Fanelli, Claudio Ferraz, Ernest Liu, Ameya Muley, Laísa Rachter, Dejanir Silva, Ludwig Straub, Iván Werning, Chong Xiang as well as seminar participant at the MIT Labor Lunch, the MIT Macro-International Seminar, and the EIIIT conference. All errors are my own.

1 Introduction

In a global economy, changes in good prices caused by shocks in one part of the world have the potential to affect factor prices in another. As shown in Panel A of Figure 1, between 1981 and 2010, increases in the world prices of basic commodities were accompanied by reductions in Brazilian wage inequality. Given the importance of the commodity sector in employment of low-income workers, this correlation suggests that changes in world demand for basic goods plausibly contributed to changes in wage inequality in Brazil.¹ Panel B reinforces this view by showing that increases in world commodity prices were also associated with increases in both the relative employment and the relative wage in the commodity sector. In this paper, I develop a new empirical strategy to quantify the causal effect of global shocks in commodity prices on Brazilian wage inequality.

My starting point is a theoretical framework where Brazil is assumed to be a collection of small open economies with segmented labor markets. Each regional economy is populated by workers of multiple demographic groups that can be employed either in the commodity or in the non-commodity sectors. The central feature of the model is a Roy's (1951) structure of within-group worker heterogeneity in terms of sector-specific productivity. Conditional on sectoral wages per efficiency unit, workers self-select into sectors according to their comparative advantages; defined as the productivity ratio in the commodity and the non-commodity sectors. In the model, workers' labor income depend on their comparative advantages as well as their absolute advantages; defined as the productivity in the non-commodity sector.

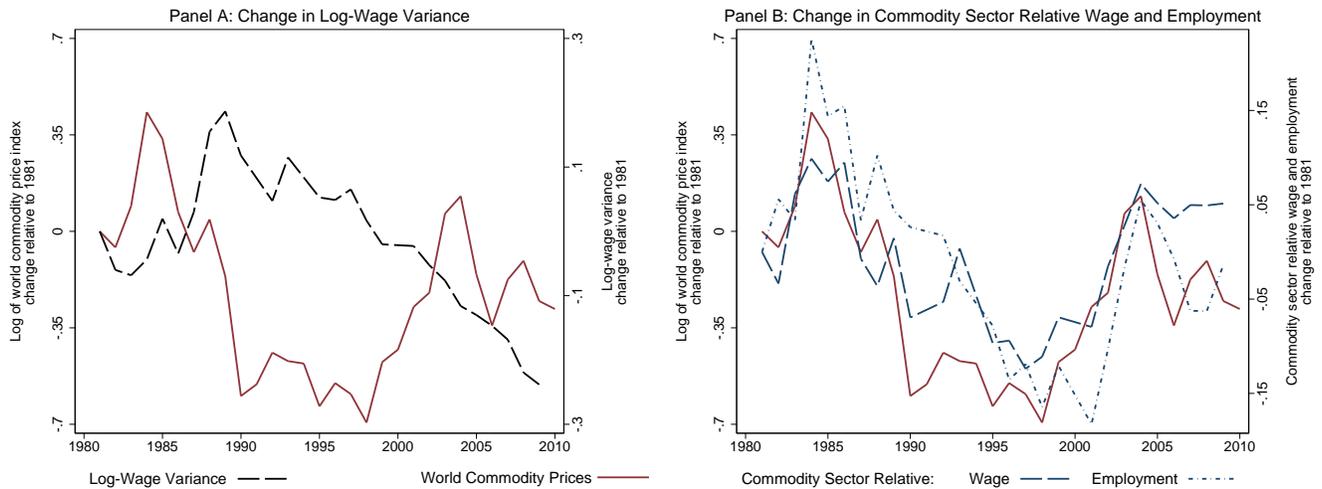
In this environment, comparative and absolute advantage have distinct roles in determining sectoral responses of employment and wages following shocks to the world prices of goods. By affecting the marginal value of labor in each sector, world price shocks induce changes in the sectoral relative wage per efficiency unit. This causes between-sector worker reallocation with magnitude regulated by the comparative advantage distribution, which I refer to as the schedule of comparative advantage. The subsequent between-sector response in average wage combines two terms. The first term is the impact of the change in the relative wage per efficiency unit for a given allocation of workers across sectors. The second term is the compositional effect stemming from the difference in the average sector-specific efficiency of sector-switchers relative to that of sector-stayers. The magnitude of this compositional effect depends on the average of the absolute advantage distribution conditional on comparative advantage, which I refer to as the schedule of absolute advantage.

These sectoral shocks trigger changes in wage inequality, both between and within worker groups. To quantify such distributional effects in the model, I focus on the shock's impact over the average and the variance of the log-wage distribution of different demographic groups. Following sectoral demand shocks, I show that responses in these outcomes are exclusively determined by the schedules of comparative and absolute advantage. Thus, knowledge of these two schedules permits a quantitative evaluation of the impact of world price shocks on wage inequality.

I then turn to the problem of recovering the schedules of comparative and absolute advantage from

¹Production of agricultural and mining products constitutes an important share of the Brazilian economy, representing, in 2010, 58.5% of exports and 19.9% of employment. Commodity sector employees earned, on average, 28.1% less than employees of other sectors in 2010. In Appendix C, I show that the component associated with workers' observable characteristics was the main driver of the movements in log-wage variance between 1981 and 2009.

Figure 1: World Commodity Prices and the Brazilian Labor Market, 1981-2010



Note: World Commodity Price is the log of the commodity price index computed with the world price of agriculture and mining products converted to Brazilian currency and deflated by the Brazilian consumer price index. Sample of full-time employed males is extracted from the National Household Sample Survey (PNAD). Commodity sector relative wage is the coefficient of the dummy for employment in commodity sector from the regression of log wage on worker attributes. Change in commodity sector relative employment is the average of the change in the log of the employment ratio in the commodity and non-commodity sectors for High School Graduates and Dropouts weighted by the group size in 1981. Details in Appendix C.

observable labor market outcomes. The challenge inherent in identifying these functions is conveyed by Heckman and Honoré’s (1990) result that, in the context of the Roy model, the sector-specific productivity distribution is not nonparametrically identified in a single cross-section of individuals. In this paper, I establish the identification of the schedules of comparative and absolute advantage in a set of regional economies. For any number of demographic worker groups, my identification result allows the two schedules to have an arbitrary shape. But it requires two central assumptions. First, I assume that observed covariates and unobserved shocks are additive shifters of the two schedules across regions. Second, given the unobserved productivity shocks, I make the standard assumption that there are excludable shifters of sector labor demand across regions. Under these assumptions, the two schedules are nonparametrically identified from cross-regional variation in sectoral responses of employment and average wages to changes in sectoral wages per efficiency unit induced by the observable sector-level demand shifter.

My nonparametric identification result is critical to inform the source of variation in the data that separately uncovers comparative and absolute advantage. An empirical application based on this result accounts for the conceptually distinct roles of these two schedules in the model. This approach contrasts with recent empirical applications of Roy-like models that build upon a productivity distribution of the Fréchet family; e.g., see Hsieh, Hurst, Jones and Klenow (2013), Burstein, Morales and Vogel (2015), and Galle, Rodriguez-Clare and Yi (2015). This distribution, although highly tractable, mixes the channels of comparative and absolute advantage with strong consequences for the model’s predictions: it implies that both sector wage differentials and log-wage variance are invariant to labor demand conditions. To incorporate these potentially important channels while maintaining tractability, my empirical application relies on a parsimonious log-linear system that strictly generalizes the system implied by the Fréchet distribution. The log-linear system contains two structural parameters that specify constant-elasticity schedules of comparative and absolute advantage. In comparison, the

Fréchet distribution restricts these two elasticities to have the same absolute value.²

Armed with these theoretical results, I apply the framework to investigate the effect of commodity price shocks on wage inequality in Brazil. To this extent, I estimate the schedules of comparative and absolute advantage in a panel of Brazilian regional economies for two demographic groups, High School Graduates and High School Dropouts. In the empirical application, two variables are needed. First, a regional shifter of sectoral demand, which I construct by interacting the change in commodities' world prices and the pre-shock participation of corresponding commodities in the region's labor payroll.³ Second, a measure of the sector wage per efficiency unit that is not immediately available in survey datasets. To estimate changes in the sector wage per efficiency unit, I propose a strategy that builds upon the model's predicted relation between wage growth and initial sector employment across quantiles of the wage distribution. For each group and region, I implement this strategy as a first-step regression using repeated cross-section data on wage and employment at the individual level.

I start by investigating the effect of exposure to commodity price shocks on sectoral labor market outcomes across Brazilian regional economies. This reduced-form exercise establishes the basic relations in the data that drive the estimation of the structural parameters of comparative and absolute advantage. For both worker groups, I find that regional economies exposed to stronger price shocks experienced stronger expansions in the commodity sector relative employment. In addition, shock exposure induced increases in the relative wage per efficiency unit of the commodity sector. The combination of these two responses determines the elasticity of the comparative advantage schedule. Lastly, I investigate the effect of shock exposure on the commodity sector wage differential, finding a positive and statistically significant response for High School Graduates and a small and statistically non-significant response for High School Dropouts. Following the commodity price shock, the change in the relative sector average wage was smaller than the change in the relative wage per efficiency unit. This wedge corresponds to the compositional effect that determines the elasticity of the absolute advantage schedule. Results are robust to the inclusion of region fixed-effects, initial region socio economic characteristics interacted with period dummies, and region-specific time trends.

Having established these reduced-form patterns in the sample of Brazilian regions, I turn to the estimation of the structural parameters separately for High School Graduates and Dropouts. Results indicate that the two groups have similar comparative advantage schedules, implying that they exhibit comparable degrees of between-sector mobility. The distinct responses in the sector wage differential for the two groups leads to different estimated coefficients of absolute advantage. Among High School Dropouts, estimates are consistent with those of a Fréchet distribution and, for this group, compositional effects completely offset the impact of price shocks on sector wage differentials. The estimated selection pattern for High School Graduates, however, differs from that implied by the Fréchet system. For this group, the log-linear system is able to replicate the estimated effect on log-wage variance as-

²In the spirit of the series estimator proposed by [Newey and Powell \(2003\)](#), the system could be augmented to include higher-order polynomials. In practice, data limitations constitute an important challenge to the implementation of a fully flexible instrumental variable estimator. As [Newey \(2013\)](#) pointed out, the estimation of nonlinear terms with instrumental variables tends to be accompanied by sharp increases in standard errors. For this reason, my benchmark specification is based on a parsimonious log-linear system with constant-elasticity schedules of comparative and absolute advantage.

³My demand shifter is implied by the assumption that production of basic commodities utilizes immobile factors like soil fertility and oil reserves whose endowment varies across regions. As a result, following world price shocks, the regional response of the commodity sector labor demand depends on the initial industry composition within the commodity sector.

sociated with exposure to higher commodity prices across Brazilian regions. Such a response is ruled out by the parametric restrictions imposed by the Fréchet distribution.

I conclude the paper by applying the framework to answer one counterfactual question: “In 1991, how would wage inequality change if commodity prices were equal to those of 2010?” To answer this question, I provide two alternative procedures to obtain changes in sectoral wages per efficiency unit stemming from shocks in world commodity price. The first relies on a reduced-form pass-through estimated from the effect of price shock exposure on the wage per efficiency unit in the sample of Brazilian regional economies. While this approach is robust to the specific production structure of the economy, it is not able to capture nationwide effects and it may not hold for shocks on other products and other periods. To address these shortcomings, the second approach relies on a fully specified general equilibrium model where I calibrate the economy’s structure of production. This procedure takes inspiration from the exact hat algebra used in recent international trade papers — see, for example, [Dekle, Eaton and Kortum \(2007\)](#) and, for a review, [Costinot and Rodríguez-Clare \(2013\)](#).

The counterfactual analysis yields similar results with both approaches, delivering two main insights. First, changes in world commodity prices have sizable distributional effects in Brazil. As a result of the 1991–2010 rise in world commodity prices, the relative wage per efficiency unit in the commodity sector increased by 8%–16%. Yet the subsequent worker reallocation created compositional effects that offset most of the shock’s impact on between-sector wage differentials. In terms of overall wage inequality, the price shock accounts for 5%–10% of the decline in Brazilian log-wage variance between 1991 and 2010. Second, flexible functional forms that separate the roles of comparative and absolute advantage are quantitatively important. For High School Graduates, the log-linear model captures 10% of the decrease in log-wage variance, but the Fréchet model implies no change in log-wage variance. In contrast, both specifications yield similar counterfactual changes in the average and the variance of the log-wage distribution for High School Dropouts, reflecting the similarity between the estimated structural parameters obtained with the two parametrizations.

This paper is related to an extensive literature on the labor market effects of international trade. Research on the topic has traditionally relied on neoclassical environments that yield stark predictions regarding the changes in relative wages across worker groups ([Stolper and Samuelson, 1941](#); [Jones, 1965](#)) and relative factor prices across industries ([Jones, 1975](#)). However, empirical studies concluded that the forces highlighted by these models were, at best, secondary drivers of the changes in wage inequality in the 1980s and early 1990s. For instance, a number of authors have documented (i) movements in wage inequality correlated in both developed and developing countries ([Goldberg and Pavcnik, 2007](#)); (ii) movements in the skill wage premium uncorrelated with changes in the relative price of skill-intensive products ([Lawrence and Slaughter, 1993](#)) while correlated with changes in the skill intensity of production within industries ([Bekman, Bound and Machin, 1998](#)); and (iii) limited between-sector responses in employment and wages following trade shocks ([Wacziarg and Wallack, 2004](#) and [Goldberg and Pavcnik, 2007](#)).

This evidence motivated departures from the neoclassic environment, giving rise to a body of work analyzing the effect of international trade on workers employed in different firms within industries (see, for example, [Verhoogen, 2008](#); [Helpman, Itskhoki and Redding, 2010](#); [Frias, Kaplan and Verhoogen, 2012](#); [Helpman, Itskhoki, Muendler and Redding, 2015](#); and [Burstein and Vogel, 2015](#)) and

on the transitional dynamics in the reallocation of workers across sectors and markets (Kambourov, 2009; Artuç, Chaudhuri and McLaren, 2010; Dix-Carneiro, 2014; Dix-Carneiro and Kovak, 2015b; and Caliendo, Dvorkin and Parro, 2015). In this paper, I build upon the neoclassical channel that emphasizes the effect of international trade on relative good prices and, consequently, on relative factor prices. Yet my framework augments traditional models with a flexible structure of sector-specific factor productivity. This idea goes back to the work of Mussa (1982) and Grossman (1983), and its implications for international trade is explored in several recent papers — see, for example, Ohnsorge and Trefler (2007), Costinot and Vogel (2010), Acemoglu and Autor (2011), and, for comprehensive reviews, Grossman (2013) and Costinot and Vogel (2014).⁴ Relative to these papers, my main contribution is to develop a novel empirical methodology that allows the model to be applied in the quantification of distributional effects of international trade shocks. In the context of the shock in world commodity prices of 1991–2010, my framework indicates sizable distributional effects in the Brazilian labor market.

Two recent papers impose a productivity distribution of the Fréchet family to quantify the portion of changes in between-group wage inequality associated with technological progress in the United States (Burstein, Morales and Vogel, 2015) and import competition in Germany (Galle, Rodriguez-Clare and Yi, 2015). My paper differs from these studies in two central aspects of methodology. First, my analysis clearly delineates the distinct roles played by comparative and absolute advantage in determining sectoral employment and sectoral wages, showing how these schedules affect both within and between-group wage inequality. Second, my nonparametric identification result sheds light on the source of variation that uncovers comparative and absolute advantage within each demographic worker group, leading to a new estimation strategy based on cross-market variation in sectoral demand shifters. The results of my counterfactual analysis suggest that the restrictive distributional assumptions imposed by these papers have the potential to significantly affect the quantitative predictions of the model.

This paper is also related to the empirical literature that examines the impact on labor market outcomes of heterogeneous exposure to import competition in terms of sector of employment (Menezes-Filho and Muendler, 2011; and Autor, Dorn, Hanson and Song, 2014), and region of residence (Topalova, 2010; Kovak, 2013; Autor, Dorn and Hanson, 2013; Costa, Garred and Pessoa, 2014; and Dix-Carneiro and Kovak, 2015a). I complement this literature by providing new evidence of sectoral responses of employment and wages using cross-regional variation in exposure to commodity price shocks in a developing country.⁵ My theoretical framework, moreover, connects these responses to structural parameters of comparative and absolute advantage. The structural estimates indicate that, due to compositional effects, the impact of world price shocks on sectoral wage per efficiency unit is larger than one would have inferred from reduced-form regressions based on sector average wages.

Lastly, this paper is related to the large literature investigating the consequences of self-selection

⁴Also, the Roy model has been recently applied in the investigation of the determinants of aggregate productivity — e.g., see Lagakos and Waugh (2013), Hsieh, Hurst, Jones and Klenow (2013), and Young (2014).

⁵Previous studies analyzing the adjustment of labor markets during trade liberalization episodes in developing countries did not find evidence of responses in employment and wages; e.g., see Wacziarg and Wallack (2004) and Goldberg and Pavcnik (2007). By contrast, evidence of sectoral responses in labor market outcomes has been documented in developed countries; e.g., see Revenga (1992), Gaston and Trefler (1997), and Autor, Dorn, Hanson and Song (2014).

based on unobservable characteristics to observable components of labor income — see [French and Taber \(2011\)](#) for a review. In the context of the Roy model, [Heckman and Honoré \(1990\)](#) offer a number of results regarding the nonparametric identification of the sector-specific productivity distribution. By focusing on the schedules of comparative and absolute advantage, my nonparametric identification result relies on weaker assumptions than those imposed by [Heckman and Honoré \(1990\)](#); in particular, I allow for cross-market variation in sectoral efficiency in the form of unobserved additive shifters of comparative and absolute advantage. In this environment, I show that the supply equations relating sector employment and sector average wages to the schedules of comparative and absolute advantage belong to the class of separable models studied by [Newey and Powell \(2003\)](#), being nonparametrically identified under the same exogeneity and completeness conditions outlined by these authors. To the extent that workers have different levels of sectoral productivity across markets, the flexibility implied by the environment in this paper is important in empirical applications of the Roy model. In fact, very different estimates of the structural parameters are obtained without sector demand shifters.

The rest of the paper is organized as follows. Section 2 presents the model and its implications for the equilibrium structure of employment and wages. Section 3 establishes the nonparametric identification of comparative and absolute advantage. Section 4 presents the estimation of these schedules in the panel of Brazilian regional labor markets differentially exposed to shocks in world commodity prices. Section 5 presents the counterfactual analysis of the effect of changes in world commodity prices on changes in Brazilian wage inequality. Section 6 offers some concluding remarks.

2 Model

My goal is to develop a framework to quantify the effect of shocks in world commodity prices on Brazilian wage inequality. For this purpose, I assume that Brazil is a collection of small open economies with segmented labor markets. Consequently, good prices are exogenously determined internationally, but factor prices are endogenously determined regionally.⁶

2.1 Environment

Each regional economy contains workers of multiple demographic groups, $g \in \{1, \dots, G\}$, and two aggregate sectors, the commodity sector ($k = C$) and the non-commodity sector ($k = N$). Within each demographic group, there is a continuum of heterogeneous individuals, $i \in \mathcal{I}_g$, endowed with a bivariate skill vector, $(L_g^C(i), L_g^N(i))$, that determines their productivity if employed in each aggregate sector of the economy. This is the core assumption of a large class of [Roy-like \(1951\)](#) models, and it is central in my analysis of the distributional effects of sectoral demand shocks.

In order to incorporate the various commodity categories in the empirical application, I assume that each aggregate sector comprises multiple perfectly competitive industries, $j \in \mathcal{J}^k$, that produce homogeneous goods freely traded in the world market at price p^j . In every industry j of the aggregate

⁶This paper abstracts from migration flows between regional labor markets in Brazil. This simplification is motivated by the empirical analysis below, where I find weak migration responses following regional shocks to the labor demand in the commodity sector. This point is carefully discussed in Section 4.

sector k , individuals have an identical level of sector-specific productivity. The production technology in industry j utilizes the total number of sector-specific efficiency units supplied by employees, L_g^j , and an industry-specific nonlabor input, X^j . Specifically, the production function is given by

$$q^j = Q^j(L_1^j, \dots, L_G^j, X^j) \quad \text{where} \quad L_g^j \equiv \begin{cases} \int_{S_g^j} L_g^C(i) di & \text{if } j \in \mathcal{J}^C \\ \int_{S_g^j} L_g^N(i) di & \text{if } j \in \mathcal{J}^N \end{cases} \quad (1)$$

and S_g^j is the set of individuals of group g employed in industry j . Function $Q^j(\cdot)$ is strictly increasing, concave, differentiable, and homogeneous of degree one. The technology allows, but does not require, the effective labor supply of workers of different groups to be imperfect substitutes in production.

This production structure determines the effect of price shocks at the product-level on the demand for labor at the sector-level. In the empirical analysis, I explore this structure to obtain a regional shifter of the commodity sector's labor demand following shocks in world commodity prices. The cross-regional variation in this shifter is implied by the limited supply of the industry-specific nonlabor factor that, in this context, corresponds to the regional endowment of natural resources necessary for production of agricultural and mining goods — e.g., fertile soil, rainfall, metal reserves, or oil reserves.⁷

The analysis is greatly simplified by working with a log-linear transformation of individuals' sector-specific productivities. Define individual i 's comparative advantage as $s_g(i) \equiv \ln[L_g^C(i)/L_g^N(i)]$, and absolute advantage as $a_g(i) \equiv \ln[L_g^N(i)]$. In a given group, suppose individuals independently draw their productivity vector from a common bivariate distribution such that, without loss of generality,

$$s_g(i) \sim F_g(s) \quad \text{and} \quad \{a_g(i) | s_g(i) = s\} \sim H_g(a|s) \quad (2)$$

where, for simplicity, $F_g(s)$ is assumed to have full support in \mathbb{R} .

2.2 Competitive Equilibrium

In the competitive equilibrium, producers maximize profits conditional on both world product prices and local factor prices. In all industries of the aggregate sector k , producers face an identical labor cost: the sector's wage per efficiency unit, w_g^k . As a result, conditional on world product prices, the labor demand in industry j of sector k is given by, for all $g = 1, \dots, G$,

$$w_g^k = p^j \cdot \frac{\partial Q^j}{\partial L_g^j} \quad \text{if } j \in \mathcal{J}^k \quad (3)$$

where $X^j = \bar{X}^j$, with \bar{X}^j denoting the economy's endowment of the industry-specific nonlabor input.

To determine the supply of efficiency units of labor in each sector, consider the employment decision of workers seeking to maximize total labor income. Individual i of group g , if employed in any industry j of sector k , receives w_g^k for each sector-specific efficiency unit supplied. Let $y_g^k(i)$ denote

⁷Alternatively, one could consider any environment with a generic sector demand for labor efficiency units. For instance, it is straight forward to allow for non-competitive product markets and other mobile factors of production. These extensions do not affect the main insights discussed in this section.

the potential log-wage of individual i in any industry of sector k . Using the log-transformation above, these potential log-wages are given by

$$y_g^N(i) \equiv \omega_g^N + a_g(i) \quad \text{and} \quad y_g^C(i) \equiv \omega_g^C + s_g(i) + a_g(i) \quad (4)$$

where $\omega_g^k \equiv \ln w_g^k$.

Because all industries of an aggregate sector yield the same labor income, individuals are indifferent between them. Yet individuals receive different wages in the two sectors and, for this reason, they self-select into the sector where their labor income is higher.⁸ Hence, the set of individuals employed in sector k , \mathcal{S}_g^k , is given by

$$\mathcal{S}_g^k \equiv \left\{ i \in \mathcal{I}_g : k = \operatorname{argmax}\{y_g^C(i), y_g^N(i)\} \right\}. \quad (5)$$

In the competitive equilibrium of this economy, sectoral wages per efficiency unit guarantee factor market clearing in the two sectors. Specifically, $\{(w_g^C, w_g^N)\}_g$ are such that, for all g and k ,

$$\sum_{j \in \mathcal{J}^k} L_g^j = \int_{\mathcal{S}_g^k} L_g^k(i) di \quad (6)$$

where, in every industry j , condition (3) determines L_g^j ; and condition (5) determines \mathcal{S}_g^k . In order to satisfy labor demand at the industry level, individuals employed in each sector are allocated across industries to satisfy conditions (1) and (3).

