WORKING PAPERS

Industrialization without Innovation

BSE Working Paper 1444| April 2024

Paula Bustos, Juanma Vincenzi, Joan Monràs, Jacopo Ponticelli

bse.eu/research

Industrialization without Innovation

Paula Bustos
ICREA, UPF, IPEG, BSE and CEPR
Juanma Castro-Vincenzi
Princeton
Joan Monras
Federal Reserve Bank of San Francisco, UPF, CREI, BSE, and CEPR
Jacopo Ponticelli*
Northwestern, NBER and CEPR

September 2023

Abstract

Labor-saving technologies in agriculture can foster structural transformation by releasing workers who find jobs in manufacturing. The traditional view is that factor reallocation towards manufacturing generates innovation and productivity growth. We document, instead, that regions more exposed to a large and exogenous increase in agricultural productivity in Brazil industrialized but experienced lower manufacturing productivity growth. Workers released from agriculture were mostly unskilled and primarily moved to the least skill-intensive manufacturing industries. This paper explores the various mechanisms that can account for the observed manufacturing productivity decline. Changes in worker composition and lower incentives to innovate within manufacturing play prominent roles.

Keywords: Agricultural Productivity, Skill-Biased Technical Change, Productivity, Labor Mobility, Genetically Engineered Soy, Brazil.

JEL codes: F16, J43, O13, O14, O33, O41.

*We received valuable comments from Thibault Fally, Manuel García-Santana, Donald Davis, Gene Grossman, Michael Peters, Diego Puga, Sergio Rebelo, Andres Rodriguez-Clare, Chris Tonetti, Chris Udry, Jose P. Vasquez, and seminar participants at CEMFI, CREI, Lugano, Columbia, Northwestern, BdE, the NBER Economic Consequences of Trade, McGill, Princeton Growth Conference, GEP/CEPR Summer Trade Workshop, and the Workshop on Structural Transformation and Macroeconomic Dynamics. This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 716388). We are also grateful to acknowledge financial support from the UK Department for International Development (DFID) and the Centre for Economic Policy Research (CEPR) (PEDL Scale-Up ERG no. 3787). Monras acknowledges financial support from the Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S). Mark He and Erick Molina provided excellent research assistance. An earlier version of this paper circulated under the title "Structural Transformation, Industrial Specialization, and Endogenous Growth". The views in this paper do not necessarily reflect those of the Federal Reserve System.

1 Introduction

Early development economists argued that the reallocation of workers from agriculture to manufacturing was fundamental to sustain long run growth (Lewis 1954, Kuznets 1973). This structural transformation process can lead to higher output because labor productivity is lower in agriculture than in the rest of the economy (McMillan and Rodrik, 2011). In addition, industrialization can lead to higher long run growth if the manufacturing sector is characterized by economies of scale and knowledge spillovers (Krugman 1987, Lucas 1988, Matsuyama 1992a).[1] However, manufacturing productivity growth depends not only on the size of the industrial sector but also on the cost of innovation (Grossman and Helpman 1991). Thus, if workers leaving the agricultural sector are mostly unskilled, the structural transformation process can lead to an increase in the skill premium, decreasing the incentives to innovate and, ultimately, slowing down productivity growth in manufacturing (**??**).

We provide direct evidence on these forces in the context of a large increase in unskilled manufacturing labor following the skill-biased and labor-saving adoption of genetically engineered (GE) soy in Brazil. For identification, we exploit the heterogeneous effects on potential yields across regions with different weather and soil characteristics.[2] Given that scale effects concentrate exclusively in low-innovation industries, we can disentangle the scale and cost of innovation channels by comparing high- and low-innovation intensive sectors within these regions. For our empirical exploration, we use firm-level data from the Brazilian Annual Industrial Survey (PIA) to measure productivity in the manufacturing sector, and the Technological Innovation Survey (PINTEC) to sort industries by innovation intensity. We also rely on detailed individual information from the Brazilian Population Census and social security data (RAIS), which allows to follow workers across sectors at fine levels of spatial aggregation. We also use the RAIS data to develop a new measure of innovation expenditures based on detailed descriptions of occupations.

We start our empirical investigation by showing that microregions facing faster agricultural technical change experienced a persistent *slowdown* in manufacturing productivity growth. This is despite the fact that the manufacturing sector in these regions absorbed workers released from agriculture (as documented in Bustos, Caprettini, and Ponticelli 2016) and attracted more capital investments. However, investments in innovation did not follow. This result contrasts to the standard theoretical argument that moving resources toward manufacturing should lead to scale economies that should encourage innovation and increase productivity, and point to the fact that keeping track of the cost of inno-

[1] Recent empirical studies analyzing these mechanisms include McCaig and Pavcnik (2013), Atkin, Costinot, and Fukui (2021), Goldberg and Reed (2020), Choi and Levchenko (2021).

[2] Our geographical units of observation are Brazilian microregions, which attempt to approximate local labor markets. The Brazilian Institute of Geography and Statistics (IBGE) defines these microregions by combining economically integrated municipalities with similar production and geographic characteristics.

vation is crucial for the development path. Our estimates imply that microregions with a one standard deviation larger increase in potential soy yields experienced a 7 percent *increase* in the size of the manufacturing sector and a corresponding 1.5 percent *lower* yearly growth rate of manufacturing productivity.

To better understand this result, we first investigate the nature of the labor reallocation from agriculture toward manufacturing. For this, we trace the flow of workers with different education levels across sectors using detailed individual information from the decadal Brazilian Population Census and social security data from RAIS. We find that the new agricultural technology led to a reallocation of mostly unskilled workers away from agriculture towards manufacturing with little reallocation towards services. Our estimates indicate that microregions with a one standard deviation higher increase in potential soy yields experienced a 2.4 percentage points larger decrease in the share of unskilled workers employed in agriculture, and a corresponding 2.1 percentage points larger increase in the share of unskilled workers employed in manufacturing. We confirm these findings using yearly information on formal workers from RAIS, which shows that the timing of this labor reallocation process corresponds to the timing of adoption of GE seeds.

From the point of view of the manufacturing sector, the reallocation of former agricultural workers amounts to an increase in the relative supply of unskilled labor. According to the classic Heckscher-Ohlin trade model, this increase in the relative supply of unskilled labor generates a comparative advantage in unskilled-labor intensive industries, which should expand by absorbing the inflow of unskilled workers and also attract other complementary factors such as capital (Rybczynski, 1955). Indeed, we find that this inflow of unskilled workers was completely absorbed by an expansion of the least skill-intensive manufacturing industries. As expected, the labor inflow was followed by an increase in capital investment. Finally, we document that the expanding industries are the least innovation-intensive as measured by expenditure in research and development (R&D) as a share of sales.

Given these empirical results, we investigate various mechanisms that could explain why an inflow of workers and capital towards manufacturing may slow down its productivity. We start by investigating whether this slowdown is simply due to a change in the composition of industries. In particular, we split the manufacturing sector by the median level of R&D as a share of sales, and label H industries the most innovation intensive manufacturing sectors (e.g. production of equipment for industrial automation, medical equipment or pharmaceutical products), and L industries the least innovation intensive manufacturing sectors (e.g. meat processing or tobacco products). Then, we decompose the manufacturing productivity slowdown into between- and within-industry components. We obtain that the between effect, which reflects the impact of factor reallocation towards low R&D industries, can explain at most 8 percent of the overall productivity decline.

The remainder is explained by a reduction of productivity growth within industries, which occurs in both H and L industries and it is larger in L industries.[3] This evidence suggests that the main driver of the productivity slowdown cannot be attributed mechanically to a change in industrial composition.

Next, we investigate whether the decline in manufacturing productivity is driven by a change in the composition of workers within manufacturing. In particular, former agricultural workers may lack the skills needed to thrive in the manufacturing sector, and thus be less productive than workers already employed in manufacturing, at least temporarily. This, in turn, could lower the productivity of the sectors where they enter. We explore this channel using individual-level data from RAIS. RAIS data shows that formal agricultural workers entering L manufacturing industries earn on average around 27% lower wages than workers already employed in the L industries – 12% lower wages when controlling for observable individual characteristics such as age and education. Using wage differences as a measure of differences in worker productivity we estimate that changes in worker composition can explain at most between 25 and 42 percent of the observed productivity slowdown in L industries, depending on the measure of productivity.[4]

While worker composition can explain part of the productivity decline in the L industries, it cannot explain why such decline occurs also in H industries, which do not absorb former agricultural workers. To understand this result, we investigate whether an increase in the relative size of the low-skill, low-innovation intensive manufacturing industries reduces the incentives to innovate within the high-skill, high-innovation intensive industries, as predicted by models of directed technical change based on market size effects à la Romer (1990), such as Acemoglu (2002). Testing this hypothesis requires a measure of innovation at fine levels of spatial aggregation. Standard innovation surveys such as PINTEC do not allow us to implement this strategy, because they are based on a sample of firms which is not representative of small geographical units (microregions).[5] Thus, we propose a new measure of innovation that is representative at any level of geographical aggregation because it can be constructed using social security data, which covers the universe of formal firms. In particular, we measure the labor input in the production of innovations using textual analysis of the task descriptions of more than 2,500 occupations in RAIS. Tasks generating innovations include, for example, developing or adapting new

[3]We use three different measures of manufacturing productivity: value added per worker, valued added per wage bill, and a measure of TFP. The measure of TFP takes into account changes in the capital stock and educational level of the workforce. In turn, the value added per wage bill measure takes into account both observed and unobserved dimensions of human capital as reflected in wages.

[4]These results are in line with wage changes among informal workers in the L industry observed across Censuses (31%), suggesting that lower wages paid to former agricultural workers are not limited to the selected sample of formal workers observed in RAIS.

[5]Alternative measures of innovation such as patents might be geographically representative but are not representative of the type of innovations which are most frequent in developing countries. According to PINTEC, only 20% of firms which introduced innovations in the period 1997-2008 filed a patent application.

products and processes, creating prototypes, or optimizing methods of production.

Using this measure, we document that in regions more exposed to agricultural technical change, the inflow of low-skill agricultural workers into L manufacturing industries was followed by lower investment in innovation in H industries. In particular, microregions with a one standard deviation larger increase in potential soy yields experienced a 20 percent larger decline in innovation expenditures in H industries, measured as the wage bill of workers in innovative occupations. We show that this decline is explained by both lower retention and lower entry of workers in innovative occupations. Next, we trace the employment path of innovation workers initially employed in H industries, and find that a quarter of lower retention is explained by reallocation to L industries. Finally, we show that a third of those former innovative workers end up in non-innovative occupations within L industries. Indeed, we do not find a significant increase in innovation expenditures within L industries. The decline in local innovation activity can rationalize at least part of the decline in manufacturing productivity growth that accompanied the industrialization process in regions subject to faster agricultural technical change.

Our findings suggest that different forces driving structural transformation can lead to different types of industrial development. In most countries, the process of labor reallocation from agriculture to manufacturing can be ascribed to one of two forces: "push" forces, such as new agricultural technologies that push workers out of agriculture, or "pull" forces, such as industrial productivity growth, that pull workers into manufacturing. We show that when labor reallocation from agriculture to manufacturing is driven by "push" forces that affect disproportionately the least skilled workers, it can generate an expansion of industries with the lowest potential contribution to aggregate productivity.

In this sense, our results are a cautionary tale on the effects of structural change on productivity growth. The adoption of new technologies in agriculture may result in static productivity gains in the agricultural sector but dynamic losses in manufacturing productivity. We think that the experience of Brazil documented in this paper is informative for low- to middle-income countries where a large share of the labor force is employed in agriculture, and who import new agricultural technologies from more developed countries.

Related Literature

Our paper is related to the classic literature arguing that labor reallocation from agriculture to manufacturing increases aggregate productivity. A first set of papers in this literature argues that, because of the sizable productivity and wage gaps between agriculture and manufacturing in developing countries, there are large static productivity gains when countries industrialize (e.g., Caselli 2005, Restuccia, Yang, and Zhu 2008, Lagakos and Waugh 2013, Gollin, Lagakos, and Waugh 2014).[6] Second, there can also be dy-

[6]More recently, Herrendorf and Schoellman (2018) measure and compare agricultural wage gaps in countries in different stages of the structural transformation process. They find that the implied barriers to labor reallocation from agriculture are smaller than usually thought in the macro-development literature,

namic productivity gains when labor reallocates towards manufacturing if manufacturing is subject to agglomeration externalities and knowledge spillovers (Krugman 1987, Lucas 1988, Matsuyama 1992a).[7] Our paper contributes to this literature by empirically showing how the impact of structural transformation on manufacturing productivity growth depends on the nature of the structural transformation process and how this affects industrial development. To make this point, we build on insights from the endogenous growth literature. In particular, the seminal work of Grossman and Helpman (1991) who study open economy endogenous growth models and show how growth may depend on industrial specialization.[8]

Recent empirical studies exploring the theoretical mechanisms highlighted above include Franck and Galor (2019) who use historical variation in the diffusion of steam engines to argue that early adoption of technologies that do not lead to human capital development hinders long-run growth. In a similar vein, McMillan and Rodrik (2011) and McMillan, Rodrik, and Sepulveda (2017) emphasize that structural change can be growth-enhancing or growth-reducing depending on the correlation between changes in employment shares and productivity levels, comparing industrialization experiences across a number of countries. Recent studies of the effects of structural transformation on productivity include McCaig and Pavcnik (2013) who document how trade-induced labor reallocation from informal to formal manufacturing increased productivity in Vietnam; and Imbert, Seror, Zhang, and Zylberberg (2020) who exploit short-run agricultural shocks in China to document how migration from rural to urban areas reduces labor costs for firms leading to an expansion in labor usage and a reduction in capital-biased technology adoption. Relative to this work, we provide evidence on how labor-saving agricultural productivity shocks can affect the patterns of industrial specialization and manufacturing productivity growth, and document the relative importance of potential mechanisms using social security data which permits to track workers across sectors and occupations.

Our paper also builds on the empirical literature studying the effects of agricultural technical change, particularly the papers that argue that technological advancements in agriculture are skill-biased. For instance, Foster and Rosenzweig (1996) study the effects of the introduction of high-yield varieties in India, and show that technological

---

and argue that labor heterogeneity and selection are important drivers of such gaps.

[7]Recent evidence suggests that this channel may be operative in some circumstances. Peters (2019) uses the displacement of Eastern Germans towards Western Germany to show that places experiencing larger population growth specialized in manufacturing and saw GDP per capita grow over the long run. This channel, however, may not be the norm. Herrendorf, Rogerson, and Valentinyi (2022) find limited evidence in support of dynamic productivity gains from labor reallocation from agriculture into manufacturing using cross-country data for the period 1990-2018. See also De Vries, Timmer, and De Vries (2015) which proposes a decomposition of the effect of structural change on the productivity growth of Sub-Saharan African countries between static and dynamic components.

[8]Some of our findings can be rationalized with a small open economy growth model with three sectors – agriculture, low- and high-skill manufacturing, and two factors, where R&D investment depends on the relative size of the high-skill industries. In this sense, our paper is also related to the classical literature on endogenous growth, which was mainly theoretical, see Romer (1990); Aghion and Howitt (2008).

innovations in agriculture increased the relative demand for skill in agriculture and, thus, returns to primary schooling.[9] We contribute to this literature by showing that the recent introduction of GE soy was also skill-biased. More importantly, we focus on the implications of skill-biased agricultural technical change for industrialization, which have not previously been explored.

Finally, it is worth noting that this paper is part of a broader research agenda that studies the effects of agricultural productivity on structural transformation in the context of the adoption of GE crops in Brazil. A first study in this agenda, Bustos et al. (2016) shows that, if agricultural technical change is labor-saving, increases in agricultural productivity can lead to a reallocation of labor towards the manufacturing sector, even in open economies.[10] Relative to Bustos et al. (2016), our paper studies how the reallocation of workers from agriculture to manufacturing can shape the growth prospects of the industrial sector. A second study in this agenda analyzes the effects of the agricultural boom in Brazil on capital markets. Bustos, Garber, and Ponticelli (2020) document that regions with faster technical change in soy experienced an increase in local savings deposits which were mostly not lent locally, leading to an increase in capital outflows. Banks redirected agricultural savings to urban areas outside soy-producing regions where they were invested in the manufacturing and service sectors. Those findings describe the effects of agricultural technical change on structural transformation through a capital supply channel and are consistent with a high level of financial integration across regions. In contrast, the current paper documents the effects of agricultural technical change through a labor supply channel which operates in local labor markets with limited migration responses. In section 5.5, we exploit this difference in the levels of labor and capital market integration to separately identify the labor and capital supply channels. Taken together, these results imply that while former agricultural workers reallocate towards local non-innovative industries, agricultural savings foster the expansion of innovative industries located in other regions, accentuating regional productivity inequalities.

[9] In related work, Bragança (2014) shows that investments in soybean adaptation in Central Brazil in the 1970s induced positive selection of labor in agriculture.

[10] As noted by Matsuyama (1992b), the mechanisms through which agricultural productivity can speed up industrial growth proposed in the classical models of structural transformation are only operative in closed economies, while in open economies high agricultural productivity induces a reallocation of labor toward agriculture, the comparative advantage sector. Bustos et al. (2016) proposes a model where the effect of agricultural productivity on industrial development in open economies depends on the factor-bias of technical change.

2 Model

3 Empirical strategy and data

Our empirical strategy aims at identifying the effects of one particular "push" factor of structural transformation: the introduction of a new labor-saving technology in agriculture. For this, we exploit the legalization of genetically engineered (GE) soy in Brazil as a natural experiment. We start by providing background information on GE soy in section 3.1. Notice that an increase in the reallocation of labor from agriculture to manufacturing in areas that adopted GE soy could be driven by a shock to labor demand in the local manufacturing sector. This would increase local wages, inducing agricultural firms to switch to less labor-intensive crops, such as soy. Thus, to establish the direction of causality, our identification strategy uses the potential increase in soy yields that can be obtained with GE seeds in each region based on its weather and soil characteristics as a plausibly exogenous measure of technical change in agriculture. We describe this strategy in detail, along with the data used to implement it, in sections 3.2 and 3.3. Finally, in section 3.4, we introduce a new measure of innovation at the microregion level that we use to study the impact of agricultural productivity on innovative activities.

3.1 Background Information on GE Soy

The purpose of GE soy seeds is to resist a specific herbicide (glyphosate). The use of these seeds allows farmers to spray their fields with glyphosate without harming soy plants, reducing labor requirements for weed control.[11] For example, the planting of traditional seeds is preceded by soil preparation in the form of tillage, the operation of removing the weeds in the seedbed that would otherwise crowd out the crop or compete with it for water and nutrients. In contrast, planting GE soy seeds requires no tillage, as the application of herbicide selectively eliminates all unwanted weeds without harming the crop. Because activities related to weed control are mostly performed by unskilled workers, the introduction of GE soy seeds tends to displace unskilled labor relatively more than skilled labor.

The first generation of GE soy seeds (Monsanto's Roundup Ready) was commercially released in the U.S. in 1996 and legalized in Brazil in 2003.[12] Prior to 2003, smuggling of GE soy seeds from Argentina was only detected in 2001 and 2002 according to the Foreign Agricultural Service of the United States Department of Agriculture (USDA, 2001). The 2006 Brazilian Agricultural Census reports that, only three years after their legalization,

[11] Other advantages of GE soy seeds are that they require fewer herbicide applications (Duffy and Smith 2001; Fernandez-Cornejo, Klotz-Ingram, and Jans 2002), allow a higher density of the crop on the field (Huggins and Reganold 2008) and reduce the time between cultivation and harvest.

[12] See Law 10.688 of 2003 and Law 11.105 – the New Bio-Safety Law – of 2005 (art. 35).

46.4% of Brazilian farmers producing soy were using GE seeds with the "objective of reducing production costs" (IBGE 2006, p.144). According to the Foreign Agricultural Service of the USDA, by the 2011-2012 harvesting season, GE soy seeds covered 85% of the area planted with soy in Brazil (USDA 2012).[13]

Panel (a) of Figure 1 documents that the legalization of GE soy seeds was followed by a fast expansion of the area planted with soy, which increased from 11 to 19 million hectares between 2000 and 2010.[14] This graph suggests that the area planted with soy started to increase very rapidly already in 2002. Panel (b) of Figure 1 documents that, in the same period, the number of workers employed in the soy sector decreased substantially. This finding is consistent with the adoption of GE seeds reducing the number of agricultural workers per hectare required to cultivate soy. Bustos et al. (2016) document that labor intensity in soy production fell from 28.6 workers per 1000 hectares in 1996 to 17.1 workers per 1000 hectares in 2006. In addition, the production of soy is less labor-intensive than all other major agricultural activities. According to the Agricultural Census, the average labor intensity of cereals in 2006 was 94.9 workers per 1,000 hectares, 129.8 for other seasonal crops, and 126.7 for permanent crops.[15] Thus, whenever soy displaced other agricultural activities, labor intensity in agriculture decreased.

