Agricultural Productivity and Structural Transformation.

Evidence from Brazil*

Paula Bustos         Bruno Caprettini         Jacopo Ponticelli‡

July 2012
Preliminary and Incomplete

Abstract

We study the effects of agricultural productivity on industrial development. Classical models of structural transformation stress that productivity growth in agriculture increases income per capita, generating demand for manufacturing goods. However, Matsuyama (1992) notes that the positive effects of agricultural productivity on industrialization occur only in closed economies, while in open economies a comparative advantage in agriculture can slow down industrial growth. In this paper we argue that this conclusion depends on the factor bias of agricultural technical change. In particular, if technical change is strongly labor saving, then increases in agricultural productivity can lead to industrialization even in an open economy. We quantify the relative importance of these forces by studying the effects of the widespread adoption of new agricultural technologies on Brazilian manufacturing firms. To establish causality, we exploit exogenous differences in soil and weather characteristics across geographical areas leading to a differential impact of the new technology on yields. We find that areas affected by labor saving technical change in agriculture experienced faster manufacturing growth.

*We received valuable comments from David Atkin, Gino Gancia, Gene Grossman, Jaume Ventura, Silvana Tenreyro, Joao Pinho de Melo, Claudio Ferraz and participants at presentations held at Pontificia Universidade Catolica de Rio de Janeiro, Fundacao Getulio Vargas, CREI, UPF, LSE and the SED annual meeting. We acknowledge financial support from the Private Enterprise Development in Low-Income Countries Project by the CEPR and UK Department for international development.
‡CREI, Universitat Pompeu Fabra and Barcelona GSE (www.crei.cat).
1 Introduction

We study the effects of agricultural productivity on industrial development. This question has a long tradition in the economic development literature since several scholars have noted that the industrial revolution in England was preceded by increases in agricultural productivity [Nurkse (1953), Rostow (1960)]. Classical models of structural transformation stress that productivity growth in agriculture increases income per capita, generating demand for manufacturing goods [Murphy, Shleifer, Vishny (1989), Golin, Parente and Rogerson (2002)]. However, Matsuyama (1992) notes that the positive effects of agricultural productivity on industrialization occur only in closed economies, while in open economies a comparative advantage in agriculture can slow down industrial growth. This is because labor reallocates towards the agricultural sector, reducing the size of the industrial sector and its scope to benefit from external scale economies.

Note that most models of the effects of agricultural productivity on industrial development usually assume that technical change is hicks-neutral. Thus, they predict that in an open economy an increase in agricultural productivity would induce increases in agricultural employment and a reduction in the size of the industrial sector as labor reallocates towards agriculture and wages increase. However, technical change in agriculture is often labor saving. In this case, increases in agricultural productivity can lead to industrialization even in an open economy. This is because if the reduction in the labor intensity of agricultural production is strong enough, employment in agriculture and wages might fall, inducing an expansion of industrial employment. Then, whether technological change in agriculture induces industrial development in an open economy depends on the factor bias of technical change.

In this paper we shed light on the effects of factor biased technical change in agriculture by studying the widespread adoption of new agricultural technologies on Brazilian manufacturing firms. In particular, we study the effects of the adoption of a new agricultural technology, namely genetically engineered soybean seeds, on industrial development in Brazil. During the ten years after the technology was invented in 1996 the output of soy doubled in Brazil, becoming the most important crop in the country. The advantage of this seeds relative to traditional ones is that they are herbicide resistant which implies that no-tillage planting techniques can be used. This new technology is then expected to save on production costs, in particular requires less labor per unit of land to yield the same output. It thus can be characterized as labor saving technical change. To identify the causal effects of this new technology, we use two sources of exogenous variation in the profitability of technology adoption. First, as the technology was invented in 1996, we use the introduction of the new technology as our source of time variation. Second, as the new technology
had a differential impact on yields depending on geographical and weather characteristics, we use differences in soil suitability across regions as our source of cross-sectional variation.

We start by reporting that during this period the municipalities where soy expanded experienced an increase in agricultural output per worker, a reduction in labor intensity in agriculture and an expansion in industrial employment. This could respond to the adoption of labor saving agricultural technologies reducing labor demand in the agricultural sector and thus inducing a reallocation of labor towards the industrial sector. Alternatively it could be due to other shocks to local labor markets. For example: an increase in labor demand in the industrial sectors could increase wages, inducing agricultural firms to switch to less labor intensive crops, like soy. To establish the direction of causality we exploit the timing of adoption and the differential impact of the new technology on potential yields across geographical areas.