2.3 Sectoral Log-Wages and Employment

To determine workers' sectoral employment decisions in the model, I consider a graphical representation of the economy where individuals are ranked according to their level of comparative advantage. For each quantile $q \in [0, 1]$, there is a set of individuals in group g whose level of comparative advantage is $\alpha_g(q) \equiv (F_g)^{-1}(q)$. By construction, $\alpha_g(q)$ is increasing in q so that individuals in higher quantiles are relatively more efficient in the commodity sector than those in low quantiles. Among individuals in quantile q , there is a conditional distribution of absolute advantage, $H_g(a|\alpha_g(q))$, with average and variance respectively denoted by $A_g(q)$ and $V_g(q)$. In the rest of this paper, $\alpha_g(\cdot)$ is the schedule of comparative advantage, and $A_g(\cdot)$ is the schedule of absolute advantage.

Figure 2 exhibits the average potential log wage in each sector for individuals of group g distributed across quantiles of comparative advantage. Immediately from expression (4), the average log-wage of workers in quantile q is $\bar{Y}_g^N(q) = \omega_g^N + A_g(q)$ if employed in the non-commodity sector. Alternatively, these workers earn an average log-wage of $\bar{Y}_g^C(q) = \omega_g^C + \alpha_g(q) + A_g(q)$ if employed in the commodity sector. In a particular quantile, the unique source of dispersion in potential sector wages

⁸This particular formulation closely follows the environment in the extensive literature inspired by the seminal work of Roy (1951). By introducing worker heterogeneity entirely on sector-specific productivity, the distributive impact of a trade shock is completely captured by the behavior of observable labor income. Notice that this model abstracts from between-sector mobility costs. In Appendix B, I explore an extension that incorporates such a feature into the model in the form of heterogeneity in non-monetary private benefits of employment across individuals. The extended model yields similar conclusions as those outlined in this section.

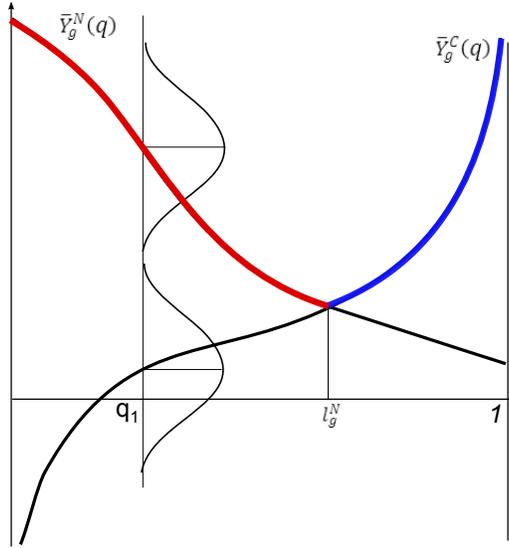


Figure 2: Sectoral Log-Wages and Employment in Equilibrium

Red: non-commodity sector employees in group g . **Blue:** commodity sector employees in group g .

is the dispersion of absolute advantage, $V_g(q)$ — illustrated by the hump-shaped curves in quantile q_1 . Lastly, it is important to notice that the two potential log-wage curves exhibit the single-crossing property, because $\alpha_g(q)$ is increasing in q .⁹

The importance of Figure 2 lies in the fact that it simultaneously illustrates sectoral employment and sectoral wages for any given level of (ω_g^C, ω_g^N) . All individuals in a particular quantile q choose to be employed in the same sector since, for all of them, the potential log-wage premium in the commodity sector is $\omega_g^C + \alpha_g(q) - \omega_g^N$.¹⁰ In high quantiles of comparative advantage, the relatively higher efficiency in the commodity sector yields a relatively higher wage in that sector, implying self-selection into the commodity sector — i.e., the blue portion of the potential average wage curve in Figure 2. In contrast, individuals in low quantiles of comparative advantage obtain a relatively lower wage in the commodity sector, finding it optimal to self-select into the non-commodity sector — i.e., the red portion of the potential average wage curve in Figure 2. Finally, the marginal individuals at the intersection of the two curves have exactly the same potential wage in the two sectors, being indifferent between them. Thus, I establish the following result.

Proposition 1. *Conditional on (ω_g^C, ω_g^N) , the allocation of individuals to sectors depends exclusively on their level of comparative advantage. In particular, individual i with $s_g(i) = \alpha_g(q)$:*

- i. self-selects into the commodity sector if $\alpha_g(q) > \omega_g^N - \omega_g^C$;*
- ii. self-selects into the non-commodity sector if $\alpha_g(q) < \omega_g^N - \omega_g^C$; and*

⁹To simplify the analysis, Figure 2 imposes that (ω_g^C, ω_g^N) are such that these curves cross at least once. Inada conditions on the production technology are sufficient for this to occur in equilibrium.

¹⁰To formalize this claim, consider individual i with comparative advantage $s_g(i) = \alpha_g(q)$. For this individual, potential sector wages in (4) correspond to vertical shifts of those of a worker with the same level of comparative advantage but a different level of absolute advantage. Consequently, the sectoral choice of individual i with $s_g(i) = \alpha_g(q)$ is identical to that of a hypothetical individual i' with $s_g(i') = \alpha_g(q)$ and $a_g(i') = A_g(q)$.

iii. is indifferent between the two sectors if $\alpha_g(q) = \omega_g^N - \omega_g^C$.

Proposition 1 indicates the central role played by comparative advantage in determining the sectoral allocation of workers in the model. In equilibrium, the sector employment composition is determined by marginal individuals with comparative advantage equal to the relative wage per efficiency unit, $\omega_g^N - \omega_g^C$. As a result, the share of individuals of group g employed in the non-commodity sector, l_g^N , is determined by the intersection of the two sectoral curves of potential average log-wages:

$$\omega_g^N - \omega_g^C = \alpha_g(l_g^N). \quad (7)$$

Given the sectoral employment decision described in Proposition 1, Figure 2 immediately yields the average log-wage of workers in each quantile of comparative advantage. Aggregating across the quantiles allocated to each sector, I obtain the sector average log-wage, \bar{Y}_g^k , which is given by

$$\bar{Y}_g^k = \omega_g^k + \bar{A}_g^k(l_g^N) \quad \text{where} \quad \bar{A}_g^k(l) \equiv \begin{cases} \frac{1}{l} \int_0^l A_g(q) dq & \text{if } k = N \\ \frac{1}{1-l} \int_l^1 [\alpha_g(q) + A_g(q)] dq & \text{if } k = C. \end{cases} \quad (8)$$

In expression (8), there are two determinants of the average sector log-wage. The first is the sector wage per efficiency unit, ω_g^k , that directly affects the log-wage of all sector employees symmetrically. The second is the sector employment composition, l_g^N , that affects the average efficiency of sector employees through the function $\bar{A}_g^k(\cdot)$. This compositional effect is generated by the variation in the average sector-specific efficiency of workers in different quantiles of comparative advantage. That is, it depends on the shape of the sectoral curves of average efficiency: $A_g(\cdot)$ in the non-commodity sector and $\alpha_g(\cdot) + A_g(\cdot)$ in the commodity sector. In Figure 2, $A_g(q)$ is decreasing and $A_g(q) + \alpha_g(q)$ is increasing. This case entails “positive selection into both sectors” because the average sector employee is *more* efficient than marginal workers indifferent between the two sectors (i.e., those in quantile l_g^N). In this case, the average sector-specific efficiency decreases as employment expands in the two sectors: $\bar{A}_g^N(l_g^N)$ is decreasing, and $\bar{A}_g^C(l_g^N)$ is increasing. The model, however, imposes only weak restrictions on the shape of $A_g(q)$ and $\alpha_g(q) + A_g(q)$ since comparative and absolute advantage can be arbitrarily related. As discussed below, the different possible shapes of these functions imply qualitatively different compositional effects in the adjustment of sector average wages to sectoral demand shocks.

Proposition 2. *Conditional on (ω_g^C, ω_g^N) , the average sector log-wage, \bar{Y}_g^k , depends on the sector employment composition, l_g^N , through the average efficiency of sector employees, $\bar{A}_g^k(l_g^N)$, in equation (8). In the non-commodity sector, this compositional effect depends on $A_g(\cdot)$; in the commodity sector, it depends on $\alpha_g(\cdot) + A_g(\cdot)$.*

2.4 Sectoral Demand Shocks and Sectoral Changes in Wages and Employment

In order to illustrate the mechanics of the model, let us analyze the adjustment of sector labor market outcomes following changes in sectoral wages per efficiency unit triggered by a positive shock in world

commodity prices. This exercise delineates the distinct roles of comparative and absolute advantage in determining sectoral responses in terms of employment and average wage.

An increase in world commodity prices translates into higher marginal value of labor in the commodity sector. To fix ideas, I consider in this section a partial equilibrium exercise in which, after the shock, ω_g^C increases and ω_g^N remains constant. Figure 3 displays the induced movements in the curves of potential sector wages in three cases. Panel (a) illustrates the case analyzed above, where $A_g(q)$ is decreasing and $A_g(q) + \alpha_g(q)$ is increasing. Panels (b) and (c) present other possible shapes for these functions that will be representative of the different qualitative patterns of compositional effects allowed in the model.

The shock causes an increase of $\Delta\omega_g^C$ in the log-wage of all commodity sector employees as represented by the upward shift of the blue curve on Figure 3. Since $\Delta\omega_g^N = 0$, the shock does not affect the wage of non-commodity sector employees and the red curve remains unchanged. Only those non-commodity sector employees who decide to switch into the commodity sector benefit from the shock. These sector-switchers are represented in green on Figure 3. Their wage gain is bounded from below by $\Delta\omega_g^N$, and from above by $\Delta\omega_g^C$. This is illustrated by the difference between the solid and dashed green curves on Figure 3.

In the model, the decision of sectoral allocation is entirely determined by each worker's comparative advantage. Thus, the mass of sector-switchers that benefit from the shock depends on the dispersion of comparative advantage among marginal workers. As implied by equation (7), this is captured by the slope of the comparative advantage schedule, $\alpha_g(\cdot)$:

$$\Delta \left[\omega_g^N - \omega_g^C \right] = \int_{l_g^N}^{l_g^N + \Delta l_g^N} \frac{\partial \alpha_g(u)}{\partial q} du \approx \frac{\partial \alpha_g(l_g^N)}{\partial q} \cdot \Delta l_g^N \quad (9)$$

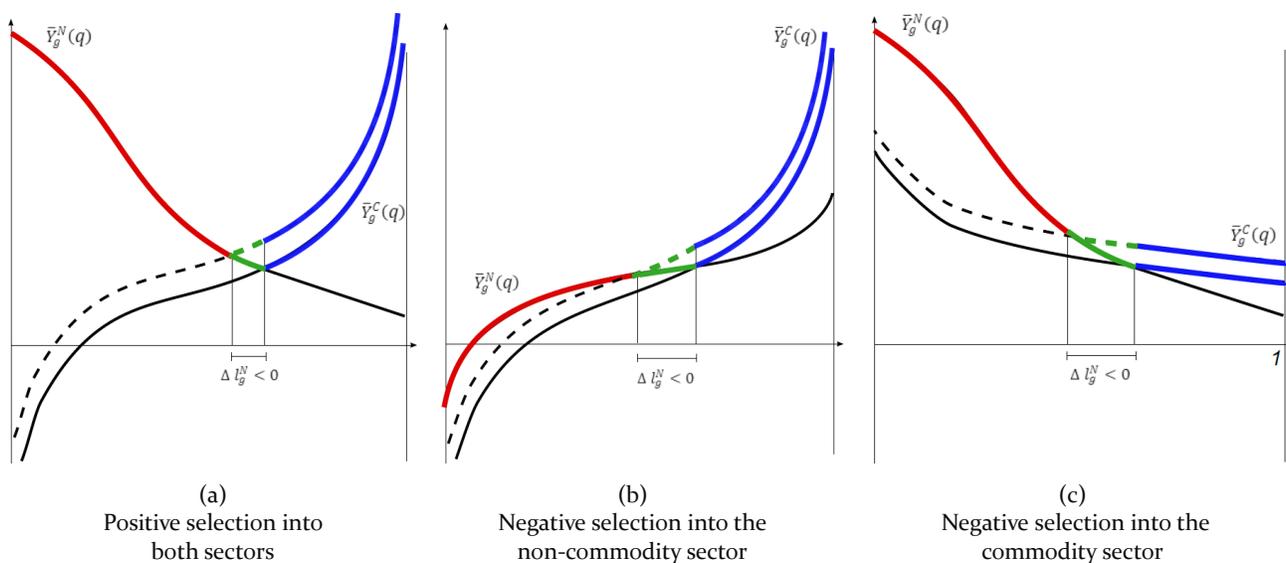


Figure 3: Comparative Statics - increase in ω_g^C
Red: sector-stayers in the non-commodity sector. **Blue:** sector-stayers in the commodity sector.
Green: switchers from the non-commodity sector to the commodity sector.

where $\frac{\partial \alpha_g(q)}{\partial q} \geq 0$ for all q .

Although the wage per efficiency unit remains constant in the non-commodity sector, the implied outflow of employees affects the sector's employment composition and, consequently, the sector's average wage. To the extent that the absolute advantage of sector-switchers differs from that of sector-stayers, the change in sector employment triggers a change in sector average efficiency. Intuitively, this is conveyed by the first-order expansion of equation (8):

$$\Delta \bar{Y}_g^N - \Delta \omega_g^N = \int_{I_g^N}^{I_g^N + \Delta I_g^N} \frac{\partial \bar{A}_g^N(u)}{\partial q} du \approx [A_g(I_g^N) - \bar{A}_g^N(I_g^N)] \cdot \Delta \ln(I_g^N). \quad (10)$$

where, by definition, $\bar{A}_g^N(I_g^N) = (1/I_g^N) \cdot \int_0^{I_g^N} A_g(q) dq$.

The right-hand side of equation (10) is the compositional effect implied by the outflow of non-commodity sector employees. This effect is proportional to the average absolute advantage of sector-switchers, $A_g(I_g^N)$, relative to that of sector-stayers, $\bar{A}_g^N(I_g^N)$. To see this, consider the three cases in Figure 3. In Panels (a) and (c), the decreasing schedule of absolute advantage $A_g(q)$ implies that non-commodity sector stayers (red) have a *higher level of absolute advantage* than sector-switchers (solid green). In this case, $A_g(I_g^N) < \bar{A}_g^N(I_g^N)$ and the compositional effect is positive. In words, the outflow of workers leaves the non-commodity sector with employees whose average absolute advantage is relatively higher than before. However, this is not the only possibility. In Panel (b), $A_g(q)$ is increasing so that $A_g(I_g^N) > \bar{A}_g^N(I_g^N)$ and the outflow of workers lowers the average wage in the non-commodity sector. In general, $A_g(\cdot)$ determines the magnitude of the compositional effect in the non-commodity sector, which can be either negative, as in Panel (b), or positive, as in Panels (a) and (c).

In the commodity sector, the shock has two effects on the average wage. First, there is an increase in the log-wage of commodity sector employees implied by $\Delta \omega_g^C > 0$ — i.e., the vertical shift of the blue curve in Figure 3. Second, there is a compositional effect driven by the inflow of new employees whose average sector-specific efficiency differs from that of original employees in the commodity sector. As in the non-commodity sector, the sign of this effect is ambiguous, and it is determined by the slope of $\bar{A}_g^C(\cdot)$. In Panels (a) and (b), new commodity sector employees (dashed green) are *less efficient* than original commodity sector employees (blue) and, therefore, the employment expansion leads to a negative compositional effect. In Panel (c), alternatively, new workers are *more efficient* than the original commodity sector employees, implying a positive compositional effect.¹¹

To summarize, an increase in world commodity prices that causes an increase in the commodity sector's relative wage per efficiency unit, $\omega_g^C - \omega_g^N$, affects both sectoral employment and sectoral wages. The increase in $\omega_g^C - \omega_g^N$ triggers an increase in the relative employment of the commodity sector whose magnitude is regulated by the schedule of comparative advantage, $\alpha_g(\cdot)$. The between-

¹¹Intuitively, the compositional effect in the commodity sector is captured by a first-order expansion of expression (8):

$$\Delta \bar{Y}_g^C - \Delta \omega_g^C = \int_{I_g^N}^{I_g^N + \Delta I_g^N} \frac{\partial \bar{A}_g^C(u)}{\partial q} du \approx [\alpha_g(I_g^N) + A_g(I_g^N) - \bar{A}_g^C(I_g^N)] \cdot \Delta \ln(I_g^C).$$

where $\bar{A}_g^C(I_g^N) = (1/(1 - I_g^N)) \cdot \int_{I_g^N}^1 \alpha_g(q) + A_g(q) dq$. Notice that the average efficiency in the commodity sector is related to the sum of the schedules of comparative and absolute advantage, $\alpha_g(\cdot) + A_g(\cdot)$.

sector worker reallocation introduces compositional effects in the response of the commodity sector's relative average wage. Such an effect may reinforce or diminish the positive impact of the increase in $\omega_g^C - \omega_g^N$. The magnitude of the compositional effect in sectoral average wages is determined, in the non-commodity sector, by $A_g(\cdot)$ and, in the commodity sector, by $\alpha_g(\cdot) + A_g(\cdot)$.

2.5 Sectoral Demand Shocks and Aggregate Changes in Wage Inequality

I now turn to movements in wage inequality stemming from the sectoral shock analyzed above. In this analysis, there are many ways of quantifying changes in wage inequality. I focus on the responses in the average and the variance of the log-wage distribution of workers in different demographic groups. The main result of this section establishes that these responses are determined by the schedules of comparative advantage, $\alpha_g(\cdot)$, and absolute advantage, $A_g(\cdot)$.

Let us first analyze the average of the log-wage distribution among workers of group g . For these workers, the average log-wage is $\bar{Y}_g \equiv l_g^N \cdot \bar{Y}_g^N + l_g^C \cdot \bar{Y}_g^C$ which, by equation (8), is equivalent to

$$\bar{Y}_g = \omega_g^C \cdot l_g^C + \omega_g^N \cdot l_g^N + \int_{l_g^N}^1 \alpha_g(q) dq + e_g$$

where $e_g \equiv \int_0^1 A_g(q) dq$.

Following a demand-driven shock in (ω_g^C, ω_g^N) , this expression implies that

$$\Delta \bar{Y}_g = \left[\Delta \omega_g^C \cdot l_g^C + \Delta \omega_g^N \cdot l_g^N \right] + \left[\alpha_g \left(l_g^N + \Delta l_g^N \right) \cdot \Delta l_g^N - \int_{l_g^N}^{l_g^N + \Delta l_g^N} \alpha_g(q) dq \right]. \quad (11)$$

In equation (11), the first term is the direct effect on the wage of sector employees if they were unable to reallocate between sectors. This direct effect depends solely on the pre-shock employment composition and the change in sectoral wages per efficiency unit. Nevertheless, this is not the only effect, because workers respond to the shock by switching sector. This composition effect is captured by the second term which, intuitively, depends on the schedule of comparative advantage, $\alpha_g(\cdot)$.¹²

Finally, I compute the log-wage variance among individuals of group g . There are two sources of wage dispersion: the between-sector average wage differential and within-sector wage dispersion. By the law of total variance, these two components imply that the log-wage variance in group g , V_g , is given by

$$V_g = l_g^N l_g^C \cdot \left(\bar{Y}_g^C - \bar{Y}_g^N \right)^2 + l_g^N \cdot V_g^N + l_g^C \cdot V_g^C$$

where V_g^k corresponds to the log-wage variance among individuals of group g employed in sector k .

As indicated in Figure 2, the within-sector wage variance, V_g^k , combines the variation in average wage of individuals distributed across the quantiles allocated to the sector, $Var[\bar{Y}_g^k(q)]$, and the absolute

¹²The compositional effect is second-order: for small shocks, sector-switchers are the marginal individuals with the same potential wage in the two sectors, so their reallocation does not affect the log-wage distribution. As a result, the change in the average log-wage, up to a first-order approximation, only depends on the initial allocation of workers across sectors. This intuition can be extended to the wage growth across quantiles of the log-wage distribution. In a first-order approximation, it depends exclusively on the pre-shock sectoral allocation of workers in each quantile. I return to this discussion in detail in Section 4.2 and in Appendix A.2.

advantage dispersion in any particular comparative advantage quantile, $V_g(q)$. Consequently, the log-wage variance of group g is given by

$$V_g = l_g^N l_g^C \cdot \left(\bar{Y}_g^C - \bar{Y}_g^N \right)^2 + l_g^N \cdot \text{Var} \left(A_g(q) | q < l_g^N \right) + l_g^C \cdot \text{Var} \left(\alpha_g(q) + A_g(q) | q \geq l_g^N \right) + v_g \quad (12)$$

where the variance is taken over the conditional uniform distribution of quantiles allocated to each sector, and $v_g \equiv \int_0^1 V_g(q) dq$ is the average dispersion in absolute advantage. Because absolute advantage affects log-wage dispersion equally in the two sectors, v_g does not depend on the sector employment composition.

Expression (12) immediately implies that the change in the log-wage variance is determined by the schedules of comparative advantage, $\alpha_g(\cdot)$, and absolute advantage, $A_g(\cdot)$. Such a change comprises two terms: the change in the sector average wage differential, and the change in the log-wage variance within each sector. Both terms are affected by the compositional effects generated by the sectoral reallocation of workers. The following proposition summarizes this discussion.

Proposition 3. *Conditional on demand-driven changes in (ω_g^C, ω_g^N) , the schedules of comparative advantage, $\alpha_g(\cdot)$, and absolute advantage, $A_g(\cdot)$, determine the changes in the average and the variance of the log-wage distribution of workers in group g .*

In the rest of the paper, I build upon Propositions 1–3 to construct an empirical strategy to quantify the distributional effects of shocks in world commodity prices. First, I show how Propositions 1–2 can be used to establish the nonparametric identification of the schedules of comparative advantage, $\alpha_g(\cdot)$, and absolute advantage, $A_g(\cdot)$. This result relies on the intuition of the comparative statics exercise in Section 2.4, where these schedules determine the magnitude of the sectoral responses in employment and average wage implied by sector demand shocks. Second, I use the model’s predicted response in the average and the variance of the log-wage distribution in Proposition 3 to quantify the effect on wage inequality of shocks in world commodity prices. This delivers an empirical framework to analyze the effect of world commodity prices on Brazilian wage inequality.

3 Identification of Comparative and Absolute Advantage

The goal of this section is to establish the nonparametric identification of the schedules of comparative and absolute advantage. The challenge inherent in identifying these functions is illustrated by [Heckman and Honoré’s \(1990\)](#) result that, in the context of the Roy model, the sector-specific productivity distribution cannot be nonparametrically identified in a single cross-section of individuals. Thus, this section represents an important first step in the empirical application of the model. The nonparametric identification result indicates the source of variation in the data that uncovers comparative and absolute advantage. Such a result does not impose additional restrictions beyond those implied by the theory. In contrast, as noted by [Matzkin \(2007\)](#), the credibility of the empirical analysis would be significantly hindered if identification could only be achieved under restrictive parametric assumptions.

To this extent, I explore the distinct roles of comparative and absolute advantage in determining sectoral employment and sectoral average wage, as described in Propositions 1 and 2. Following a

sector demand shock, the schedule of comparative advantage determines the between-sector response of employment. Simultaneously, the schedule of absolute advantage determines the compositional effects embedded in the response of sector average wages. Reflecting these conceptually different effects, the main result of this section establishes that the schedules of comparative and absolute advantage are nonparametrically identified from cross-regional variation in the sectoral responses of employment and wages to observable sector-level demand shifters.

3.1 Assumptions

In order to establish identification of comparative and absolute advantage, I make additional assumptions regarding observable labor market outcomes, as well as their relation to unobservable variables.

Segmented Labor Markets. Consider the set of regional economies with segmented labor markets generated by the model in Section 2. Each regional market is indexed by m . For workers in a demographic group g , I assume that there is observable information on sector employment composition, $l_{g,m}^k$, sector average wages, $\bar{Y}_{g,m}^k$, and sector wage per efficiency unit, $\omega_{g,m}^k$.¹³ In addition, I assume that the productivity distribution in every market m satisfies the following conditions.