Figure 1 goes around here

In Panel (c) of Figure 1, we decompose the decrease in employment in the soy sector between skilled and unskilled workers, where workers are considered skilled if they have completed at least the 8^{th} grade. As shown, the decrease in employment in the soy sector is entirely driven by low-skilled workers, while the skilled ones were retained. This finding is consistent with GE soy seeds being an unskilled labor saving technology. Notice that in addition to being less labor intensive, soy production is also more skill intensive than most other agricultural activities. As shown in Panel (d) of Figure 1, the share of skilled workers (those completed at least the 8^{th} grade) employed in soy is above 20 percent, while in most other agricultural activities this share ranges between 5 and 15 percent. Thus, whenever

[13]Note as well that although the initial patent of GE soy seeds was filed in the US by the multinational corporation Monsanto, the final product available in the Brazilian market was the outcome of an adaptation process that involved a Brazilian firm. The year after patenting the Roundup Ready^{TM} (RR) soy seeds in the US in 1996, Monsanto started a collaboration with Embrapa – the Brazilian Research Institute for new agricultural technologies – to develop a version of the GE soy seeds adapted to the agro-ecological conditions of Brazil. Under this agreement, Embrapa started conducting crossings between the herbicide tolerant variety developed by Monsanto for the US market and seeds previously developed by Embrapa itself for the Brazilian climate. Hence, it necessarily took a few years before GE soy seeds adapted to the Brazilian climate were available.

[14]According to the two most recent agricultural censuses, the area planted with soy increased from 9.2 to 15.6 million hectares between 1996 and 2006 (IBGE 2006, p.144).

[15]According to the 2006 Agricultural Census, even cattle ranching uses more workers per unit of land than soy production (30.6 per 1000 hectares).

soy displaced other agricultural activities, the skill intensity of agriculture increased along with the decrease in labor intensity.

3.2 Identification strategy

Our identification strategy builds on Bustos et al. (2016): we exploit the legalization of GE soy seeds in Brazil as a source of time variation and differences in the potential increase in soy yields from the introduction of the new technology across regions as a source of cross-sectional variation. This approach allows us to identify how changes in agricultural technology lead to structural transformation and to study its consequences on local economies.

The potential increase in soy yields due to GE soy seeds is constructed using data on potential soy yields sourced from the FAO-GAEZ database. This dataset reports the maximum attainable yield for a specific crop in a given geographical area. In addition, it reports the maximum attainable yields of each crop under different technologies or input combinations. Yields under the *low* technology are described as those obtained from planting traditional seeds, with no use of chemicals or mechanization. Yields under the *high* technology are obtained using improved high-yielding varieties, with optimum application of fertilizers, herbicides, and mechanization.

Following Bustos et al. (2016), we define technical change in soy production as the difference in potential yields between *high* and *low* technology. This measure aims at capturing the theoretical change in soy yields obtained by switching from traditional soy production to the use of improved seeds and optimum weed control, among other characteristics. Technical change in soy production in microregion k is therefore defined as:

$$\Delta A_k^{soy} = A_k^{soy,High} - A_k^{soy,Low}$$

where $A_k^{soy,Low}$ is equal to the potential soy yield under the *low* technology and $A_k^{soy,High}$ is equal to the potential soy yield under the *high* technology.[16] ΔA_k^{soy} is our exogenous measure of agricultural technical change in agriculture.

Figure A.1 in the Appendix, shows the geographical variation in this measure of technical change across microregions. The map suggests large variation in agricultural technical change across Brazilian microregions. Some regions, most notably the regions around the Amazon river, and near the South-East coast, experienced little changes in soy productivity. Instead, the regions of the Center-West and South gained substantially from the

[16]Although soy farming in certain areas of Brazil was already using relatively advanced techniques before the introduction of GE soybeans, our conversations with researchers in charge of the FAO-GAEZ dataset show that GE soy seeds are, in fact, the improved seed varieties used to compute predicted soy yields for Brazil under *high* inputs. The predictive power of the instrument on GE soy seeds adoption documented in what follows supports this.

introduction of the new seed.

With decennial data, we use the following specification to estimate the effect of soy technical change on (long-run) changes in outcomes of interest:

$$\Delta Y_{k(r)} = \alpha + \beta \Delta A^{soy}_{k(r)} + \varphi X_{k(r)} + \delta_r + \varepsilon_k \tag{1}$$

where $\Delta Y_{k(r)}$ is the change in the outcome of interest in microregion k (located in macroregion r) between 2000 and 2010 – the years of the last two Population Censuses –, and $X_{k(r)}$ is a vector of controls of microregion k. δ_r indicates macroregion fixed effects that account flexibly for trends across the five major geographical regions of the country: North, Northeast, South, Southeast and Central-West. Our identification strategy relies on the fact that the new GE soybeans seeds were introduced around 2001 or 2002 and legalized in Brazil in 2003, and that this new technology disproportionately favored microregions with certain soil and weather characteristics (as captured by $\Delta A^{soy}_{k(r)}$), something that was not anticipated as of 2000. In all our specifications we include the share of rural population, the initial level of income per capita, the alphabetization rate, and population density at the microregion level, all observed in 1991 and sourced from the Population Census, and the measure of maize technical change (discussed further below and presented in Table A.1 of the Appendix). These controls are meant to flexibly capture differential trends across microregions with different initial levels of income and human capital.

When we analyze the manufacturing sector in detail, we use annual data from RAIS and PIA. This allows us to trace the timing of the effect more precisely by estimating two types of equations. First, to provide visual support to our evidence, we estimate the following dynamic difference-in-differences specification:

$$\ln y_{k(r),t} = \delta_t + \delta_k + \delta_{rt} + \sum_{\substack{j=1998 \\ j\neq 2000}}^{2009} \beta_j 1[j=t] \Delta A^{soy}_{k(r)} + \gamma X_{k(r),t} + t \times X'_{k(r),1991}\omega + \varepsilon_{k(r),t} \tag{2}$$

where $\Delta A^{soy}_{k(r)}$ is the long-run change in our exogenous measure of technical change in soy in microregion k, and $\ln y_{k(r),t}$ is an outcome of interest in microregion k at time t. β_j estimates the effect of the change in the productivity of soy in each year between 1998 and 2009 (using 2000 as our reference year). Thus, we flexibly allow β_j to capture the effect of soy technical change on the outcomes of interest in each year. This type of specification is informative of the timing and persistence of the effects. δ_k and δ_t are microregion and year fixed effects, respectively. δ_{rt} are macro-region times year fixed effects. $X_{k(r),t}$ are time-varying controls and $X_{k(r),1991}$ are the baseline controls discussed above, interacted with a time trend.

With annual data, we estimate the effect of agricultural technical change on manufacturing outcomes using the following specification:

$$\ln y_{k(r),t} = \delta_t + \delta_k + \delta_{rt} + \beta A^{soy}_{k(r),t} + \gamma X_{k(r),t} + t \times X'_{k(r),1991}\omega + \varepsilon_{k(r),t} \quad (3)$$

where $A^{soy}_{k(r),t}$ is defined as potential soy yield under high inputs for the years between 2003 and 2009, and the potential soy yield under low inputs for the years between 1999 and 2002 in microregion k. δ_k and δ_t are microregion and year fixed effects, respectively, δ_{rt} are macro-region flexible trends, and $X_{k(r),t}$ are time-varying controls and $X_{k(r),1991}$ are baseline controls interacted with a time trend. Hence, β is the (continuous) difference-in-differences estimate obtained from comparing microregions before and after 2003.[17]

Table A.1 in the Appendix reports a set of results aimed at validating our measure of soy technical change using data from the 1996 and 2006 Agricultural Censuses. First, in Panel A, we show that our measure of soy technical change strongly predicts variation in the actual adoption of GE seeds by Brazilian farmers across microregions (columns 1 and 2). Importantly, it does not predict the expansion of area farmed with traditional soy (columns 3 and 4). This indicates that this measure of the effect of technical change on potential soy yields is a good proxy of the actual benefits of GE soy adoption given soil and weather characteristics of different areas. Second, in Panel B, we show that our measure of soy technical change predicts the expansion of agricultural area farmed with soy, but not the one farmed with maize, the other main temporary crop which experienced significant technological innovation in this period (columns 1 and 2).[18] If we build a measure of maize technical change using the same methodology, we find that such measure predicts the expansion in maize area between 1996 and 2006, but not the expansion of soy area (columns 3 and 4). This indicates that our measure of technical change is a good proxy of technological innovation at the crop level. Note that the results reported in Table A.1 effectively replicate the results presented in Bustos et al. (2016) at a larger level of aggregation (microregion instead of municipality).

3.3 Data sources

In this section, we describe the main datasets used in the empirical analysis. We obtain information on employment from two different sources: the Population Census and RAIS, the social security records dataset of the Ministry of Labor. The Population Census has the advantage of covering both formal and informal workers, and it is available at 10-year intervals. RAIS covers only formal employees, but it has the advantage of being available at the yearly frequency. We also use data from two different manufacturing surveys: PIA and PINTEC. We use data from PIA – the Brazilian manufacturing survey – to construct measures of manufacturing productivity and capital. We use data from

[17]In these specifications, we use a balanced panel of microregions that includes all the microregions for which we have observations in each year of the decade.

[18]See Bustos et al. (2016) for a detailed discussion of second-season maize.

PINTEC – the Brazilian Innovation survey – to classify industries by innovation intensity. In what follows, we describe these four data sources in more detail.

We use the Censuses of 2000 and 2010 to obtain detailed information on employment and wages in all sectors. We focus on individuals with strong labor force attachment. In particular, we include individuals aged between 25 and 55 that work more than 35 hours a week.[19] Differently from social security data, the Population Census covers both formal and informal workers, which makes it well suited to study movements of workers in the agricultural sector – whose labor force is largely informal – as well as any effect on informal employment in manufacturing. For each individual, we define the sector of occupation as the sector of their main job during the reference week of the census. The Population Census also provides information on the number of hours worked during the reference week and the monthly wage.[20] We use information on education to categorize individuals as unskilled or skilled. We define workers as skilled if they have completed at least the 8^{th} grade, although our results are robust to alternative definitions of this threshold. This level should be attained when an individual is 14 or 15 years old, and is equivalent to graduating from middle school in the US. We also use data from the Population Census to compute "composition-adjusted" wages (i.e., wages net of observable workers' characteristics). To this end, we estimate a Mincerian regression of log hourly wages on observable characteristics for the two census years of 2000 and 2010, as explained in Appendix B.

The Annual Social Information System (RAIS) is an employer-employee dataset that provides individual information on the universe of formal workers in Brazil.[21] We use RAIS to study movements of workers across industries within manufacturing at yearly level from 1998 to 2009. As in the Population Census, we focus on individuals aged between 25 and 55 that work more than 35 hours a week.[22] RAIS contains detailed information on workers' occupations, which we use to construct the new spatial measure of the labor input in innovation activities described below.

We use data on number of workers, capital, value added, and wage bill from the Annual

[19] In order to deal with extreme observations, we focus on individuals whose absolute and hourly wages are between the 1st and the 99th percentile for the distribution of wages in their respective year, and who work less than the 99th percentile of hours. Moreover, we only consider individuals not enrolled in the education system at the time of the survey.

[20] We compute hourly wages as the monthly wage divided by 4.33 times the hours worked reference week.

[21] Employers are required by law to provide detailed worker information to the Ministry of Labor. See Decree n. 76.900, December 23^{rd} 1975. Failure to report can result in fines. RAIS is used by the Brazilian Ministry of Labor to identify workers entitled to unemployment benefits (*Seguro Desemprego*) and federal wage supplement program (*Abono Salarial*).

[22] Following Helpman, Itskhoki, Muendler, and Redding (2017), our data cleaning procedure includes: (i) restricting to workers employed as of December 31st in each year; (ii) restricting to the highest-paying job for each worker that appears more than once in the data during one year (randomly dropping ties).

Industrial Survey (PIA). The PIA survey is constructed using two strata: the first includes a sample of firms with 5 to 29 employees (*estrato amostrado*), and it is representative at the sector and state level. The second includes all firms with 30 or more employees (*estrato certo*). We restrict the analysis to firms with 30 or more employees so that our outcomes are representative at the microregion and industry level for those larger firms. We define employment as the end-of-year number of workers, and value added as the difference between value of production and expenditure in intermediate inputs. The PIA survey does not report information on the capital stock. Thus, we use data on investment, depreciation, and the book value of assets in the first year a firm appears on the sample to construct a firm-level measure of capital stock using the perpetual inventory method. For multiplant firms, we allocate capital stock to each of their plants using employment shares. We focus on firms operating in manufacturing as defined by the CNAE 1.0 classification (codes between 15 and 37) and on the period between 2000 and 2009.

Finally, we use data from the Survey of Innovation PINTEC to classify manufacturing industries into H and L industries. We think of H industries as industries that use relatively more skilled labor and dedicate more resources to innovations that can generate knowledge spillovers for other sectors. On the other hand, we think of L industries as traditional, unskilled-labor intensive industries in which the scope for process innovation is lower and that are less likely to generate knowledge spillovers toward other sectors. This split is useful for investigating how the soy shock affected innovation incentives, as we discuss below. The PINTEC survey is designed to capture innovation activities of Brazilian firms and it is available every 3 years starting in 2000. Using this data, we construct a measure of R&D intensity at the industry level, computed as the monetary value of R&D expenditures divided by total sales in the baseline year 2000. The measure of R&D expenditure encompasses expenditure in both internal R&D and external R&D, as well as expenditure in external know-how, machinery and equipment, training, and expenditures related to introducing innovation in the market. Because this measure subsumes expenditure in components of innovation that might be cataloged as intermediate inputs, we normalize it by total value of output in the industry (sales) rather than value added.[23] We define H industries as those above the median level of R&D intensity, weighting industries by their employment at baseline. Table A.2 reports the full list of manufacturing industries by R&D intensity and skill intensity.[24] R&D intensity and

[23]Other papers in the innovation literature that define R&D Intensity as R&D expenditures over sales include, but are not limited to, the seminal papers on the exploration and characterization of industry and firm R&D Intensity of Pakes and Schankerman (1984), Cohen, Levin, and Mowery (1987), Jaffe (1988) and Cohen and Klepper (1992), and more recent works such as Acemoglu, Akcigit, Alp, Bloom, and Kerr (2018) and Autor, Dorn, Hanson, Pisano, and Shu (2020).

[24]The 60 manufacturing industries reported in Table A.2 correspond to the industry classification CNAE-Domiciliar used in the Population Census. Our original measure of R&D intensity at industry level constructed using PINTEC data is based on the 4-digit CNAE 1.0 industry classification, which defines 267 different manufacturing industries (PIA and RAIS datasets use the same industry classification).

skill intensity at the industry level are highly correlated, as can be seen in Figure A.2 in the Appendix.[25] Indeed, the production function estimates presented in Table B.10 that we use to compute Total Factor Productivity (TFP) imply that H industries are more skill-intensive

Table 1 reports summary statistics of individual level characteristics observed in the Population Census for workers operating in agriculture, L manufacturing, H manufacturing and services.[26] As shown, there is large heterogeneity in skill intensity of workers across these broad sectors. Almost 90 percent of workers in agriculture had not completed the 8^{th} grade in 2000, while this number is around 50 percent for manufacturing and services. Large differences are also present within manufacturing, where the share of high-skill workers tends to be higher in H industries, particularly in 2010.

Table 1 goes around here

Table 2 provides summary statistics for the main variables used in the empirical analysis at the microregion level. Microregions are statistical units defined by the IBGE and consist of a group of municipalities. Brazil has 557 microregions, with an average population of around 300,000 inhabitants. We use microregions as an approximation of the local labor market of a Brazilian worker. They can be thought of as small, open economies that trade in agricultural and manufacturing goods but where production factors are immobile.[27] For outcomes sourced from the Population Census, which are observed in 2000 and 2010, we report the mean and standard deviation of their level in the baseline year (2000) and of their change between 2000 and 2010.

Table 2 goes around here

3.4 A new measure of innovation across space

One of the mechanisms that we explore when trying to understand the decline in manufacturing productivity is whether the inflow of low-skill workers into manufacturing

To map the 267 industries in PINTEC with the 60 industries reported in Table A.2 we use the official conversion tables provided by the IBGE (`https://concla.ibge.gov.br/`).

[25]Notice that data on R&D expenditure from the PINTEC survey is not representative at the microregion level. Thus, to construct a measure of innovation that is representative at any geographical level, we use the description of occupations reported in the social security records, as described in section 3.4.

[26]We define agriculture, manufacturing and services by following the classification of the CNAE Domiciliar of the 2000 census. Agriculture includes Sections A and B (agriculture, cattle, forestry, and fishing). Manufacturing includes Section D, which corresponds to the transformation industries. Services include: construction, commerce, lodging and restaurants, transportation, finance, housing services, domestic workers, and other personal services. We exclude the following sectors because they are mostly under government control: public administration, education, health, international organizations, extraction, and public utilities.

[27]In Table A.5 of the Appendix we show that internal migration did not respond to the shock. This is in line with evidence from Brazil's lack of internal migration responses documented also in Dix-Carneiro and Kovak (2019) and Costa, Garred, and Pessoa (2016). Migration in Brazil seems to have been more central in the 1950s, as documented in Pellegrina and Sotelo (2022).

changed the incentives to innovate. This requires to observe innovation at the microregion level, our unit of observation of the empirical analysis. For this purpose, we develop a new measure of innovation which is representative at any level of geographical disaggregation, using the description of occupations in RAIS. More specifically, we propose a new measure of the labor input in innovation activities based on textual analysis of the task descriptions of more than 2,500 occupations. Tasks generating innovations include, for example, developing new products and processes, creating prototypes, or optimizing methods of production. An important advantage of this measure is that it allows us to track innovation workers across sectors and regions. This is because the social security data covers the universe of formal firms. In contrast, standard manufacturing innovation surveys, such as PINTEC, are based on a sample of firms that is not representative at low levels of geographical disaggregation, and do not allow to trace workers' movements across firms.

In what follows, we describe our methodology to identify workers in innovative occupations. As a first step, we digitized the text containing the official description of the tasks associated with each occupation as provided in the "Brazilian Classification of Occupations" published by the Ministry of Labor. In the second step, we defined a set of 39 keywords or combination of keywords capturing tasks related to innovative activities. To generate this list, we identified a set of words that are used to define activities related to innovation either in the task description of occupations provided by the Ministry of Labor, or in the technical documentation of PINTEC, the Survey of Innovation of Brazilian firms. The list of keywords used is reported in Appendix Table A.3. As shown, most entries are a combination of a verb and a noun describing a task associated with innovation. These combinations can be grouped in those capturing innovation of products (e.g. "develop/improve product/s"), innovation of processes (e.g. "develop/improve/test process/es"), innovation of machinery and equipment (e.g. "develop device/s", "develop equipment"). We also include single nouns, combinations of nouns, or combinations of nouns and adjectives that are often found in the description of innovation intensive tasks (e.g. "innovation", "prototypes", "research and development", "new technologies"). Finally, in the last step, we run a text analysis that identifies all occupations whose description contains at least one of the keywords listed in Appendix Table A.3. This methodology identifies 251 occupations, which we define as innovation-intensive.[28]

Figure A.3 shows the total number and the share of manufacturing workers in innovation-intensive occupations in Brazil. According to our measure, the number of workers in innovation-intensive occupations increased from approximately one hundred thousand in 2000 to three hundred thousands in 2014, and started falling afterward when Brazil entered into a severe recession. Workers in innovation intensive occupations constitute

[28]See Lagaras (2017) for an application of this methodology at the firm-level in order to explore the impact of corporate acquisitions on labor reorganization and firm-level innovation.

between 3 and 4 percent of total manufacturing formal employment. This share has been increasing during the period under study from 2.5 percent in the early 2000s to slightly above 4 percent in most recent years.[29] Figure A.4 reports the share of local manufacturing employment engaged in innovation intensive activities in each microregion of Brazil in the baseline year 2000. As shown, the share of innovation workers ranges from 0 to almost 20 percent of formal manufacturing employment, with higher shares observed in the coastal regions of the South and South-east of Brazil, but also in several microregions encompassing large cities in the North and Center-West regions of the country.

We perform a set of consistency tests on our measure of innovation. Figure A.5 shows the correlation between employment share in innovation-intensive occupations and other measures of innovation that are available at the industry level from the PINTEC survey. We include measures that capture the amount of inputs devoted to the innovation process – such as R&D expenditure per worker – as well as measures capturing the output of the innovation process – such as the share of firms in a given sector that have filed patents and the share of firms that have introduced new processes or products. As shown, the share of innovation-intensive workers is highly correlated with all these alternative measures, with the additional advantage of being available not only at the sector level but also at fine levels of geographical aggregation. Table A.4 reports the magnitude of the correlations between the share of innovation-intensive workers in each industry and the alternative measures reported in Figure A.5. The estimates indicate that a 1 percentage point increase in the share of innovation-intensive workers in a given industry is associated with a 6 percent increase in R&D expenditure over sales, a 1.6 percentage point increase in the share of firms filing for patents, and a 1.6 percentage points increase in the share of firms introducing either a new product or a new process.

It is perhaps also useful to discuss in some detail the differences between the measure of innovation based on workers' task description proposed in this paper and the main alternative measure of innovation used in the literature: patenting. One advantage of patent data is that it captures the output of the innovation process, and – by using patent citations – it allows researchers to make statements about the quality of the innovation produced (Carlino and Kerr 2015). However, an important disadvantage of patent data is that, in many instances, firms introduce new products or processes without patenting them. Data from PINTEC, shows that, in the decade 1997 to 2008, 34 percent of surveyed firms introduced new processes or products. However, only 7 percent of those firms have

[29] The Brazilian Ministry of Labor has updated its classification of occupations in 2002. RAIS uses the new classification (CBO2002) starting from 2003. We identify innovation intensive occupations using the the description of tasks provided for the CBO2002 classification. To extend our analysis to the pre-2003 years we match the old classification (CBO 1994) and new classification (CBO 2002) using the official correspondences provided by the Ministry of Labor. Whenever one occupation in the old classification is matched with multiple occupations in the new one, we weight the number of workers in that occupation by the share of innovation workers observed in the first year in which the new classification is used (2003).

filed a patent application or have an approved patent for such innovation.[30] This fact is visible also in panel (f) of Figure A.5, which shows how in many sectors with a high share of firms introducing new processes and products, no firms report patenting activity. The fact that many firms decide not to patent their innovations has been documented also in other countries. For example, Cohen, Nelson, and Walsh (2000) analyze survey data from approximately 1,500 R&D labs of manufacturing firms in the US, and show that patenting is used less frequently than other approaches to protect the return from invention, as patent applications require firms to disclose to competitors a large amount of information. According to the same survey, smaller firms tend not to apply for patents due to their legal costs, and are also more likely to consider patents ineffective.