We construct a municipality-level measure of the potential profitability of technology adoption using data on potential soil yields from the FAO-GAEZ database. These yields are calculated by incorporating local soil and weather characteristics into a model that predicts yields for each crop given certain climate and soil conditions. Potential yields are a source of exogenous variation in agricultural productivity because they are a function of weather and soil characteristics, not of actual yields in Brazil. In addition, the database reports potential yields under different technologies. Thus, we exploit the predicted differential impact of the high technology on yields across geographical areas in Brazil as our source of cross-sectional variation in agricultural productivity. This research design allows us to investigate whether exogenous shocks to local agricultural productivity lead to changes in the size and efficiency of the local industrial sectors. Note that this identification strategy relies on the assumption that although Brazil is an open economy, the existence of transport costs implies that local markets are important sources of labor for industrial firms. This research design allows us to investigate whether exogenous shocks to local agricultural productivity lead to changes in the size and efficiency of the local industrial sector.

In a preliminary analysis of the data, we find that municipalities where the new technology is predicted to have a higher effect on potential yields of soy did experience a higher increase in the area planted with GE soy. These preliminary findings show that our instrument (the impact of new agricultural technologies on yields) is a good predictor of the profitability of GE soy adoption. In addition, these regions experienced increases in the value of agricultural output per worker and reductions in labor intensity measured as employment per hectare. Finally, these regions experienced faster employment growth and wage reductions in the industrial sector. Interestingly, the effects of technology adoption are very different for maize, a labor intensive crop that also experienced technical change during this periods as new seeds were introduced by the Brazilian government’s
agency of agricultural technology development (EMBRAPA). Regions where the FAO potential yields are predicted to increase the most when switching from the low to the high technology, did indeed experience a higher increase in the area planted with maize. In addition, they experienced increases in the value of agricultural output per worker and labor intensity. Finally, they experienced increases in wages in the industrial sector.

The differential effects of technological change in agriculture documented for GE soy and maize indicate that the factor bias of technical change is a key factor in the relationship between agricultural productivity and industrial growth in open economies. Our purpose in this paper is to investigate these effects further to isolate the channels and mechanisms through which structural transformation takes place.

The remaining of the paper is organized as follows. Section 2 describes the data. Section 3 presents the empirical strategy and results. Section 4 concludes.

2 Data

In this paper we use three main data sources: the Agricultural Census for data on agriculture, the Yearly Industrial Survey (PIA) for the data on manufacturing and the FAO Global Agro-Ecological Zones database for potential yields of soy and other crops.

The Agricultural Census is released at intervals of 10 years by the IBGE, the Brazilian National Statistical Institute. We use data from the last two rounds of the census that have been carried out in 1996 and in 2006. This allows us to observe agricultural variables both before and after the introduction of genetically engineered soybean seeds, that were legalized in Brazil in 2003). The census data is collected through direct interviews with the managers of each agricultural establishment and is made available online by the IBGE aggregated at municipality level. The main variables we use from the Census are: the value of agricultural production, the number of agricultural workers and the area devoted to agriculture in each municipality. Out of the area devoted to agriculture in each municipality we are able to disentangle the area devoted to each crop in a given Census year. This allows us to monitor how land use has changed between 1996 and 2006. The upper panel of Table 1 reports the total area reaped (in millions hectares) for the three major crops by cultivated area produced in Brazil: soybean, maize and sugar. Among these three crops, soybean registered the largest absolute increase in terms of area reaped between the last two Census years. The area devoted to soybean increased from 9.2 to 15.7 millions hectares between 1996 and 2006, more than half of the total increase in all seasonal crops. The lower panel of Table 1 reports total employment (in millions workers) in seasonal crops production and in agriculture.
as a whole. As reported in the last row, there are around 17 millions Brazilians - around 20% of the Brazilian active population - whose main working activity is agriculture. Notice that between 1996 and 2006, although the area cultivated with seasonal crops has increased by a third (from 36.8 to 48.2 millions hectares), the number of workers employed in seasonal crops’ production decreased from 6.8 to 6.4 millions. This might be due to 2 reasons: technological change in the production of single crops (within-crop effect) and the switch from more to less labor intensive crops (across-crops effect). The introduction of genetically engineered soybean, whose production requires less workers per hectare than normal soybean, is an example of a within-crop effect. The across-crop effect derives from the fact that the production of some crops is less labor intensive to start with. Data from the 1996 Agricultural Census suggests that soy production in Brazil employs on average 42 workers per thousand hectares, while maize and sugar production employ on average respectively 106 and 138 workers per thousand hectares. Part of the reduction in the number of workers employed in seasonal crops production could be due to agricultural establishment switching from more to less labor intensive crops, e.g. from maize to soy production.