Assumption 1. Individual i in market m , $i \in \mathcal{I}_{g,m}$, independently draws $(s_g(i), a_g(i))$ as follows.

i. *Comparative Advantage:*

$$s_g(i) = \tilde{s}_g(i) + \tilde{u}_{g,m} \quad \text{and} \quad \{\tilde{s}_g(i)\} \sim F_g(s)$$

where $\tilde{u}_{g,m}$ is a group-market shifter of comparative advantage, and $\alpha_g(q) \equiv (F_g)^{-1}(q)$.

ii. *Absolute Advantage:*

$$\{a_g(i) | \tilde{s}_g(i) = s\} \sim \mu H_g^a(a|s) + (1 - \mu) H_{g,m}^e(a)$$

where $H_{g,m}^e(a) \equiv H^e(a | \tilde{u}_{g,m}, \theta_{g,m})$ is a group-market mixing distribution of absolute advantage such that

$$A_g(q) \equiv \mu \int a dH_g^a(a | \alpha_g(q)) \quad \text{and} \quad \tilde{v}_{g,m} \equiv (1 - \mu) \int a dH_{g,m}^e(a).$$

Assumption 1 imposes no restrictions on the shape of the productivity distribution, allowing it to vary arbitrarily between worker groups. However, the productivity distribution is assumed to only vary across markets with respect to market-specific shifters in comparative and absolute advantage. Specifically, $\tilde{u}_{g,m}$ represents a shock to the relative efficiency of workers in the commodity sector. Also, $\tilde{v}_{g,m}$ is a shifter of the average absolute advantage of workers in the market, capturing supply shocks to workers' productivity in the non-commodity sector. Assume that these supply shocks combine observable and unobservable components as follows.

¹³In this section, I treat $\omega_{g,m}^k$ as observable variables determined in the competitive equilibrium of each region. Section 4.2 provides a methodology to estimate changes in the wage per efficiency unit, $\Delta\omega_{g,m}^k$, based on the model's predicted relation between wage growth and initial sector employment across quantiles of the log-wage distribution.

Assumption 2. The shifters of comparative and absolute advantage, $(\tilde{u}_{g,m}, \tilde{v}_{g,m})$, are given by

$$\tilde{u}_{g,m} = \mathbf{X}_{g,m} \gamma_g^u + u_{g,m}, \quad \text{and} \quad \tilde{v}_{g,m} = \mathbf{X}_{g,m} \gamma_g^v + v_{g,m}$$

where $\mathbf{X}_{g,m}$ is an observable vector of group-market variables; and $(u_{g,m}, v_{g,m})$ is an unobservable vector of group-market supply shocks. These shifters are normalized such that $E[\tilde{u}_{g,m}] = E[\tilde{v}_{g,m}] = 0$.

Under Assumptions 1 and 2, the supply equations determining sector employment composition in (7) and sector average wage in (8) are equivalent to

$$\left[\omega_{g,m}^N - \omega_{g,m}^C \right] = \alpha_g \left(l_{g,m}^N \right) + \mathbf{X}_{g,m} \gamma_g^u + u_{g,m} \quad (13)$$

$$\left[\bar{Y}_{g,m}^N - \omega_{g,m}^N \right] = \bar{A}_g^N \left(l_{g,m}^N \right) + \mathbf{X}_{g,m} \gamma_g^v + v_{g,m} \quad (14)$$

$$\left[\bar{Y}_{g,m}^C - \omega_{g,m}^C \right] = \bar{A}_g^C \left(l_{g,m}^N \right) + \mathbf{X}_{g,m} (\gamma_g^u + \gamma_g^v) + (u_{g,m} + v_{g,m}). \quad (15)$$

Equations (13)–(15) highlight the importance of Assumption 1: it implies identical patterns of selection across markets in the form of the common schedules of comparative advantage, $\alpha_g(\cdot)$, and absolute advantage, $A_g(\cdot)$. In this context, the pair of unobservable productivity shifters, $(u_{g,m}, v_{g,m})$, generates variation in the sector-specific productivity distribution across markets. Accordingly, Assumptions 1 and 2 are weaker than Heckman and Honoré’s (1990) restriction of an identical sector-specific productivity distribution in every market. To the extent that workers of a particular group are different in terms of sectoral labor efficiency across markets, the flexibility implied by the unobserved productivity shifters is important in the empirical application of the model. In fact, variation in these shifters translates into variation in the effective labor supply in the two sectors, generating simultaneous general equilibrium responses in sector wage per efficiency unit, $\omega_{g,m}^k$, and sector employment composition, $l_{g,m}^N$. As a result, identification of $\alpha_g(\cdot)$ and $A_g(\cdot)$ based on the supply equations (13)–(15) requires a sector demand shock that is orthogonal to the productivity shifters in the cross-section of markets.¹⁴

Instrument: Sector Demand Shifter. Consider an observable vector, $\mathbf{Z}_{g,m}^k$, of sector demand shifters across markets. To be a valid instrument in the supply equations (13)–(15), this sector demand shifter has to be mean independent from unobserved shocks to the productivity distribution, $(u_{g,m}, v_{g,m})$. Thus, I assume that $\mathbf{Z}_{g,m}^k$ satisfies the following exogeneity restriction.

Assumption 3. $E \left[u_{g,m} | \mathbf{Z}_{g,m}^k, \mathbf{X}_{g,m} \right] = E \left[v_{g,m} | \mathbf{Z}_{g,m}^k, \mathbf{X}_{g,m} \right] = 0$.

Additionally, $\mathbf{Z}_{g,m}^k$ has to induce enough exogenous variation in the endogenous sector composition

¹⁴Thorough the lens of Assumption 1, Heckman and Honoré’s (1990) restriction is equivalent to imposing $\tilde{u}_{g,m} = 0$ and $\mu = 1$. In this case, markets are not subject to unobserved supply shocks and, therefore, any cross-market variation in sector employment composition is necessarily generated by sectoral demand shocks. Consequently, the cross-market variation in wage per efficiency unit leads to the identification of $\alpha_g(\cdot)$ and $A_g(\cdot)$ from equations (13)–(15) with $u_{g,m} = v_{g,m} = 0$.

tion $l_{g,m}^N$ to uniquely discriminate the underlying productivity distribution of the economy. In the environment introduced in Section 2, this shifter must affect sectoral labor demand differentially across markets. Formally, the instrument has to satisfy the equivalent of a rank requirement in the context of nonparametric models. As shown by [Newey and Powell \(2003\)](#), the necessary and sufficient completeness condition that guarantees identification of the class of models covering equations (13)–(15) is described as follows.

Assumption 4. For any $f(\cdot)$ with finite expectation, $E \left[f(l_{g,m}^N, \mathbf{X}_{g,m}) | \mathbf{Z}_{g,m}^k, \mathbf{X}_{g,m} \right] = 0$ implies $f(l_{g,m}^N, \mathbf{X}_{g,m}) = 0$ almost surely.

3.2 Nonparametric Identification of Comparative and Absolute Advantage

With Assumptions 1–4, I now establish the identification of the schedules of comparative and absolute advantage. Under Assumptions 3–4, the observable shifter $\mathbf{Z}_{g,m}^k$ generates exogenous variation in the sector composition $l_{g,m}^N$ that can be used to identify equations (13)–(15). To formalize this intuition, I demonstrate in Appendix A.1 the following particular case of the general result in [Newey and Powell \(2003\)](#) regarding the nonparametric identification of separable models with endogenous variables.

Lemma 1. [[Newey and Powell \(2003\)](#)] Consider a model of the form

$$y_{g,m} = \Phi_g \left(l_{g,m}^N \right) + \mathbf{X}_{g,m} \gamma_g + u_{g,m},$$

and a vector $\mathbf{Z}_{g,m}^k$ satisfying Assumptions 3–4. Then, the function $\Phi_g(\cdot)$ is identified up to a constant. With the normalization $E[\mathbf{X}_{g,m} \gamma_g] = 0$, the constant in $\Phi_g(\cdot)$ is also identified.

Notice that, under Assumptions 1–2, the supply equations in (13)–(15) belong to the class of models covered by Lemma 1. Thus, the instrument $\mathbf{Z}_{g,m}^k$ satisfying Assumptions 3–4 identifies $\alpha_g(\cdot)$ from equation (13). Similarly, Lemma 1 establishes the identification of $\bar{A}_g^N(\cdot)$ and $\bar{A}_g^C(\cdot)$ respectively from equations (14) and (15). This leads to the identification of $A_g(\cdot)$ and $A_g(\cdot) + \alpha_g(\cdot)$ since, by the definition in (8),

$$A_g(q) = \frac{\partial}{\partial q} \left[q \cdot \bar{A}_g^N(q) \right] \quad \text{and} \quad \alpha_g(q) + A_g(q) = -\frac{\partial}{\partial q} \left[(1-q) \cdot \bar{A}_g^C(q) \right].$$

Hence, I establish the following theorem.

Theorem 1. Consider a set of segmented markets, m , subject to a sector demand shifters, $\mathbf{Z}_{g,m}^k$, such that Assumptions 1–4 hold. For each worker group g ,

- i. $\alpha_g(\cdot)$ is identified from equation (13);
- ii. $A_g(\cdot)$ is identified from equation (14); and
- iii. $\alpha_g(\cdot) + A_g(\cdot)$ is identified from equation (15).

Theorem 1 is directly related to the comparative statics exercise in Section 2.4. Figure 4 illustrates a situation where the demand shifter $Z_{g,m}^k$ induces a change in the commodity sector's wage per efficiency unit. The vertical shift in the commodity sector curve of potential wage triggers between-sector worker reallocation represented by $\Delta l_{g,m}^N$. Conditional on $\Delta\omega_{g,m}^C - \Delta\omega_{g,m}^N$, the magnitude of the change in sector employment composition is determined by the difference between the slopes of the two curves of potential sector wages: the schedule of comparative advantage, $\alpha_g(\cdot)$, in equation (13). The subsequent change in the composition of sector employees triggers an observable response in the measured sector average efficiency, $\Delta[\bar{Y}_{g,m}^k - \omega_{g,m}^k]$. In each sector, the magnitude of this compositional effect is determined by the function $\bar{A}_g^k(\cdot)$ in equations (14)–(15). Because the compositional effect corresponds to the difference between the average sector-specific efficiency of switchers and stayers, $A_g(\cdot)$ is identified from the average efficiency change in the non-commodity sector, and $\alpha_g(\cdot) + A_g(\cdot)$ is identified from the average efficiency change in the commodity sector.

As a corollary of Theorem 1, the schedules of comparative advantage, $\alpha_g(\cdot)$, and absolute advantage, $A_g(\cdot)$, are identified with only two out of the three supply equations in (13)–(15). In other words, the model is overidentified whenever employment and average wages are available for the two sectors of the economy. The model's overidentification relies on the fact that sector-specific efficiency is the sole determinant of both sectoral wages and sectoral employment. Accordingly, the presence of non-monetary employment benefits breaks overidentification. Appendix B establishes the nonparametric identification of the extended model, where workers have heterogeneous private values of employment. In this case, I define a generalized notion of comparative advantage that includes these private benefits. The implied schedule of comparative advantage is identified from equation (13). In addition, the schedules of sector-specific average efficiency are identified from equations (14)–(15).

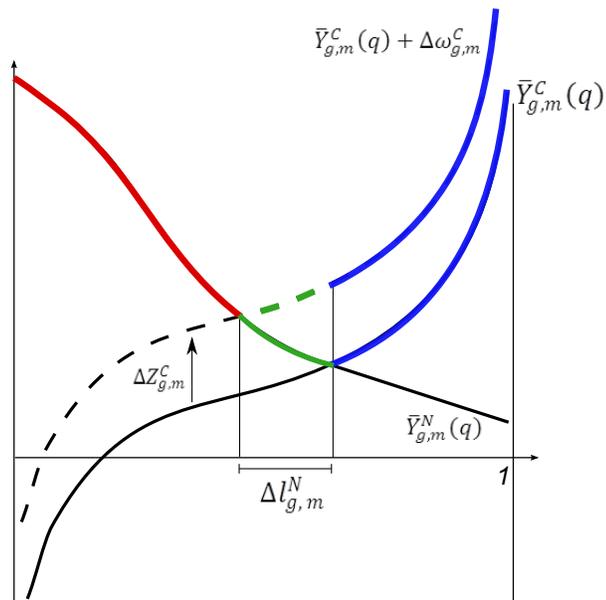


Figure 4: Identification of Comparative and Absolute Advantage

Red: sector-stayers in the non-commodity sector. **Blue:** sector-stayers in the commodity sector.
Green: switchers from the non-commodity sector to the commodity sector.

4 Empirical Application

The above result establishes the nonparametric identification of the schedules of comparative and absolute advantage using cross-market variation in observable shifters of sectoral labor demand. Armed with this theoretical result, I now estimate these schedules in a sample of Brazilian regional labor markets differentially exposed to shocks in world commodity prices. I then use these estimates to investigate the effect on Brazilian wage inequality of shocks in world commodity prices.

4.1 Sample of Regional Labor Markets and Exposure to World Commodity Price Shocks

The empirical application relies on wage and employment data from the Brazilian Census collected by the Brazilian Institute of Geography and Statistics (IBGE) for 1991, 2000, and 2010. In order to implement the identification strategy outlined above, it is necessary to construct a sample of segmented labor markets. To this end, I use Brazilian regional labor markets as implied by the microregion concept in the Census. The IBGE defines these microregions by aggregating economically integrated municipalities with similar production and geographic characteristics. For each microregion, I select a sample of full-time white employed males aged 16–64. Workers in the sample have strong labor force attachment, diminishing the importance of endogenous responses in total labor supply. I allocate individuals to a group of education (High School Graduates and High School Dropouts) and a sector of employment (commodity and non-commodity).¹⁵ Industries specialized in the production of agricultural and mining products are included in the commodity sector. All manufacturing and service industries are included in the non-commodity sector. In 1991, the commodity sector accounted for 5.2% of employment among High School Graduates (HSG) and 26.2% among High School Dropouts (HSD). In the analysis, I only consider those microregions with positive employment in the commodity sector for all years and groups. As a result, the final sample contains 518 microregions that represented 98.4% of the country’s population in 1991. Appendix C discusses details on the construction and measurement of labor market outcomes.

As a sectoral demand shifter, I construct a regional measure of exposure to shocks in world commodity prices separately for HSG and HSD. Specifically, the exposure vector of group g in microregion r to commodity price shocks at year t is given by

$$\Delta \mathbf{Z}_{g,r,t}^C = \left\{ \phi_{g,r}^{C,j} \cdot \Delta \ln p_t^j \right\}_{j \in \mathcal{J}^C} \quad (16)$$

where $\Delta \ln p_t^j$ is the log-change in the international price of product j between years $t - 1$ and t ; and $\phi_{g,r}^{C,j}$ is the share of industry j in total labor payments of the commodity sector to individuals of group g in microregion r on the initial year of 1991.

I construct the exposure measure in equation (16) using world prices of five major commodity

¹⁵These two educational groups are representative of the Brazilian workforce: among male workers, the High School graduation rate was 22.2% in 1991 and 44.5% in 2010. I restrict the benchmark sample to include only white and male individuals because of the strong declines in gender and race wage differential between 1995 and 2010; see [Ferreira, Firpo and Messina \(2014\)](#). The model in this paper does not speak directly to these components of wage inequality and, therefore, I exclude their behavior from the baseline empirical analysis. In robustness exercises, I extend the benchmark sample to also include female and non-white workers.

groups: Grains, Soft Agriculture, Livestock, Mining, and Energy. As described in Appendix C, I compute price indices for each category with data on commodity transactions in the main exchange markets of the United States. To replicate relative prices faced by producers in Brazil, I convert world commodity prices to Brazilian currency and deflate by the Brazilian consumer price index.

The exposure measure in (16) is based on the intuition that the response of the commodity sector’s labor demand is stronger in regions that specialize in the production of basic products experiencing stronger international price gains. In the model of Section 2, this demand shifter is generated by the limited supply of natural resources specific to commodity production — e.g, fertile soil, or oil and metal reserves. In such an environment, an increase in the world product price triggers an increase in the labor demand of firms producing that product whose effect on the commodity sector’s labor demand is proportional to the product’s importance in local employment.¹⁶

Given the shock to world commodity prices, the cross-regional variation in $\Delta Z_{g,r,t}^C$ depends entirely on the cross-regional variation in initial industry composition among workers of a particular group. As shown in Table C3, the great extent of such variation in Brazil implies large variation in shock exposure across regions. This is illustrated in Figure 5, which exhibits the total shock exposure across microregions for HSG (left panel) and HSD (right panel). As a consequence of the difference in industry allocation for the two groups, shock exposure differs significantly between HSG and HSD — specifically, the correlation in group exposure is .493.

Exogeneity Assumption. In the empirical application, the regional shifter of sector labor demand must satisfy the central exogeneity restriction imposed in Section 3: $\Delta Z_{g,r,t}^C$ has to be uncorrelated with regional shocks to sectoral worker efficiency. This requirement is likely to hold for the following three reasons.

First, Brazilian regions are small relative to the world market of basic commodities, implying that local supply shocks are unlikely to affect international prices. Any national shock correlated across microregions is captured by the time fixed effect included in the specification below. Furthermore, the 1991–2010 period was marked by strong growth in Chinese imports of agriculture and mining products. A growth which, arguably, represents an exogenous demand shock to the relative price of raw materials.¹⁷

Second, the exogeneity restriction requires regional shock exposure to not affect the productivity distribution of workers. This requirement would be violated if the pool of workers in the market varies in response to commodity price shocks because of changes in the labor supply of either native

¹⁶Recent empirical papers have built on related measures of local shock exposure in order to investigate the labor market effect of import competition — e.g., see Topalova (2010), Kovak (2013), and Autor, Dorn and Hanson (2013). In my model, the fixed supply of industry-specific factors guarantees the finiteness of the elasticity of industry labor demand to product price shocks. Thus,

$$\frac{\partial \log \left(\sum_{j \in \mathcal{J}^C} L_{g,r}^j \right)}{\partial \log p^j} = \phi_{g,r}^{C,j} \cdot \frac{\partial \log L_{g,r}^j}{\partial \log p^j}.$$

¹⁷Between 1992 and 2010, the average annual growth rate of Chinese imports was 17.2% for all products, 16.2% for Agriculture, and 28.3% for Mining. Over the period, Hanson (2012) provides a careful discussion of the transformation in the profile of international trade of emerging economies and, in particular, China. To the extent that this transformation was mainly driven by internal changes in the production structure of China, this large demand shock represented an exogenous impulse to world commodity prices in the period.

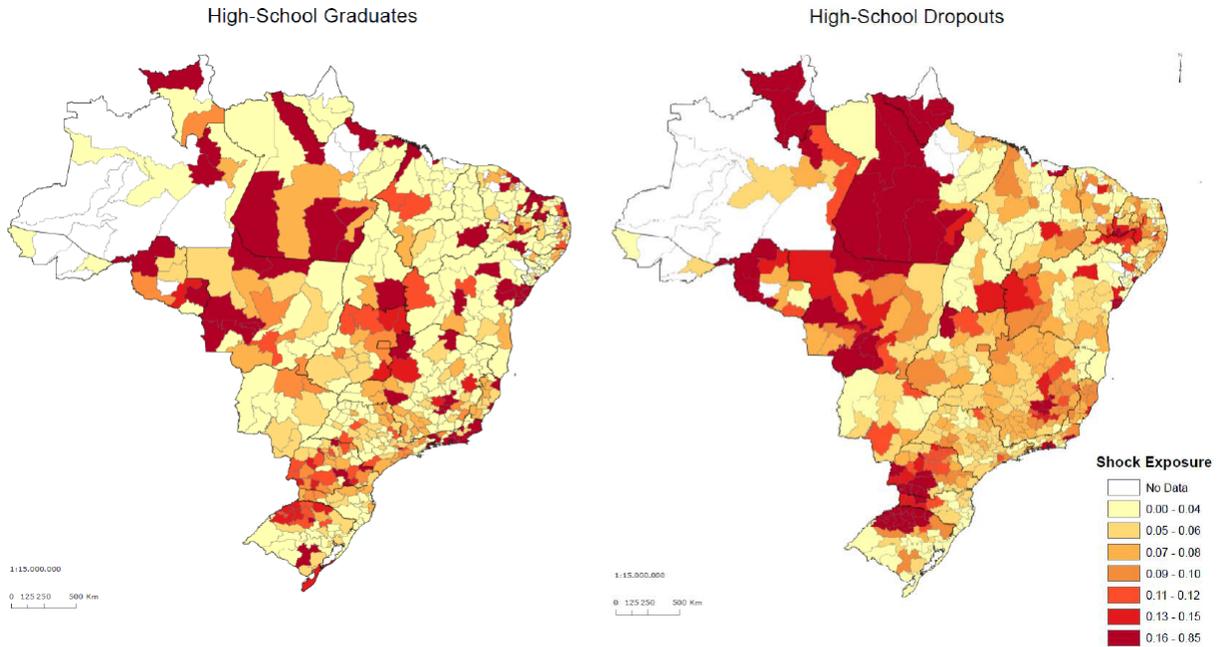


Figure 5: Heterogeneity in Exposure to Commodity Price Shock, 1991-2010

Note. For each microregion, the map presents the total exposure to the commodity price shock between 1991 and 2010: $\sum_{j \in \mathcal{J}^C} \phi_{g,r}^{C,j} \cdot \Delta \ln p_t^j$ where $\Delta \ln p_t^j$ is the log-change in the world price of product j in 1991–2010.

or immigrant workers. In the empirical application, such a concern is unlikely to be important, because the correlation between regional shock exposure and changes in the labor supply of both native and immigrants is small and nonsignificant. This result is partially driven by the inclusion of only full-time prime-aged males in the benchmark sample. See Table D5 in Appendix D.2.¹⁸

Third, the empirical application includes a variety of controls intended to capture changes in the productivity distribution potentially correlated with exposure to higher commodity prices. In particular, the control vector includes region-group fixed effects and period dummies interacted with initial regional characteristics (e.g., sector composition and socio economic variables). In this context, identification relies exclusively on the cross-region variation in the exposure to shocks in relative product prices within the commodity sector, allowing for arbitrary shocks to relative prices products in the non-commodity sector.

4.2 Estimation of Sector Wage per Efficiency Unit

The identification strategy of Section 3 requires information on sector wage per efficiency unit, $\omega_{g,r,t}^k$ for each triple of group-region-period. In this section, I propose a methodology to estimate $\Delta \omega_{g,r,t}^k$ using available information on labor income and employment at the individual level.

In the model, comparative advantage determines worker allocation across sectors, implying that the wage of sector employees is only exposed to changes in the wage per efficiency unit of their own

¹⁸Recently, Dix-Carneiro and Kovak (2015b) also find weak responses in migration flows across Brazilian regional labor markets differentially exposed to the tariff reductions during the trade liberalization of 1990–1995.

sector of employment. Following changes in $\omega_{g,r,t}^k$, this observation implies that, across different parts of the wage distribution, variation in the pre-shock sector employment composition translates into variation in the growth of wages. Intuitively, if all individuals at the bottom of the distribution are employed in the commodity sector, then the wage gain at the bottom is entirely attributed to the change in the commodity sector's wage per efficiency unit. In such case, an increase in the non-commodity sector's wage per efficiency unit has no impact on the wage of individuals at the bottom of the wage distribution.