Finally, it may also be useful to clarify that what our measure of innovation intends to capture not just investment in R&D that pushes the world technology frontier, but also investment in adapting innovations developed elsewhere to the Brazilian market or to firm-specific production processes. Indeed, aggregate data from the PINTEC suvey indicate that most innovations introduced by Brazilian firms happen through adaptation of technologies that are new to the firm but already in use elsewhere. In particular, about 84% of new products and 94% of new processes introduced by Brazilian firms surveyed in PINTEC are an innovation for the firm but already exist in some form either in Brazil or in the rest of the World.[31] In this sense, we think of investments aimed at "adapting" a new technology developed elsewhere to the Brazilian market, and making it usable for local firms, as an investment in innovation.

4 The industrialization process

4.1 Industrialization without productivity growth

We start by studying the effect of soy technical change on the reallocation of workers and capital towards manufacturing and its impact on manufacturing productivity growth. To this end, we use data on employment from the Population Census and social security records (RAIS), and data on capital from the annual manufacturing survey (PIA).

The results are reported in Table 3. In Panel A, we study the effect of soy technical change on labor reallocation across sectors using Census data and the ten year

[30]These statistics are based on Table 6497 of the PINTEC surveys run in 2000, 2003, 2005 and 2008. Each PINTEC survey captures the innovative activities in the previous three years, so they effectively cover the decade 1997 to 2008. The statistics reported are averages across the four waves.

[31]The case of GE soy is illustrative in this respect. While the initial patent of GE soy seeds was deposited in the US by the multinational corporation Monsanto, the final product available in the Brazilian market was the outcome of an adaptation process that involved Embrapa – the Brazilian Research Institute for new agricultural technologies. In particular, Embrapa conducted a series of crossings between the herbicide tolerant variety developed by Monsanto for the US market and seeds previously developed by Embrapa itself to develop a version of the GE soy seeds adapted to the agro-ecological conditions of Brazil.

first-difference specification explained in section 3.2, equation (1). We find that microregions with higher exposure to soy technical change experienced a decrease in the share of workers employed in agriculture and an increase in the share of workers employed in manufacturing and services.[32] The magnitude of the estimates indicates that agricultural workers displaced by the new technology relocated mostly into manufacturing: microregions with a one standard deviation larger increase in soy technical change experienced a 2.4 percentage points larger decline in the agricultural employment share, a 1.8 percentage points increase in the manufacturing employment share, and a 0.6 percentage points increase in the services employment share. Overall, these results indicate that soy technical change was labor-saving and led to structural transformation, which are the main findings documented in Bustos et al. (2016).[33]

Table 3 goes around here

In Panel B, we move to data from the manufacturing survey (PIA), which allows us to use a yearly panel of microregions. We estimate the specification discussed in section 3.2, equation (3), using as outcome variable the total number of workers in manufacturing in a given microregion and year (in logs). The estimated coefficient shows that microregions more exposed to soy technical change experienced a larger increase in manufacturing employment. The magnitude of the coefficient indicates that a one standard deviation differential change in soy technical change leads to a 7 percent larger increase in manufacturing labor. Next, we investigate whether soy technical change also affected capital investment by manufacturing firms. The results are reported in column (2), and show that capital also moved towards manufacturing. The estimates suggest that a one standard deviation differential change in soy technical change leads to an increase in capital in the manufacturing sector of around 17.6 percent.

In columns (3) to (5), we study the effect of soy technical change on manufacturing productivity. We construct three measures of productivity using data from the manufacturing survey PIA: value added per worker, valued added over wage bill, and total

[32] Soy technical change had only small and not significant effects on total employment. Thus, the employment changes that we document in what follows are not driven by migration between microregions or by changes in the total number of workers employed, but by movement of workers across sectors within microregions. Table A.5 provides evidence on the effect of soy technical change on total employment and migration.

[33] Bustos et al. (2016) find that soy technical change had a positive and significant effect on the employment share in manufacturing but no significant effect on the employment share in the services sector. Panel A of Table 3 in this paper documents that microregions more exposed to soy technical change experienced an increase in employment share in both manufacturing and services. There are two reasons behind this difference in results when the outcome is the employment share in the services sector. The first is that, in this paper, we focus on remunerated labor – i.e. workers receiving a wage – whereas Bustos et al. (2016) also included workers who helped household members without receiving a payment or worked in subsistence agriculture. The second is the unit of observation, which is a microregion in this paper, a municipality in Bustos et al. (2016).

factor productivity.[34] The results show that, although both labor and capital reallocated towards manufacturing, regions more exposed to soy technical change experienced a relative decline in manufacturing productivity. The magnitude of the coefficient in column (3) indicates that labor productivity declined by about 10 percent for a standard deviation differential change in soy technical change, which correspond to about 1.5 percent lower growth rate in manufacturing productivity in the post GE legalization period. We find similar magnitudes for alternative measures of productivity computed as value added per wage bill or total factor productivity.

Taken together, the results presented in Table 3 indicate that, despite the fact that soy technical change drove factors of production from agriculture towards manufacturing, productivity in manufacturing slowed down in the years following GE soy legalization. In the next sections, we explore why we find an empirical result that seems to contradict existing theoretical predictions that moving resources toward manufacturing should boost productivity, both statically and dynamically. We first investigate the nature of the agricultural shock in more detail by documenting the skill composition of the workers moving into manufacturing and the patterns of industrial specialization. Next, in Section 5, we investigate potential mechanisms that can explain the evolution of manufacturing productivity.

4.2 Unskilled-labor saving agricultural technical change and the skill premium

In this subsection, we study the impact of GE soy technical change on workers with different skills using data from both the Population Census and social security records (RAIS). Table 4 presents the results using Census data. We estimate the ten year first-difference specification presented in equation (1), and use as outcome variables the changes in the share of unskilled and skilled workers in agriculture, manufacturing and services between 2000 and 2010.

Columns (1) to (3) focus on unskilled workers. We find that microregions more exposed to soy technical change experienced a reallocation of unskilled workers from agriculture to manufacturing. The magnitude of the estimated coefficients indicates that microregions with a standard deviation higher increase in soy technical change experienced a 2.4 per-

[34] We compute total factor productivity as the Solow residual of a Cobb-Douglas production function that combines skilled labor, unskilled labor, and capital in constant returns to scale fashion. We compute the factor shares for skilled and unskilled labor by combining the share of the aggregate wage bill that corresponds to each type of labor in each type of industry with measures of the labor share at the industry-level for the US retrieved from Becker, Gray, and Marvakov (2021). The capital share is calibrated by leveraging the constant returns to scale assumption. These assumptions imply that for a given industry the production technology is the same across microregions and periods, and thus, changes in the TFP are dictated by changes in the allocation of production factors. See Appendix B.1 for a more detailed explanation of how the TFP measure is computed.

centage points larger decrease in the share of low-skilled workers employed in agriculture, and a corresponding 2.2 percentage points larger increase in the share of low-skilled workers employed in manufacturing. These magnitudes correspond to a 7.2 percent decrease in the initial share of low-skilled workers employed in agriculture and a 16.1 percent increase in the share of those employed in manufacturing. Combined with the fact that soy technical change had no differential effect on total employment (see Table A.5 in the Appendix), these results are consistent with a decline in the absolute demand for low-skill labor in agriculture in response to skilled labor-augmenting technical change.

Columns (4) to (6) focus instead on skilled workers. We find that microregions more exposed to soy technical change experienced a larger decrease in the share of high-skill workers in agriculture, and a larger increase in the share of high-skill workers employed in manufacturing, as expected if low and high-skill workers are to some extent complementary in production. In terms of magnitude, the effect of soy technical change on low-skill labor is about twice as large as the effect on high-skill labor.

Table 4 goes around here

Next, we explore in more detail the labor reallocation process described above using yearly social security data from RAIS. Although RAIS data captures only formal employment, its annual frequency allows us to check whether the employment changes documented with Census data occurred right after GE soy was introduced in Brazil. For this, we plot the interaction of year dummies with our measure of soy technical change as explained in Section 3.2, see equation (2). As can be seen in Figure 3 (a), low-skilled labor started to move towards manufacturing in microregions more exposed to soy technical change around 2002, while there is no systematic difference in the trends leading to this year. When focusing on formal employment captured by social security data, we find no differential increase in skilled labor moving towards manufacturing, as shown in Figure 3 (b). The timing of the effect suggests that changes were permanent. Reallocation of unskilled labor towards manufacturing started around 2002, one year after the first reported smuggling of the GE soy seeds in Brazil and the year when the area planted with soy started expanding at a faster rate (Figure 1). The reallocation then accentuated around 2004, one year after the formal legalization of GE soy in Brazil, and stabilized during the second half of the decade.

Figure 3 goes around here

Taken together, the estimates presented in Table 4 and Figure 3 show that the agricultural sector experienced a decrease in its employment share of both low-skill and high-skill labor, while the manufacturing sector experienced an increase in employment driven mainly by low-skill labor. These findings indicate that labor-saving technical change in agriculture driven by the adoption of GE soy was skill-biased and led mainly low-skill workers to reallocate towards manufacturing.

5 Scale and cost of innovation

In this section, we investigate the possible causes that can explain the manufacturing productivity decline documented in section 4.1. We start by showing that manufacturing productivity decline is similar across sectors. Next, we document how part of the productivity decline in L industries may be explained by worker composition. Finally, we explore whether changes in the incentives to innovate can be behind the decline in productivity in H (and potentially L) industries.

5.1 Industrial specialization

From the point of view of the manufacturing sector, the inflow of former agricultural workers documented in Section 4.2 amounts to an increase in the (relative) supply of unskilled labor. In closed economies, this would be accommodated by an increase in the use of low-skill workers across industries. In open economies, an inflow of unskilled labor generates a comparative advantage in unskilled-labor intensive industries, which should expand by absorbing the inflow of unskilled workers and also attract other complementary factors such as capital and skilled labor (Rybczynski, 1955). To empirically investigate the effects of soy technical change on industrial specialization we use the classification of manufacturing in H and L industries based on PINTEC data described in section 3.3. In this split, which will be particularly useful in our exploration of the various mechanisms in Section 5, H (resp. L) industries are defined as those with higher (resp. lower) than median R&D intensity and tend to use skilled (resp. low-skilled) labor more intensively.

We start by documenting the effect of soy technical change on industrial specialization using population census data using equation (1). Panel A of Table 5 shows that the labor inflow into manufacturing is concentrated in L industries whose employment expanded by around 17 percent for a one standard deviation differential increase in potential soy yields. In contrast, H industries did not experience any differential change in employment. In Panel B, we investigate industrial specialization using social security data and the yearly panel regression introduced in equation (3). The results are in line with the ones obtained with Population Census data: labor was absorbed by L industries. Point estimates are smaller possibly due to the fact that social security data only includes formal labor. As the level of informality is higher in agriculture than manufacturing, it is possible that former agricultural workers were more likely to accept informal contracts. In this case, their employment in the manufacturing sector is captured by the Population Census but not by social security data.

Next, we investigate the effects of agricultural technical change on capital investment in manufacturing. Agricultural technical change could affect capital investment in two ways. First, a labor supply channel: the labor inflow in the L industry documented above would tend to increase the marginal product of capital, which should attract more investment into this industry. Second, a capital supply channel: high agricultural productivity increased local savings, as documented in Bustos et al. (2020). In turn, this increase in local capital supply could have increased capital investment in capital-intensive industries. Panel C of Table 5 shows that capital inflows to manufacturing concentrated in L industries, where capital increased by 26.7 percent for a one standard deviation differential change in soy technical change. In contrast, H industries did not experience these capital inflows. Hence, this evidence favors the labor supply explanation at the local level.[35]

Table 5 goes around here

To investigate in more detail the timing of these labor and capital inflows into manufacturing, we estimate the dynamic difference-in-differences specification described in equation (2). The coefficient estimates are presented in Figure 4, which shows that the adoption of GM soy induced the reallocation of labor towards L industries starting in 2002. This timing coincides with the first reports of large scale smuggling of GE soy seeds and the expansion of the area planted with soy (Figure 1) in 2002, which led to the legalization of the new soy seeds in 2003. In addition, the figure shows that labor inflows into manufacturing precede capital inflows by one year, which is further evidence in support of the labor supply mechanism discussed above.

Figure 4 goes around here

In sum, we find evidence of local industrial specialization into L industries. This finding is consistent with Rybczynski-type forces prevalent in small open economies (Rybczynski, 1955), since the labor released from agriculture was mainly low-skilled and that L industries tend to use low-skilled labor more intensively. In what follows, we investigate how this pattern of industrial specialization can explain the decline in manufacturing productivity.

[35]Bustos et al. (2020) show that the capital supply channel occurs across regions. We return to this point in Section 5.5 where we investigate the capital supply mechanism in more detail.

5.2 Industrial composition

Given the evidence shown in section 4, one natural candidate to explain the overall manufacturing productivity decline is the change in industrial composition. If the industries that are able to absorb former agricultural workers are less productive to start with, when resources move toward such industries, overall manufacturing productivity should mechanically decline. To investigate this channel we decompose the total effect of soy technical change on manufacturing productivity into three components: changes in productivity *within* the L industry, changes in productivity *within* the H industry, and the change in productivity driven by reallocation of factors *between* industries. Specifically, we use the following decomposition:

$$\Delta \log TFP_{it} = \underbrace{s_{iLt} \times \Delta \log TFP_{iLt}}_{\text{Change within } L \text{ industry}} + \underbrace{(1 - s_{iLt}) \times \Delta \log TFP_{iHt}}_{\text{Change within } H \text{ industry}} +$$

$$+ \underbrace{s_{iLt} \times \Delta \log \omega_{it} \times \frac{(TFP_{iLt} - TFP_{iHt})}{TFP_{iLt}}}_{\text{Composition effects}} + \varepsilon_i$$

where $s_{iLt} = \frac{\omega_i \times TFP_{iLt}}{\omega_i \times TFP_{iLt} + (1-\omega_i) \times TFP_{iHt}}$ corresponds to the share of TFP in the L industry weighted by $\omega_i = \frac{VA_{iL,2000}}{VA_{iL,2000} + VA_{iH,2000}}$ which is the value added share of each sector in 2000, before the adoption of the new GE soy seeds.[36] The first and second terms of this equation reflect changes in manufacturing productivity within the L and H industries, respectively. The last term captures changes in overall manufacturing productivity due to composition effects, driven by changes in the relative size of each industry.

The overall decline in manufacturing productivity is shown in panels (a) and (b) of Figure 5, while the decompositions into *between* and *within* components, using value added per worker and our measure of TFP, are reported in the same Figure 5, graphs (c) and (d). These graphs show that most of the reduction in the level of manufacturing productivity is driven by the *within* components, the *between* component being small. In fact, the estimates imply that the "between" component can explain at most 8 percent of the overall decline in manufacturing productivity. Hence, most of the manufacturing productivity decline is not driven by a change in the relative size of each industrial sector.

Figure 5 goes around here

Table 6 quantifies further the results shown in Figure 5. The estimated coefficients in columns (1), (2), and (3) indicate that microregions with a one standard deviation

[36] Notice that ε_{it} is a residual that comes from the fact that we use pre-shock weights and aggregate differences in TFP between sectors, rather than microregion specific ones.

faster technical change in soy experienced a decline in manufacturing productivity in the L industry of between 10 and 11 percent when using value added per worker or value added over wage bill as measures of productivity, and of about 18 percent when using TFP. The declines in manufacturing productivity in the H industry are between 9 and 11 percent for a standard deviation difference in soy technical change across all measures of productivity.

Table 6 goes around here

The large reductions in productivity within the L industry reported above, in the context of an overall increase in the size of this industry, have important implications for interpreting the evidence. First, this finding confirms that capital and labor are not pulled into the L industry by increases in its productivity but pushed by labor-saving technical change in agriculture. Second, the resulting larger scale of this industry does not appear to generate increasing returns, as some previous work might have predicted. Third, the evidence in this section suggests that industrial composition effects do not explain substantially the manufacturing productivity decline.

5.3 Worker composition

A second mechanism that can explain the decline in manufacturing productivity is a change in the composition of workers entering the manufacturing sector. Workers entering manufacturing who were previously employed in agriculture may lack the skills necessary to thrive in the manufacturing sector. This may result in substantial (average) productivity declines in the industries where they enter, at least temporarily.

Some of the measures of productivity used in section 4.1 such as value added over wage bill are already designed to account not only for observable differences in workers' human capital like education, but also for unobservable differences, to the extent that workers' productivity is passed on to wages. To see this, note that we can decompose labor productivity in L industries as valued added per wage bill and average wages: $\ln VA/L = \ln VA/wL + \ln w$. In perfectly competitive labor markets, average wages should reflect the average marginal product of labor, which can be tied almost one to one to productivity (less than one to one if there are fixed factors of production). In this case, negative worker selection would manifest in a decline in average wages and, hence, large differences in our estimates of the effect of soy technical change on manufacturing productivity measured as value added per worker and as value added per wage bill. Comparing the estimates in columns (1) and (2) of Table 6 imply that around 11% of the decline in manufacturing productivity in the L industries can be attributed to worker composition.[37]

[37]This number results from the following calculation: $(0.151 - 0.135)/0.151 = 0.106$.

At the micro-level, we can investigate the impact of soy technical change on worker composition using data from RAIS, which allows us to track the trajectory of workers initially employed in agriculture. It is important to note that these workers are necessarily formal, and hence, a relatively small and potentially selected sample. In fact, according to Census data, over 75% of agricultural workers in 2000 are informal. Still, when comparing the maximum level of education attained by informal agricultural workers observed in the Population Census and the one attained by formal agricultural workers observed in RAIS in 2000, we find similar shares of unskilled labor in both samples (87% among informal workers in the Census, 83% in RAIS). Overall, we think that following formal workers across sectors in RAIS is informative about the patterns of worker selection.

We start by studying the employment trajectory of individuals observed in agriculture in the pre-2003 period. As in the case of Census data, we focus this analysis on full-time male employees between 25 and 55 years of age. We observe 1,304,659 unique individuals with such characteristics that are formally employed in agriculture in the pre-2003 period. We then categorize their employment trajectories starting in 2003, which include: remaining in agriculture, moving into sectors other than agriculture, or moving out of sample (including informality, unemployment or self-employment). We are particularly interested in studying the effect of soy technical change on the probability of transitioning into H vs L manufacturing industries. The results are reported in the first two columns of Table 7. Column (1) shows that formal workers who were employed in agriculture prior to 2003 in areas with higher soy technical change are significantly more likely to move toward L manufacturing industries (3.7 percentage point for a standard deviation higher increase in potential soy yields). Column (2) shows that such workers are also significantly less likely to enter H industries. These patterns confirm that formal workers relocated across sectors in a way that is similar to how all workers relocated – which we captured with Census data.

In columns (3) and (4) we compare the wages paid to former agricultural workers when moving toward the L industry with the wages of incumbent workers in the L industry. For this analysis, we restrict our sample to workers observed in the L industry starting in 2003, and that were previously employed either in agriculture or in the L industry. The estimate in column (3) indicates that, unconditionally, former agricultural workers are paid 27 percent less than incumbent workers in the L industry, which is an indication that they may be as much as 27 percent less productive.[38] In column (4) we use as outcome a measure of wages that takes into account differences in age and education

[38] In unreported results we investigate if the potential patterns of selection of workers moving from agriculture to L industries differs across regions with different soy technical change. The estimated coefficient on the interaction term between workers moving from agriculture and soy technical change is 0.022 with a standard error of 0.030. The fact that the interaction coefficient is small and not distinguishable from zero indicates that the negative selection of movers toward L industries does not change significantly with the soy shock.

across workers. Conditioning on these observable characteristics, the wage penalty for former agricultural workers is 12 percent.