Figures 1 to 3 compare the distributions of average actual yields (tons per hectare) across Brazilian municipalities in 1996 and 2006 for, respectively, soy, maize and sugar. As for soy and maize, there was a clear shift to the right in the distribution of average yields, indicating some type of technological improvement taking place. As for sugar, on the contrary, the distribution of average yields looks very similar in 1996 and 2006.

Our second source of data is the Yearly Industrial Survey (PIA), produced by the IBGE, that monitors the performance of Brazilian firms in the extractive and manufacturing sectors. We focus on the manufacturing sector as defined by the Brazilian sector classification CNAE 1.0 (sectors 15 to 37). We use yearly data from 1996 to 2007. The population of firms eligible for the survey is composed by all firms with more than 5 employees registered in the national firm registry (CEMPRE, Cadastro Central de Empresas). The survey is constructed using two strata: the first includes a representative sample of firms having between 5 and 29 employees (estrato amostrado), the second includes all firms having 30 or more employees (estrato certo). For all firms in the survey the data is available both at firm and at plant level (when firms are composed by more than one plant). Our unit of observation is, for most outcomes, the plant. At plant level the survey includes information on: number of employees, wage bill, revenues, costs, capital investment and gross value added.

Finally, we use data on potential yields for soy and other crops from the Global Agro-Ecological Zones database produced by the FAO. Potential yields are the maximum yields attainable for a crop
in a certain geographical area. They depend on the climate and soil conditions of that geographical area, and the level of technology available. The FAO-GAEZ database provides estimates of potential yields under three levels of technology: low, intermediate and high. Each of these levels is captured by the availability of certain inputs like machines and fertilizers. When the level of technology is assumed to be low, agriculture is aimed at subsistence. It is mostly labor-intensive, it uses traditional cultivars and does not use nutrients or chemicals for pest and weed control. When the level of technology is assumed to be intermediate, agriculture is partly market oriented. Production is partly mechanized, it uses improved varieties and some fertilizers and chemicals for pest and weed control. When the level of technology is high, agriculture is market oriented. Production is fully mechanized, it uses improved or high yielding varieties and "optimum" application of nutrients and chemical pest, disease and weed control. The database reports potential yields for each crop under low, medium and high technological levels available in agriculture for a worldwide grid at a resolution of 9.25 x 9.25 km. Figure 4, 5 and 6 show the potential yields for soybean in Brazil under, respectively, low and high technology. The same type of maps are also available for maize and sugar. In order to match the potential yields data with agriculture and industry variables we superimposed each of the potential yields' maps with a political map of Brazil reporting the boundaries of each municipality. Then we took the average of the potential yield, weighted by the area, within each municipality. We repeated this operation per each crop and per each of the three levels of technology. Finally, we measure technological change within each municipality by computing the difference between yields under the high and the low technology. Figure 7 illustrates the resulting measure of technological change in soy.

3 Empirics

In this section we study effect of the adoption of a new agricultural technology, genetically engineered (GE) soybean seeds. The GE soy seeds were first commercially introduced in the U.S. in 1996 and legalized in Brazil in 2003. The advantage of this seeds relative to traditional ones is that they are herbicide resistant which implies that no-tillage planting techniques can be used. The planting of traditional seeds is preceded by soil preparation in the form of tillage to kill the weeds in the seedbed that would crowd out the crop or compete with it for water and nutrients. In contrast, the planting of herbicide resistant GE soy seeds requires no tillage as the herbicide kills

---

1Genetic engineering (GE) techniques allow a precise alteration of a plant’s traits and permit targeting a single plant trait. This facilitates the development of characteristics not possible through traditional plant breeding and decrease the number of unintended characteristics. In the case of herbicide resistant GE soy seeds their genes were altered to include those of a bacteria that was herbicide resistant.
the weeds. Then, the GE soy seeds can be applied directly on last season’s crop residue. This new technology is then expected to save on production costs, in particular requires less labor per unit of land to yield the same output. Then, the adoption of GE soy seeds can be characterized as labor saving technical change.