To formalize this intuition, let $Y_{g,r,t}(\pi)$ denote the π -quantile of the log-wage distribution of group g in region r at year t . For small shocks, I show in Appendix A.2 that the wage growth between periods t_0 and t in quantile π of the log-wage distribution is given by

$$\Delta Y_{g,r,t}(\pi) = \Delta \omega_{g,r,t}^C + \left[\Delta \omega_{g,r,t}^N - \Delta \omega_{g,r,t}^C \right] \cdot l_{g,r,t_0}^N(\pi) + \mu_{g,r,t} \cdot \tilde{\mathbf{X}}_{g,r,t}(\pi) + \Delta v_{g,r,t}(\pi) \quad (17)$$

where, at quantile π of the log-wage distribution, $l_{g,r,t_0}^N(\pi)$ is the initial employment share of the non-commodity sector and $\tilde{\mathbf{X}}_{g,r,t}(\pi)$ is a set of observable controls. In equation (17), $\Delta v_{g,r,t}(\pi)$ is a shock to the absolute advantage of workers in quantile π of the log-wage distribution.¹⁹

For each group-region-period, equation (17) implies that $\Delta \omega_{g,r,t}^k$ can be consistently estimated from the relation between the initial sector composition, $l_{g,r,t_0}^N(\pi)$, and the wage growth, $\Delta Y_{g,r,t}(\pi)$, across quantiles π of the log-wage distribution. In this context, an estimator of $\Delta \omega_{g,r,t}^k$ based on equation (17) relies on the assumption that, conditional on the set of controls $\tilde{\mathbf{X}}_{g,r,t}(\pi)$, pre-shock variation in sector employment composition is uncorrelated with variation in labor efficiency shocks among individuals with different levels of labor income in a particular group-region-period. This estimator hinges on a central feature of the Roy model embedded in equation (17): the indifference of marginal individuals between the two sectors. For small price shocks, sector-switchers are the marginal individuals with an identical potential wage in the two sectors, implying that their reallocation has no first-order impact on the group's log-wage distribution.²⁰

Armed with the model's prediction in equation (17), I proceed to estimate $(\Delta \omega_{g,r,t}^C, \Delta \omega_{g,r,t}^N)$ by regressing wage growth between two consecutive years of the Census, $\Delta Y_{g,r,t}(\pi)$, on the initial year's sector employment composition, $l_{g,r,t_0}^N(\pi)$, in a set of wage distribution quantiles. For each of the 2,072 group-region-period triples, I implement this regression with 88 bins of 1 p.p. width between the 6th and the 94th percentiles of the wage distribution. The baseline specification contains a set of controls intended to capture potential confounding effects related to differential efficiency growth across workers of various levels of income. These controls include dummies for wage distribution ranges (bottom, middle, top), and dummies for earnings below the minimum wage. Thus, $\Delta \omega_{g,r,t}^k$ is identified from the

¹⁹In Appendix A.2, I show that equation (17) is generated by a first-order expansion of the implicit equation defining $Y_{g,r,t}(\pi)$. In this context, $\Delta v_{g,r,t}(\pi)$ is a shock to the absolute advantage of individuals spread across quantiles of the log-wage distribution. It is introduced by shocks to $(\tilde{u}_{g,r,t}, \boldsymbol{\theta}_{g,r,t})$ that affect the market-specific mixing distribution of absolute advantage, $H_{g,r,t}^e(a) \equiv H^e(a | \tilde{u}_{g,r,t}, \boldsymbol{\theta}_{g,r,t})$ in Assumption 1. The change in the mixing distribution of absolute advantage has consequences for the labor efficiency of individuals at different income levels. As a result, the process generating innovations in $(\tilde{u}_{g,r,t}, \boldsymbol{\theta}_{g,r,t})$ creates, through $\Delta v_{g,r,t}(\pi)$, idiosyncratic shocks to wage growth across quantiles of the wage distribution.

²⁰Expression (17) is modified whenever there exists a wedge in sector potential wage of sector-switchers. This is the case in the presence of non-monetary benefits of employment. To the extent that sector-switchers are spread over the wage distribution, the wedge affects wage gains across quantiles. Consistent with this intuition, there is a new term in equation (17) that is proportional to the fraction of sector-switchers among individuals at the π -quantile of the log-wage distribution.

variation in pre-shock sector employment across small neighborhoods of the log-wage distribution in a group-region-period. Appendix D.1 provides details regarding the implementation of this methodology along with an investigation of the robustness of estimates to implementation choices.

To evaluate the impact of exposure to commodity price shocks on changes in sectoral wage per efficiency unit, consider the following regression:

$$\Delta\omega_{g,r,t}^k = \beta_g^k \cdot \left[\sum_{j \in \mathcal{J}^C} \phi_{g,r}^{C,j} \cdot \Delta \ln p_t^j \right] + \Delta \mathbf{X}_{g,r,t} \gamma_g^k + \Delta e_{g,r,t} \quad (18)$$

where $\Delta\omega_{g,r,t}^k$ is the estimated wage per efficiency unit; and $\mathbf{X}_{g,r,t}$ is a control vector of group-region characteristics potentially correlated with the exposure measure. In the baseline specification, I include period dummies interacted with five macroregion dummies, and I weight microregions by their 1991 share in the national population.²¹ Also, I cluster standard errors by microregion to account for serially correlated shocks.

Table 1 reports the estimation of equation (18) in the sample of Brazilian microregions in 1991–2000 and in 2000–2010. The positive and statistically significant coefficients in column (1) indicate that, for both HSG and HSD, regional exposure to higher commodity prices triggers an increase in the commodity sector’s wage per efficiency unit. With the aim of eliminating potentially confounding effects, I augment the model with a set of flexible controls for the initial sector composition in the region. In this case, estimation relies on cross-regional variation in exposure to higher relative product prices within the commodity sector. Although these additional controls absorb a large part of the cross-section variation in $\Delta\omega_{g,r,t}^C$, they do not substantially alter evaluated estimates, which actually become more precisely estimated.

Lastly, column (3) includes period dummies interacted with initial labor market conditions and non-commodity sector composition controls. These controls represent period-specific effects projected on initial region characteristics, capturing, for example, effects related to the introduction of cash transfer programs and secular differences in sector productivity growth. In column (3), the response of the commodity sector’s wage per efficiency unit to shock exposure is economically large: a 10% increase in commodity prices induces an increase in the commodity sector’s wage per efficiency unit of 9.6% for HSG and 14% for HSD.

Columns (4)–(6) of Table 1 present the estimation of equation (18) for the non-commodity sector’s wage per efficiency unit. The estimated coefficients indicate that exposure to higher commodity prices entails a much weaker effect on the wage per efficiency unit in the non-commodity sector. Consequently, there is an increase in the relative wage per efficiency unit of the commodity sector following an increase in world commodity prices. This movement in $\Delta\omega_{g,r,t}^N - \Delta\omega_{g,r,t}^C$ is inconsistent with per-

²¹As discussed in Appendix D.1, the precision of the estimated sectoral wage per efficiency unit is related to the number of individuals in the microregion. For efficiency purposes, I follow the standard approach of weighting regressions by the population size of the microregion. Alternatively, regions could be weighted by the inverse of the standard error of the estimated wage per efficiency unit. In the baseline specification, I adopted the simple weight by population share for two reasons. First, sectoral regressions would entail different regional weights due to the difference in standard errors in the estimate of each sector’s wage per efficiency unit. Second, inference in equation (17) is nonstandard, requiring a computationally burdensome bootstrap procedure for each of the 2,072 triples of group-period-region.

Table 1: Exposure to Commodity Price Shocks and Sector Wage per Efficiency Unit

Dependent Variable: change in wage per efficiency unit	Commodity sector			Non-commodity sector		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: High School Graduates						
Commodity price shock	0.960+ (0.530)	1.369** (0.405)	0.962** (0.359)	0.410** (0.088)	0.351** (0.072)	0.282** (0.062)
R^2	0.200	0.550	0.598	0.552	0.560	0.592
Panel B: High School Dropouts						
Commodity price shock	1.977* (0.835)	1.651* (0.732)	1.381* (0.624)	-0.239 (0.167)	-0.028 (0.108)	-0.021 (0.087)
R^2	0.272	0.646	0.673	0.193	0.484	0.575
Baseline Controls						
Initial commodity sector size controls	No	Yes	Yes	No	Yes	Yes
Initial commodity sector size controls x period dummy	No	No	Yes	No	No	Yes
Initial labor market conditions x period dummy	No	No	Yes	No	No	Yes
Initial size of manufacturing sector x period dummy	No	No	Yes	No	No	Yes

Note. Stacked sample of 518 microregions in 1991–2000 and in 2000–2010. All regressions are weighted by the microregion share in national population in 1991 and include ten macroregion-period dummies. Commodity sector size controls: quadratic polynomial of commodity sector share in group labor income and dummy for commodity sector share in group labor income in the bottom and top deciles of national distribution. Labor market conditions: quadratic polynomial of per-capita income, share of white employees, share of employed individuals, share of formal sector employees, share of individuals earning less than the federal minimum wage. Standard Errors clustered by microregion ** $p < 0.01$, * $p < 0.05$, + $p < 0.10$

fect substitutability of workers in the two sectors of the economy. Therefore, it cannot be generated in traditional trade models in which workers are perfectly exchangeable between sectors. As discussed below, the magnitude of $\Delta\omega_{g,r,t}^N - \Delta\omega_{g,r,t}^C$ is central in the estimation of the schedule of comparative advantage from the subsequent response in sector employment composition.²²

4.3 Parametric Restrictions: Log-Linear System

The nonparametric identification result in Section 3 is critical to inform the source of variation in the data that recovers comparative and absolute advantage. In practice, data limitations are severe and they may prevent the implementation of a fully flexible estimator capable of nonparametrically recovering the functions of interest. In such cases, auxiliary functional form assumptions on $\alpha_g(\cdot)$ and $A_g(\cdot)$ are particularly useful to increase estimation precision. It is important, however, that these parametric assumptions do not impose artificial restrictions on the model. In the particular case analyzed in this paper, it is particularly relevant that functional forms allow for separate roles for comparative

²²In the general equilibrium of the model in Section 2, the commodity sector demand shifter affects the wage per efficiency unit in the two sectors of the economy. First, there is a response in the commodity sector's wage per efficiency unit implied by the shift in the sector's labor demand. Second, the wage per efficiency unit in the non-commodity sector responds because, as workers move to the commodity sector, the lower supply of effective labor units in the non-commodity sector can only be an equilibrium if firms in the sector face a higher wage per efficiency unit.

and absolute advantage, since they are related to distinct predictions of the model. Accordingly, the benchmark specification in the empirical application is based on the following parametric assumption.

Assumption 5. *Suppose that the schedules of comparative and absolute advantage are given by*

$$\alpha_g(q) \equiv \alpha_g \cdot [\ln(q) - \ln(1 - q)] \quad \text{and} \quad A(q) \equiv \tilde{A}_g + A_g \cdot \ln(q)$$

where $\alpha_g \geq 0$.

Assumption 5 commands constant-elasticity schedules of comparative and absolute advantage. Following the discussion in Section 2, the positive parameter α_g controls the dispersion of comparative advantage; alternatively, the parameter A_g controls the pattern of variation in average absolute advantage of individuals distributed across quantiles of comparative advantage.²³ In the empirical application, the parametric restrictions in Assumption 5 are useful for its dimensionality reduction: there are only two parameters to capture separately comparative and absolute advantage.²⁴

The system in Assumption 5 is a strict generalization of the system obtained under a Fréchet distribution of sector-specific productivity. Because of its tractability, this distributional assumption is the basis of numerous recent empirical applications of the Roy model — see, for example, Hsieh, Hurst, Jones and Klenow (2013); Burstein, Morales and Vogel (2015); and Galle, Rodriguez-Clare and Yi (2015). As discussed in Appendix E, the Fréchet distribution leads to a similar log-linear system, but it contains a single parameter to control both comparative and absolute advantage. In terms of the system above, the Fréchet distribution requires that $\alpha_g = -A_g$ where $\alpha_g < 1$.

The Fréchet distribution mixes the distinct roles of comparative and absolute advantage emphasized in this paper, with strong consequences for the model's predictions. Namely, it imposes constraints not only on the magnitude of the between-sector reallocation, but also on the pattern of selection into both sectors. In fact, $\alpha_g = -A_g$ implies that the sector wage differential is constant, being unable to replicate the positive correlation between world commodity prices on the relative average wage of commodity sector employees documented below. Also, the Fréchet distribution implies that the log-wage variance is constant among workers of the same demographic group. These implications of the Fréchet distribution are not generated by robust features of the model, and they may prevent the model from capturing the full extent of wage inequality movements observed in the data.²⁵

²³In Assumption 5, the distributions of comparative and absolute advantage have finite moments for every $\alpha \geq 0$ and $A_g \in \mathbb{R}$. But this is not necessarily true for its moment generating function. As discussed in Appendix E.3, finite moment generating functions can be guaranteed with bounds on the support of comparative and absolute advantage. Alternatively, one could impose parameter restrictions, $0 \leq \alpha_g < 1$ and $A_g > -1$.

²⁴In the spirit of the series estimator proposed by Newey and Powell (2003), the log-linear system could be augmented to include higher-order polynomials. In the limit, such an expansion would recover nonparametrically functions $\alpha_g(\cdot)$ and $A_g(\cdot)$. Yet, as pointed out by Newey (2013), the estimation of nonlinearities tends to be accompanied by sharp increases in standard errors, requiring multiple strong instruments. The application in this paper is no exception and, for this reason, the constant-elasticity specification in Assumption 5 is particularly attractive.

²⁵Appendix E provides a detailed discussion on the pattern of sector selection implied by the Fréchet distribution. While the restriction of $\alpha_g = -A_g$ is a direct implication of assuming a Fréchet distribution, the restriction of $\alpha_g < 1$ is necessary to guarantee a finite effective labor supply in each sector. Appendix E also discusses the system implied by normally distributed sector-specific productivities — as in Heckman and Sedlacek (1985) and Ohnsorge and Trefler (2007). Although the normal distribution leads to distinct functional forms, the implied system also entails two parameters that parametrize the slopes of the schedules of comparative and absolute advantage.

In contrast, the more general log-linear system in Assumption 5 contains parameters that separately control comparative and absolute advantage. This additional degree of freedom enhances the model's ability to capture movements in wage inequality. In particular, the parameters α_g and A_g allow for much more flexible patterns of selection, generating responses in both sector wage differentials and log-wage variance that would not emerge under the Fréchet distribution.

4.4 Estimation Procedure

Now we are ready to propose an estimator for the schedules of comparative and absolute advantage directly related to the identification result in Theorem 1. Towards this goal, I take advantage of the parametric restrictions in Assumption 5 to construct a consistent GMM procedure with moment conditions that use the differential exposure of Brazilian microregions to the variation in international commodity prices in the two period windows of 1991–2000 and 2000–2010.

To this end, let us combine equations (13)–(15) with the functional forms in Assumption 5 to write the following first-difference system:

$$\Delta\omega_{g,r,t}^N - \Delta\omega_{g,r,t}^C = \alpha_g \cdot \Delta \ln \left(l_{g,r,t}^N / l_{g,r,t}^C \right) + \Delta\mathbf{X}_{g,r,t} \gamma_g^u + \Delta u_{g,r,t} \quad (19)$$

$$\Delta\bar{Y}_{g,r,t}^N - \Delta\omega_{g,r,t}^N = A_g \cdot \Delta \ln \left(l_{g,r,t}^N \right) + \Delta\mathbf{X}_{g,r,t} \gamma_g^v + \Delta v_{g,r,t} \quad (20)$$

$$\Delta\bar{Y}_{g,r,t}^C - \Delta\omega_{g,r,t}^C = -(\alpha_g + A_g) \cdot \Delta \left[\frac{l_{g,r,t}^N}{l_{g,r,t}^C} \ln l_{g,r,t}^N \right] - \alpha_g \cdot \Delta \ln \left(l_{g,r,t}^C \right) + \Delta\mathbf{X}_{g,r,t} \gamma_g^e + \Delta e_{g,r,t} \quad (21)$$

where $\mathbf{X}_{g,r,t}$ is a control vector of group-microregion-period variables that include group-microregion fixed effects; and $\Delta\omega_{g,r,t}^k$ is the change in the wage per efficiency unit of sector k estimated with the procedure described in Section 4.2.

Conditional on the parameter vector $\Theta_g \equiv (\alpha_g, A_g, \gamma_g^u, \gamma_g^v, \gamma_g^e)$, equations (19)–(21) immediately allow the computation of the vector of structural errors: $\mathbf{e}_g(\Theta_g) \equiv [\Delta u_{g,r,t}, \Delta v_{g,r,t}, \Delta e_{g,r,t}]_{r,t}$. I combine this error vector with the matrix of instruments in (16), $\mathbf{W}_g \equiv [\Delta\mathbf{Z}_{g,r,t}^C, \Delta\mathbf{X}_{g,r,t}]_{r,t}$, to obtain moment conditions that allow the consistent estimation of Θ_g . Specifically, I use the following GMM estimator:

$$\hat{\Theta}_g = \arg \min_{\Theta_g} \mathbf{e}_g(\Theta_g)' \mathbf{W}_g \Phi \mathbf{W}_g' \mathbf{e}_g(\Theta_g) \quad (22)$$

where Φ is a matrix of moment weights.²⁶ As above, microregions are weighted by their share in the national population of 1991, and standard errors are clustered by microregion.

²⁶In the baseline specification, I use the optimal weights implied by the two-stage GMM estimator. Below, I attest that similar results are obtained using other matrices of moment weights.

4.5 Results

4.5.1 Reduced-Form Estimates

Before turning to the estimates of the comparative and absolute advantage, I investigate the effect of exposure to commodity price shocks on sectoral employment and wages with the following specification:

$$\Delta Y_{g,r,t} = \beta_g \cdot \left[\sum_{j \in \mathcal{J}^C} \phi_{g,r}^{C,j} \cdot \Delta \ln p_t^j \right] + \Delta \mathbf{X}_{g,r,t} \gamma_g + \Delta e_{g,r,t} \quad (23)$$

where $\Delta Y_{g,r,t}$ is the change in a labor market outcome for individuals of group g in microregion r between years $t - 1$ and t .

Columns (1)–(3) of Table 2 present the estimation of equation (23), with the dependent variable being the commodity sector employment share. Panel A presents estimates for HSG and Panel B for HSD. In line with the correlation documented in Figure 1, the positive and statistically significant coefficients indicate that, for both groups, exposure to higher commodity prices induces workers to reallocate from the non-commodity to the commodity sector.²⁷ In the structural estimation below,

Table 2: Exposure to Commodity Price Shocks and Sector Employment and Wages

Dependent Variable:	Change in commodity sector employment share			Change in commodity sector average log wage premium		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: High School Graduates						
Commodity price shock	0.039** (0.010)	0.031** (0.011)	0.035** (0.010)	0.369** (0.097)	0.302** (0.096)	0.441** (0.108)
R^2	0.217	0.291	0.413	0.115	0.145	0.217
Panel B: High School Dropouts						
Commodity price shock	0.187** (0.068)	0.067* (0.030)	0.061* (0.029)	-0.158 (0.149)	-0.127 (0.150)	0.030 (0.155)
R^2	0.236	0.515	0.561	0.148	0.170	0.215
Baseline Controls						
Initial commodity sector size controls	No	Yes	Yes	No	Yes	Yes
Initial commodity sector size controls x period dummy	No	No	Yes	No	No	Yes
Initial labor market conditions x period dummy	No	No	Yes	No	No	Yes
Initial size of manufacturing sector x period dummy	No	No	Yes	No	No	Yes

Note. Stacked sample of 518 microregions in 1991–2000 and in 2000–2010. All regressions are weighted by the microregion share in national population in 1991 and include ten macroregion-period dummies. Commodity sector size controls and labor market conditions as in Table 1. Standard Errors clustered by microregion ** $p < 0.01$, * $p < 0.05$, + $p < 0.10$

²⁷In Table D5 of Appendix D.2, I show that regional exposure to higher commodity prices does not induce higher labor supply of both native and migrant workers. Thus, the expansion in the commodity sector employment share is driven by the between-sector reallocation of individuals in the market.

the parameter of comparative advantage is implied by the combination of the response in sectoral employment in Table 2 and the response in the relative wage per efficiency unit in Table 1.

Turning to the impact of commodity prices on sectoral wages, I estimate equation (23), with the dependent variable being the commodity sector’s relative average wage. Results in columns (4)–(6) indicate different qualitative responses for the two worker groups. The price shock triggers a significant positive response of the commodity sector wage differential for HSG; in contrast, there is only a small and imprecisely estimated response for HSD. For both groups, however, these estimated responses are much smaller than the estimated response in the relative wage per efficiency unit presented in Table 1. In the model, the difference between the commodity sector’s response in terms of relative average wage and relative wage per efficiency unit corresponds to the compositional effect induced by worker reallocation between sectors. In fact, the magnitude of this difference determines the magnitude of the structural parameters of comparative and absolute advantage presented below.

Lastly, it is important to notice that the positive response in sectoral wages for HSG is inconsistent with the selection pattern implied by a Fréchet distribution. Below, this leads to the rejection of the parametric restrictions required by the Fréchet model in the HSG’s structural estimates. For HSD, the weaker response of sectoral wages yields a selection pattern similar to that implied by the Fréchet distribution.

In Appendix D.2, I investigate the robustness of these results. In particular, Table D3 shows that results are similar if the baseline specification is extended to include the additional period of 1980–1991 and microregion-specific time trends. In addition, Table D4 reports similar qualitative patterns of sectoral responses in employment and wages for additional demographic groups, including female and nonwhite workers.

4.5.2 Estimated Parameters of Comparative and Absolute Advantage

I now present the estimates of the comparative and absolute advantage parameters obtained with the procedure described in Section 4.4. These estimates are reported in Table 3 together with their standard errors clustered by microregion. Column (1) reports the structural parameters implied by the estimation of equations (19)–(21) under the parametric restriction imposed by the Fréchet distribution, $\alpha_g = -A_g$. Estimated parameters indicate that an increase of 1% in the relative wage per efficiency unit of the commodity sector triggers an increase in the relative employment in the commodity sector of approximately 1.2% for both groups (i.e., the inverse elasticity $1/\alpha_g$).

Column (2) presents the estimates obtained under the unrestricted log-linear system in (19)–(21). In this case, comparative advantage parameters indicate between-sector employment reallocation whose magnitude is similar to that of the Fréchet model in column (1) for both groups. Nevertheless, the additional degree of freedom is important for HSG, as the estimated parameter of absolute advantage changes substantially. Among HSG, the strong response of sectoral wage differentials documented above yields an absolute advantage parameter that indicates negative selection into the non-commodity sector. This parameter implies curves of potential sector wages similar to those displayed in case (b) of Figure 3. In contrast, the weak response of sector wage differentials for HSD drives an absolute advantage parameter that indicates a pattern of selection similar to that implied by the Fréchet

Table 3: Estimated Parameters of Comparative and Absolute Advantage

	Fréchet model $\alpha_g = A_g$	Log-linear model
	(1)	(2)
Panel A: High School Graduates		
α_{HSG}	0.819** (0.192)	0.835** (0.212)
A_{HSG}	-0.819** (0.192)	1.966* (0.935)
Test of Fréchet restriction (p-value)	-	0.005
Panel B: High School Dropouts		
α_{HSD}	0.856** (0.140)	0.916* (0.399)
A_{HSD}	-0.856** (0.140)	-0.727** (0.142)
Test of Fréchet restriction (p-value)	-	0.644

Note. Stacked sample of 518 microregions in 1991–2000 and 2000–2010. Two-Step GMM estimator with microregions weighted by their share in the 1991 national population. All equations include macroregion-period dummies, initial commodity sector size controls, and initial labor market conditions as in Table 1. Excluded instruments: quadratic polynomial of regional exposure to world product prices. Standard Errors clustered by microregion ** p<0.01, * p<0.05

model. Indeed, the Fréchet restriction cannot be rejected at usual significance levels. For HSD, there is positive selection into both sectors with curves of potential sector wages similar to those shown in case (a) of Figure 3.

Appendix D.3 investigates the robustness of the results presented in Table 3. To increase confidence in the baseline GMM estimator, Table D6 presents the estimated structural parameters obtained with the separate 2SLS estimation of equations (19) and (20). Such an estimator is less efficiency, because it ignoring the overidentification restriction provided by the response in the commodity sector’s average wage in (21). Despite this fact, point estimates are not only similar in magnitude, but also qualitative conclusions are similar with inference methods robust to weak instruments. Also, I find similar estimated parameters when, as in the reduced-form regressions, the unique instrument is the aggregate exposure to commodity price shocks. Table D7 reports that similar results are implied by the GMM estimator with restricted vectors of excluded instruments and alternative matrices of moment weights. Finally, Table D8 shows that similar results are obtained with alternative specifications in the estimation of sectoral wages per efficiency unit.