Table 7 goes around here

We can use the wage penalty estimates of formal worker moving from agriculture to L manufacturing to assess how much of the productivity decline observed in L manufacturing can be explained by changes in worker composition. For this exercise, we compute how many workers moved from agriculture to L industries and assume that informal workers experience similar wage penalty as formal workers when moving across these sectors. Microregions with one standard deviation higher soy technical change experience a 17% higher increase in employment in L manufacturing (Table 5), which is driven by a reallocation of labor from agriculture (Table 3). If the wage penalty that former agricultural workers face in L industries is due to productivity disadvantages, then worker selection can explain at most between 25 and 42 percent of the productivity decline in L industries, depending on the measure of productivity (TFP vs value-added per worker).[39]

The main caveat with this analysis is that RAIS data only covers formal workers, and yet, a big share of agricultural workers in Brazil are informal. To investigate the patterns of selection of agricultural workers we turn to evidence on how the soy technical change affected wages of informal workers across sectors. In particular, in Table A.6, we focus on decadal changes in the average wage of unskilled informal workers in manufacturing.[40] We see that, on average, wages of unskilled informal workers declined somewhat in manufacturing, particularly in L industries (3.1 percent for a one standard deviation in soy technical change). If all of this decline reflects changes in productivity, then this estimate suggests that worker selection can account for at most 31% of the decline in manufacturing productivity.[41]

Overall, the results presented in this section indicate that former agricultural workers are, on average, less productive than incumbent workers in the L industry. The estimates presented in this section imply that changes in worker composition can contribute to

[39]These numbers are obtained using the wage penalty implied by column (3) of Table 7. First, we multiply the percentage increase in employment in the L manufacturing industry for one standard deviation in soy technical change by the wage penalty estimated in Table 7, that is: $0.271 \times 0.17 = 0.047$. This gives us the productivity decline due to worker composition for a standard deviation difference in soy technical change. Next, we compare the 0.047 with the productivity decline in the L industry for a standard deviation of soy technical change documented in Table 6. The range refers to the log TFP measure ($0.047/0.0186 = 0.251$) and the log value added per worker measure ($0.047/0.112 = 0.417$). Notice that if we were to use the wage penalty conditional on worker characteristics (column (4) of Table 7), then worker selection could explain between 11 and 18% of the productivity decline, for the same two measures of productivity.

[40]In this case we take all workers, including those who work part-time and females, to try to be inclusive on all potentially informal workers in the economy. See more details on how we compute composition adjusted wages in the data section.

[41]This number comes from dividing our wage estimate in Table A.6 by the smallest estimate of manufacturing productivity decline shown in Table 6, i.e. 0.042/0.135.

explain about one third of the decline in productivity within the L industry (25 to 43% when using formal workers' wages, and 31% when using informal workers' wages). This result can help to rationalize the larger declines in manufacturing productivity observed in L industries relative to H industries in Table 6. However, worker composition is unlikely to be a driver of the productivity decline within the H industry, because the latter does not absorb agricultural workers. Hence, we need to investigate alternative potential drivers of manufacturing productivity decline across industries. The endogenous growth literature suggests one natural candidate – innovation – which we discuss in the next section.

5.4 Innovation

In this section, we investigate whether part of the simultaneous reduction in productivity within L and H industries can be explained by changes in innovation activities. A major challenge to studying the response of industrial innovation to agricultural technical change is the lack of a measure of investment in R&D which is representative at low levels of spatial aggregation such as micro-regions. Thus, as detailed in Section 3.3, we use the description of workers' occupations in social security data to develop a new measure of investment in innovation that varies both across regions and sectors. We measure the labor input in innovation as the total wage bill of workers in innovation-intensive occupations, which are defined as those effectively producing new ideas – such as new products and processes – within each industry.

This measure allows us to estimate the effect of agricultural technical change on the allocation of innovative activities across industries. In particular, we estimate equation (2) using as outcome variable the total wage bill of workers employed in innovation and non-innovation intensive occupations and report estimates in Figure 6. The top panels (a) and (b) confirm that regions more exposed to soy technical change experienced an increase in the wage bill in non-innovative activities within L industries. In turn, panels (c) and (d) report the effect of agricultural technical change on investment in innovation-intensive activities. Estimates reported in panel (c) show that, if anything, innovative activities in L industries increased slightly. In turn, estimates in panel (d) show that regions more exposed to soy technical change experienced a sharp decline in investment in innovative activities within H industries, whose timing corresponds with the legalization of GE soy in 2003.

Figure 6 goes around here

Table 8 quantifies the effects documented in Figure 6. The coefficients reported in columns (1) and (2) confirm that microregions with a one standard deviation larger increase in potential soy yields experienced an 11 percent higher increase in the wage bill of non-innovative labor in L industries and no increase in H industries. In turn, column (3) shows a positive but not statistically significant increase in the wage bill of innovative labor in L industries. Finally, column (4) shows that microregions with one standard deviation larger increase in soy technical change experienced a 20 percent larger decline in the wage bill of innovative labor in H industries. Hence, these results suggest that R&D expenditures and innovation activities moved toward L industries, which were expanding as a result of the soy shock. It is worth emphasizing that this decline in potentially higher spillover type industries can explain, at least part, of the manufacturing productivity decline in both industries.

Table 8 goes around here

To understand in more detail how this relocation of innovation workers occurred, we perform two additional exercises. First, we investigate whether the decline of innovation in the H industry is driven by lower entry or lower retention of workers that might then be attracted to the L industry. For this, we use social security data to classify innovation workers employed in the H industry into five different categories: stayers (i.e. innovation workers already observed in the previous year in the H industry), entrants in the labor market, entrants from informality/self-employment/unemployment, switchers from L industries and switchers from sectors other than the L industry. We then estimate the impact of soy technical change on the wage bill of each of these categories separately. The results, reported in Figure 7, show that in regions more exposed to soy technical change, innovation workers are both less likely to enter and to stay in the H industry. The magnitude of the effect on the different components indicates that about half of the total effect is driven by a lower probability of innovation workers remaining in the H industry. The other half is driven by lower entry into the H industry in regions more exposed to soy technical change.

Figure 7 goes around here

Second, we track whether innovation workers initially in H industries moved differentially into L industries and whether they continued to be employed in innovation tasks or moved into other occupations. Figure 8 (a) reports that in microregions with a one standard deviation higher exposure to soy technical change, 2.4 percentage points of the wage bill share of innovation workers is lost due to reallocation of workers towards the L industry over the whole post-period. Hence, innovation workers seem to have followed other factors of production such as non-innovation labor and capital. Figure 8 (b), moreover, shows that a significant fraction of these workers moved into non-innovating activities.[42]

[42]See Figure A.6 for the same figures using labor instead of wage bills.

We find that more than one-third of innovation workers moving from H to L industries also changed occupation, switching from an innovation-intensive to a non-innovation intensive job.

Figure 8 goes around here

Taken together, the results presented in this section show a decline in innovative activities performed in H industries in regions more exposed to agricultural technical change, and a reallocation of such activities from H to L industries. In addition, the results indicate that a significant fraction of innovation workers moving to L industries also switched from innovation intensive to non-innovation intensive occupations. These results can explain the decline in manufacturing productivity growth that accompanied the industrialization process in such regions. In particular, the results suggest that the decline in innovation within H industries reduced local knowledge production, potentially causing manufacturing productivity growth to decline in both industries. Note that lower innovation in the H industry appears to have a direct effect on its productivity and (potentially) an indirect effect on the productivity of the L industry through local knowledge spillovers.

5.5 The role of capital

In section 5.1, we documented that the reallocation of agricultural workers into low-R&D intensive manufacturing industries was followed by an inflow of capital. We evaluated two potential channels through which agricultural productivity growth can lead to capital investment in manufacturing. First, a labor supply channel: labor-saving technical change in agriculture generated an inflow of workers in the L industry. Larger employment is expected to increase the marginal product of capital and thus attract more investment into this industry. Second, a capital supply channel: high agricultural productivity increased local savings, as documented in Bustos et al. (2020). In turn, this increase in local capital supply could have increased capital investment in capital-intensive industries. We concluded that the evidence presented in that section favoured the labor supply mechanism for two reasons. First, capital inflows into manufacturing lagged labor inflows by one year. Second, capital inflows were concentrated in L industries while both industries have a similar capital-intensity. Still, in the current section we conduct a more detailed exploration of the capital supply channel.

We build on previous work studying the effects of the agricultural boom in Brazil on capital markets. Bustos et al. (2020) document that regions with faster technical change in soy experienced an increase in local savings deposits which were not lent locally, leading to an increase in capital outflows. They use detailed credit registry data to track the destination of capital flows and find that banks capturing deposits in soy boom areas redirected them to other regions where they had branches. This increase in bank lending was concentrated outside of soy-producing regions and in the manufacturing and service sectors. Note that this finding stands in contrast to the findings on the effect of agricultural technical change on structural transformation through the labor supply channel, which operates in local labor markets with limited migration responses (see Table A.5). Instead, the capital supply channel operates across regions financially integrated with soy boom areas through the bank branch network.

We can simultaneously estimate the effect of the labor and capital supply mechanisms on structural transformation by adding to our baseline specification, described in equation (1), a measure of exposure to capital inflows from soy boom areas through the bank branch network following Bustos et al. (2020). In this analysis, we use the change in employment share in manufacturing between 2000 and 2010 as our measure of structural transformation. Estimation results are reported in Appendix Table A.7. Two key findings emerge. First, we confirm that the local effect of soy technical change is concentrated in soy-producing regions, and in low-R&D intensive manufacturing – what we label L industries in this paper. On the other hand, the indirect effect of soy technical change via bank exposure to soy boom areas has the following characteristics: it affects manufacturing in destination regions, it is concentrated in non-soy producing regions, and it is stronger for high-R&D intensive manufacturing – what we label H industries in this paper.

These results have relevant implications for the role of technical change in agriculture for industrialization and growth. In particular, they suggest that while agricultural savings can foster the growth of productivity-enhancing manufacturing in urban regions financially integrated via the bank network, the reallocation of former agricultural workers towards low-skill intensive industries can slow down productivity growth at the local level. Thus, our findings indicate that structural transformation obtained through unskilled labor-saving technical change in agriculture – which may be quite common when developing countries adopt agricultural technologies from more developed ones – can attenuate the standard gains from reallocation into manufacturing emphasized by the existing literature, and potentially accentuate regional productivity inequalities.

5.6 Robustness Tests

In this section, we address some additional concerns regarding the interpretation of our estimates. First, we investigate whether the expansion of L manufacturing industries in areas experiencing technical change in soy could be the result of larger local demand for agricultural inputs or larger local supply of agricultural outputs for further processing. In this case, the expansion of manufacturing employment would be driven by sectors connected to soy production via input-output linkages. This includes sectors using soy as an input – such as the food processing industry – or that produce inputs for the soy sector – such as the production of fertilizers, herbicides or other agricultural inputs. To identify the sectors linked to soy production via input-output linkages we use the Input-Output tables computed by the IBGE.[43] Although the majority of the output of the soy sector is exported, two downstream manufacturing sectors report using soy as an input: "Slaughtering and preparation of meat and fish" (SNA code 1091) and "Other food products" (SNA code 1093).[44] Upstream industries of the soy sector include: "Fertilizers and other inorganic chemicals" (2412, 2413, 2419) and "Refined petroleum" (232).

In Table A.8 in the Appendix we replicate the main results of the paper excluding upstream and downstream industries. As shown, the estimated coefficients that we obtain are similar in magnitude to those obtained in the main tables of the paper. For example, the coefficient capturing the effect of soy technical change on employment in the L manufacturing industry has a magnitude of 0.121 in Table 5, and a magnitude of 0.124 when excluding sectors connected via input-output linkages in Table A.8. We interpret these results as indicating that the effect of soy technical change on local employment is not driven by local demand effects in manufacturing industries related to soy via production networks.

Next, we investigate to what extent the reduction in innovation within high R&D industries has the potential to reduce productivity not only in this industry but also in the local low R&D industry through local knowledge spillovers. Note that this is key for our empirical identification strategy that compares microregions differently affected by soy technical change, and thus requires that knowledge spillovers are stronger within than across regions. Evidence from the existing literature suggests that spillovers exist and tend to be local, as shown, for example, in Greenstone, Hornbeck, and Moretti (2010) and Giroud, Lenzu, Maingi, and Mueller (2021) in the context of the US. Greenstone et al. (2010) document that the construction of large manufacturing plants in a given county generates productivity spillovers for existing plants in the same county. Importantly, they show that such spillovers occur across manufacturing industries, and are stronger across industries with larger technological linkages, as measured by industry-to-industry R&D flows and industry-to-industry patent citations (Ellison, Glaeser, and Kerr, 2010). Building on the same experiment, Giroud et al. (2021) show that across firms and across-industry spillovers are very local, and only travel across regions within multi-plant firms.

[43]The tables are publicly available on the IBGE website: https://www.ibge.gov.br/en/statistics/economic/national-accounts/. This IO Tables use the SNA sector classification, which include 67 sectors.

[44]These 2 SNA sectors correspond to the following sectors in the CNAE 1.0 sector classification at 4-digits used in the paper: (1511)-(1514), (1521)-(1523), (1531)-(1533), (1541)-(1543), (1571), (1572), (1551)-(1556), (1559), (1581)-(1586), and (1589). To identify which of these 29 CNAE 1.0 sectors use soy as an input we looked at the description of the activities classified in each sector according documentation provided by the National Commission of Classifications (CONCLA). We identified 5 sectors whose description indicate they use soy as an input: 1531, 1532, 1533, 1586, 1589. We also added to the downstream industries the biofuels sector.

Still, it is worth exploring whether spillovers are also local in our context. First, we turn to survey evidence from PINTEC. This survey indicates, for each firm acquiring innovation externally, the location of the external firm that developed the innovation. In particular, the survey asks respondents to indicate the Brazilian state in which the external innovating firm is located. Survey responses show that around 70% of external innovation is developed by firms located in the same state as the respondent, consistent with important local innovation spillovers. Second, we directly test for spillovers across microregions, our unit of observation. We augment our main specification of the effect of soy technical change on local manufacturing productivity with three additional controls capturing changes in soy technical change in neighboring microregions. More specifically, we include soy technical change in the five closest, in the five to ten closest, and in the ten to twenty closest microregions. Each group captures geographical spillovers at different distances, starting from adjacent microregions. Table A.9 reports the results for our key measures of manufacturing productivity. As shown, we find that changes in agricultural productivity in nearby microregions do not affect local manufacturing productivity.

6 Conclusions

The reallocation of labor from agriculture into manufacturing is generally regarded as positive in the economic development literature. Several studies have documented that the manufacturing sector has, on average, higher productivity and pays higher wages. However, little is known about which type of workers are released from the agricultural sector and which manufacturing industries absorb them during the process of structural transformation. Our paper contributes to the literature by showing that the forces driving structural transformation can shape the type of industries in which a country specializes. In particular, we show that when labor reallocation from agriculture to manufacturing is driven by agricultural productivity growth that displaces unskilled labor, it can generate an expansion in less innovation-intensive manufacturing sectors, which can reduce investment in innovation and slow down aggregate manufacturing productivity growth.

References

Acemoglu, D. (2002, October). Directed Technical Change. *The Review of Economic Studies 69*(4), 781–809.

Acemoglu, D., U. Akcigit, H. Alp, N. Bloom, and W. Kerr (2018, November). Innovation, reallocation, and growth. *American Economic Review 108*(11), 3450–91.

Aghion, P. and P. Howitt (2008). *The Economics of Growth*. The MIT Press.

Atkin, D., A. Costinot, and M. Fukui (2021). Globalization and the ladder of development: Pushed to the top or held at the bottom? Technical report, National Bureau of Economic Research.

Autor, D., D. Dorn, G. H. Hanson, G. Pisano, and P. Shu (2020, September). Foreign competition and domestic innovation: Evidence from us patents. *American Economic Review: Insights 2*(3), 357–74.

Becker, R., W. Gray, and J. Marvakov (2021). Nber-ces manufacturing industry database: Technical notes. Technical report, National Bureau of Economic Research.

Bragança, A. (2014). *Three Essays on Rural Development in Brazil*. Ph. D. thesis, PUC-Rio.

Bustos, P., B. Caprettini, and J. Ponticelli (2016). Agricultural Productivity and Structural Transformation: Evidence from Brazil. *American Economic Review*.

Bustos, P., G. Garber, and J. Ponticelli (2020). Capital accumulation and structural transformation. *The Quarterly Journal of Economics 135*(2), 1037–1094.

Carlino, G. and W. R. Kerr (2015). Agglomeration and innovation. In *Handbook of regional and urban economics*, Volume 5, pp. 349–404. Elsevier.

Caselli, F. (2005). Chapter 9 Accounting for Cross-Country Income Differences. Volume 1, Part A of *Handbook of Economic Growth*, pp. 679 – 741. Elsevier.

Choi, J. and A. A. Levchenko (2021). The long-term effects of industrial policy. Technical report, National Bureau of Economic Research.

Cohen, W. M. and S. Klepper (1992). The anatomy of industry r&d intensity distributions. *The American Economic Review 82*(4), 773–799.

Cohen, W. M., R. C. Levin, and D. C. Mowery (1987). Firm size and r& d intensity: A re-examination. *The Journal of Industrial Economics 35*(4), 543–565.

Cohen, W. M., R. R. Nelson, and J. P. Walsh (2000). Protecting their intellectual assets: Appropriability conditions and why us manufacturing firms patent (or not).

Costa, F., J. Garred, and J. P. Pessoa (2016, September). Winners and losers from a commodities-for-manufactures trade boom. *Journal of International Economics 102*, 50–69.

De Vries, G., M. Timmer, and K. De Vries (2015). Structural transformation in africa: Static gains, dynamic losses. *The Journal of Development Studies 51*(6), 674–688.

Dix-Carneiro, R., P. K. Goldberg, C. Meghir, and G. Ulyssea (2021, January). Trade and informality in the presence of labor market frictions and regulations. Working Paper 28391, National Bureau of Economic Research.

Dix-Carneiro, R. and B. K. Kovak (2019, March). Margins of labor market adjustment to trade. *Journal of International Economics 117*, 125–142.

Duffy, M. and D. Smith (2001). "Estimated Costs of Crop Production in Iowa". *Iowa State University Extension Service FM1712*.

Ellison, G., E. L. Glaeser, and W. R. Kerr (2010). What causes industry agglomeration? evidence from coagglomeration patterns. *American Economic Review 100*(3), 1195–1213.

Fernandez-Cornejo, J., C. Klotz-Ingram, and S. Jans (2002). "Estimating Farm-Level Effects of Adopting Herbicide-Tolerant Soybeans in the USA". *Journal of Agricultural and Applied Economics 34*, 149–163.

Ferreira, F., S. Firpo, and J. Messina (2017). Ageing Poorly? Accounting for the Decline in Earnings Inequality in Brazil, 1995-2012. *IDB Working Paper Series*.

Foster, A. D. and M. R. Rosenzweig (1996). Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution. *The American Economic Review 86*(4), 931–953.

Franck, R. and O. Galor (2019). Flowers of evil? industrialization and long run development. *Journal of Monetary Economics*.

Giroud, X., S. Lenzu, Q. Maingi, and H. Mueller (2021). Propagation and amplification of local productivity spillovers. Technical report, National Bureau of Economic Research.

Goldberg, P. K. and T. Reed (2020). Income distribution, international integration, and sustained poverty reduction. Technical report, National Bureau of Economic Research.

Gollin, D., D. Lagakos, and M. E. Waugh (2014). The Agricultural Productivity Gap. *Quarterly Journal of Economics*.

Greenstone, M., R. Hornbeck, and E. Moretti (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy 118*(3), 536–598.

Grossman, G. and E. Helpman (1991). *Innovation and Growth in the Global Economy.* The MIT Press.

Helpman, E., O. Itskhoki, M.-A. Muendler, and S. J. Redding (2017). Trade and inequality: From theory to estimation. *The Review of Economic Studies 84*(1), 357–405.

Herrendorf, B., R. Rogerson, and Á. Valentinyi (2022). New evidence on sectoral labor productivity: Implications for industrialization and development. Technical report, National Bureau of Economic Research.

Herrendorf, B. and T. Schoellman (2018). Wages, Human Capital and Structural Transformation. *American Economic Journal: Macroeconomics*.

Huggins, D. R. and J. P. Reganold (2008). “No-Till: the Quiet Revolution”. *Scientific American 299*, 70–77.

IBGE (2006). *“Censo Agropecuário 2006”*. Rio de Janeiro, Brazil: Instituto Brasileiro de Geografia e Estatística (IBGE).

Imbert, C., M. Seror, Y. Zhang, and Y. Zylberberg (2020). Migrants and firms: Evidence from china.

Jaffe, A. B. (1988). Demand and supply influences in r & d intensity and productivity growth. *The Review of Economics and Statistics 70*(3), 431–437.

Krugman, P. (1987). The Narrow Moving Band, the Dutch Disease, and the Competitive Consequences of Mrs. Thatcher: Notes on Trade in the Presence of Dynamic Scale Economies. *Journal of Development Economics 27*(1-2), 41–55.

Kuznets, S. (1973). Modern Economic Growth: Findings and Reflections. *American Economic Review*.

Lagakos, D. and M. E. Waugh (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review 103*(2), 948–80.

Lagaras, S. (2017). Corporate takeovers and labor restructuring. *Available at SSRN 3222950*.

Lewis, A. W. (1954). Economic Development with Unlimited Supplies of Labor. *The Manchester School*.

Lucas, R. (1988). On the mechanics of economic development. *Journal of Monetary Economics*.

Matsuyama, K. (1992a). A Simple Model of Sectoral Adjustment. *The Review of Economic Studies*, 375–388.

Matsuyama, K. (1992b, December). Agricultural Productivity, Comparative Advantage, and Economic Growth. *Journal of Economic Theory 58*(2), 317–334.

McCaig, B. and N. Pavcnik (2013). Moving out of agriculture: structural change in vietnam. Technical report, National Bureau of Economic Research.

McMillan, M. and D. Rodrik (2011). Globalization, Structural Change and Productivity Growth. In *NBER Working Paper 17143*. ILO/WTO.

McMillan, M., D. Rodrik, and C. Sepulveda (2017). Structural Change, Fundamentals and Growth: A Framework and Case Studies. *National Bureau of Economic Research Working Paper Series No. 23378*.