4.6 Model Fit

In order to build confidence in the model, I investigate the model’s ability to generate responses in the log-wage distribution that are consistent with those observed in the data. Thus, I estimate equation (23)

using both actual data and the model’s predictions regarding changes in the average and the variance of log wages across Brazilian microregions. Since the estimation of the structural parameters relied on sectoral responses in terms of employment and average wages, this exercise constitutes a test of the model’s goodness of fit.

To implement this test, I compute the model’s predicted changes in the average and the variance of the log-wage distribution using, respectively, equations (11) and (12) derived in Section 2. Expressions (11)–(12) require the changes in sectoral wages per efficiency unit generated by the shock in world commodity prices. I obtain these responses directly from the predicted changes implied by the estimates in columns (3) and (6) of Table 1.

Table 4 presents the results of this exercise. Let us first analyze the response in the average log-wages presented on the top row of each panel. In this case, both the Fréchet and the log-linear models deliver responses whose cross-regional relation to shock exposure is consistent with the cross-regional relation in the data. The similar responses with the two specifications follow from the similar estimated parameters of comparative advantage reported in columns (1) and (2) of Table 3. To see this, recall that the change in the average log-wage in equation (11) only depends on the schedule of comparative advantage.

When we turn to the variance of log-wages within each group, we see in Table 4 that the two specifications yield very different responses. For HSG, the log-linear model implies a negative relation

Table 4: Average and Variance of Log-Wages, Model Predictions and Actual Data

	Predicted change Fréchet model	Predicted change Log-linear model	Actual data
	(1)	(2)	(3)
Panel A: High School Graduates			
Change in Log-Wage Average	0.331	0.331	0.262** (0.054)
Change in Log-Wage Variance	0.000	-0.121	-0.117* (0.049)
Panel B: High School Dropouts			
Change in Log-Wage Average	0.140	0.138	0.166* (0.077)
Change in Log-Wage Variance	0.000	0.075	-0.005 (0.071)
Baseline Controls			
Controls in Table 1	Yes	Yes	Yes

Note. Estimated coefficient of the regression of the dependent variable on shock exposure using the stacked sample of 518 microregions in 1991–2000 and 2000–2010. Regressions are weighted by the microregion share in national population in 1991 and include the baseline controls in Table 1 and the initial wage dispersion (log-wage variance regressions). Dependent variables in columns (1)–(2) are counterfactual changes implied by the model. Dependent variables in column (3) are actual data. Standard Errors clustered by microregion ** p<0.01, * p<0.05

whose magnitude is similar to the negative and statistically significant relation in the data. In contrast, the Fréchet model is unable to generate this relation, since it entails a constant log-wage variance. For HSD in Panel B, the cross-region response of log-wage variance to shock exposure is small and imprecisely estimated. This is consistent with the prediction of the Fréchet model. In this case, the log-linear model yields a small positive response.

5 Counterfactual Simulation: Effect of World Commodity Prices on Brazilian Wage Inequality

To conclude, I use the estimated schedules of comparative and absolute advantage to investigate the consequences to the Brazilian wage distribution of shocks in world commodity prices. Precisely, I ask: “In 1991, how would wage inequality change if commodity prices were equal to those of 2010?” In order to answer this question, I proceed in two steps. In the first step, I compute the change in sectoral wages per efficiency unit implied by the shock to world commodity prices. In the second, I use the model’s predictions to compute the counterfactual change in the average and the variance of the log-wage distribution implied by the change in wage per efficiency unit. While the second step is a straight forward application of the sufficiency result in Proposition 3 of Section 2, the first step is not. In this section, I present the counterfactual changes in Brazilian wage inequality obtained with two alternative procedures to compute changes in sectoral wages per efficiency unit.

As in Section 4.6, the first approach uses the estimated pass-through in Table 1 to compute the effect of regional exposure to price shocks on the sector wage per efficiency unit. This methodology has the advantage of being robust to parametric restrictions on the economy’s structure of production, but it is subject to two shortcomings. First, it does not capture nationwide effects on sectoral wages per efficiency unit, since these are absorbed by period fixed effects included in the regressions. Second, the estimated pass-through is a reduced-form relation that may not hold for different price shocks and different periods. To address these deficiencies, the second approach relies on a fully specified general equilibrium model to obtain the endogenous change in the wage per efficiency unit following exogenous shocks in world commodity prices. This procedure takes inspiration from the exact hat algebra used in recent quantitative papers in international trade — see, e.g., Dekle, Eaton and Kortum (2007) and Costinot and Rodríguez-Clare (2013). It illustrates how the structural parameters of comparative and absolute advantage can be combined with specific assumptions regarding the production and market structures to investigate the distributional effects of sectoral shocks.

5.1 Counterfactual Simulation: Reduced-Form Pass-Through from World Commodity Prices to Sector Wage per Efficiency Unit

In this section, I present the main results of the paper regarding the counterfactual change in Brazilian wage inequality implied by the shock in world commodity prices. I compute changes in sectoral wages per efficiency unit from the reduced-form pass-through in columns (3) and (6) of Table 1. With these variables, I then compute the predicted changes in the average and the variance of the log-wage

distribution using equations (11) and (12) and the estimated parameters of comparative and absolute advantage in Table 3.

5.1.1 Counterfactual Changes in Between-Sector Wage Differentials

Figure 6 reports the changes in between-sector wage differentials across Brazilian microregions. The top panel shows that there are large changes in the relative sectoral wages per efficiency unit. Such a change corresponds to the effect of the world price shock on the wage differential of those who do not switch sectors; that is, sector-stayers in the commodity sector versus sector-stayers in the non-commodity sector. At the national level, the change in the relative sectoral wages per efficiency unit is 8% for HSG and 16% for HSD. Reflecting the strong compositional effects implied by the sectoral reallocation of workers, the bottom panel shows that the response in the commodity sector average wage premium is much smaller, with a national average of 1% for both groups. Following the shock, the predicted expansion of the commodity sector in terms of relative employment is 9.2% for HSG and

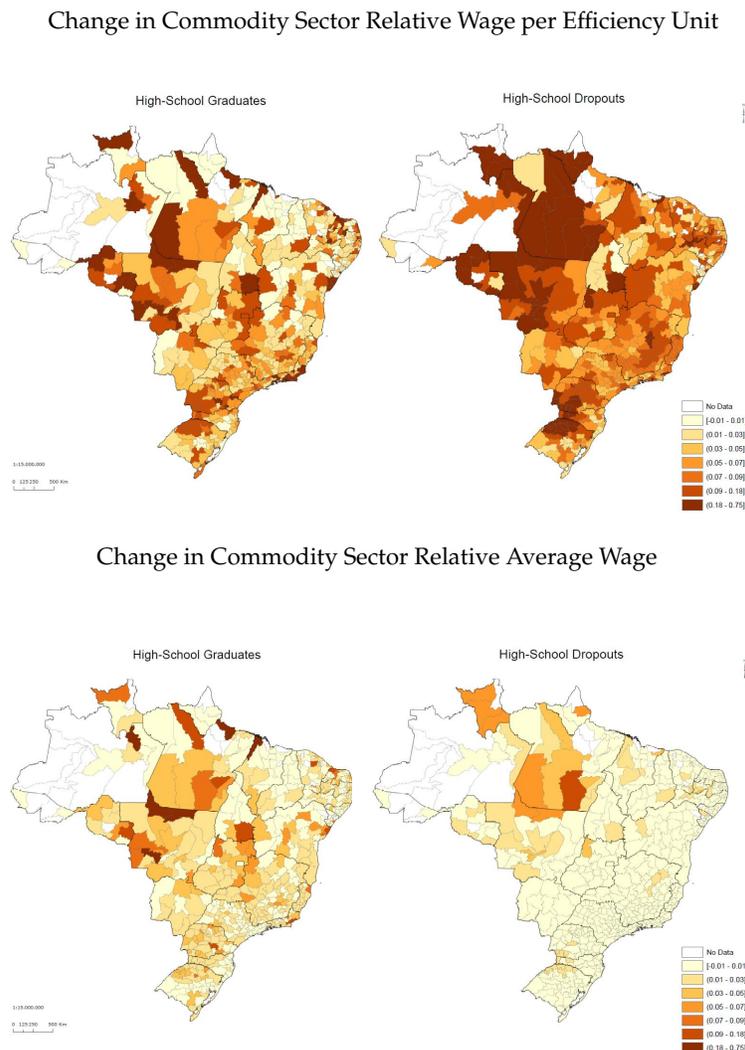


Figure 6: Counterfactual Change in Between-Sector Wage Differentials

13.7% for HSD. This implies an average increase in the commodity sector employment share from 5.2% to 5.7% among HSG, and from 26.2% to 28.9% among HSD.

5.1.2 Counterfactual Changes in Brazilian Wage Inequality

Figure 7 presents the counterfactual change in average log-wage implied by the rise in commodity prices between 1991 and 2010. The positive price shock triggers average wage gains for the two worker groups. Yet the wage gain is more pronounced among HSD, due to their higher employment share in the commodity sector. Consequently, the shock causes a decrease in the HSG-HSD wage premium of approximately 1.1%.

Table 5 reports these counterfactual responses at the national level, where the aggregate log-wage variance is computed with the total variance formula and the microregion's employment share in 1991. Column (2) shows that the price shock triggers a decrease in the log-wage dispersion in the two worker groups. Such a response arises for two reasons. First, regions that specialized in commodity production had lower wages initially, but they experienced stronger wage growth following the positive price shock. Second, the estimated schedules of comparative and absolute advantage imply that, within groups and regions, the shock affects the log-wage variance due to movements in sectoral wage differentials and sectoral employment composition. At the national level, this effect is reinforced by the reduction in the HSG-HSD wage premium triggered by the shock. Panel C shows that 5.6% of the fall in Brazilian log-wage variance is related to the increase in world commodity prices.

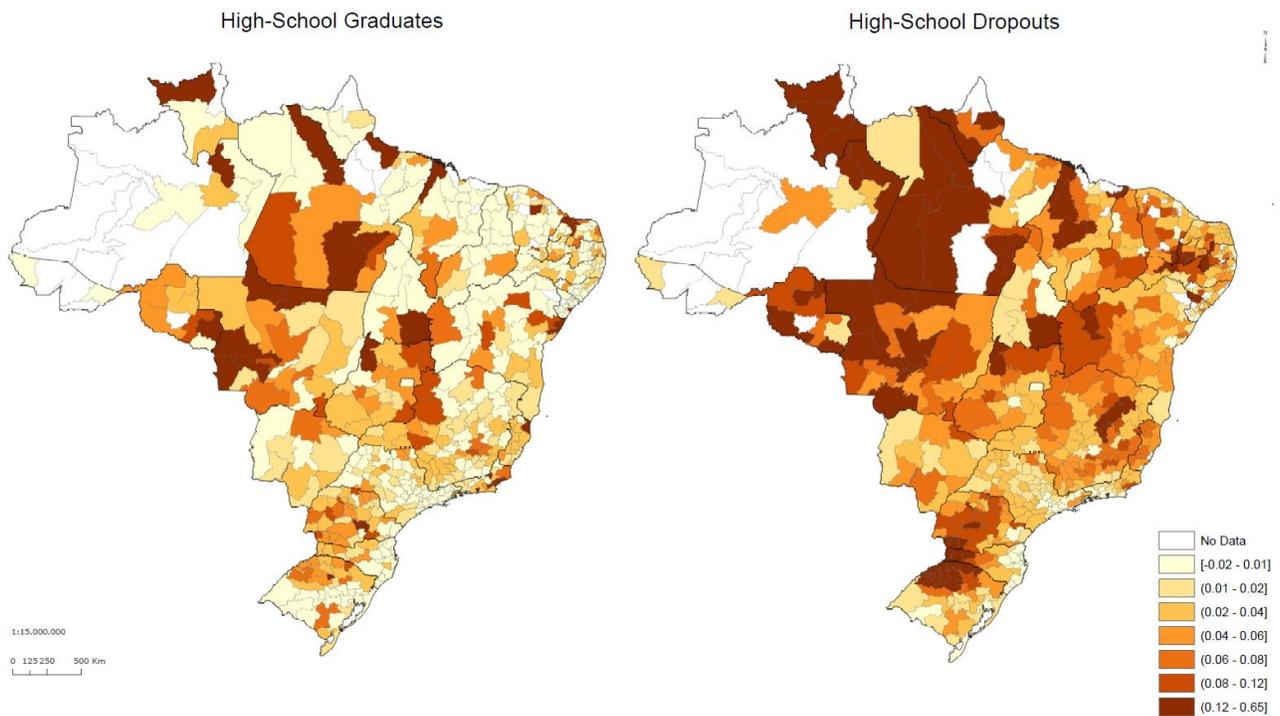


Figure 7: Counterfactual Change in Average Log-Wage

Table 5: Effect of 1991–2010 Rise in World Commodity Prices on Brazilian Wage Inequality

Change in Log-Wage Average	Change in Log-Wage Variance (percent of 1991-2010 change)
(1)	(2)
Panel A: High School Graduates	
0.039	-0.014 7.91%
Panel B: High School Dropouts	
0.050	-0.012 3.97%
Panel C: All Workers	
0.047	-0.017 5.55%

Note. Estimated parameters of comparative and absolute advantage of the log-linear model in column (2) of Table 3.

5.2 Counterfactual Simulation: General Equilibrium Model

In this section, I introduce additional functional form assumptions in the economy presented in Section 2. These assumptions allow the computation of changes in sectoral wages per efficiency unit implied by changes in final product prices in a fully specified general equilibrium model. I denote by a “hat” the change in each variable between the initial and the counterfactual equilibrium.

5.2.1 Additional Parametric Assumptions

In the environment of Section 2, I assume that there are two demographic groups, High School Graduates and High School Dropouts. In each industry, production utilizes the effective labor supplied by employees of both groups and an industry-specific input. In particular, assume that the production function has the following nested CES structure:

$$q^j = (L^j)^{\eta^j} (X^j)^{1-\eta^j} \text{ where } L^j \equiv \left[\beta_{HSG}^j (L_{HSG}^j)^{\frac{\rho-1}{\rho}} + \beta_{HSD}^j (L_{HSD}^j)^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}},$$

η^j is the labor share in total revenue of industry j , and X^j is an industry-specific input with fixed supply. In this production function, High School Graduates and High School Dropouts are imperfect substitutes with ρ denoting the constant elasticity of substitution between the effective labor supplied by the two groups.²⁸

²⁸This particular production structure is imposed mainly due to the limited availability of production data for Brazilian microregions. First, I introduce an industry-specific factor as a simplifying device to generate curvature in sector labor demand in every region. Similarly, one could allow, as in Costinot, Donaldson and Smith (2012), regions to have a continuum of land units with heterogeneous productivity in various industries. Second, I use a Cobb-Douglas function without

With this production technology, shocks in product and factor prices cause responses in the labor demand of industry j of sector k that are given by

$$\hat{L}_g^j = (\hat{w}_g^k)^{-\rho} (\hat{W}^j)^{\rho - \frac{1}{1-\eta^j}} (\hat{p}^j)^{\frac{1}{1-\eta^j}} \quad \text{and} \quad \hat{W}^j \equiv \left[\psi_{HSG}^j (\hat{w}_{HSG}^k)^{1-\rho} + \psi_{HSD}^j (\hat{w}_{HSD}^k)^{1-\rho} \right]^{\frac{1}{1-\rho}} \quad (24)$$

and ψ_g^j is the share of group g in the total wage bill of industry j at the initial equilibrium.

In addition, I assume that changes in sectoral supply of effective labor units are given by

$$\hat{L}_g^N = \frac{\int_0^{I_g^N \hat{I}_g^N} e^{A_g(q)} dq}{\int_0^{I_g^N} e^{A_g(q)} dq} \quad \text{and} \quad \hat{L}_g^C = \frac{\int_{I_g^N}^1 e^{\alpha_g(q) + A_g(q)} dq}{\int_{I_g^N}^1 e^{\alpha_g(q) + A_g(q)} dq} \quad (25)$$

where the change in sector employment composition is determined by equation (7):

$$\ln \hat{w}_g^N - \ln \hat{w}_g^C = \alpha_g \left(I_g^N \hat{I}_g^N \right) - \alpha_g \left(I_g^N \right). \quad (26)$$

The schedules of comparative and absolute advantage govern the sector-specific efficiency of workers employed in each sector, determining the sector supply of effective labor units. In general equilibrium, the sector effective labor supply has to be finite and, therefore, the integrals in equation (25) have to be well defined.²⁹

In this environment, the counterfactual changes in the wage per efficiency unit, $\{\hat{w}_g^C, \hat{w}_g^N\}_g$, have to guarantee labor market clearing for all sectors and groups:

$$\sum_{j \in \mathcal{J}^k} \phi_g^{k,j} \cdot \hat{L}_g^j = \hat{L}_g^k \quad (27)$$

where \hat{L}_g^j is the change in the labor demand of industry j given by equation (24), and \hat{L}_g^k is the change in sector labor supply given by equation (25). At the initial equilibrium, $\phi_g^{k,j}$ denotes the share of industry j in total labor payments of sector k to individuals of group g .

Calibration. The change in sectoral labor demand in equation (24) requires additional parameters that are calibrated as follows. First, the Cobb-Douglas production function implies that the parameter η^j corresponds to the share of labor in the total revenue of industry j . Thus, I calibrate η^j using information on the cost structure of industries in Brazil computed by the IBGE in the 2009 national accounts. Second, sectoral labor demand requires the elasticity of substitution between skilled and unskilled

other mobile factors of production because of the lack of data on the cost structure of industries at the regional level. Third, an analysis of inter-regional trade linkages as in [Caliendo, Parro, Rossi-Hansberg and Sarte \(2014\)](#) cannot be implemented due to insufficient data on cross-regional trade in Brazil.

²⁹In the model, the effective labor supply in the non-commodity sector is $\bar{L}_g^N = \int_0^{I_g^N} E[e^{a(i)} | \alpha_g(q)] dq$. To obtain expression (25), it is sufficient that $E[e^{a(i)} | \alpha_g(q)] = \kappa_g \cdot e^{A_g(q)}$ for some constant κ_g . As discussed in Appendix E.3, such a relation holds under a variety of assumptions regarding the conditional distribution of absolute advantage, including the normal distribution and the Gumbel distribution. A similar argument holds for the effective labor supply in the commodity sector. For all possible parameters, I show in Appendix E.3 that the finiteness of the integrals in (25) is guaranteed by bounds on the productivity support.

workers, ρ . I calibrate this parameter using the estimated elasticity in [Katz and Murphy \(1992\)](#) for the U.S.: $\rho = 1.8$. Third, I use labor market data from the Census to compute the initial cost structure in each region: (i) the share of industry j in total labor payments of sector k to workers of group g , $\phi_{g,r}^{k,j}$; and (ii) the share of group g in total wage bill of industry j , $\psi_{g,r}^j$.

5.2.2 Counterfactual Changes in Brazilian Wage Inequality

Starting from the initial equilibrium in each microregion, I use equations (24)-(27) to compute the counterfactual changes in the wage per efficiency unit, $\{\hat{\omega}_{HSD,r}^C, \hat{\omega}_{HSD,r}^N, \hat{\omega}_{HSG,r}^C, \hat{\omega}_{HSG,r}^N\}$, implied by shocks to final product prices, \hat{p}^j . I then proceed as above and compute the counterfactual change in Brazilian wage inequality.

Table 6 presents the counterfactual results computed with the general equilibrium calibrated model in columns (1) and (2) and compares them to the results computed with the reduced-form pass-through in columns (3) and (4). The two approaches yield similar counterfactual changes in log-wage variance both in the aggregate and within the two groups. In the context of the shock in world commodity prices, the general equilibrium model does not deliver additional insights beyond those obtained with the reduced-form pass-through. Nevertheless, the general equilibrium model provides a methodology by which the structural parameters of comparative and absolute advantage can be used to evaluate the distributional effects of other shocks in the economy — for example, shocks to tariffs in manufactured good produced by industries in the non-commodity sector.

However, columns (1) and (2) of Table 6 indicate that the distinct roles of comparative and absolute advantage are quantitatively important in determining the predicted change in log-wage vari-

Table 6: Effect of 1991–2010 Rise in World Commodity Prices on Brazilian Wage Inequality

Effect of World Commodity Prices on Sector Wage Per Efficiency Unit:	Calibrated General Equilibrium Model		Reduced-Form Pass-Through	
	Fréchet model	Log-linear model	Fréchet model	Log-linear model
	(1)	(2)	(3)	(4)
Panel A: High School Graduates				
	-0.002	-0.018	-0.002	-0.014
	1.24%	10.06%	1.02%	7.91%
Panel B: High School Dropouts				
	-0.014	-0.009	-0.023	-0.012
	4.60%	2.81%	7.65%	3.97%
Panel C: All Workers				
	-0.022	-0.019	-0.023	-0.017
	7.22%	5.99%	7.52%	5.55%

Note. Table reports the counterfactual change in Brazilian log-wage variance, along with the percentage of the actual change in log-wage variance, between 1991 and 2010. Computation uses the estimated parameters of comparative and absolute advantage on Table 3.

ance. With the log-linear model shown in column (2), the shock in world commodity prices triggers a decrease in wage inequality among HSG driven by movements in sectoral wage differentials and within-sector wage dispersion. Such an effect accounts for 10% of the reduction in wage inequality among HSG. In contrast, these responses are ruled out by the Fréchet model and, therefore, column (1) reports almost no change in log-wage variance for HSG. Among HSD, this pattern is inverted: the fall in log-wage variance predicted by the log-linear model is weaker than that predicted by the Fréchet model. The small magnitude of this difference reflects the similar estimates of the structural parameters obtained with the two parametrizations for HSD.

At the national level, the two specifications yield similar predicted changes in log-wage variance. This similarity reflects the fact that the differences in predicted changes go in opposite directions within the two groups. Also, the more pronounced difference among HSG is attenuated in the aggregate since this group represented only 22.2% of the labor force in 1991. Compared to the initial log-wage variance in 1991, the rise in world commodity prices triggers an inequality fall of 2.0% and 1.7% according, respectively, to the Fréchet model and the log-linear model. These figures correspond, respectively, to 7.2% and 6.0% of the total fall in Brazilian wage inequality between 1991 and 2010.

6 Conclusion

This paper starts from one observation: movements in world commodity prices tend to be accompanied by changes in Brazilian wage inequality. Motivated by this aggregate correlation, I developed a unified theoretical and empirical framework to quantify the causal effect of shocks to relative good prices in the world market on the wage distribution within Brazil.

I proposed a model featuring workers' heterogeneity in comparative and absolute advantage with respect to their productivities in the two sectors of the economy. In this environment, I clearly delineated the distinct roles played by the schedules of comparative and absolute advantage in determining sectoral employment and sectoral wages, which allowed me to establish their nonparametric identification in a sample of regional economies. Building on this result, I estimated the schedules of comparative and absolute advantage for High School Graduates and Dropouts using the differential exposure of Brazilian regions to shocks in world commodity prices. Because these schedules are sufficient to compute changes in the average and variance of the log-wage distribution, I was able to use the structural estimates in a quantitative investigation of the effect of shocks in world commodity prices on Brazilian log-wage variance. I concluded that the rise in world commodity prices accounted for 5% to 10% of the decline in Brazilian log-wage variance between 1991 and 2010.

To put my results in perspective, I compare them to the distributional effects of the Brazilian trade liberalization estimated by two recent papers. [Dix-Carneiro and Kovak \(2015a\)](#) investigate the effect of cross-regional variation in exposure to the tariff reduction on the HSG-HSD wage premium, finding that 11% of the 1991–2010 drop in this variable can be attributed to the trade liberalization. Focusing on a different channel, [Helpman, Itskhoki, Muendler and Redding \(2015\)](#) conclude that heterogeneous worker exposure to firms differentially affected by the trade liberalization caused log-wage dispersion to increase by 2% between 1986 and 1994. In contrast, I study an alternative source of international trade shocks in Brazil: the variation in world commodity prices. When I compare my results to those in

these two papers, I find that the shocks in world demand for basic products generate changes in wage inequality with magnitude similar to that created by tariff shocks. My results indicate the distributional importance of shocks in world commodity prices in other developing countries where a large fraction of the workforce is employed in the commodity sector.

In this paper, distributional effects arise from worker heterogeneity in terms of comparative and absolute advantage. My analysis highlighted the different roles of these two economic forces, demonstrating the potential harm of ignoring their distinct implications for changes in wage inequality. This is illustrated by the counterfactual change in wage inequality for HSG, which is much lower under the parametric restrictions imposed by the Fréchet distribution. As a result, I developed a flexible methodology that can be readily applied to investigate changes in between- and within-group wage inequality stemming from a variety of sectoral shocks, including higher competition of foreign manufacturing imports and reductions in the trade cost of services.

In order to dispense with parametric assumptions, my methodology restricted the dimensionality of worker heterogeneity to a bivariate vector of sector-specific productivities. In Roy-like models, it remains an open question how to generalize the insights in this paper to an environment with higher heterogeneity dimensionality while achieving both tractability and flexibility in the analysis of wage inequality movements. Such an extension would enhance the range of questions that could be addressed by the model, being particularly useful to quantify distributional effects of sectoral shocks in countries where workers present high spatial mobility.

References

- Acemoglu, Daron and David H. Autor**, "Skills, tasks and technologies: Implications for employment and earnings," *Handbook of Labor Economics*, 2011, 4, 1043–1171.
- Artuç, Erhan, Shubham Chaudhuri, and John McLaren**, "Trade Shocks and Labor Adjustment: A Structural Empirical Approach," *The American Economic Review*, 2010, pp. 1008–1045.
- Autor, David H, David Dorn, and Gordon H Hanson**, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *The American Economic Review*, 2013, 103 (6), 2121–2168.
- , —, —, —, and **Jae Song**, "Trade Adjustment: Worker-Level Evidence," *The Quarterly Journal of Economics*, 2014, 129 (4), 1799–1860.
- , **Lawrence F Katz, and Melissa S Kearney**, "Trends in US wage inequality: Revising the revisionists," *The Review of Economics and Statistics*, 2008, 90 (2), 300–323.
- Bekman, Eli, John Bound, and Stephen Machin**, "Implications of Skill-Biased Technological Change: International Evidence," *The Quarterly Journal of Economics*, 1998, 113 (4), 1245–1279.
- Borjas, George J**, "Self-Selection and the Earnings of Immigrants," *The American Economic Review*, 1987, 77 (4), 531–553.
- Burstein, Ariel and Jonathan Vogel**, "International trade, technology, and the skill premium," *Manuscript, Columbia University and UCLA*, 2015.

- , **Eduardo Morales, and Jonathan Vogel**, “Accounting for Changes in Between-Group Inequality,” Technical Report, Working Paper 2015.
- Caliendo, Lorenzo, Fernando Parro, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte**, “The impact of regional and sectoral productivity changes on the US economy,” Technical Report, National Bureau of Economic Research 2014.
- , **Maximiliano Dvorkin, and Fernando Parro**, “The Impact of Trade on Labor Market Dynamics,” Technical Report, National Bureau of Economic Research 2015.
- Costa, Francisco, Jason Garred, and Joao Paulo Pessoa**, “Winners and Losers from a Commodities-for-Manufactures Trade Boom,” Technical Report, Centre for Economic Performance, LSE 2014.
- Costinot, Arnaud and Andrés Rodríguez-Clare**, “Trade Theory with Numbers: Quantifying the Consequences of Globalization,” in “In: Helpman, E.(Ed.), Handbook of international economics” Citeseer 2013.
- **and Jonathan E Vogel**, “Matching and Inequality in the World Economy,” *Journal of Political Economy*, 2010, 118 (4), 747–786.
- **and Jonathan Vogel**, “Beyond Ricardo: Assignment Models in International Trade,” Technical Report, National Bureau of Economic Research 2014.
- , **Dave Donaldson, and Cory Smith**, “Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from a 9 Million-Field Partition of the Earth,” *Journal of Political Economy*, Forthcoming, 2012.
- Dekle, Robert, Jonathan Eaton, and Samuel Kortum**, “Unbalanced Trade,” *The American Economic Review*, 2007, pp. 351–355.
- Dix-Carneiro, Rafael**, “Trade liberalization and labor market dynamics,” *Econometrica*, 2014, 82 (3), 825–885.
- **and Brian K Kovak**, “Trade Liberalization and the Skill Premium: A Local Labor Markets Approach,” *American Economic Review*, 2015, 105 (5), 551–57.
- **and —** , “Trade reform and regional dynamics: evidence from 25 years of Brazilian matched employer-employee data,” Technical Report, National Bureau of Economic Research 2015.
- Ferreira, Francisco HG, Sergio Firpo, and Julian Messina**, “A more level playing field? Explaining the decline in earnings inequality in Brazil, 1995-2012,” 2014.
- French, Eric and Christopher Taber**, “Identification of Models of the Labor Market,” *Handbook of Labor Economics*, 2011, 4, 537–617.
- Frias, Judith A, David S Kaplan, and Eric Verhoogen**, “Exports and Within-Plant Wage Distributions: Evidence from Mexico,” *American Economic Review*, 2012, 102 (3), 435–40.
- Galle, Simon, Andres Rodriguez-Clare, and Moises Yi**, “Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade,” *UC Berkeley*, 2015, *mimeo*.
- Gaston, Noel and Daniel Trefler**, “The labour market consequences of the Canada-US Free Trade Agreement,” *Canadian Journal of Economics*, 1997, 30 (1), 18–41.

- Goldberg, Pinelopi Koujianou and Nina Pavcnik**, "Distributional Effects of Globalization in Developing Countries," *Journal of Economic Literature*, 2007, 45 (1), 39–82.
- Grossman, Gene M**, "Partially mobile capital: A general approach to two-sector trade theory," *Journal of International Economics*, 1983, 15 (1), 1–17.
- , "Heterogeneous workers and international trade," *Review of World Economics*, 2013, 149 (2), 211–245.
- Hanson, Gordon H**, "The Rise of Middle Kingdoms: Emerging Economies in Global Trade," *The Journal of Economic Perspectives*, 2012, 26 (2), 41–63.
- Heckman, James J and Bo E Honoré**, "The Empirical Content of the Roy Model," *Econometrica*, 1990, pp. 1121–1149.
- and **Guilherme Sedlacek**, "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *The Journal of Political Economy*, 1985, 93 (6), 1077.
- Helpman, Elhanan, Oleg Itskhoki, and Stephen Redding**, "Inequality and unemployment in a global economy," *Econometrica*, 2010, 78 (4), 1239–1283.
- , —, **Marc-Andreas Muendler, and Stephen J Redding**, "Trade and inequality: From theory to estimation," Technical Report, National Bureau of Economic Research 2015.
- Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow**, "The Allocation of Talent and US Economic Growth," Technical Report, National Bureau of Economic Research 2013.
- Jones, Ronald W**, "The structure of simple general equilibrium models," *The Journal of Political Economy*, 1965, pp. 557–572.
- , "Income distribution and effective protection in a multicommodity trade model," *Journal of Economic Theory*, 1975, 11 (1), 1–15.
- Kambourov, Gueorgui**, "Labour market regulations and the sectoral reallocation of workers: The case of trade reforms," *The Review of Economic Studies*, 2009, 76 (4), 1321–1358.
- Katz, Lawrence F and Kevin M Murphy**, "Changes in relative wages, 1963–1987: supply and demand factors," *The quarterly journal of economics*, 1992, 107 (1), 35–78.
- Kovak, Brian K**, "Regional Effects of Trade Reform: What is the Correct Measure of Liberalization?," *The American Economic Review*, 2013, 103 (5), 1960–1976.
- Lagakos, David and Michael E Waugh**, "Selection, Agriculture, and Cross-Country Productivity Differences," *The American Economic Review*, 2013, 103 (2), 948–980.
- Lawrence, Robert Z and Matthew J Slaughter**, "International trade and American wages in the 1980s: giant sucking sound or small hiccup?," *Brookings papers on economic activity. Microeconomics*, 1993, pp. 161–226.
- Matzkin, Rosa L**, "Nonparametric Identification," *Handbook of Econometrics*, 2007, 6, 5307–5368.
- Menezes-Filho, Naércio Aquino and Marc-Andreas Muendler**, "Labor reallocation in response to trade reform," Technical Report, National Bureau of Economic Research 2011.
- Mulligan, Casey B and Yona Rubinstein**, "Selection, investment, and women's relative wages over time," *The Quarterly Journal of Economics*, 2008, 123 (3), 1061–1110.

- Mussa, Michael**, "Imperfect factor mobility and the distribution of income," *Journal of International Economics*, 1982, 12 (1), 125–141.
- Newey, Whitney K**, "Nonparametric Instrumental Variables Estimation," *The American Economic Review*, 2013, 103 (3), 550–556.
- **and James L Powell**, "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 2003, 71 (5), 1565–1578.
- Ohnsorge, Franziska and Daniel Trefler**, "Sorting It Out: International Trade with Heterogeneous Workers," *Journal of Political Economy*, 2007, 115 (5), 868–892.
- Reventa, Ana L**, "Exporting jobs? The impact of import competition on employment and wages in US manufacturing," *The Quarterly Journal of Economics*, 1992, 107 (1), 255–284.
- Roy, A. D.**, "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 1951, 3 (2), pp. 135–146.
- Stolper, Wolfgang and Paul A Samuelson**, "Protection and Real Wages," *Review of Economic Studies*, 1941, 9 (1), 58–73.
- Topalova, Petia**, "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India," *American Economic Journal: Applied Economics*, 2010, 2 (4), 1–41.
- Verhoogen, Eric A**, "Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector," *The Quarterly Journal of Economics*, 2008, 123 (2), 489–530.
- Wacziarg, Romain and Jessica Seddon Wallack**, "Trade liberalization and intersectoral labor movements," *Journal of International Economics*, 2004, 64 (2), 411–439.
- Young, Alwyn**, "Structural Transformation, the Mismeasurement of Productivity Growth, and the Cost Disease of Services," *The American Economic Review*, 2014, 104 (11), 3635–3667.

A Proofs

A.1 Proof of Lemma 1

By Assumption 3, $E[u_{g,m} | \mathbf{Z}_{g,m}^k, \mathbf{X}_{g,m}] = 0$ so that $E[\Phi_g(l_{g,m}^N) + \mathbf{X}_{g,m}\gamma_g | \mathbf{Z}_{g,m}^k, \mathbf{X}_{g,m}] = E[y_{g,m} | \mathbf{Z}_{g,m}^k, \mathbf{X}_{g,m}]$. Now let us proceed by contradiction. Suppose there exist $\tilde{\Phi}_g(l_{g,m}^N)$ and $\tilde{\gamma}_g$ such that

$$E[\tilde{\Phi}_g(l_{g,m}^N) + \mathbf{X}_{g,m}\tilde{\gamma}_g | \mathbf{Z}_{g,m}^k, \mathbf{X}_{g,m}] = E[y_{g,m} | \mathbf{Z}_{g,m}^k, \mathbf{X}_{g,m}]$$

By Assumption 4,

$$\Phi_g(l_{g,m}^N) - \tilde{\Phi}_g(l_{g,m}^N) + \mathbf{X}_{g,m}(\gamma_g - \tilde{\gamma}_g) = 0 \text{ almost surely.}$$

Take markets m and n such that $\mathbf{X}_{g,m} = \mathbf{X}_{g,n}$. The condition above implies that

$$\Phi_g(l_{g,m}^N) - \Phi_g(l_{g,n}^N) = \tilde{\Phi}_g(l_{g,m}^N) - \tilde{\Phi}_g(l_{g,n}^N) \text{ for all } l_{g,m}^N \text{ and } l_{g,n}^N.$$

Thus, $\Phi_g(\cdot)$ is identified up to a constant. To determine this constant, we can use the normalizations $E[u_{g,m}] = E[\mathbf{X}_{g,m}\gamma_g] = 0$, which imply that $E[\Phi_g(l_{g,m}^N)] = E[y_{g,m}]$. \square

A.2 Derivation of Equation (17)

Consider shocks to endogenous and exogenous variables in a particular market m where the sector-specific productivity distribution satisfies Assumption 1. To simplify notation, I drop the index m .

Recall that individual i 's log-wage is given by $y_g(i) = \max\{\omega_g^C + s_g(i) + a_g(i); \omega_g^N + a_g(i)\}$. Under Assumption 1, this implies that the log-wage distribution of group g is given by

$$\begin{aligned} Pr[y_g(i) \leq y] &= Pr[y_g(i) \leq y; s_g(i) \leq \omega_g^N - \omega_g^C] + Pr[y_g(i) \leq y; s_g(i) > \omega_g^N - \omega_g^C] \\ &= \int_0^{l_g^N} Pr[a_g(i) \leq y - \omega_g^N | \alpha_g(q)] dq + \int_{l_g^N}^1 Pr[a_g(i) \leq y - \omega_g^C - \alpha_g(q) - \tilde{u}_g | \alpha_g(q)] dq \end{aligned} \quad (28)$$

where $Pr[a_g(i) \leq a | \tilde{s}_g(i) = s] \equiv \mu H_g^a(a|s) + (1 - \mu) H_g^e(a | \tilde{u}_g, \theta_g)$.

By construction, $Y_g(\pi)$ solves $Pr[y_g(i) \leq Y_g(\pi)] = \pi$. Taking a first-order expansion,

$$\left[f_g^N(Y_g(\pi)) + f_g^C(Y_g(\pi)) \right] \Delta Y_g(\pi) = f_g^N(Y_g(\pi)) \cdot \Delta \omega_g^N + f_g^C(Y_g(\pi)) \cdot \Delta \omega_g^C - \Delta v_g'(\pi)$$

where $f_g^N(y) \equiv \frac{\partial Pr[y_g(i) \leq y; s_g(i) \leq \omega_g^N]}{\partial y}$, $f_g^C(y) \equiv \frac{\partial Pr[y_g(i) \leq y; s_g(i) > \omega_g^N]}{\partial y}$, and $\omega_g \equiv \omega_g^N - \omega_g^C$.

In this expression, the first-order impact of the endogenous change in l_g^N is eliminated by the employment condition (7), reflecting the fact that marginal workers are indifferent between the two sectors. The term $v_g'(\pi)$ incorporates changes to other exogenous parameters of the productivity distri-

bution:

$$\begin{aligned} \Delta v'_g(\pi) &\equiv (1 - \mu) \left[\int_0^{l_g^N} \nabla_{\theta} H^e(Y_g(\pi) - \omega_g^N | \tilde{u}_g, \theta_g) dq + \int_{l_g^N}^1 \nabla_{\theta} H^e(Y_g(\pi) - \omega_g^C - \alpha_g(q) - \tilde{u}_g | \tilde{u}_g, \theta_g) dq \right] \cdot \Delta \theta_g \\ &+ \left[f_g^C(Y_g(\pi)) + (1 - \mu) \left(\int_0^{l_g^N} \nabla_u H^e(Y_g(\pi) - \omega_g^N | \tilde{u}_g, \theta_g) dq + \int_{l_g^N}^1 \nabla_u H^e(Y_g(\pi) - \omega_g^C - \alpha_g(q) - \tilde{u}_g | \tilde{u}_g, \theta_g) dq \right) \right] \Delta \tilde{u}_g. \end{aligned}$$

To obtain equation (17), notice that, by definition, $l_g^N(\pi) \equiv P[s_g(i) \leq \omega_g^N - \omega_g^C | y_g(i) = Y_g(\pi)]$. This is equivalent to

$$l_g^N(\pi) = \frac{\Pr[y_g(i) = Y_g(\pi); s_g(i) \leq \omega_g^N - \omega_g^C]}{\Pr[y_g(i) = Y_g(\pi)]} = \frac{f_g^N(Y_g(\pi))}{f_g^N(Y_g(\pi)) + f_g^C(Y_g(\pi))}.$$

Thus,

$$\Delta Y_g(\pi) = \Delta \omega_g^C \cdot l_g^C(\pi) + \Delta \omega_g^N \cdot l_g^N(\pi) + \Delta v''_g(\pi)$$

where $\Delta v''_g(\pi) \equiv -\Delta v'_g(\pi) / [f_g^N(Y_g(\pi)) + f_g^C(Y_g(\pi))]$.

Finally, equation (17) is obtained by projecting $\Delta v''_g(\pi)$ on observable covariates and unobservable variables such that

$$\Delta v''_g(\pi) \equiv \mu_g \cdot \tilde{\mathbf{X}}_g(\pi) + \Delta v_g(\pi).$$

B Model Extension: Non-monetary Employment Benefits

This section extends the model of Section 2 by incorporating non-monetary employment benefits — a reduced-form for work conditions and switching cost. The environment of Section 2 remains the same except for workers' preference structure. If employed in sector k , assume that individual $i \in \mathcal{I}_{g,m}$ obtains utility $\tau_g^k(i) \cdot u(c)$ from consuming bundle c where $u(\cdot)$ is homogeneous of degree one. Thus, individual i 's payoff of employment in sector k is given by

$$U_g^k(i) = \tau_g^k(i) \cdot \frac{w_{g,m}^k L_g^k(i)}{P_m} \quad (29)$$

where $\tau_g^k(i)$ is individual i 's private benefit of being employed in sector k ; and P_m is the price index in market m .

In the presence of non-monetary employment benefits, I extend the notion of comparative advantage to also include relative sectoral preferences. Accordingly, define individual i 's comparative advantage as

$$s_g(i) \equiv \ln[L_g^C(i)/L_g^N(i)] + \ln[\tau_g^C(i)/\tau_g^N(i)],$$

and individual i 's efficiency in sector k as

$$a_g^k(i) \equiv \ln[L_g^k(i)] \quad \text{for } k = C, N.$$

In a given group and market, consider the following distribution of preferences and productivities.

Assumption 6. Suppose individual i in market m , $i \in \mathcal{I}_{g,m}$, independently draws $(s_g(i), a_g^C(i), a_g^N(i))$ as follows.

i. **Comparative Advantage:**

$$s_g(i) = \tilde{s}_g(i) + \tilde{u}_{g,m} \quad \text{and} \quad \{\tilde{s}_g(i)\} \sim F_g(s)$$

where $\tilde{u}_{g,m}$ is a group-market shifter of comparative advantage, and $\alpha_g(q) \equiv (F_g)^{-1}(q)$.

ii. **Sector-Specific Efficiency:**

$$a_g^k(i) \equiv \tilde{a}_g^k(i) + \tilde{v}_g^k \quad \text{s.t.} \quad \left\{ (a_g^C(i), a_g^N(i)) \mid \tilde{s}_g(i) = s \right\} \sim H_g^a(a^C, a^N \mid s)$$

where $\tilde{v}_{g,m}^k$ is a group-market shifter of sector efficiency and

$$A_g^k(q) \equiv \int \int a^k dH_g^a(a^C, a^N \mid \alpha_g(q)).$$

The preference structure in (29) immediately implies that utility maximizing individuals choose to be employed in the non-commodity sector if, and only if, $s_g(i) \leq \omega_{g,m}^N - \omega_{g,m}^C$. Thus,

$$\omega_{g,m}^N - \omega_{g,m}^C = \alpha_g(l_{g,m}^N) + \tilde{u}_{g,m}. \quad (30)$$

Given the allocation of workers to sectors, the average log-wage in the non-commodity sector is $\bar{Y}_g^N \equiv E[\bar{Y}_g^k(q) \mid q < l_{g,m}^N]$, which is equivalent to

$$\bar{Y}_{g,m}^N = \omega_{g,m}^N + \bar{A}_g^N(l_{g,m}^N) + \tilde{v}_{g,m}^N \quad \text{s.t.} \quad \bar{A}_g^N(l) \equiv \frac{1}{l} \int_0^l A_g^N(q) dq. \quad (31)$$

Also, the average log-wage in the commodity sector is $\bar{Y}_g^C \equiv E[\bar{Y}_g^C(q) \mid q \geq l_{g,m}^N]$ and, therefore,

$$\bar{Y}_{g,m}^C = \omega_{g,m}^C + \bar{A}_g^C(l_{g,m}^N) + \tilde{v}_{g,m}^C \quad \text{s.t.} \quad \bar{A}_g^C(l) \equiv \frac{1}{1-l} \int_l^1 A_g^C(q) dq. \quad (32)$$

With a sector demand shifter satisfying Assumptions 3–4, Lemma 1 establishes the identification of $\alpha_g(\cdot)$ from equations (30). Also, Lemma 1 establishes that $\bar{A}_g^N(\cdot)$ and $\bar{A}_g^C(\cdot)$ are respectively identified from equations (31)–(32). To recover $A_g^C(\cdot)$ and $A_g^N(\cdot)$, notice that

$$A_g^N(q) = \frac{\partial}{\partial q} \left[q \cdot \bar{A}_g^N(q) \right] \quad \text{and} \quad A_g^C(q) = -\frac{\partial}{\partial q} \left[(1-q) \cdot \bar{A}_g^C(q) \right].$$

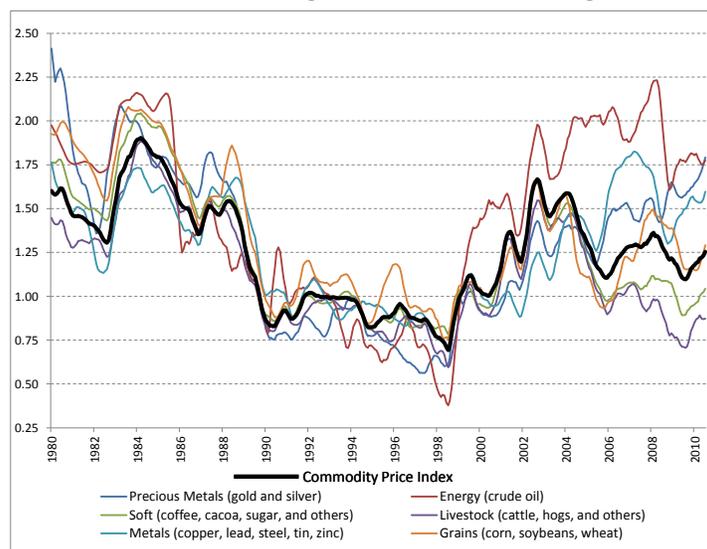
Theorem 2. Consider a set of segmented markets, m , subject to sector demand shifters, $\mathbf{Z}_{g,m}^k$, such that Assumptions 2–4 and 6 hold. For each worker group g , $\alpha_g(\cdot)$ is identified from equation (30), $A_g^N(\cdot)$ is identified from equation (31); and $A_g^C(\cdot)$ is identified from equation (32).

C Data Construction and Measurement

C.1 World Price of Agriculture and Mining Commodities

To capture Brazil's exposure to world prices of basic products, I build price indices for each commodity category. The first source of international commodity prices is the Commodity Research Bureau, which publishes price indices by commodity group based on product spot prices in the main exchange markets in the United States. In the paper, I use those groups with sizable employment participation in Brazil: Grains (corn, soybeans, and wheat), Soft Agriculture (cocoa, coffee, sugar, orange juice, and others), Livestock (hides, hogs, lard, steers, tallow, and others), and Metals (copper scrap, lead scrap, steel scrap, tin, zinc, and others). In addition, I build price indices for two commodity groups using future prices in the New York Mercantile Exchange: Precious Metals (gold and silver) and Energy (crude oil). Due to their small employment importance in Brazil, I aggregate Metals and Precious Metals into a single Mining category. These series of international nominal prices were converted into local currency using the nominal exchange rate and deflated by the Brazilian consumer price index (IPCA).³⁰ To avoid short-term price volatility, I use the average price in the six months preceding the process of data collection of the Census; that is, the average price between March and August of each year of the Census.

Figure C1: World Price of Agriculture and Mining Commodities



C.2 Industry Composition

Individuals in the sample are allocated to sectors according to their self-reported industry of employment. Table C1 shows the industry classification used in this paper together with corresponding industry codes used by the IBGE in each year of the Census and the PNAD. I use crosswalk tables publicly

³⁰All commodity price series were downloaded from the Global Financial Database. In the end of 2008, the soft and grains indices were unified under the foodstuff index. Thus, I build these series for 2009–2010 using each index description. Series of nominal exchange rate and IPCA were downloaded from the IPEADATA.

provided by the IBGE to link the different activity codes across years. The division of industries in the commodity sector accommodates available information on international prices as described above.