Muendler, M.-A. (2002). Definitions of brazilian mining and manufacturing sectors and their conversion. Technical report.

Pakes, A. and M. Schankerman (1984). *An Exploration into the Determinants of Research Intensity*, pp. 209–232. University of Chicago Press.

Pellegrina, H. and S. Sotelo (2022). Migration, specialization, and trade: Evidence from brazil's march to the west.

Peters, M. (2019). Market size and spatial growth - evidence from germany's post-war population expulsions. *Unpublished manuscript*.

Reinbold, B. and P. Restrepo-Echavarria (2018). Measuring Labor Share in Developing Countries. *The Regional Economist 26*(1).

Restuccia, D., D. T. Yang, and X. Zhu (2008). Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis. *Journal of Monetary Economics 55*(2), 234 – 250.

Romer, P. (1990). Endogenous Technological Change. *Journal of Political Economy 98*(5).

Rybczynski, T. (1955). Factor endowments and relative commodity prices. *Economica*.

Ulyssea, G. (2018, August). Firms, informality, and development: Theory and evidence from brazil. *American Economic Review 108*(8), 2015–47.

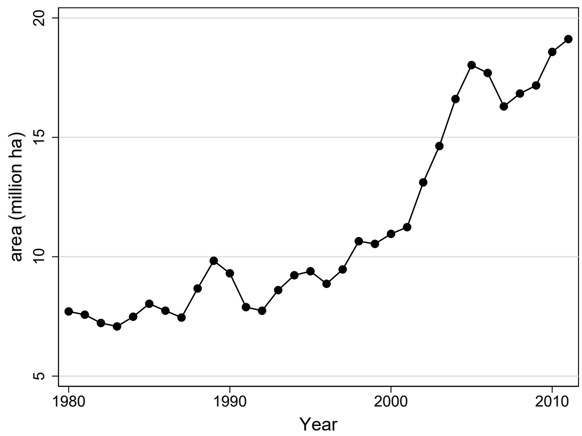
USDA (2001). "Agriculture in Brazil and Argentina: Developments and Prospects for Major Field Crops". United States Department of Agriculture, Economic Research Service.

USDA (2012). "Agricultural Biotechnology Annual". United States Department of Agriculture, Economic Research Service.

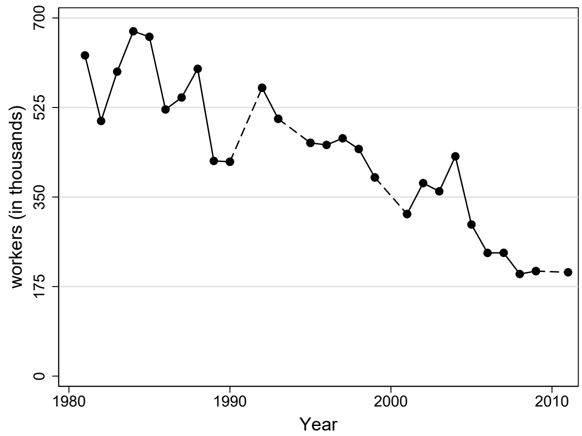
7 Figures and Tables

Figure 1: Soy Production and Employment

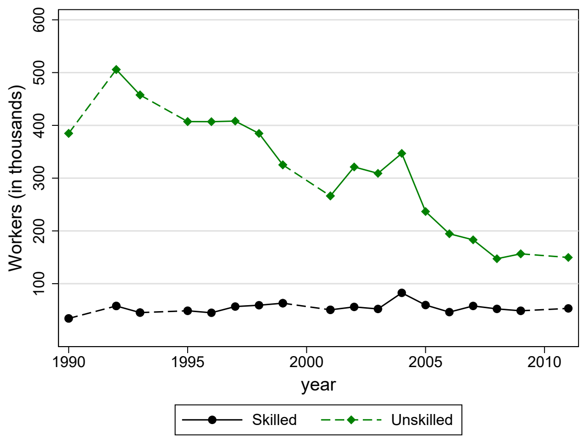
(a) Soy: Area planted



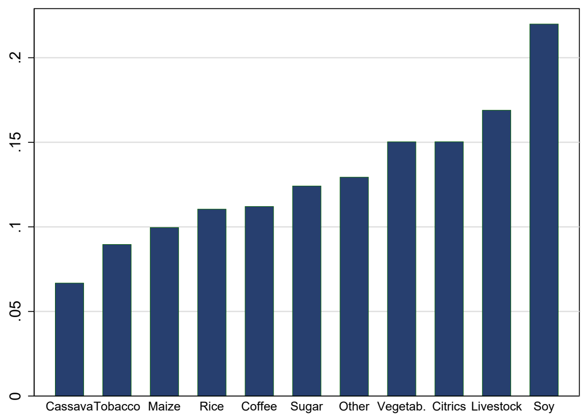
(b) Soy: Employment



(c) Soy: Employment by Skill Group



(d) Share of Skilled Workers by Agricultural Activity

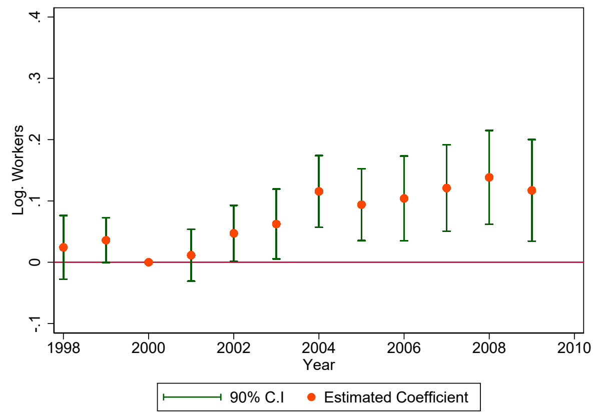


Notes: Figures in Panels (a) and (b) are from Bustos et al. (2016). Data sources are CONAB (Panel A), PNAD (Panel B and C) and 2000 Population Census (Panel D). CONAB is the Companhia Nacional de Abastecimento, an agency within the Brazilian Ministry of Agriculture, which runs surveys of farmers and agronomists to monitor the annual harvests of major crops in Brazil. PNAD is the Brazilian National Household Sample Survey. The states of Rondonia, Acre, Amazonas, Roraima, Pará, Amapá, Tocantins, Mato Grosso do Sul, Goias, and Distrito Federal are excluded due to incomplete coverage by PNAD in the early years of the sample. In Panels C and D, an individual is classified as skilled if she has completed at least the 8^{th} grade.

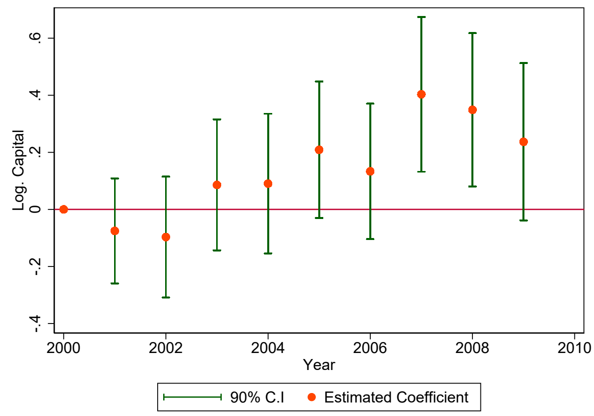
Figure 2: Effect of agricultural technical change on industrialization

Yearly data

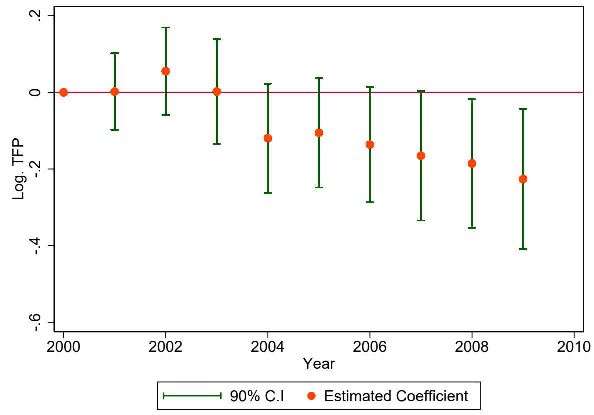
(a) Labor



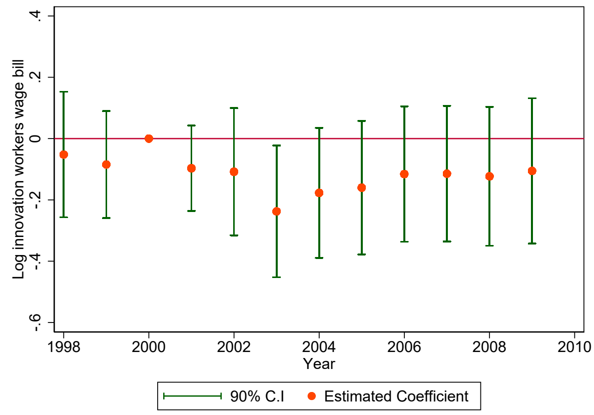
(b) Capital



(c) TFP



(d) Innovation

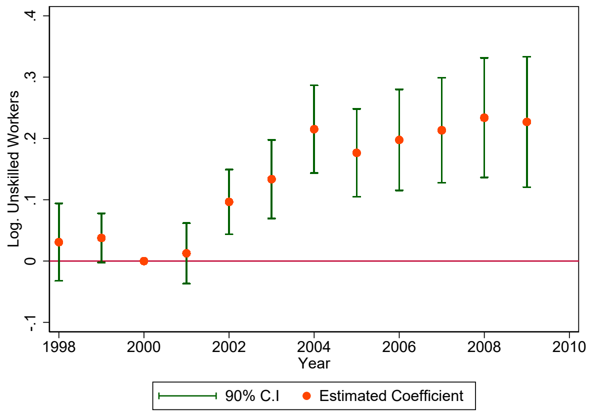


S **Notes**: The figure shows the point estimates and the 90% confidence intervals for the estimates of the β_j of equation (2) where $\ln y_{k,r,t}$ corresponds to aggregate log employment of unskilled and skilled labor in microregion k located in region r at the end of year t in manufacturing. An individual is classified as skilled if she has completed at least the 8^{th} grade. (Source: RAIS). Standard errors are clustered at the microregion level.

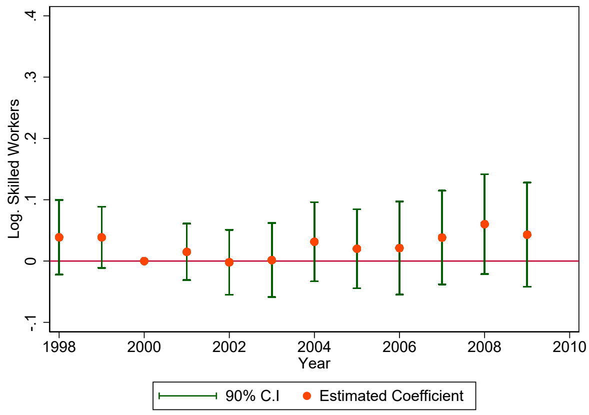
Figure 3: Effect of agricultural technical change on manufacturing employment, by skill

Yearly Social Security Data (1998-2009)

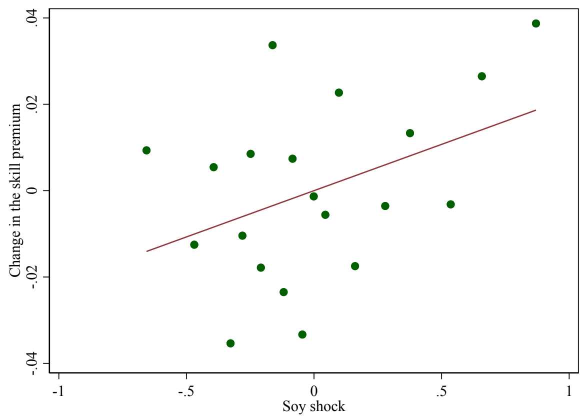
(a) **Unskilled Labor**



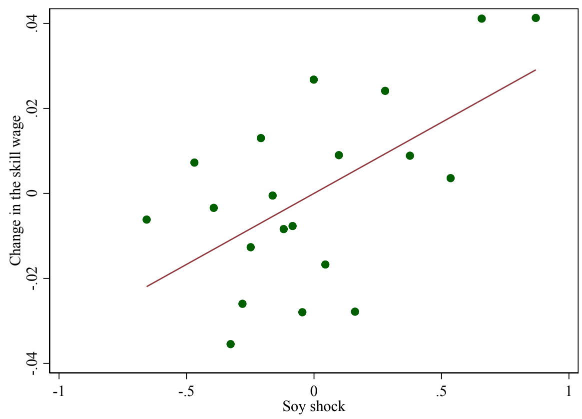
(b) **Skilled Labor**



(c) **Skill Premium**



(d) **Skill Wage**



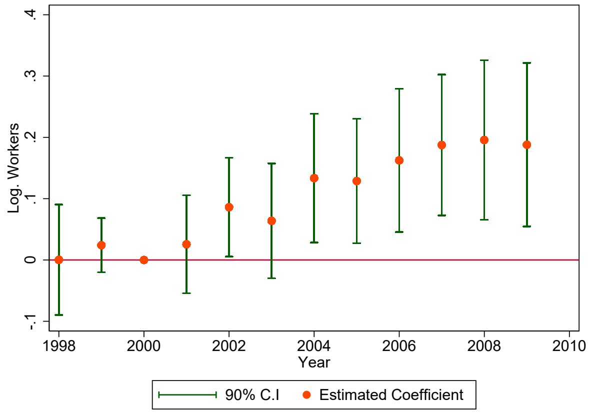
Notes: The figure shows the point estimates and the 90% confidence intervals for the estimates of the β_j of equation (2) where $\ln y_{k,r,t}$ corresponds to aggregate log employment of unskilled and skilled labor in microregion k located in region r at the end of year t in manufacturing. An individual is classified as skilled if she has completed at least the 8^{th} grade. (Source: RAIS). Standard errors are clustered at the microregion level.

Figure 4: Effect of agricultural technical change on labor and capital allocation within manufacturing

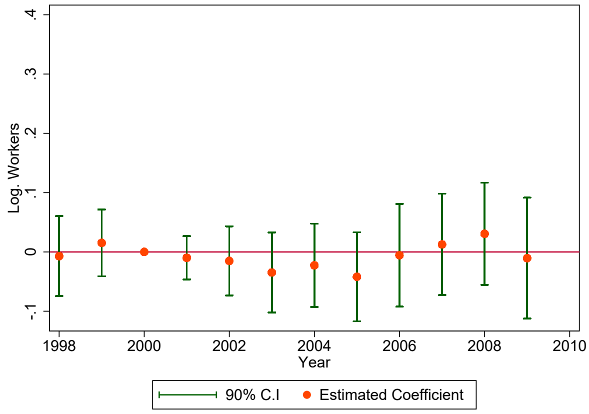
Yearly Social Security Data (1998-2009) and Annual Manufacturing Survey(2000-2009)

Labor

(a) **L Industry**

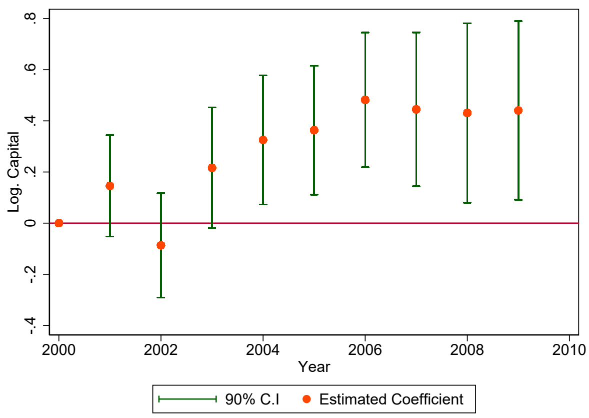


(b) **H Industry**

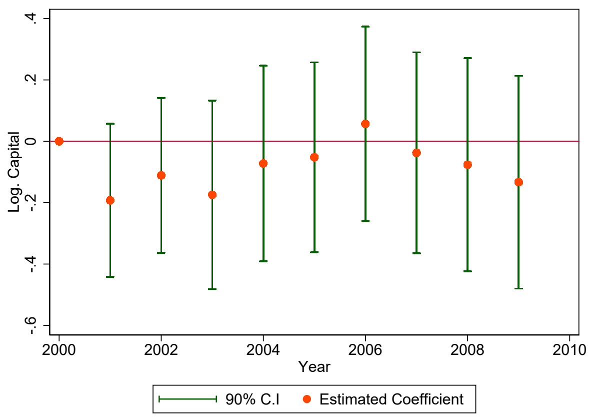


Physical Capital

(c) **L Industry**

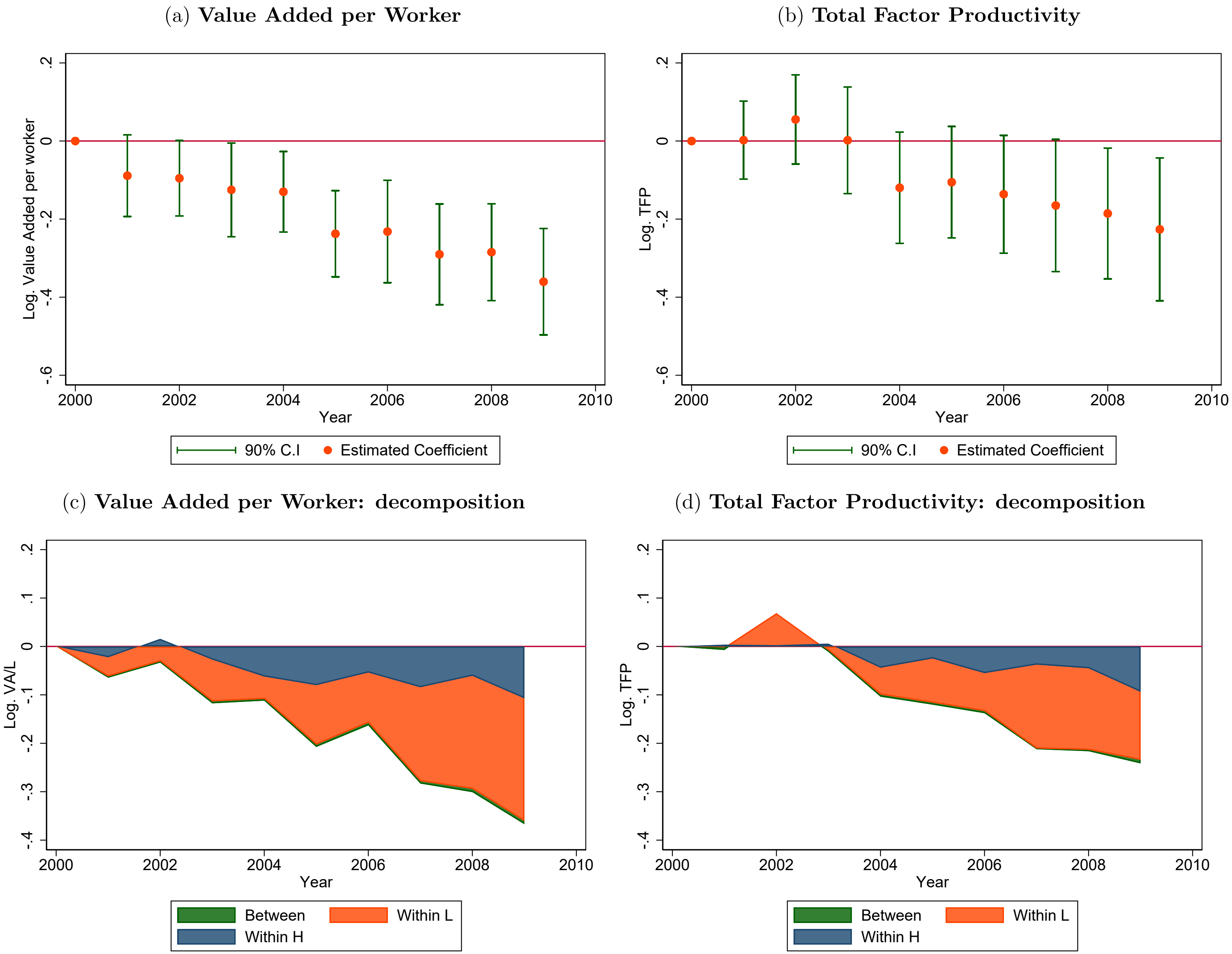


(d) **H Industry**



Notes: The figure shows the point estimates and the 90% confidence intervals for the estimates of the β_j coefficients of equation (2) where $\ln y_{k,r,t}$ corresponds to aggregate log employment and log capital in microregion k located in region r at the end of year t for each type of manufacturing industry (Source: PIA and RAIS). Manufacturing industries are classified as L or H depending on whether their R&D intensity is below or above the median in 2000 (weighting industries by number of employees so that each group captures around 50 percent of total manufacturing employment). We define R&D intensity as R&D expenditure as a share of total sales at baseline and we source it from from the 2000 *Pesquisa de Inovação Tecnológica* (PINTEC). Standard errors are clustered at the microregion level.

Figure 5: The effect of agricultural technical change on manufacturing productivity



(a) **Value Added per Worker**

(b) **Total Factor Productivity**

(c) **Value Added per Worker: decomposition**

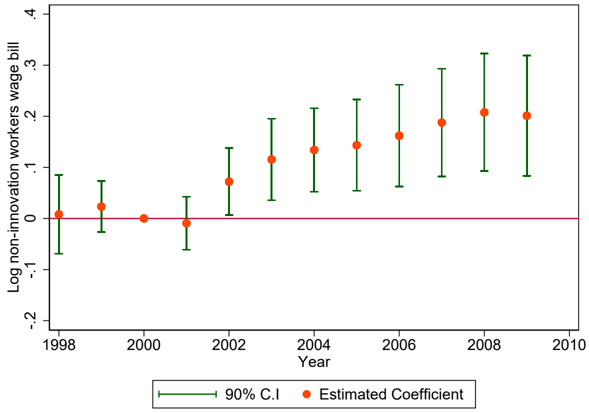
(d) **Total Factor Productivity: decomposition**

Notes: Graphs (a) and (b) show the point estimates and 90% confidence intervals for the estimates of the β_j coefficients of equation (2) using two measures of manufacturing productivity as outcomes: log value added per worker and log TFP (Source: PIA). Graphs (c) and (d) show the decomposition of the total effect into the within components for each of the two industries and the between component across industries. Standard errors are clustered at the microregion level.

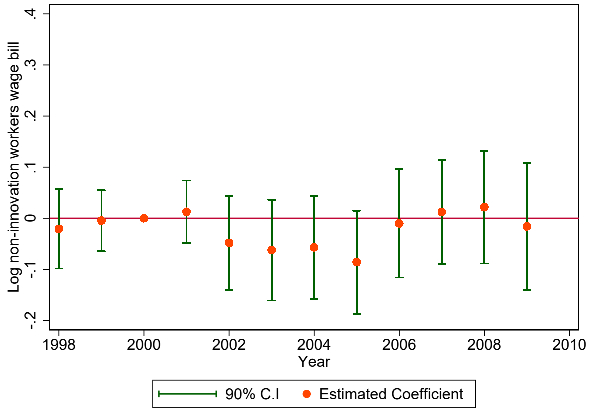
Figure 6: Effect of agricultural technical change on expenditure on non-innovative and innovative occupations
Yearly Social Security Data (1998-2009)

Non-innovative labor wage bill

(a) **L Industry**

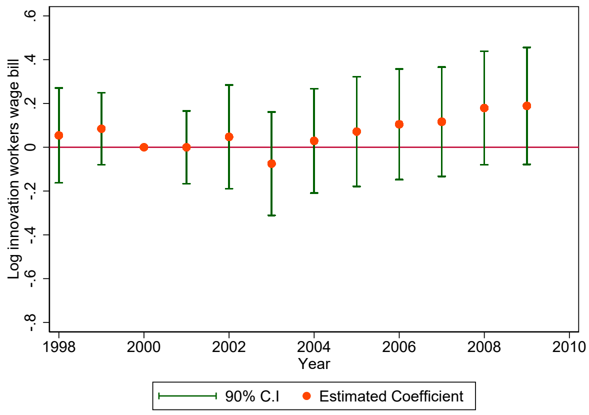


(b) **H Industry**

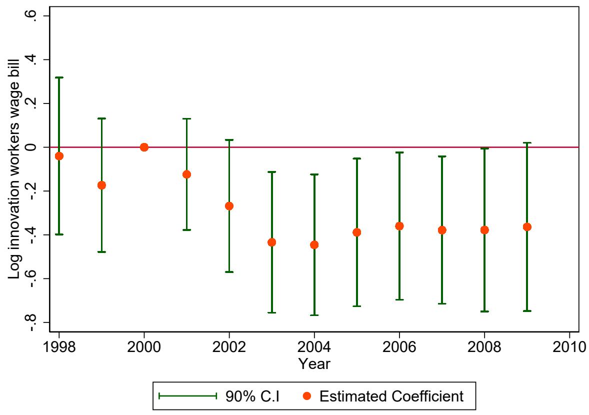


Innovative labor wage bill

(c) **L Industry**



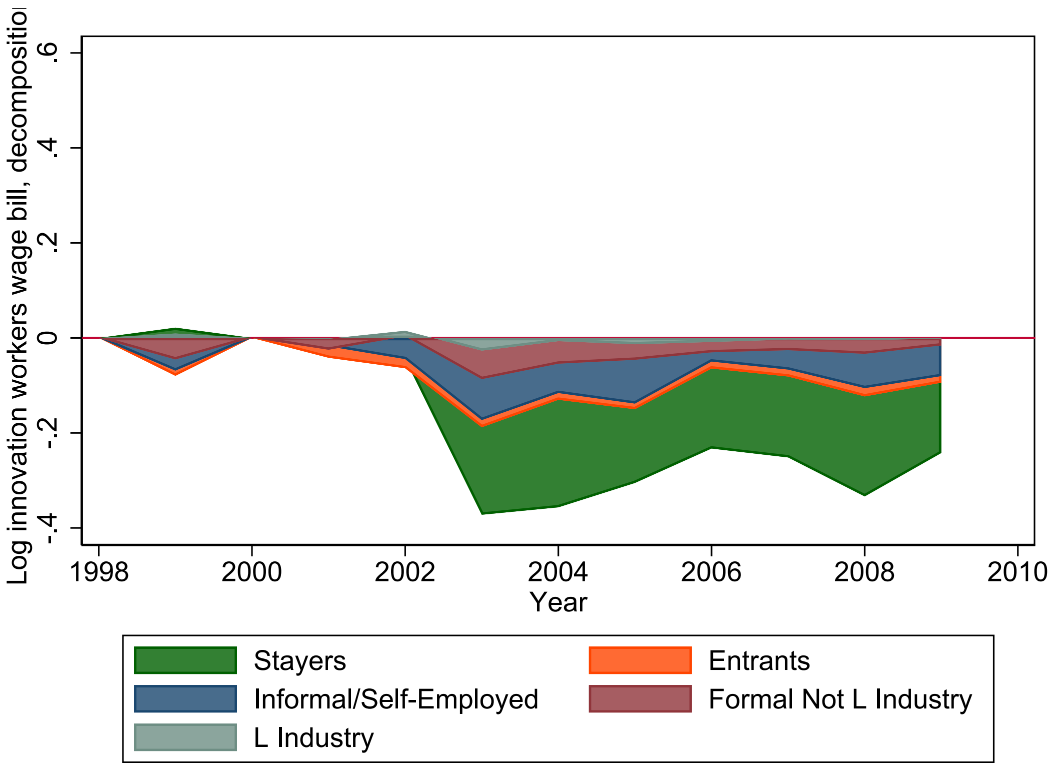
(d) **H Industry**



Notes: The figure shows the point estimates and the 90% confidence intervals for the estimates of the β_j coefficients of equation (2) where $\ln y_{k,t}$ corresponds to the log wage bill of non-innovative and innovative labor in microregion k located in region r at the end of year t for L and H manufacturing industries (Source: RAIS). An occupation is classified as innovative following the methodology outlined in Section 3.4. Manufacturing industries are classified as L or H depending on whether their R&D intensity is below or above the median in 2000 (weighting industries by number of employees so that each group captures around 50 percent of total manufacturing employment). We define R&D intensity as R&D expenditure as a share of total sales at baseline and we source it from from the 2000 *Pesquisa de Inovação Tecnológica* (PINTEC). Standard errors are clustered at the microregion level.

Figure 7: Decomposition of effect of agricultural technical change on innovation activities in the *H* industry

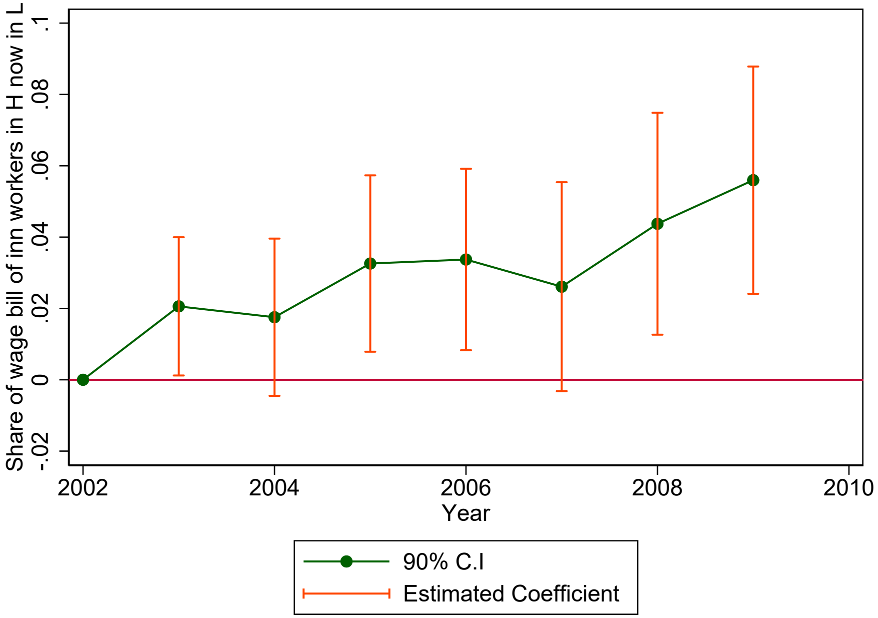
Yearly Social Security Data (1998-2009)



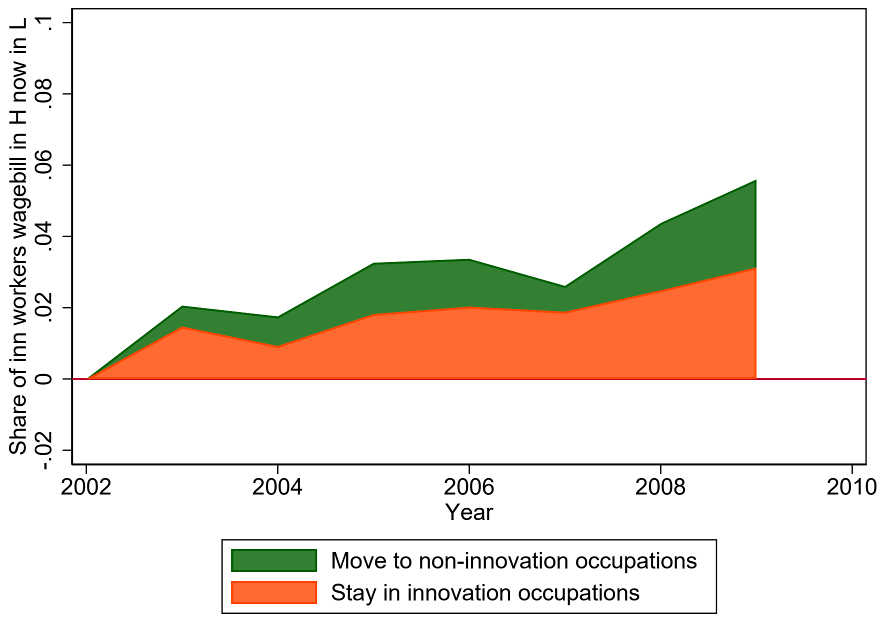
Notes: The graph shows the decomposition of the total effect of agricultural technical change on log wage bill of innovative workers in the H manufacturing industry into components coming from workers who stay in the industry, workers transitioning from either informality or self employment, workers transitioning from the L industry, workers who are entering the labor force, and workers who transition from formal sectors other than the L industry. Manufacturing industries are classified as L or H depending on whether their R&D intensity is below or above the median in 2000 (weighting industries by number of employees so that each group captures around 50 percent of total manufacturing employment). We define R&D intensity as R&D expenditure as a share of total sales at baseline and we source it from from the 2000 *Pesquisa de Inovação Tecnológica* (PINTEC). Standard errors are clustered at the microregion level.

Figure 8: Agricultural technical change and the reallocation of innovation activities across industries and occupations

(a) **Reallocation of innovative workers' wage bill from H to L**



(b) **Decomposition by new occupation in the L industry**



Notes: Graph (a) shows the point estimates and the 90% confidence intervals for the estimates of the β_j coefficients of equation (2) where the outcome variable is the share of the wage bill of workers who were employed in innovative occupations in the H sector during a year between 1998 and 2002 that moved to the L industry in microregion k located in region r at the end of year t (Source: RAIS). Graph (b) decomposes Graph (a) by the new occupation in the L industry of workers who were employed in innovative occupations in the H sector during a year between 1998 and 2002. An occupation is classified as innovative following the methodology outlined in Section 3.4. Manufacturing industries are classified as L or H depending on whether their R&D intensity is below or above the median in 2000 (weighting industries by number of employees so that each group captures around 50 percent of total manufacturing employment). We define R&D intensity as R&D expenditure as a share of total sales at baseline and we source it from from the 2000 *Pesquisa de Inovação Tecnológica* (PINTEC). Standard errors are clustered at the microregion level.

Table 1: Summary Statistics of the Sample of Individuals by Sector

	2000	**2010**
Agriculture		
Age	38.0	39.0
Male (% of the Total)	89.3	81.2
White (% of the Total)	55.4	48.6
Education level (highest degree obtained)		
Less than Middle School (% of the Total)	86.1	72.7
Completed Middle School (% of the Total)	7.4	13.8
High School Graduates (% of the Total)	5.2	11.4
University Graduates (% of the Total)	1.3	2.1
Average log real hourly wage	0.81	1.06
For skilled labor	1.39	1.38
For unskilled labor	0.71	0.95
Manufacturing L Industry		
Age	36.8	37.3
Male (% of the Total)	61.6	58.7
White (% of the Total)	65.0	55.6
Education level (highest degree obtained)		
Less than Middle School (% of the Total)	52.2	36.8
Completed Middle School (% of the Total)	20.4	21.5
High School Graduates (% of the Total)	21.9	35.2
University Graduates (% of the Total)	5.5	6.6
Average log real hourly wage	1.23	1.51
For skilled labor	1.73	1.63
For unskilled labor	1.15	1.23
Manufacturing H Industry		
Age	36.28	36.9
Male (% of the Total)	80.6	76.2
White (% of the Total)	63.0	55.2
Education level (highest degree obtained)		
Less than Middle School (% of the Total)	49.8	31.3
Completed Middle School (% of the Total)	20.0	19.8
High School Graduates (% of the Total)	23.4	39.8
University Graduates (% of the Total)	6.8	9.1
Average log real hourly wage	1.58	1.66
For skilled labor	1.92	1.81
For unskilled labor	1.24	1.35
Services		
Age	37.1	37.8
Male (% of the Total)	67.3	62.1
White (% of the Total)	58.9	50.8
Education level (highest degree obtained)		
Less than Middle School (% of the Total)	51.1	36.0
Completed Middle School (% of the Total)	17.9	19.3
High School Graduates (% of the Total)	23.4	34.3
University Graduates (% of the Total)	7.6	10.4
Average log real hourly wage	1.42	1.51
For skilled labor	1.77	1.67
For unskilled labor	1.01	1.24

Notes: The data comes from the Population Censuses for years 2000 and 2010. Summary statistics refer to our final sample of individuals as detailed in Section 3.3. An individual is classified as skilled if she has at least completed the 8^{th} grade. This level should be attained when an individual is 14 or 15 years old and is equivalent to graduating from middle school. Manufacturing industries are classified as L or H intensive depending on whether their R&D intensity is below or above the median in 2000 (weighting industries by number of employees so that each group captures around 50 percent of total manufacturing employment). We define R&D intensity as R&D expenditure as a share of total sales at baseline and we source it from from the 2000 *Pesquisa de Inovação Tecnológica* (PINTEC).

Table 2: Summary Statistics of the Sample of Microregions

		Panel A: Decadal Variables				
		2000		Δ2000-2010		
	Source:	Mean	SD	Mean	SD	Observations
Potential Yields	*FAO-GAEZ*					
Soy		0.286	0.135	1.787	0.740	557
Maize		1.847	0.9984	3.082	1.639	557
Employment Shares	*Population Census*					
Agriculture		0.279	0.140	-0.050	0.055	557
Manufacturing L Industry		0.081	0.055	0.007	0.033	557
Manufacturing H Industry		0.067	0.043	-0.001	0.025	557
Services		0.573	0.118	0.044	0.057	557
Log. Employment	*Population Census*					
Agriculture		8.268	0.890	0.122	0.249	557
Manufacturing L Industry		7.076	1.569	0.358	0.400	557
Manufacturing H Industry		6.897	1.485	0.309	0.394	557
Services		9.194	1.887	0.404	0.175	557

		Panel B: Yearly Variables		
	Source:	Mean	SD	Observations
Manufacturing Employment	*RAIS (1998-2009)*			
Log. Employment				
Manufacturing L Industry		7.753	1.315	3,816
Manufacturing H Industry		7.509	1.384	3,816
Log. Non-Innovative Wage Bill				
Manufacturing L Industry		16.103	2.206	3,816
Manufacturing H Industry		15.883	2.304	3,816
Log. Innovative Wage Bill				
Manufacturing L Industry		12.781	2.869	3,816
Manufacturing H Industry		12.523	3.273	3,816
Manufacturing Productivity	*PIA (2000-2009)*			
Log. Value Added per Worker				
Manufacturing L Industry		10.692	0.866	3,070
Manufacturing H Industry		10.536	0.944	3,070
Log. Value Added per Wage Bill				
Manufacturing L Industry		1.537	0.593	3,070
Manufacturing H Industry		1.360	0.613	3,070
Log. Total Factor Productivity				
Manufacturing L Industry		5.623	0.788	3,035
Manufacturing H Industry		5.722	0.752	2,950
Log. Capital	*PIA (2000-2009)*			
Manufacturing L Industry		17.812	2.216	3,037
Manufacturing H Industry		16.359	2.580	2,971

Notes: The data sources are the Population Census (2000, 2010), RAIS and PIA. Manufacturing industries are classified as Low-R&D or High-R&D intensive depending on whether their R&D intensity is below or above the median in 2000 (weighting industries by number of employees so that each group captures around 50 percent of total manufacturing employment). We define R&D intensity as R&D expenditure as a share of total sales at baseline and we source it from from the 2000 *Pesquisa de Inovação Tecnológica* (PINTEC). A worker is classified as skilled if she has completed at least the 8^{th} grade (completed middle school).

Table 3: Effect of agricultural technical change on sectoral employment shares, manufacturing growth and manufacturing productivity

Panel A: Decadal Population Census Data (2000 and 2010)

Outcome:	Change in employment shares by sector		
	Agriculture (1)	Manufacturing (2)	Services (3)
ΔA_{soy}	-0.033***	0.025***	0.008*
	[0.005]	[0.005]	[0.004]
Observations	557	557	557
R-squared	0.246	0.166	0.359
Region FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Panel B: Yearly Manufacturing Survey Data (2000-2009)

Outcome:	Manufacturing Outcomes				
	Labor (1)	Capital (2)	VA/L (3)	VA/WL (4)	TFP (5)
A_{soy}	0.095***	0.257**	-0.141***	-0.133***	-0.173***
	[0.035]	[0.092]	[0.043]	[0.040]	[0.066]
Observations	3,070	3,069	3,070	3,070	3,069
R-squared	0.977	0.913	0.876	0.735	0.542
Microregion FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Controls × Linear trends	Yes	Yes	Yes	Yes	Yes
Region x Year FEs	Yes	Yes	Yes	Yes	Yes

Notes: Panel A shows coefficient estimates corresponding to equation (1). Changes in dependent variables are calculated over the years 2000 and 2010. The unit of observation is the microregion. These regressions include as controls the share of rural population, income per capita (in logs), population density (in logs), literacy rate, all observed in the 1991 Population Census, a measure of technical change in maize and region fixed effects. Panel B shows coefficient estimates corresponding to equation (3). The dependent variables correspond to the total labor (in logs), total capital (in logs), total value added divided by employment (in logs), total value added divided by total wage bill (in logs) and total factor productivity for manufacturing in each microregion. We include only those microregions that have positive employment for all the years in the sample. Controls include the share of rural population, income per capita (in logs), population density (in logs), literacy rate, all observed in 1991, all interacted with a linear trend, a measure of technical change in maize and region times year fixed effects. Robust standard errors are reported in brackets in Panel A, and standard errors clustered at the microregion level are reported in brackets in Panel B. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 4: Effect of agricultural technical change on sectoral employment shares by skill group

Decadal Population Census Data (2000 and 2010)

Outcome:	Change in employment shares of unskilled workers by sector			Change in employment shares of skilled workers by sector		
Sector:	Agriculture	Manufacturing	Services	Agriculture	Manufacturing	Services
	(1)	(2)	(3)	(4)	(5)	(6)
ΔA_{soy}	-0.033***	0.030***	0.004	-0.015***	0.014***	0.001
	[0.006]	[0.005]	[0.004]	[0.004]	[0.005]	[0.005]
Observations	557	557	557	557	557	557
R-squared	0.126	0.157	0.208	0.047	0.112	0.103
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Changes in dependent variables are calculated over the years 2000 and 2010 (source: Population Censuses). The unit of observation is the microregion. All the regressions include as controls the share of rural population, income per capita (in logs), population density (in logs), literacy rate, all observed in the 1991 Population Census, a measure of technical change in maize and region fixed effects. Robust standard errors reported in brackets. Significance levels: $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table 5: Effect of agricultural technical change on industrial specialization within manufacturing

Panel A: Decadal Population Census Data (2000 and 2010)

Outcome:	Change in employment by manufacturing industry	
Industry:	L Industry	H Industry
	(1)	(2)
ΔA_{soy}	0.233***	0.015
	[0.038]	[0.036]
Observations	557	557
R-squared	0.129	0.099
Region FE	Yes	Yes
Controls	Yes	Yes

Panel B: Yearly Social Security Data (1998-2009)

Outcome:	Employment by manufacturing industry	
Industry:	L Industry	H Industry
	(1)	(2)
A_{soy}	0.121**	-0.008
	[0.050]	[0.037]
Observations	5,640	5,640
R-squared	0.455	0.410
Microregion FE	Yes	Yes
Year FE	Yes	Yes
Controls $\times$ Linear trends	Yes	Yes
Region x Year FEs	Yes	Yes

Panel C: Yearly Manufacturing Survey Data (2000-2009)

Outcome:	Capital by manufacturing industry	
Industry:	L Industry	H Industry
	(1)	(2)
A_{soy}	0.360***	0.027
	[0.112]	[0.115]
Observations	3,037	2,969
R-squared	0.891	0.894
Microregion FE	Yes	Yes
Year FE	Yes	Yes
Controls $\times$ Linear trends	Yes	Yes
Region x Year FEs	Yes	Yes

Notes: In Panel A the dependent variables are changes in total employment (in logs) calculated over the years 2000 and 2010 (source: Population Censuses). The unit of observation is the microregion. Controls include: share of rural population in 1991, income per capita (in logs), population density (in logs), and literacy rate, all observed in the 1991 Population Census, as well as a measure of technical change in maize and region fixed effects. In Panel B, the dependent variable is total employment (in logs) for each manufacturing industry in each microregion. We use aggregate information from RAIS at the microregion-industry level for the time period 1998-2009. We include only those microregions that have positive employment for all the years in the sample. In Panel C, the dependent variable is capital (in logs) for each manufacturing industry in each microregion. We use aggregate information from PIA at the microregion level for the time period 2000-2009. In Panels B and C, A_{soy} is defined as potential soy yield under high inputs for the years between 2003 and 2009, and the potential soy yield under low inputs for the years between 2000 and 2002. Controls include the share of rural population, income per capita (in logs), population density (in logs), literacy rate, all observed in 1991, all interacted with a linear trend, a measure of technical change in maize and region year fixed effects. In these regressions, manufacturing industries are classified as L or H depending on whether their R&D intensity is below or above the median in 2000 (weighting industries by number of employees so that each group captures around 50 percent of total manufacturing employment). We define R&D intensity as R&D expenditure as a share of total sales at baseline and we source it from from the 2000 *Pesquisa de Inovação Tecnológica* (PINTEC). Robust standard errors are reported in Panel A, and standard errors clustered at the microregion level are reported in Panels B and C. Significance levels: $^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.1$.