Table C1: Industry Classification and IBGE Activity Codes

Industry	Atividades	CNEA-Dom	CNEA-Dom 2.0
	PNADs of 1981-2001 1980 Census and 1991 Census	PNADs of 2002-2009 2000 Census	- 2010 Census
Commodity Sector			
Grains (corn, soybeans, and wheat)	20; 21; 22	1102; 1103; 1107	1102; 1103; 1107
Soft (coffee, cocoa, sugar, and others)	11; 12; 14-17; 23; 24	1104; 1105; 1110-1116; 2001; 2002; 15022; 15042	1104; 1105; 1110-1116; 10022; 10093
Livestock (cattle, hogs, and others)	26; 27; 41; 42	1201-1205; 1208; 1209; 1300; 1402; 5001; 5002; 15010; 15030	1201-1205; 1208; 1209; 1402; 1999; 3001; 3002; 10010; 10030
Metals (copper, lead, steel, zinc, and others)	58	13002	7002
Precious Metals (gold and silver)	55	13001	7001
Energy (crude oil)	51	11000	6000
Other agriculture and mining	13; 18-19; 25; 28; 29; 31-37; 50; 52-54; 56; 57; 59; 581	1101; 1106; 1108; 1109; 1117; 1118; 1206; 1207; 1401; 10000; 12000; 14001-14004	1101; 1106; 1108; 1109; 1117-1119; 1206; 1207; 1401; 5000; 8001-9000
Non-Commodity Sector			
Manufacturing	100-300	15021; 15041; 15043; 15050; 16000- 37000	10021; 10091; 10092; 10099-32999; 38000
Non-Tradable Goods and Services	340-901	1500; 40010-99000	1500; 33001-37000; 39000-99000

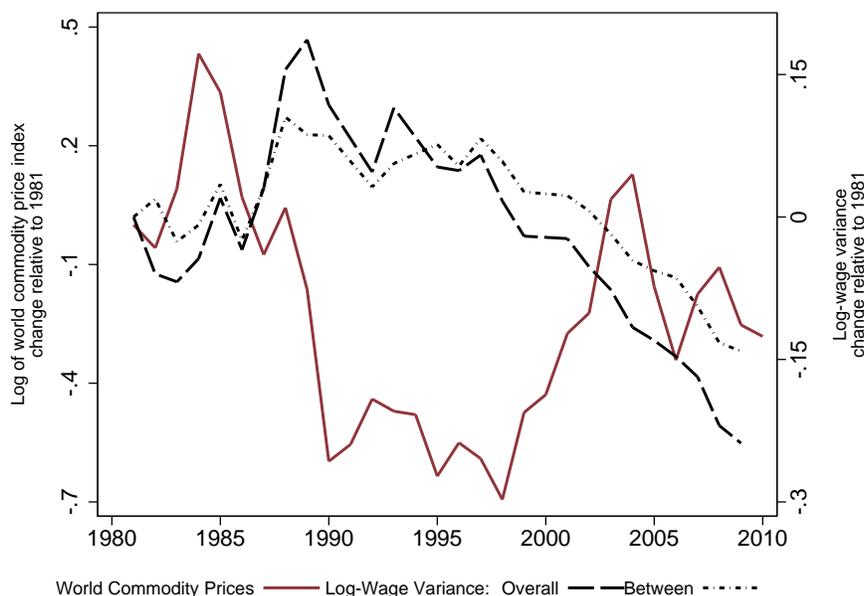
C.3 Trends in Brazilian Wage Dispersion

I obtain annual data on Brazilian labor market outcomes from the National Household Sample Survey (PNAD) collected by the Brazilian Institute of Geography and Statistics (IBGE) between 1981 and 2009.³¹ To focus on individuals with strong labor force attachment, I consider a benchmark sample of full-time male employed individuals aged 16–64. I decompose the movement in log-wage variance into observable and residual components by regressing log wages on a full set of dummies for years of experience (0–39 years), years of education (0–16 years), state of residence (27 states), race (white dummy), and sector of employment (commodity sector dummy). Figure C2 presents the trends in Brazilian wage inequality between 1981 and 2009. Throughout the period, observable worker attributes account for a large share of the change in log-wage variance: 73% of the increase in 1981–1990, and 65% of the decline in 1990–2009.

Table C2 presents the full decomposition of log-wage variance. In 1981–1990, the increase in the between-component of wage variance was mainly driven by the covariance term with additional contributions of the terms related to sector and education dummies. In the 1990–2009, the sharp drop in the between-component of wage dispersion is distributed across all terms, with the largest contribution coming from the education dummies. Conclusions in Table C2 are related to results reported elsewhere in the literature. In particular, [Ferreira, Firpo and Messina \(2014\)](#) highlight the importance of falling educational and state wage gaps for the decrease in Brazilian wage inequality between 1995 and 2012. Using administrative data for formal sector employees, [Helpman, Itskhoki, Muendler and Redding \(2015\)](#) conclude that observable worker attributes account for roughly half of the increase

³¹The National Household Sample Survey is not available for the years in which the IBGE published the Brazilian Demographic Census (1980, 1991, 2000, and 2010).

Figure C2: World Commodity Prices and Brazilian Log-Wage Variance, 1981-2010



Note. World Commodity Prices correspond to the log of the commodity price index computed with the world price of agriculture and mining products converted to Brazilian currency and deflated by the Brazilian consumer price index. Sample of male full-time workers extracted from the PNAD. Between component of log-wage variance computed with the predicted values of the regression of log wage on a full set of dummies for years of experience (0–39 years), years of education (0–16 years), state of residence (27 states), race (white dummy), and sector of employment (commodity sector).

in log-wage variance between 1986 and 1995. In Table C2, the residual component of log-wage variance presents a lower contribution to log-wage variance movements, due to the inclusion of a more comprehensive set of dummies for state of residence, years of education, and years of experience.

Table C2: Decomposition of Brazilian Log-Wage Variance, 1981–2009

	1981	1986	1990	1995	1999	2005	2009
Overall	0.935	0.900	1.053	0.987	0.915	0.805	0.697
Residual	0.459	0.449	0.492	0.444	0.418	0.390	0.366
Between	0.475	0.451	0.561	0.544	0.496	0.415	0.331
Sector	0.024	0.013	0.031	0.033	0.027	0.013	0.012
Education	0.257	0.263	0.278	0.260	0.248	0.232	0.190
State	0.053	0.044	0.052	0.052	0.043	0.045	0.034
Race	-	-	0.005	0.005	0.006	0.003	0.004
Experience	0.109	0.111	0.107	0.078	0.075	0.067	0.053
Covariance	0.034	0.020	0.088	0.115	0.097	0.055	0.037

Note. Sample of male full-time workers extracted from the PNAD. Wage decomposition implied by a regression of log wage on a full set of dummies for years of experience (0–39 years), years of education (0–16 years), state of residence (27 states), race (white dummy), and sector of employment (commodity sector).

C.4 Empirical Application: Data Construction

Labor Market Data. I obtain data on labor market outcomes from publicly available long versions of the Brazilian Census collected by the Brazilian Institute of Geography and Statistics (IBGE) for the years of 1980, 1991, 2000, and 2010. From the Census, I extract a sample of full-time workers aged between 16 and 64. Full-time workers are defined as those reporting more than 35 weekly worked hours. I restrict the sample to workers with calculated experience between 0 and 39 years. Experience is defined as the individual's age minus a predicted initial working age that equals 23 for college graduates, 18 for High School graduates, and 15 for those with only primary education. The benchmark sample is further restricted to include only white male workers. This restriction allows us to focus on individuals with strong labor force. In addition, it excludes individuals directly affected by the strong declines in gender and race wage differential between 1995 and 2010 (Ferreira et al., 2014). Such movements in wage differentials are not the directly related to the model in this paper. In robustness exercises, I extend the benchmark sample to also include female and non-white workers.

Regional Labor Markets. I use the microregion concept created by the IBGE in the 1991 Census as a regional labor market unit. Each of the 558 microregions corresponds to a set of economically integrated municipalities with interconnected labor markets. This definition was used in a series of recent papers analyzing the response of local labor markets to aggregate trade shocks (e.g., Kovak, 2013 and Dix-Carneiro and Kovak, 2015a,b). The microregion concept in Brazil is similar to the Commuting Zones in the United States used by Autor, Dorn and Hanson (2013). Despite the sharp increase in the number of municipalities between 1991 and 2010, the IBGE maintained the same microregion definition in the Censuses of 1991, 2000, and 2010.³² In the 1980 Census, the microregion variable does not exist, so I created it from existing municipalities in 1980. Because of the change in municipality borders between 1980 and 1991, it is only possible to replicate a subset of the microregions using historical administrative borders. To be more precise, I recover 540 microregions in the 1980 Census compared to the 558 microregions in the 1991 Census.³³

Sample Selection. In the empirical application, I select a baseline sample of 518 microregions with positive employment in the commodity sector for all groups in the 1991–2010 period, covering 98.4% of the country's population in 1991. The 1980–1991 period is excluded from the baseline sample mainly because of the turbulent economic environment in Brazil during the 1980s. The decade was marked by hyperinflationary episodes, suspension of foreign currency convertibility, and the adoption of re-

³²There were 4,491 municipalities in 1991 and 5,565 in 2010. Out of the 1074 municipalities created in the period, 998 municipalities had parent municipalities in a single microregion and, therefore, they were allocated to this microregion. The other 76 municipalities had parent municipalities in more than one microregion. These municipalities, which represented .33% of employment in 2000, were allocated to the microregion of the parent municipality that ceded the highest population share to the new municipality. This procedure adopted by the IBGE minimizes any measurement error implied by the border change. In fact, all results in the paper are robust to using a sample of 491 microregions built by aggregating microregions such as to keep borders unchanged in the period.

³³Out of the 3,991 municipalities in the 1980 Census, I am able to link 3,938 municipalities to at least one of the 4,491 municipalities in the 1991 Census. With these linked municipalities, I construct microregions in 1980 using the microregion assigned to corresponding municipalities in 1991. The main problem of this method is the existence of new municipalities in 1991 that belonged to a different microregion than their parent municipalities in 1980. This is the case for 85 of the 500 municipalities created between 1980 and 1991, accounting for .67% of total employment in 2000.

strictive internal controls on prices and wages. In this environment, it is not clear that relative international prices were very informative about relative prices faced by domestic producers when deciding resource allocation. More normal economic conditions returned after the series of structural reforms implemented in 1993–1994 that brought monetary stabilization, eliminated price controls, and restored full currency convertibility. Robustness exercises attest that similar results hold in the extended sample spanning the entire 1980–2010 period.

Sector Demand Shifter. To build the group-region exposure to international commodity prices, I compute total labor income by industry. To this end, I consider the weighted sum of monthly wages of individuals reporting to hold their main job in the industry using Census sampling weights. Denote $Y_{g,r,t}^j$ as the total labor payments of industry j to workers of group g in microregion r at year t . The initial participation of industry j in the labor payments of sector k to group g in microregion r is

$$\phi_{g,r}^{kj} \equiv \frac{Y_{g,r,1991}^j}{\sum_{j' \in \mathcal{J}^k} Y_{g,r,1991}^{j'}} \quad \text{where } j \in \mathcal{J}^k.$$

Table C3 reports summary statistics of industry composition in the sample of 518 microregions in 1991. Columns (1) and (3) indicate that regions, on average, have a large fraction of their work force allocated to the commodity sector, with agriculture accounting for the bulk of the sector’s labor expenditure. Importantly, columns (2) and (4) document great heterogeneity in industry composition across microregions. Comparing columns (1) and (3), it is possible to identify different exposure patterns for the two groups. While HSD are more likely to be employed in the production of grains and soft agricultural items, HSG are more likely to be employed in the production of livestock and crude oil. Due to their small employment share, I aggregate the Metals and the Precious Metals groups into a single Mining category.

Table C3: Summary Statistics: Labor Income Share by Industry in Brazil, 1991

Industry	High School Graduates		High School Dropouts	
	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)
1. Commodity Sector	9.0%	9.6%	21.6%	19.7%
Grains (corn, soybeans, and wheat)	4.7%	11.6%	10.1%	16.3%
Soft (coffee, cocoa, sugar and other)	13.0%	16.1%	19.5%	18.0%
Livestock (cattle, hogs, and others)	35.5%	21.1%	26.8%	15.6%
Metals (copper, lead, steel, zinc, and others)	3.0%	7.2%	1.6%	4.2%
Precious Metals (gold and silver)	1.0%	4.0%	1.8%	4.7%
Energy (crude oil)	8.4%	17.1%	2.3%	6.2%
Other agriculture and mining	34.3%	20.7%	37.9%	19.8%
2. Manufacturing	16.1%	10.7%	18.1%	11.3%
3. Non-Tradable Goods and Services	74.9%	10.8%	60.4%	14.8%

Note. Sample of male white full-time workers extracted from the Brazilian Census of 1991. Statistics weighted by the microregion share in the 1991 national population.

Regional Labor Market Outcomes. To calculate wage outcomes, I estimate wage regressions separately for each year using the entire sample of workers in the country. Specifically, I regress the log monthly wage on a full set of experience dummies (0–39 years) interacted with dummies for female and white workers. The residual of this regression corresponds to a wage measure adjusted for variation in these demographic characteristics across groups and microregions. For each triple of group-region-year, the sector average log wage is the weighted average of the adjusted wage among individuals reporting to hold their main job in the sector. Lastly, the average log wage is the weighted average of the adjusted wage among all individuals in the triple of group-region-year. In both cases, the computation uses Census sampling weights.

Following closely [Autor, Katz and Kearney \(2008\)](#), I use an efficiency-adjusted measure of total hours. With this measure, I compute sectoral employment using the sum of efficiency-adjusted hours supplied by sector employees in a group-region-year. I perform the efficiency adjustment by multiplying individual weekly hours by a time-invariant measure of relative wage for each cell of sex-race-education-experience.³⁴ I then compute the total sector employment, $H_{g,r,t}^k$, as the weighted sum of efficiency-adjusted hours of individuals reporting to have their main job in the sector. This computation uses the Census sampling weights. Finally, sector employment share is defined as $l_{g,r,t}^k \equiv H_{g,r,t}^k / (H_{g,r,t}^C + H_{g,r,t}^N)$.

To obtain total labor supply, I use the aggregate amount of efficiency-adjusted hours in a group-region-year: $\bar{H}_{g,r,t} \equiv H_{g,r,t}^C + H_{g,r,t}^N$. Lastly, the labor supply of immigrants is computed exactly as above in a restricted sample of individuals identified as non-native residents of each microregion.³⁵

D Empirical Application

D.1 Estimation of Sector Wage per Efficiency Unit

Data. In the structural exercise, the estimation of sector wage per efficiency unit requires initial sector employment and wage growth across percentiles of the wage distribution for each of the 2,072 group-region-period triples. To create this dataset, I compute percentiles of log hourly wage from the sample of individuals in a group-region-period using Census sampling weights. Individuals are distributed across percentile bins according to their log hourly wage. The sector employment share in each percentile bin corresponds to the fraction of efficiency-adjusted hours reported by sector employees in that percentile bin. Additionally, the wage growth in each percentile corresponds to the difference in the average hourly wage of individuals in that percentile bin between two consecutive years. Since

³⁴I consider 48 cells based on two sex groups, two race groups (white and non-white), three educational groups (high school dropouts, high school graduates, and college graduates), and four experience groups (0–9, 10–19, 20–29, and 30–39 years). For each cell, the relative wage is the average hourly wage divided by the average wage of female non-white High School dropouts with 0–9 years of experience. The cell weight is the average relative wage across regions and years (1991, 2000, 2010). In the 1980 Census, weekly hours are only reported in ranges. To compute efficiency-adjusted hours, I assign 45 and 54 weekly hours to individuals reporting, respectively, 40–48 and 49+ hours.

³⁵Given the information in the Census, I can only identify as microregion natives those individuals satisfying one out of two conditions. First, they were born in the same municipality in which they currently live. Second, if they were born in a different municipality, then I also consider microregion natives those that moved into the current municipality from another municipality in the same microregion during the previous ten years.

extreme wage values are more likely to be generated by measurement error, I ignore the wage distribution tails by restricting the estimation to bins between the 6th and the 94th percentiles.

Baseline Specification. In principle, equation (17) can be implemented with any division of individuals into quantiles; in practice, however, this choice entails a tradeoff. On the one hand, a coarse discretization yields a low number of quantiles with potentially little variation in initial sector employment to precisely estimate the wage per efficiency unit. On the other hand, a refined discretization exacerbates measurement error of sector employment in each quantile because of the low number of sampled individuals in each sector. With these considerations in mind, I implement the estimation with 88 percentile bins of 1 p.p. width between the 6th and the 94th percentiles. Below, I show that similar results are obtained with bins of 2 p.p. width.

The implementation of expression (17) allows for a vector of observable variables that vary with the position in the wage distribution. Accordingly, the baseline specification includes the following dummy variables as nonparametric controls: (i) indicator that wage percentile is at the bottom (P6-P30) or middle (P30-P75) of the log-wage distribution; and (ii) indicator that wage percentile is below the federal minimum wage (pre-year and post-year). These dummies capture, for example, differential efficiency gains for workers in distant parts of the wage distribution, and income gains generated by bunching around the minimum wage. In this specification, sectoral wages per efficiency unit are identified from the variation in pre-shock sector employment in small neighborhoods of the log-wage distribution of workers in the same group-region-period.

Results. Table D1 presents the summary statistics of estimated wages per efficiency unit implied by the baseline specification for each of the 2,072 group-region-period triples. Columns (1)–(2) display statistics of the estimated wage per efficiency unit in the commodity sector, $\Delta\omega_{g,r,t}^C$, and columns (3)–(4) of the estimated relative wage per efficiency unit in the non-commodity sector, $\Delta\omega_{g,r,t}^N - \Delta\omega_{g,r,t}^C$. The commodity sector’s wage per efficiency unit presented robust growth in both periods. Between 1991 and 2010, the average increase was 47.0 log-points for HSG and 96.4 log-points for HSD. Simultaneously, the relative wage per efficiency unit in the commodity sector increased sharply. Lastly, column (5) reports the average R^2 of the estimation in the sample of microregions. A large fraction of the variation in wage growth across quantiles of the earnings distribution is captured by equation (17); in the two periods, the average R^2 is above 55% for HSG and 71% for HSD.

To address robustness to implementation choices, Table D2 presents the correlation between the estimates of sectoral wages per efficiency unit implied by different specifications of equation (17) and those implied by the baseline specification. Columns (1)–(2) and (4)–(5) indicate a high correlation between estimates obtained with different control sets. Notice that, when minimum wage controls are omitted, estimated wages per efficiency unit are very similar to those of the baseline specification. This suggests that quantile range controls absorb much of the variation captured in the minimum wage dummies. Columns (3) and (6) attest that the particular choice of bin width has little impact on estimates: the correlation is above .88 between baseline estimates and those obtained with a coarser discretization of 2 p.p. bins.

Table D1: Summary Statistics: Estimated Change in Wage per Efficiency Unit, 1991–2010

	Commodity sector wage per efficiency unit		Commodity sector relative wage per efficiency unit		R^2
	Mean	SD	Mean	SD	Mean
	(1)	(2)	(3)	(4)	(5)
Panel A: High School Graduates					
1991 - 2000	0.320	0.370	0.151	0.347	0.555
2000 - 2010	0.150	0.645	0.306	0.609	0.758
Panel B: High School Dropouts					
1991 - 2000	0.524	0.579	0.364	0.619	0.715
2000 - 2010	0.440	0.579	0.360	0.634	0.830

Note. Sample of 518 microregions in 1991–2000 and 2000–2010. Statistics are weighted by the microregion share in national population in 1991. Baseline estimates based on the discretization of the wage distribution in 88 bins of 1 p.p. width, including indicator dummies of percentile bins below the federal minimum wage (pre and post years); and percentile bins in bottom, middle, or top of the wage distribution (P6-P30 and P30-P75).

Table D2: Estimated Change in Wage per Efficiency Unit, Correlation with Benchmark Specification

	Commodity sector			Non-commodity sector		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: High School Graduates						
Correlation with baseline estimates	0.855	0.973	0.926	0.874	0.969	0.916
Panel B: High School Dropouts						
Correlation with baseline estimates	0.914	0.960	0.886	0.912	0.960	0.893
Baseline Controls						
Percentile below federal minimum wage	Yes	No	Yes	Yes	No	Yes
Percentile in bottom, middle or top	No	Yes	Yes	No	Yes	Yes
Discretization of wage distribution						
Bins of 1 p.p. (N = 88)	Yes	Yes	No	Yes	Yes	No
Bins of 2 p.p. (N = 44)	No	No	Yes	No	No	Yes

Note. Sample of 518 microregions in 1991–2000 and 2000–2010. Statistics are weighted by the microregion share in national population in 1991. Baseline estimates based on the discretization of the wage distribution in 88 bins of 1 p.p. width, including indicator dummies of percentile bins below the federal minimum wage (pre and post years); and percentile bins in bottom, middle, or top of the wage distribution (P6-P30 and P30-P75).

D.2 Reduced-Form Evidence: Sensitivity Analysis

This section investigates the robustness of the reduced-form results reported in Section 4.5.1. To this end, I estimate model (23) with additional periods, additional worker groups, and additional labor market outcomes.

Additional Period. Table D3 estimates the model in the extended sample spanning the entire period of 1980–2010. As argued above, the peculiar economic conditions in Brazil could potentially weaken the connection between domestic and international commodity prices during the 1980s. Yet column (2) indicates very similar responses in terms of commodity sector employment. Differences arise for the response of the commodity sector wage differential in column (8). In this case, the coefficient for HSG falls by 40%, moving towards the lower bound of the baseline confidence interval. For HSD, we obtain a higher and more precise coefficient compared to the nonsignificant coefficient implied by the baseline specification.

Compared to the period of 1991–2010, the 1980s exhibit another important difference: commodity prices experienced strong losses in the decade. Taking advantage of this qualitatively different price behavior, columns (3) and (6) estimate the model with microregion-specific time trends. Such a specification relies exclusively on differential exposure within-microregion across periods. For this reason, it addresses concerns that shock exposure is picking up secular trends in microregions specialized in the commodities with larger price gains in 1991–2010. Although these additional variables absorb much of

Table D3: Exposure to Commodity Price Shocks and Sector Employment and Wages, Additional Period

Dependent Variable:	Change in commodity sector employment share			Change in commodity sector average log wage premium		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: High School Graduates						
Commodity price shock	0.035** (0.010)	0.055** (0.013)	0.063** (0.018)	0.441** (0.108)	0.271* (0.105)	0.256+ (0.138)
R^2	0.413	0.336	0.430	0.217	0.214	0.314
Panel B: High School Dropouts						
Commodity price shock	0.061* (0.029)	0.065** (0.021)	0.061* (0.028)	0.030 (0.155)	0.322** (0.095)	0.402** (0.128)
R^2	0.561	0.498	0.612	0.215	0.236	0.403
Baseline Controls						
Controls in Table 1	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls						
Microregion-specific time trend	No	No	Yes	No	No	Yes
Sample Period						
Baseline sample: 1991-2000 and 2000-2010	Yes	Yes	Yes	Yes	Yes	Yes
Extended sample: including 1980-1991	No	Yes	Yes	No	Yes	Yes

Note. Stacked sample of 518 microregions in baseline sample and 503 microregions in extended sample. All regressions are weighted by the microregion share in national population in 1991. Regressions include macroregion-period dummies and the baseline controls in Table 1. Industry composition measured in the initial period of 1991 for baseline sample and of 1980 for extended sample. Commodity sector size controls in extended sample: share in group labor income of other agriculture and mining industries in the commodity sector. Standard Errors clustered by microregion ** p<0.01, * p<0.05, + p<0.10

the cross-section variation in labor market outcomes, they have little effect on estimated coefficients.³⁶

Additional Worker Groups. In Table D4, I extend the sample to include female and non-white individuals. With this exercise, I evaluate whether these additional worker groups exhibit similar behaviors in the labor market. This possibility is especially relevant given the large changes in gender and race wage gaps in the period (Ferreira, Firpo and Messina, 2014). In columns (2) and (6), I include female white individuals without significant changes in estimated coefficients. The inclusion of non-white male individuals entails a more intricate change in estimated coefficients, as shown in columns (3) and (7). Responses of sector employment and wages became weaker for HSG in Panel A, but the opposite is true for HSD in Panel B. These different estimated responses are likely related to differences between white and non-white individuals in terms of unobservable characteristics driving their sectoral allocation. This is particularly important among HSG because of the extremely low High School graduation rate among non-white individuals in Brazil.