Table 6: Effect of agricultural technical change on manufacturing productivity

Yearly Manufacturing Survey Data (2000-2009)

Outcomes:	L Industry Productivity			H Industry Productivity		
Measure:	Log Value Added per Worker	Log Value Added per Wage Bill	Log TFP	Log Value Added per Worker	Log Value Added per Wage Bill	Log TFP
	(1)	(2)	(3)	(4)	(5)	(6)
A_{soy}	-0.151**	-0.135**	-0.251***	-0.119*	-0.109*	-0.128*
	[0.059]	[0.054]	[0.077]	[0.071]	[0.057]	[0.073]
Observations	3,070	3,070	3,035	3,070	3,070	2,949
R-squared	0.796	0.627	0.590	0.799	0.635	0.572
Region x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variables are: total value added divided by employment (in logs), total value added divided by total wage bill (in logs) and total factor productivity for each type of manufacturing industry in each microregion as a proxy for productivity. We include only those microregions that have positive employment for all the years in the sample. A_{soy} is defined as potential soy yield under high inputs for the years between 2003 and 2009, and the potential soy yield under low inputs for the years between 2000 and 2002. Controls include the share of rural population, income per capita (in logs), population density (in logs), literacy rate, all observed in 1991, all interacted with a linear trend, a measure of technical change in maize and region year fixed effects. The unit of observation is a microregion. In these regressions, manufacturing industries are classified as L or H depending on whether their R&D intensity is below or above the median in 2000 (weighting industries by number of employees so that each group captures around 50 percent of total manufacturing employment). We define R&D intensity as R&D expenditure as a share of total sales at baseline and we source it from from the 2000 *Pesquisa de Inovação Tecnológica* (PINTEC). Standard errors clustered at the microregion level reported in parentheses. Significance levels: $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table 7: Worker-level trajectories from agriculture to manufacturing, and wages of former agricultural workers

Sample:	Employed in Agriculture pre-2003		Employed in L industry from 2003, in Agriculture or L-industry pre-2003	
Outcomes:	Move to L industry	Move to H industry	log(wage)	
				composition-adjusted
	(1)	(2)	(3)	(4)
ΔA_{soy} in pre-2003 microregion	0.050***	-0.008***		
	[0.009]	[0.002]		
1(Employed in Agriculture pre-2003)			-0.271***	-0.118***
			[0.020]	[0.017]
Observations	1,304,659	1,304,659	951,453	951,453
R-squared	0.025	0.003	0.205	0.155
Microregion Controls	Yes	Yes	Yes	Yes

Notes: Microregion Controls include the share of rural population, income per capita (in logs), population density (in logs), literacy rate, the share of individuals earning minimum wage, all observed in 1991, and a measure of technical change in maize. Standard errors clustered at micro-region level reported in brackets. Significance levels: $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table 8: Effect of agricultural technical change on innovation in manufacturing

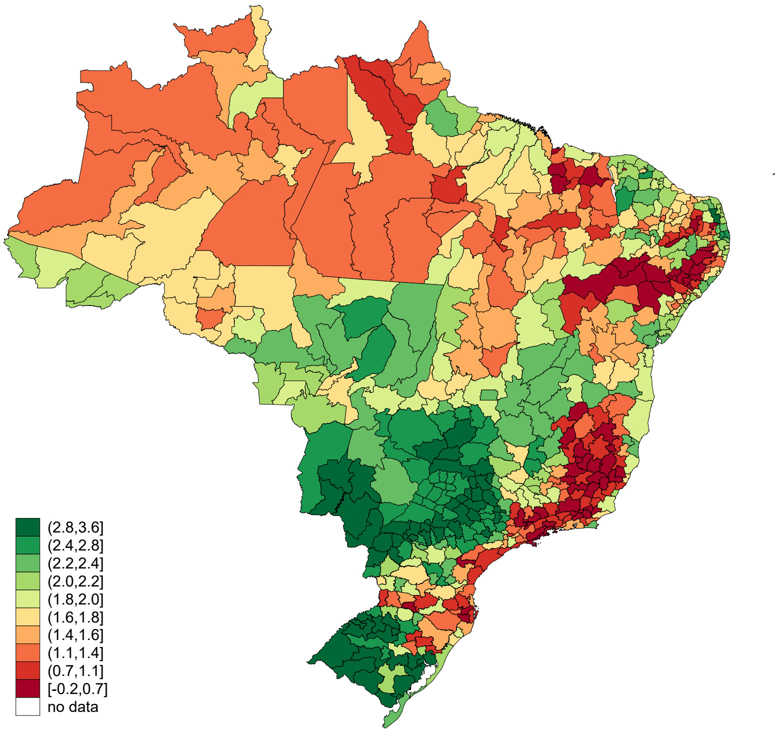
Yearly Social Security Data (1998-2009)

VARIABLES	(1) $\log wL_m^{non-inno}$ Low R&D	(2) $\log wL_m^{non-inno}$ High R&D	(3) $\log wL_m^{inno}$ Low R&D	(4) $\log wL_m^{inno}$ High R&D
A_soy	0.144***	-0.019	0.048	-0.274*
	[0.047]	[0.046]	[0.107]	[0.151]
Observations	3,828	3,828	3,828	3,828
R-squared	0.969	0.978	0.910	0.905
Baseline Controls	Yes	Yes	Yes	Yes
All Controls	Yes	Yes	Yes	Yes

Notes: The dependent variables in columns (1) and (2) are the total wage bill of non-innovation workers (in logs) for each manufacturing industry in each microregion as a proxy for industry size, and in columns (3) and (4) are the total wage bill of innovation workers (in logs) for each type of industry in every microregion as a proxy for expenditure in innovation. We use aggregate information from RAIS at the microregion-industry level for the time period 1998-2009. We include only those microregions that have positive employment for all the years in the sample. A^{soy} is defined as potential soy yield under high inputs for the years between 2003 and 2009, and the potential soy yield under low inputs for the years between 1998 and 2002. Controls include the share of rural population, income per capita (in logs), population density (in logs), literacy rate, all observed in 1991, all interacted with a linear trend, a measure of technical change in maize and region year fixed effects. The unit of observation is a microregion. In these regressions, manufacturing industries are classified as Low-R&D or High-R&D intensive depending on whether their R&D intensity is below or above the median in 2000 (weighting industries by number of employees so that each group captures around 50 percent of total manufacturing employment). We define R&D intensity as R&D expenditure as a share of total sales at baseline and we source it from from the 2000 *Pesquisa de Inovação Tecnológica* (PINTEC). Standard errors clustered at the microregion level reported in parentheses. Significance levels: $^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.1$.

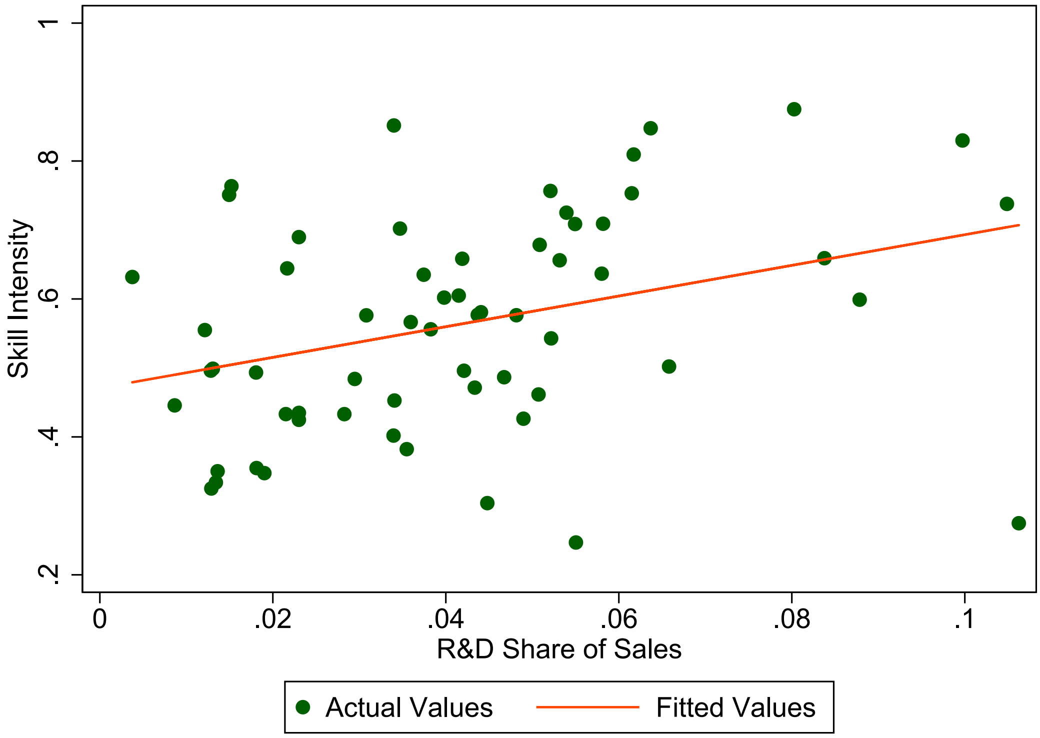
A Appendix: Figures and Tables

Figure A.1: Δ in Potential Soy Yield 2000-2010



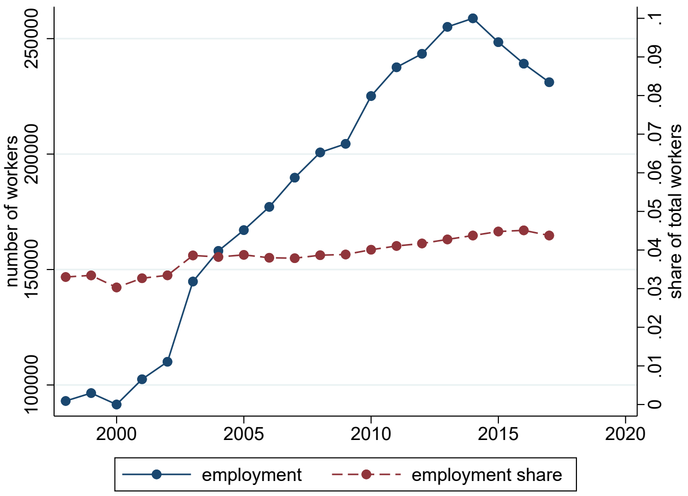
Notes: Authors' calculations from FAO-GAEZ data. Technical change in soy production for each microregion is computed by deducting the average potential yield under low inputs from the average potential yield under high inputs.

Figure A.2: Skill Intensity and R&D Intensity at Industry Level



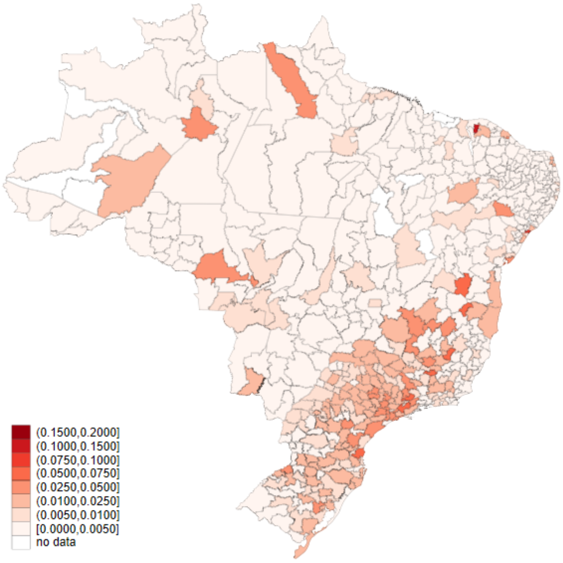
Notes: We define skill intensity as the share of skilled individuals in a particular industry in Brazil at baseline and we source it from the 2000 Population Census. Our measure of R&D activity is R&D expenditure as a share of total sales at baseline and we source it from from the 2000 *Pesquisa de Inovação Tecnológica](PINTEC)*. The correlation between these variables is approximately 0.34.

Figure A.3: Manufacturing Employment in Innovation Intensive Occupations



Notes: Authors' calculations using RAIS data. Innovation intensive occupations are defined using the methodology described in Section 3.4.

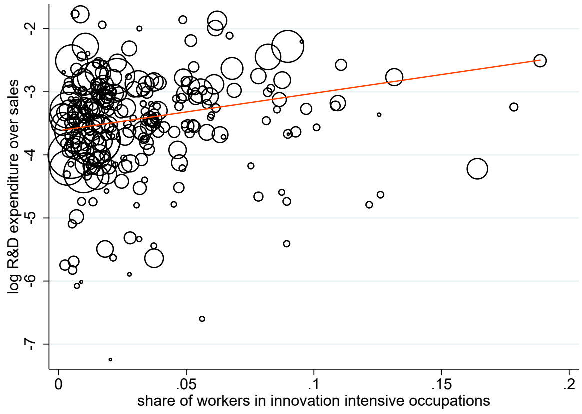
Figure A.4: Geographical distribution of share of manufacturing workers in innovation intensive occupations in 2000



Notes: The Figure reports the share of innovation intensive workers over total workers in the manufacturing sector in the year 2000 by microregion. Innovation intensive occupations are defined using the methodology described in Section 3.4.

Figure A.5: Correlations between share of workers in innovation intensive occupations and industry-level measures of innovation

(a) R&D expenditure/sales

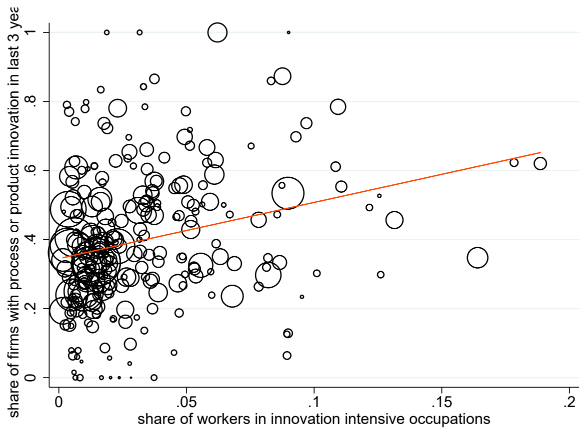


(b) Patent share

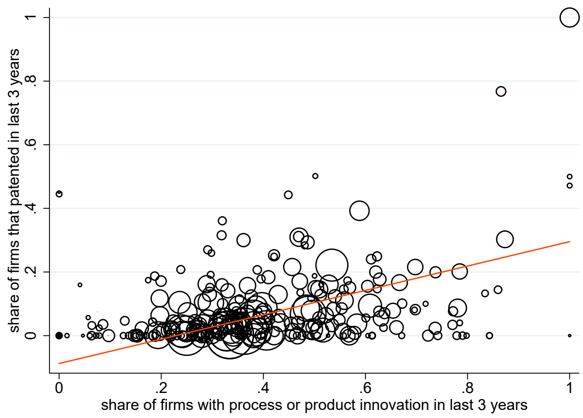
(c) Product innovation share

(d) Process innovation share

(e) Product or Process innovation share



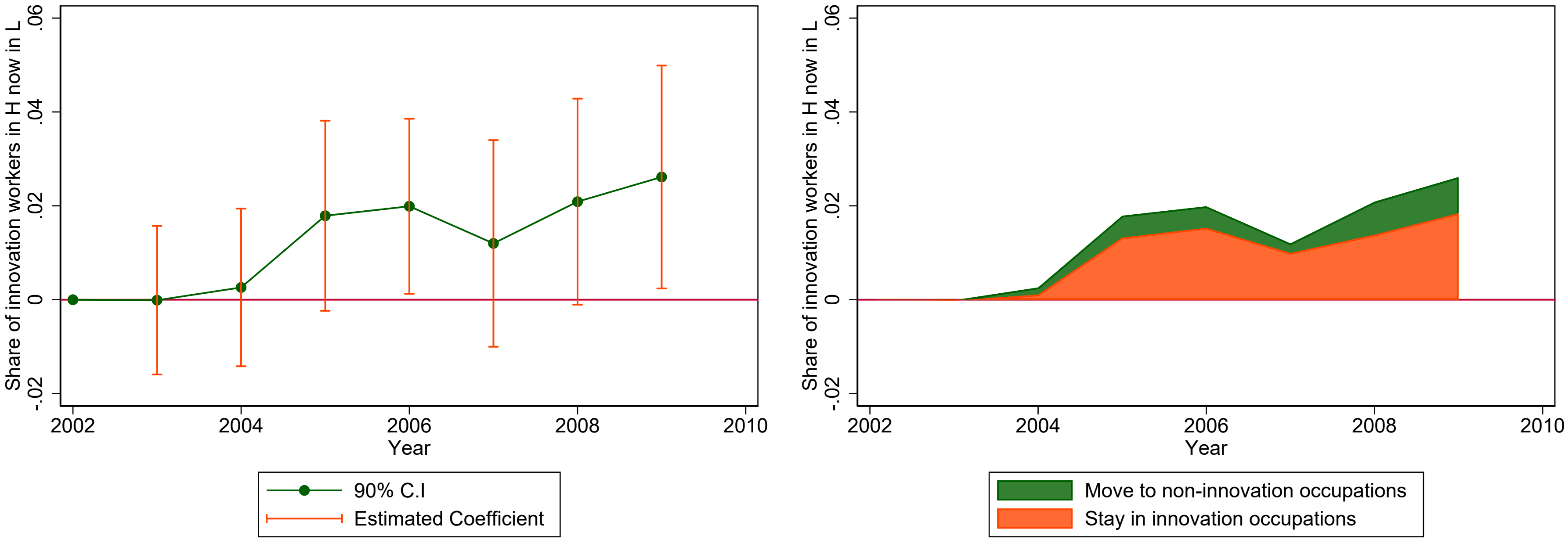
(f) Patents vs Product or Process innovation share



Notes: The share of workers in innovation intensive occupations in each sector is constructed using RAIS data for the year 2000 and the methodology described in Section 3.4. All measures of innovation at industry level (R&D expenditure over sales, patent share, product and process innovation share) are computed using the 2000 *Pesquisa de Inovação Tecnológica (PINTEC)*. Dot size captures size of the industry in terms of number of employees in 2000.

Figure A.6: Agricultural technical change and the reallocation of innovation workers across industries and occupations

(a) **Reallocation of innovative workers from H to L Industry** (b) **Decomposition by new occupation in the L industry**



Notes: Graph (a) shows the point estimates and the 90% confidence intervals for the estimates of the β_j coefficients of equation (2) where the outcome variable is the share of workers who were employed in innovative occupations in the H sector during a year between 1998 and 2002 that moved to the L industry in microregion k located in region r at the end of year t (Source: RAIS). Graph (b) decomposes Graph (a) by the new occupation in the L industry of workers who were employed in innovative occupations in the H sector during a year between 1998 and 2002. An occupation is classified as innovative following the methodology outlined in Section 3.4. Manufacturing industries are classified as L or H depending on whether their R&D intensity is below or above the median in 2000 (weighting industries by number of employees so that each group captures around 50 percent of total manufacturing employment). We define R&D intensity as R&D expenditure as a share of total sales at baseline and we source it from from the 2000 *Pesquisa de Inovação Tecnológica* (PINTEC). Standard errors are clustered at the microregion level.

Table A.1: Effect of Agricultural Technical Change on GE Soy Adoption

Panel A	Δ GE-soy area share (1)	Δ GE-soy area share (2)	Δ non-GE soy area share (3)	Δ non-GE soy area share (4)
ΔA^{soy}	0.022***	0.020***	-0.007*	-0.008**
	[0.005]	[0.004]	[0.004]	[0.004]
Share rural population	0.034***	0.117***	-0.009	-0.057**
	[0.010]	[0.023]	[0.009]	[0.023]
Log Income per capita		-0.009		-0.002
		[0.006]		[0.007]
Literacy rate		0.162***		-0.043
		[0.034]		[0.035]
Log population density		0.005***		-0.006***
		[0.001]		[0.001]
Observations	557	557	557	557
R-squared	0.094	0.208	0.013	0.053

Panel B	Δ Soy area share (1)	Δ Soy area share (2)	Δ Maize area share (3)	Δ Maize area share (4)
ΔA^{soy}	0.022***	0.016***	-0.006	0.000
	[0.004]	[0.004]	[0.004]	[0.004]
ΔA^{maize}	-0.003**	-0.001	0.005***	0.003*
	[0.001]	[0.001]	[0.002]	[0.002]
Share rural population	0.029***	0.064***	0.020***	0.012
	[0.007]	[0.013]	[0.008]	[0.015]
Log Income per capita		-0.010*		-0.011
		[0.006]		[0.007]
Literacy rate		0.122***		-0.002
		[0.018]		[0.023]
Log population density		-0.001		0.003***
		[0.001]		[0.001]
Observations	557	557	556	556
R-squared	0.135	0.245	0.041	0.066

Notes: Changes in dependent variables are calculated over the years 1996 and 2006 (source: Agricultural Census). The unit of observation is the microregion. Robust standard errors reported in brackets. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.2: Classification of Manufacturing Industries by R&D Intensity

IBGE Code	Description	R&D Share of Sales	Skill Intensity
26091	Ceramic products	0.106	0.275
34001	Manufacturing and assembly of motor vehicles	0.105	0.738
23030	Production of nuclear fuels	0.100	0.830
31002	Electrical material for vehicles	0.088	0.599
27001	Steel products	0.084	0.659
35030	Construction, assembly and repair of airplanes	0.080	0.875
28002	Foundries, stamping shops, powder metallurgy and metal treatment services	0.066	0.502
33003	Machines, equipment for electronic systems for industrial automation, and control	0.064	0.848
24020	Pharmaceutical products	0.062	0.809
33001	Medical equipment	0.061	0.753
29002	Appliances	0.058	0.709
34002	Cabins, car bodies, trailers and parts for motor vehicles	0.058	0.637
20000	Wooden products	0.055	0.247
33004	Equipment, instruments and optical, photographic and cinematographic material	0.055	0.709
33002	Measuring, testing and control equipment - except for controlling industrial processes	0.054	0.725
24010	Paints, dyes, varnish, enamels and lacquers	0.053	0.656
25020	Plastic products	0.052	0.543
32000	Electronic material and communications equipment	0.052	0.757
31001	Machines, equipment and miscellaneous electric material - except for vehicles	0.051	0.678
27003	Foundries	0.051	0.462
15043	Other food products	0.049	0.426
36090	Miscellaneous products	0.048	0.576
23010	Coke plants	0.047	0.487
37000	Recycling	0.045	0.304
35090	Miscellaneous transportation equipment	0.044	0.581
21002	Corrugated cardboard, packaging, and paper and cardboard objects	0.044	0.577
17001	Processing of fibers, weaving and cloth making	0.043	0.471
28001	Metal products - except machines and equipment	0.042	0.496
24030	Soap, detergents, cleaning products and toiletries	0.042	0.658
29001	Machines and equipment - except appliances	0.041	0.605
21001	Pulp, paper and smooth cardboard, poster paper and card paper	0.040	0.602
34003	Reconditioning or restoration of engines of motor vehicles	0.038	0.556
24090	Miscellaneous chemical products	0.037	0.635
25010	Rubber products	0.036	0.567
26092	Miscellaneous products of non-metallic minerals	0.035	0.382
22000	Editing, printing and reproduction of recordings	0.035	0.702
19012	Leather objects	0.034	0.453
30000	Office machines and data-processing equipment	0.034	0.852
36010	Pieces of furniture	0.034	0.402
26010	Glass and glass products	0.031	0.576
15021	Preserves of fruit, vegetables and other vegetable products	0.029	0.484
17002	Manufacturing of textile objects based on cloth - except for garments	0.028	0.433
18001	Making of clothing articles and accessories - except on order	0.023	0.425
18002	Making clothing articles and accessories - on order	0.023	0.435
18999	Making of clothing articles and accessories - on order or not	0.023	0.690
27002	Non-ferrous metals	0.022	0.644
15030	Dairy products	0.022	0.433
19020	Footwear	0.019	0.348
15010	Slaughtering and preparation of meat and fish	0.018	0.355
35010	Construction and repair of boats	0.018	0.493
23020	Products in oil refining	0.015	0.763
33005	Chronometers, clocks and watches	0.015	0.751
23400	Alcohol production	0.014	0.350
15041	Manufacturing and refining of sugar	0.013	0.334
15042	Roasting and grinding of coffee	0.013	0.499
19011	Tanning and other preparations of leather	0.013	0.325
16000	Tobacco products	0.013	0.496
15050	Beverages	0.012	0.555
15022	Vegetable fat and oil	0.009	0.446
35020	Construction and assembly of locomotives, cars and other rolling stock	0.004	0.632
	Median	0.041	0.432

Notes: The industry codes correspond to the CNAE-Domiciliar, the industry classification used in the 2000 Population Census. Industries are sorted by their R&D intensity at baseline. We measure R&D intensity as R&D expenditure as a share of total sales at baseline and we source it from from the 2000 *Pesquisa de Inovação Tecnológica* (PINTEC). We define skill intensity as the share of skilled individuals in a particular industry in Brazil at baseline and we source it from the 2000 Population Census. The correlation between these variables is approximately 0.34. Industries below the median are classified as low and the ones above the median as high.

Table A.3: Keywords Used to Identify Innovative Occupations

Panel A: Nouns or combination of nouns from task description of occupations

Portuguese	English
pesquisa e desenvolvimento	research and development
inovação	innovation
p&d	R&D
desenvolvimento de produtos	product development
desenvolvimento de processos	process development
pesquisador	researcher
novas tecnologias	new technologies
protótipos	prototypes
pesquisas tecnologicas	technological research
automaçao de processos	process automation

Panel B: Actions (verb + noun) from task description of occupations

Portuguese	English
desenvolvem produtos	develop products
desenvolvem pesquisas	develop research
desenvolvem equipamentos	develop equipment
desenvolvem processos	develop processes
desenvolvem dispositivos	develop devices
otimizam métodos	optimize methods
otimizam os meios	optimize means
aperfeiçoam sistemas	improve systems
aperfeiçoam processos	improve processes
aperfeiçoam produtos	improve products
aperfeiçoam dispositivos	improve devices
implementam dispositivos de automaçao	implement automation devices
desenvolvem, testam e supervisionam sistemas, processos e métodos produtivos	develop, test and supervise systems, processes and production methods

Panel C: Nouns or combinations of nouns from PINTEC survey

Portuguese	English
produto novo / novo produto	new product
produtos novos / novos produtos	new products
produto aprimorado	improved product
produtos aprimorados	improved produts
inovação de produto	product innovation
aperfeiçoamento de produto	product improvement
processo novo / novo processo	new process
processos novos / novos processos	new processes
processo aprimorado	improved process
processos aprimorados	improved processes
inovação de processo	process innovation
aperfeiçoamento de processo	process improvement

Notes: The Table reports the keywords used to identify innovation intensive occupations and their English translation. Keywords reported in Panels A and B are sourced from the "Brazilian Classification of Occupations", Ministry of Labor, 3rd Edition (2010). Keywords reported in Panel C are sourced from the Technical Appendix of the 2008 PINTEC Survey.

Table A.4: Industry-level Measures of Innovation

Outcome:	Measure of Industrial Innovation				
Measure:	log R&D expenditure over sales	Share of patenting firms	Share of product innovation firms	Share of process innovation firms	Share of product or process innovation firms
	(1)	(2)	(3)	(4)	(5)
Share of innovation	5.929**	1.557***	2.178***	0.885**	1.628***
	(2.502)	(0.411)	(0.406)	(0.427)	(0.473)
Observations	271	274	274	274	274
R-squared	0.062	0.166	0.172	0.037	0.098

Notes: Dependent variables are calculated from 2000 *Pesquisa de Inovação Tecnológica* (PINTEC). The unit of observation is the 4-digit CNAE industry. These regressions compute the OLS coefficient of a number of outcomes on the share of innovation workers. The share of innovation workers in each industry are computed for the year 2000 using the methodology described in Section 3.4. All regressions are weighted by number of workers in each industry in 2000. Robust standard errors reported in brackets. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.5: Internal migration

Skill Group:	All				Skilled			Unskilled		
Outcomes	$\Delta \log L$	Net Migration	In-Migration	Out-Migration	Net Migration	In-Migration	Out-Migration	Net Migration	In-Migration	Out-Migration
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ΔA_{soy}	-0.014	0.004	0.002	-0.003	-0.001	-0.004	-0.003	0.012	0.011**	-0.002
	[0.013]	[0.009]	[0.005]	[0.006]	[0.010]	[0.005]	[0.007]	[0.008]	[0.005]	[0.006]
Observations	557	557	557	557	557	557	557	557	557	557
R-squared	0.171	0.553	0.401	0.592	0.507	0.380	0.593	0.582	0.407	0.566
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Dependent variables are calculated for 2010 (source: Population Censuses). The unit of observation is the microregion. These regressions compute the 5 year internal migration rate between 2005 and 2010, using the microregion of residence 5 years prior to the Census 2010. All the regressions include the baseline specification controls which are the share of rural population in 1991, a measure of technical change in maize and region fixed effects. The regressions with all controls also include income per capita (in logs), population density (in logs), literacy rate, all observed in the 1991 Population Census. Robust standard errors reported in brackets. Significance levels: $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table A.6: Effect of technical change in soy on informal workers' wages

VARIABLES	(1) Manufacturing Unskilled	(2) L-Manufacturing Unskilled	(3) H-Manufacturing Unskilled
ΔA_{soy}	-0.027	-0.042**	-0.020
	[0.019]	[0.017]	[0.024]
Observations	556	556	556
R-squared	0.180	0.140	0.129
Baseline Controls	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
All Controls	Yes	Yes	Yes

Notes: Changes in informal wages are calculated over the years 2000 to 2010. Informal workers are defined based on the position held in the occupation of the main job. Controls include the share of rural population, income per capita (in logs), population density (in logs), literacy rate, all observed in 1991, a measure of technical change in maize, and the change in the share of workers below the minimum wages. Dependent variables are computed from taking the average log hourly wage. Robust standard errors reported in brackets. Significance levels: $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table A.7: Labor (direct) and Capital (indirect) Channels of Structural Transformation

Outcome:	Change in manufacturing employment share				
Region/Industry:	All	Soy regions	Non-soy regions	*L* industries	*H* industries
	(1)	(2)	(3)	(4)	(5)
ΔA_{soy}	0.015***	0.020***	-0.002	0.014***	0.001
	[0.004]	[0.005]	[0.006]	[0.003]	[0.003]
Δ Exposure to capital inflows	0.023*	0.001	0.042	0.004	0.019*
	[0.014]	[0.016]	[0.028]	[0.010]	[0.011]
Observations	540	385	155	540	540
R-squared	0.253	0.099	0.431	0.292	0.076

Notes: Changes in dependent variables are calculated over the years 2000 and 2010 (source: Population Censuses). The unit of observation is the microregion. All the regressions include as controls the share of rural population, income per capita (in logs), population density (in logs), literacy rate, all observed in the 1991 Population Census, a measure of technical change in maize and region fixed effects. Observations are weighted by total employment in a given microregion in 2000. Robust standard errors reported in brackets. Significance levels: $^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.1$..

Table A.8: Effect of agricultural technical change on manufacturing outcomes excluding industries linked to soy

Yearly Social Security Data (1998-2009) and Yearly Manufacturing Survey Data (2000-2009)

Panel A

Outcome:	Employment by manufacturing industry		Capital by manufacturing industry	
Industry:	L Industry	H Industry	L Industry	H Industry
	(1)	(2)	(3)	(4)
A_{soy}	0.124**	-0.006	0.363***	0.043
	[0.050]	[0.041]	[0.118]	[0.122]
Observations	5,640	5,640	3,001	2,942
R-squared	0.448	0.385	0.900	0.894
Region x Year FEs	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Panel B

Outcomes:	L Industry Productivity			H Industry Productivity		
Measure:	Log Value Added per Worker	Log Value Added per Wage Bill	Log TFP	Log Value Added per Worker	Log Value Added per Wage Bill	Log TFP
	(1)	(2)	(3)	(4)	(5)	(6)
A_{soy}	-0.150**	-0.135**	-0.234***	-0.126*	-0.118**	-0.144*
	[0.061]	[0.054]	[0.081]	[0.072]	[0.057]	[0.074]
Observations	3,055	3,055	2,999	3,069	3,069	2,922
R-squared	0.770	0.568	0.594	0.797	0.633	0.585
Region x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Panel C

Outcomes:	Wage Bill of Non-Innovation Workers		Wage Bill of Innovation Workers	
Industry:	L Industry	H Industry	L Industry	H Industry
	(1)	(2)	(3)	(4)
A_{soy}	0.153***	-0.005	0.034	-0.284*
	[0.047]	[0.049]	[0.111]	[0.167]
Observations	3,816	3,816	3,796	3,815
R-squared	0.983	0.987	0.935	0.927
Controls	Yes	Yes	Yes	Yes

Notes: This table replicates the results presented in Table 5 (Panel A), Table 6 (Panel B) and Table 7 (Panel C) excluding sectors directly linked to soy via input-output linkages. Such sectors include: "Slaughtering and preparation of meat and fish" (SNA code 1091), "Other food products" (SNA code 1093), "Fertilizers and other inorganic chemicals" (2412, 2413, 2419) and "Refined petroleum" (232). Standard errors clustered at the microregion level reported in parentheses. Significance levels: $^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.1$.

Table A.9: Geographical spillovers of the effect of technical change in soy on manufacturing outcomes

Outcomes:	Log Value Added per Worker			Log Value Added per Wage Bill			Log TFP		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A_{soy}	-0.141***	-0.120*	-0.120*	-0.133***	-0.107*	-0.108*	-0.173***	-0.178*	-0.178*
	[0.043]	[0.067]	[0.068]	[0.040]	[0.061]	[0.062]	[0.066]	[0.095]	[0.096]
A_{soy} N5		-0.028	-0.023		-0.034	-0.017		0.006	0.008
		[0.061]	[0.092]		[0.056]	[0.079]		[0.091]	[0.117]
A_{soy} N5-N10			-0.006			-0.021			-0.002
			[0.071]			[0.062]			[0.097]
Observations	3,070	3,070	3,070	3,070	3,070	3,070	3,069	3,069	3,069
R-squared	0.876	0.876	0.876	0.735	0.735	0.735	0.542	0.542	0.542
Region x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variables correspond to the total value added divided by employment (in logs), and total factor productivity for each type of manufacturing industry in each microregion as a proxy for productivity. We use aggregate information from PIA at the microregion level for the time period 2000-2009. We include only those microregions that have positive employment for all the years in the sample. A_{soy} is defined as potential soy yield under high inputs for the years between 2003 and 2009, and the potential soy yield under low inputs for the years between 2000 and 2002. A_{soy} N5, A_{soy} N5-N10 and A_{soy} N10-20 are defined as the average potential soy yield for the five closest neighbors, the fifth to the tenth closest neighbors and the tenth to twentieth closest neighbors of a particular microregion weighted by the inverse distance. Controls include the share of rural population, income per capita (in logs), population density (in logs), literacy rate, all observed in 1991, all interacted with a linear trend, a measure of technical change in maize and region year fixed effects. Standard errors clustered at the microregion level reported in parentheses. Significance levels: $^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.1$.

B Appendix: data

B.1 Total Factor Productivity

In this Appendix, we describe how we compute the measure of total factor productivity at the microregion-industry level that is used to understand the impact on manufacturing productivity of the reallocation of workers to manufacturing induced by the introduction of GE soy. Concretely, we compute total factor productivity in a manufacturing industry as the Solow residual of a Neoclassical Cobb-Douglas production function that combines skilled labor, unskilled labor, and capital and features constant returns to scale, i.e we start by assuming the following Cobb-Douglas value-added function for each industry j located in microregion i in a period t,

$$VA_{ijt} = e^{z_{ijt}} S_{ijt}^{\alpha_{s,j}} U_{ijt}^{\alpha_{u,j}} K_{ijt}^{1-\alpha_{s,j}-\alpha_{u,j}} \tag{4}$$

where i indexes microregions, j indexes industries and t refers to time. Notice that for a given industry the production technology is the same across microregions and periods. As is well known in the Growth Accounting literature, the growth rate in value added can be decomposed into components associated to factor accumulation and technological progress. Assuming industries are perfectly competitive and price takers in factor markets, one can recover the parameters of the production function using data on factor shares. It is worth noticing that in this formulation we are assuming that industries in different microregions share the same primitive technological parameters and the differences in value added are the result of either different productivity levels or differences in the number of production factors used as inputs.

We calibrate the factor shares for skilled and unskilled labor using a two-step approach. In the first step, we compute the labor share out of value added for L and H industries. To do this, we use the NBER-CES Manufacturing Database (Becker et al. 2021) to compute the labor share as the fraction of value added that corresponds to labor payments for each NAIC-1997 industry. We match each NAICS industry to the corresponding CNAE industry in the database using the crosswalk provided in Muendler (2002). Then, we compute the labor share in L and H industries as the simple average of the corresponding labor shares of the industries that belong to each category. Doing this we compute a labor share of 0.40 for L industries and of 0.42 for H industries.

The reason for using an external source to compute the labor share as opposed to computing them directly in PIA is that the values obtained from computing the labor share in the PIA dataset are much smaller than the other estimates of the labor share for Brazil (e.g. Reinbold and Restrepo-Echavarria 2018) and for similar industries in other countries. This is related to the high degree of informality in the manufacturing sector in Brazil which makes observing the true wage bill, and thus, computing the labor share challenging as approximately 50% of the Brazilian labor force is informally employed. (see Ulyssea 2018 and Dix-Carneiro, Goldberg, Meghir, and Ulyssea (2021)).

In the second step, we split the calibrated labor share values between skilled and unskilled labor. Since PIA does not differentiate between skilled and unskilled labor, we compute wage bill shares by type of labor in RAIS and, then, apportion the labor share previously computed in PIA to skilled and unskilled labor. Therefore, in practice $\alpha_{s,j} = \frac{W_{s,j}^{RAIS} S_j^{RAIS}}{W_{s,j}^{RAIS} S_j^{RAIS} + W_{u,j}^{RAIS} U_j^{RAIS}} \frac{W_j^{PIA} L_j^{PIA}}{VA_j^{PIA}}$ and $\alpha_{u,j} = \frac{W_{u,j}^{RAIS} U_j^{RAIS}}{W_{s,j}^{RAIS} S_j^{RAIS} + W_{u,j}^{RAIS} U_j^{RAIS}} \frac{W_j^{PIA} L_j^{PIA}}{VA_j^{PIA}}$.

Finally, we leverage the constant returns to scale assumption to compute the capital share $\alpha_{k,j}$ as $1 - \alpha_{s,j} - \alpha_{u,j}$. In Table B.10, we describe the factor shares for the manufacturing L industry and the H industry.

Once we have computed the shares for the three types of factors, we compute $\log TFP$ in microregion i, industry j and time t as

$$\log TFP_{ijt} = \log VA_{ijt} - \alpha_{kj} \log(p_t K_{ijt}) - \alpha_{1j} \log(L_{1,ijt}) - \alpha_{2j} \log(L_{2,ijt}) \tag{5}$$

Table B.10: Factor Shares

L Industry			H Industry		
α_s	α_u	α_k	α_s	α_u	α_k
0.29	0.11	0.60	0.33	0.09	0.58

B.2 Wages

To compute composition-adjusted wages we estimate the following Mincerian regressions:

$$ln(w_{ikt}) = \gamma_{kt} + H_{ikt}\beta_{Ht} + \varepsilon_{ikt} \quad \textit{for t=2000, 2010} \tag{6}$$

where $ln(w_{ijkt})$ is the log hourly wage of individual i, working in sector j in microregion k at time t, and γ_{kt} is a microregion fixed effect, while H_{ijkt} is a vector of individual characteristics, which includes dummies for sector, skill group, age group, race, and all the interactions between these variables. We estimate the previous Mincerian regression for each microregion and for each broad sector separately. Also, we estimate these regressions constraining the sample to either unskilled or skilled labor only, recovering the unit price of labor in each microregion for each type of labor in both cross sections. Since the existing literature documented how Brazil has experienced a considerable reduction in its gender pay gap (Ferreira, Firpo, and Messina 2017), we estimate equation (6) only for male workers. Observations are weighted by their corresponding population census weight. Next, we use the microregion fixed effects estimated above as the unit price of labor for a given skill group in a given microregion, and we compute the change in unit prices of labor in microregion k between 2000 and 2010 as $\Delta\gamma_k = \gamma_{k,2010} - \gamma_{k,2000}$, which gives us the change in the composition-adjusted wages at the microregion level.