Table D4: Exposure to Commodity Price Shocks and Sector Employment and Wages, Additional Groups

Dependent Variable:	Change in commodity sector employment share				Change in commodity sector average log wage premium			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: High School Graduates								
Commodity price shock	0.035** (0.010)	0.029** (0.008)	0.022* (0.009)	0.017* (0.008)	0.441** (0.108)	0.499** (0.112)	0.318** (0.103)	0.405** (0.103)
R^2	0.413	0.431	0.418	0.452	0.217	0.273	0.257	0.340
Panel B: High School Dropouts								
Commodity price shock	0.061* (0.029)	0.078** (0.030)	0.076* (0.031)	0.085* (0.033)	0.030 (0.155)	0.059 (0.142)	0.168+ (0.095)	0.150+ (0.087)
R^2	0.561	0.594	0.627	0.657	0.215	0.259	0.281	0.313
Baseline Controls								
Controls in Table 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Worker Groups								
Baseline: Male / White	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Including female	No	Yes	No	Yes	No	Yes	No	Yes
Including non-white	No	No	Yes	Yes	No	No	Yes	Yes

Note. Stacked sample of 518 microregions in 1991–2000 and 2000–2010. All regressions are weighted by the microregion share in national population in 1991. All regressions include macroregion-period dummies and the baseline controls in Table 1. Standard Errors clustered by microregion ** p<0.01, * p<0.05, + p<0.10

³⁶In omitted exercises, I have estimated the model with microregion fixed effects in the baseline sample of 1991–2010. In this case, standard errors become five to ten times higher. This increase is related to the high correlation in shock exposure between 1991–2000 and 2000–2010 — the autocorrelation of shock exposure is 0.734. Consequently, there is little within-microregion exposure variation to precisely estimate the coefficient of interest. When the 1980–1991 period is included, there is a significant increase in exposure variation within microregions, leading to the more precise results in Table D3.

Additional Labor Market Outcomes. Table D5 investigates the effect of shock exposure on the total quantity of hours supplied by workers in a microregion. Such a response is potentially related to changes in the labor supply of native workers and/or changes in the labor supply of immigrant workers. For HSG and HSD, shock exposure presents a small and statistically nonsignificant relation with the total labor supply of both native and non-native workers. This result is consistent with the assumptions required for identification of comparative and absolute advantage: following the commodity price shock, it is unlikely that a market experienced changes in the productivity distribution due to inflow of new workers from either the home sector or other regions.

Table D5: Exposure to Commodity Price Shocks and Total Labor Supply

	Change in Log of Total Labor Supply	Change in Log of Immigrants' Labor Supply
	(1)	(2)
Panel A: High School Graduates		
Commodity price shock	0.110 (0.141)	0.082 (0.161)
R^2	0.798	0.738
Panel B: High School Dropouts		
Commodity price shock	0.180 (0.151)	0.120 (0.191)
R^2	0.864	0.852
Baseline Controls		
Controls in Table 1	Yes	Yes

Note. Stacked sample of 518 microregions in 1991–2000 and 2000–2010. All regressions are weighted by the microregion share in national population in 1991. All regressions include macroregion-period dummies, initial relative group size, and the baseline controls in Table 1. Standard Errors clustered by microregion ** p<0.01, * p<0.05, + p<0.10

D.3 Structural Estimation: Sensitivity Analysis

This section investigates the robustness of the estimates of the structural parameters reported in Section 4.5.2. To this end, I present results obtained with alternative estimators, alternative specifications of the GMM estimator, and alternative estimates of sectoral wage per efficiency unit.

Alternative Estimator. Table D6 investigates the robustness of results to the particular choice of estimator. Column (1) replicates the baseline specification obtained from the estimation of equations (19)–(21), with the Two-Step GMM estimator and the full vector of disaggregated exposure to price shocks. The remaining columns present the estimates of α_g and A_g obtained, respectively, from the separate esti-

mation of equations (19) and (20). This procedure is less efficient than the baseline specifications, since it does not use the full structure of the model. That is, the estimation equation-by-equation ignores the overidentification restriction provided by the response in the commodity sector's average wage in (21). Nevertheless, this estimator clearly delineates the source of variation driving the structural estimates.

Column (2) shows that the OLS estimation of these equations yields very different results. In this case, OLS is a biased estimator of the structural parameters because supply shocks generate endogenous responses in sectoral wage per efficiency unit and sector employment composition. For the parameter of comparative advantage, the difference in results between columns (1) and (2) has the expected sign. In equation (19), a positive shock in workers' comparative advantage in the commodity sector is equivalent to a negative shock to the relative supply of labor in the non-commodity sector, giving rise to a negative bias in the OLS estimator. The sign of the bias in the absolute advantage parameter is less clear, because it depends on the pattern of selection into the two sectors.

Column (3) presents the 2SLS estimation equation-by-equation using the same set of excluded variables of the baseline specification. This estimator yields point estimates that are similar to the baseline but, as expected, estimates have higher standard errors. Because F-stats are low in column (3), weak instruments are a potential concern that I address in two ways. First, I report the 95% confidence intervals computed by conditional likelihood-ratio (CLR), which are similar to those obtained with the usual asymptotic distribution of the 2SLS estimator. Thus, the qualitative selection patterns inferred from the structural parameters are robust to weak instruments. Second, I estimate the same equations where, as in the reduced-form regressions, the unique instrument is the aggregate exposure to commodity price shocks. In this case, the model is just-identified, and the 2SLS is "unbiased." Column (4) shows that this procedure yields similar point estimates, but standard errors are even higher — especially for α_{HSD} that entails a low F-stat. In general, columns (3) and (4) indicate that these restricted estimators yield similar estimated structural parameters as those obtained with the baseline specification in column (1). In this sense, the joint estimation of equations (19)–(21) by GMM provides efficiency gains that translate into more precise estimates.

Table D6: Parameters of Comparative and Absolute Advantage, Alternative Estimator

Estimator:	Baseline - GMM	OLS	2SLS	2SLS
Equations used in estimation:	(19)-(21)	(19)-(20)	(19)-(20)	(19)-(20)
	(1)	(2)	(3)	(4)
Panel A: High School Graduates				
α_{HSG}	0.835** (0.212)	0.087 (0.073)	1.088** (0.234)	1.510** (0.468)
CLR - 95% Confidence Interval	-	-	[0.503, 2.007]	[0.036, 3.303]
F of excluded variables	-	-	1.83	6.62
A_{HSG}	1.966* (0.935)	0.050 (0.155)	2.056* (0.971)	1.714+ (0.985)
CLR - 95% Confidence Interval	-	-	[0.480, 5.861]	[-0.355, 5.602]
F of excluded variables	-	-	2.22	10.38
Panel B: High School Dropouts				
α_{HSD}	0.916* (0.399)	-0.134* (0.059)	1.655+ (0.919)	2.212+ (1.242)
CLR - 95% Confidence Interval	-	-	[0.455, 5.255]	[0.097, 7.081]
F of excluded variables	-	-	1.83	2.78
A_{HSD}	-0.727** (0.142)	-0.442** (0.032)	-0.814** (0.150)	-0.955** (0.293)
CLR - 95% Confidence Interval	-	-	[-1.401,-0.560]	[-1.778,-0.433]
F of excluded variables	-	-	6.63	14.79
Excluded Instruments				
Disaggregated exposure to price shocks	Yes	No	Yes	No
Aggregate exposure to price shocks	No	No	No	Yes

Note. Stacked sample of 518 microregions in 1991–2000 and 2000–2010. All equations are weighted by the microregion share in national population -n 1991 and include the baseline controls in Table 3. Disaggregated exposure to price shocks: quadratic polynomial of regional exposure to world product prices. Aggregated exposure to price shocks: sum of regional exposure to world product prices. Standard Errors clustered by microregion ** p<0.01, * p<0.05, + p<0.10

Alternative Specification. In Table D7, I evaluate the robustness of the structural results to specific choices of the estimation procedure regarding the moment weighting matrix and the set of excluded instruments. Again, column (1) replicates the baseline specification obtained with the Two-Step GMM estimator and the full vector of disaggregated exposure to price shocks.

Columns (2)–(3) estimate the model with alternative moment weighting matrices. Specifically, column (2) imposes that structural errors in the three equations are independent and, in addition, column (3) imposes that structural errors are homoskedastic (i.e., 2SLS weights). Although point estimates are similar, both estimators yield more imprecise estimates. Such a result is expected since these alternative specifications are less efficient under a general structure of error correlation.

In columns (4)–(5), I estimate the model with restricted sets of excluded instruments. The instrument vector in column (3) is restricted to contain only the exposure to commodity price shocks by category. In this case, estimates are similar for HSG, but the comparative advantage parameter for

HSD is lower and less precise. Similar conclusions are obtained when the vector of instruments is further restricted to include only the aggregate exposure to price shocks in Agriculture and Mining.

Table D7: Parameters of Comparative and Absolute Advantage, Alternative Specification

	Baseline	Matrix of Moment Weights		Vector of Excluded Variables	
	(1)	(2)	(3)	(4)	(5)
Panel A: High School Graduates					
α_{HSG}	0.835** (0.212)	0.879** (0.185)	0.997** (0.237)	0.924* (0.359)	1.198** (0.453)
A_{HSG}	1.966* (0.935)	1.759* (0.834)	2.032* (0.978)	1.501+ (0.873)	1.730+ (0.980)
Panel B: High School Dropouts					
α_{HSD}	0.916* (0.399)	1.302* (0.701)	1.475+ (0.879)	0.394 (0.536)	0.538 (0.988)
A_{HSD}	-0.727** (0.142)	-0.640** (0.134)	-0.795** (0.147)	-0.811** (0.190)	-1.072** (0.318)
Matrix of Moment Weights (Optimal weights):					
Two-Step GMM weights	Yes	No	No	Yes	Yes
Independent equations	No	Yes	No	No	No
Independent equations / Homoskedasticity	No	No	Yes	No	No
Excluded Instruments					
Disaggregated exposure to price shocks (linear)	Yes	Yes	Yes	Yes	No
Disaggregated exposure to price shocks (quadratic)	Yes	Yes	Yes	No	No
Aggregated exposure to price shocks	No	No	No	No	Yes

Note. Stacked sample of 518 microregions in 1991–2000 and 2000–2010. All equations are weighted by the microregion share in national population in 1991 and include the baseline controls in Table 3. Disaggregated Excluded Instruments: quadratic polynomial of regional exposure to world product prices. Aggregated exposure to price shocks: sum of regional exposure to world product prices (Agriculture, Mining). Standard Errors clustered by microregion ** p<0.01, * p<0.05, + p<0.10

Alternative Estimates of Wage per Efficiency Unit. The estimation of the structural parameters of comparative and absolute advantage relied on estimated dependent variables — the changes in sector wage per efficiency unit. To address concerns regarding the implementation choices adopted in the estimation of these variables, Table D8 presents structural parameters estimated with alternative measures of the sector wage per efficiency unit. Column (2) shows that the particular choice of bin width has little impact on estimates. A coarser discretization of 2 p.p. bins yields similar point estimates with higher standard errors. These more imprecise results reflect the higher measurement error in dependent variables due to fewer data points used in the estimation of $(\Delta\omega_{g,r,t}^C, \Delta\omega_{g,r,t}^N)$ for each group-region-period. Columns (3) and (4) display results with estimates of $(\Delta\omega_{g,r,t}^C, \Delta\omega_{g,r,t}^N)$ obtained from equation (17) using different control sets. These controls capture potential shocks in labor efficiency across workers in various ranges of income. Column (3) indicates estimated coefficients are

very similar without the minimum wage controls. This similarity reflects the high correlation shown in Table D2. However, column (4) shows that the percentile range controls are important in the estimation of structural parameters of absolute advantage.

Table D8: Parameters of Comparative and Absolute Advantage, Alternative Estimates of Wage per Efficiency Unit

	Baseline	Alternative estimates of wage per efficiency unit		
	(1)	(2)	(3)	(4)
Panel A: High School Graduates				
α_{HSG}	0.835** (0.212)	0.650 (0.438)	0.658** (0.359)	1.576** (0.489)
A_{HSG}	1.966* (0.935)	1.552* (0.962)	2.020* (0.951)	0.281 (0.506)
Panel B: High School Dropouts				
α_{HSD}	0.916* (0.399)	0.950* (0.495)	1.515** (0.504)	0.688 (0.568)
A_{HSD}	-0.727** (0.142)	-0.623** (0.155)	-0.716** (0.161)	-0.321* (0.171)
Baseline Controls				
Percentile below federal minimum wage	Yes	Yes	No	Yes
Percentile in bottom, middle or top	Yes	Yes	Yes	No
Discretization of wage distribution				
Bins of 1 p.p. (N = 88)	Yes	No	Yes	Yes
Bins of 2 p.p. (N = 44)	No	Yes	No	No

Note. Stacked sample of 518 microregions in 1991–2000 and 2000–2010. Two-Step GMM estimator with microregions weighted by their share in the 1991 national population. All equations include the baseline controls in Table 3. Excluded instruments: quadratic polynomial of regional exposure to world product prices. Standard Errors clustered by microregion ** p<0.01, * p<0.05, + p<0.10

E Parametric Restrictions on the Distribution of Comparative and Absolute Advantage

This section discusses prominent distributional assumptions that determine the form of $\alpha_g(\cdot)$ and $A_g^k(\cdot)$. To simplify notation, I omit subscripts for groups, regions and years.

E.1 Normal Distribution

Particularly important in the selection literature is the case of log-normally distributed sector-specific productivity (Roy, 1951; Heckman and Sedlacek, 1985; Borjas, 1987; Ohnsorge and Trefler, 2007; and

Mulligan and Rubinstein, 2008). In my model, this is equivalent to assuming that the sector-specific productivity vector is independently drawn from a bivariate log-normal distribution:

$$\left(\ln L^C(i), \ln L^N(i) \right) \sim \mathcal{N} \left(\begin{bmatrix} \mu_C \\ \mu_N \end{bmatrix}; \begin{bmatrix} \sigma_C^2 & \sigma_{CN} \\ \sigma_{CN} & \sigma_N^2 \end{bmatrix} \right).$$

Because the comparative advantage of individual i is defined as $s(i) = \ln L^C(i) - \ln L^N(i)$, it is straight forward to conclude that $s(i) \sim \mathcal{N}(\mu, \sigma^2)$ where $\mu \equiv \mu_C - \mu_N$ and $\sigma^2 \equiv \sigma_C^2 + \sigma_N^2 - 2\sigma_{CN}$. Thus, $(s(i), a(i))$ is jointly normal with covariance of $\text{Cov}(s(i), a(i)) = \sigma_{CN} - \sigma_N^2$ and the distribution of $a(i)$ conditional on $s(i) = s$ is normal with conditional mean given by

$$E[a(i)|s(i) = s] = \bar{\mu} + \rho \cdot s \quad \text{s.t.} \quad \bar{\mu} \equiv (1 + \rho)\mu_N - \rho\mu_C, \quad \rho \equiv \frac{\sigma_{CN} - \sigma_N^2}{\sigma_C^2 + \sigma_N^2 - 2\sigma_{CN}};$$

and conditional variance given by

$$V[a(i)|s(i) = s] = \sigma_N^2 - \frac{(\sigma_{CN} - \sigma_N^2)^2}{\sigma_C^2 + \sigma_N^2 - 2\sigma_{CN}}.$$

By definition, $F(s) \equiv \Phi\left(\frac{s-\mu}{\sigma}\right)$ where $\Phi(\cdot)$ is the CDF of the standard normal distribution. Thus,

$$\alpha(q) \equiv F^{-1}(q) = \mu + \sigma \cdot \Phi^{-1}(q). \quad (33)$$

Also, notice that

$$\bar{A}^N(l^N) \equiv \frac{1}{l^N} \int_0^{l^N} E[a(i)|s(i) = \alpha_s(q)] dq = \bar{\mu} + \frac{\rho}{l^N} \int_0^{l^N} \alpha(q) dq = (\bar{\mu} + \rho\mu) + \frac{\rho\sigma}{l^N} \int_0^{l^N} \Phi^{-1}(q) dq.$$

Because $\int_0^{l^N} \Phi^{-1}(q) dq = \phi(\Phi^{-1}(l^N))$,

$$\bar{A}^N(l) = \bar{\mu} - (\rho\sigma) \cdot \frac{\phi(\Phi^{-1}(l^N))}{l^N} \quad (34)$$

where $\bar{\mu} \equiv (\bar{\mu} + \rho\mu) = \mu_N$.

For completeness, consider the average efficiency in the commodity sector:

$$\bar{A}^C(l^N) \equiv \frac{1}{1-l} \int_l^1 \alpha(q) + E[a(i)|s(i) = \alpha(q)] dq = (\mu + \bar{\mu}) + \sigma(1 + \rho) \cdot \frac{\phi(\Phi^{-1}(l^N))}{1-l^N}.$$

Equations (33)-(34) illustrate the connection between the parameters governing the productivity distribution and the schedules of comparative and absolute advantage. First, the dispersion of comparative advantage, σ , controls the magnitude of the between-sector reallocation of individuals in response to changes in the relative wage per efficiency unit. Second, the sensitivity of the mean absolute advantage to the comparative advantage, ρ , controls the compositional effect of employment on sector average wage.

E.2 Extreme Value Distribution

Recent papers have adopted a productivity distribution of the Fréchet family (Hsieh, Hurst, Jones and Klenow, 2013; Burstein, Morales and Vogel, 2015; Galle, Rodriguez-Clare and Yi, 2015). The main advantage of this distribution is its tractability in the multi-dimensional problem of sectoral choice, allowing for an analytical characterization of the equilibrium with an arbitrary number of sectors. As discussed below, this tractability comes at a price: it imposes a restrictive pattern of selection across sectors.

Specifically, assume that sector-specific productivity is independently drawn from a Fréchet distribution:

$$\left(L^C(i), L^N(i) \right) \sim \exp \left[- \sum_{k=C,N} \left(L^k \right)^{-\kappa} \right]$$

where I assume that $\kappa > 1$ to guarantee finiteness of first-order moments.

First, consider the distribution of comparative advantage:

$$F(s) \equiv \Pr [s(i) < s] = \int_{-\infty}^{\infty} e^{-e^{-\kappa(a+s)}} \kappa e^{-\kappa a} e^{-e^{-\kappa a}} da = \int_{-\infty}^{\infty} \kappa e^{-\kappa a} e^{-(1+e^{-\kappa s})e^{-\kappa a}} da.$$

Define $x \equiv (1 + e^{-\kappa s})e^{-\kappa a}$ such that $dx = -\kappa(1 + e^{-\kappa s})e^{-\kappa a} da$. Thus, $F(s) = \frac{1}{1+e^{-\kappa s}}$ and, therefore,

$$\alpha(q) \equiv F^{-1}(q) = \frac{1}{\kappa} \ln \left(\frac{q}{1-q} \right). \quad (35)$$

Second, consider the joint distribution of absolute and comparative advantage:

$$\Pr [a(i) < \bar{a}; s(i) < s] = \int_{-\infty}^{\bar{a}} \kappa e^{-\kappa a} e^{-(1+e^{-\kappa s})e^{-\kappa a}} da = \frac{1}{1+e^{-\kappa s}} e^{-(1+e^{-\kappa s})e^{-\kappa \bar{a}}}.$$

To obtain the average efficiency, notice that the productivity distribution in the non-commodity sector is

$$\Pr \left[a(i) < a | s(i) < \alpha \left(l^N \right) \right] = e^{-\left(1 + e^{-\kappa \alpha \left(l^N \right)} \right) e^{-\kappa a}} = e^{-\frac{1}{l^N} e^{-\kappa a}} = e^{-e^{-\kappa \left(a + \frac{1}{\kappa} \ln l^N \right)}}$$

where the second equality follows from the definition of $\alpha(\cdot)$.

Since this is a Gumbel distribution with parameters $\beta \equiv 1/\kappa$ and $\mu \equiv -\frac{1}{\kappa} \ln l^N$, the average efficiency in the non-commodity sector is

$$\bar{A}^N(l) = \frac{\gamma}{\kappa} - \frac{1}{\kappa} \ln l^N$$

where γ is the Euler-Mascheroni constant. From this expression, we obtain

$$A(q) = \frac{\gamma - 1}{\kappa} - \frac{1}{\kappa} \ln q. \quad (36)$$

Analogously, the productivity distribution in the commodity sector is

$$Pr \left[\ln T^C(i) < a^C | s(i) > \alpha \left(l^N \right) \right] = e^{-e^{-\kappa \left(a^C + \frac{1}{\kappa} \ln(1-l^N) \right)}} \Rightarrow \bar{A}^C \left(l^N \right) = \frac{\gamma}{\kappa} - \frac{1}{\kappa} \ln \left(1 - l^N \right).$$

The schedules of comparative and absolute advantage in equations (35)-(36) are fully characterized by the dispersion parameter, κ . If productivity dispersion is low (i.e., κ is high), then a small variation in the relative wage per efficiency unit is associated with a large response of sector employment. In addition, a sector employment expansion causes a decrease in the average sector efficiency whose magnitude is also controlled by the productivity dispersion. In other words, the extreme value distribution only allows for positive selection in both sectors. This very particular pattern of selection has strong implications for the log-wage distribution, implying that both sectors exhibit the same distribution of labor earnings. Specifically, the log-wage distribution in sector k is

$$G^N(y) = e^{-e^{-\kappa \left(y - \omega^N + \frac{1}{\kappa} \ln l^N \right)}} = e^{-e^{-\kappa \left(y - \omega^C + \frac{1}{\kappa} \ln(1-l^N) \right)}} = G^C(y)$$

where the second equality follows from the employment equation in (7).

Finally, the log-wage distribution belongs to the Gumber family and, therefore, the log-wage variance is given by $\pi^2/6\kappa$. Thus, this distributional assumption implies that the dispersion of log wages in a demographic group is constant.

E.3 Log-Linear System: An Example

In this section, I describe a distribution that delivers the log-linear functional forms in Assumption 5. To guarantee finite supply of effective labor units for all parameters, assume that the quantile function of comparative advantage is bounded with the following form:

$$\alpha(q) = \begin{cases} \underline{\alpha} & \text{if } 0 \leq q < \varepsilon \\ \alpha \ln [q/(1-q)] & \text{if } \varepsilon \leq q < 1 - \varepsilon \\ \bar{\alpha} & \text{if } 1 - \varepsilon \leq q \leq 1 \end{cases}$$

where $\varepsilon \geq 0$, $\bar{\alpha} \equiv \alpha \frac{\varepsilon-1}{\varepsilon} \ln(1-\varepsilon) - \alpha \ln(\varepsilon)$, and $\underline{\alpha} \equiv \alpha \ln[\varepsilon/(1-\varepsilon)]$.

Although the comparative advantage distribution has finite moments for every $\varepsilon \geq 0$ and $\alpha > 0$, this is not necessarily true for its moment generating function. Accordingly, the upper bound in the support implies a well defined moment generating function for all $\varepsilon > 0$ and, therefore, a finite supply of effective labor units. For ε arbitrarily small, there is positive employment in both sectors and the empirically relevant portion of the quantile function is that presented in Assumption 5.

Also, assume that the conditional distribution of absolute advantage is normal with a linear conditional mean:

$$\{a(i) | \tilde{s}(i) = \alpha(q)\} \sim \mathcal{N}(A_g(q), \sigma^2) \text{ where } A(q) \equiv \begin{cases} \underline{A} & \text{if } 0 \leq q \leq \varepsilon \\ \bar{A} + A \ln q & \text{if } \varepsilon < q \leq 1 \end{cases}$$

with $A \in \mathbb{R}$, and $\underline{A} \equiv (\bar{A} - A) + A \ln \varepsilon$.

Thus,

$$\bar{A}^N(l^N) \equiv \frac{1}{l^N} \int_0^{l^N} A(q) dq = A^N + A \cdot \ln l^N$$

where $A^N = (\bar{A} - A)$.

By assuming that $\varepsilon < l_g^N < 1 - \varepsilon$,

$$\bar{A}^C(l^N) \equiv \frac{1}{1 - l^N} \int_{l^N}^1 \alpha(q) + A(q) dq = A^C - (\alpha + A) \cdot \frac{l^N}{1 - l^N} \ln l^N - \alpha \cdot \ln(1 - l^N).$$

where $A^C \equiv (\bar{A} - A)$.