Note that traditional models of the effects of agricultural productivity on industrial development like Matsuyama (1992) focus on hicks neutral technical change. They predict that in an open economy an increase in agricultural productivity would induce increases in agricultural employment and a reduction in the size of the industrial sector as labor reallocates towards agriculture and wages increase. The type of technical change we study is instead labour saving. As a result, new forces emerge: if the reduction in the labor intensity of agricultural production is strong enough, employment in agriculture and wages might fall, inducing an expansion of industrial employment. Then, whether technological change in agriculture induces industrial development in an open economy depends on the factor bias of technical change. In this section we exploit the adoption of GE soy seeds to assess the relative importance of these two forces.

For this purpose, we first study the effect of the adoption of GE soy on the factor intensity of agricultural production and agricultural labor markets. Next, we assess its impact on industrial employment. We start by reporting simple correlations between the expansion of the planted area with soy relative to other crops and agricultural and industrial labor market outcomes. Next, to establish causality, we exploit the timing of legalization and the differential impact of the new technology on potential yields across geographical areas, which depends on local weather and soil characteristics.

3.1 Basic Correlations in the Data (OLS)

We start by documenting how soy and maize expansion during the 1996-2006 period relates to changes in the agricultural production and in the industrial labour market. Here we present a set of OLS regressions in which agricultural and industrial outcomes are regressed on the percentage of farm land cultivated with soy and maize, inserted one at a time in separate regressions. Note that these results are intended to introduce to the basic correlations in the data, but do not claim to uncover any causal relation between variables.

The basic form of our equation is:

\[ y_{jt} = \alpha_j + \alpha_t + \beta \left( \frac{\text{Crop Area}}{\text{Agricultural Area}} \right)_{jt} + \varepsilon_{jt} \]  

(1)
where \( \left( \frac{\text{Crop Area}}{\text{Agricultural Area}} \right)_{jt} \) is total area reaped with either soy or maize divided by total farm land\(^2\) and \( y_{jt} \) are agricultural and industrial outcomes of interest. The units of observation are Brazilian smallest comparable areas (AMC: Área Mínima Comparável in Portuguese)\(^3\), and we wish to control for both AMC and year fixed effects (\( \alpha_j \) and \( \alpha_t \)). Our source for agricultural variables is the decennal Agricultural Census, which means that for both the independent variable and the agricultural outcomes we have only two observations over the last 20 years: one in 1996 and the other in 2006. With only two periods, fixed effects and first difference estimates are identical, so we estimate (1) in first differences:

\[
\Delta y_j = \Delta \alpha + \beta \Delta \left( \frac{\text{Crop Area}}{\text{Agricultural Area}} \right)_j + \Delta \varepsilon_j
\]

where changes are defined between 1996 and 2006 throughout.

Table 2 reports correlations between changes in area reaped and changes in agricultural production: the first panel presents the results using changes in area reaped with soy while the second panel presents results with changes in area reaped with maize. Together they suggest that although technological progress was fast in both crops, soy production became much less labor intensive and that this in turn drove labour out of agriculture.

The first column of table 2 reports the relationship with productivity, and shows that in places where soy and maize cultivation expanded, also the value of agricultural production per worker increased. Value per worker is defined here as the total value of crop production divided by total number of workers and refers only to seasonal crop production\(^4\). Column 2 confirms that labour intensity in agriculture decreased where soy cultivation expanded and increased where maize cultivation did. Agricultural labour intensity is measured here as number of workers per acre in seasonal crops cultivation. These results support the notion that changes in agricultural productivity have come with major changes in the mode of production, and in particular that in 2006 soy is being produced with relatively less labour input per acre and maize with relatively more. This evidence in turn, is consistent with the qualitative description of technological change in the two crops: in 2006 much of soybean was produced with GE seeds that need much less labour, while in the same year many establishments that cultivated maize had started doing so over two different seasons.

\(^2\)Total farm land includes areas devoted to crop cultivation (both permanent and seasonal crops), animal breeding and logging.

\(^3\)Municipalities are the smallest administrative areas for which many Census variables are recorded in Brazil. Over the years however, many municipalities split and reorganized to accommodate population growth and migration, with the result that municipalities are not always comparable across years. Smallest comparable area is the smallest geographical breakdown for which consistency is warranted overtime.

\(^4\)Both soy and maize are seasonal crops.
every year, a process that in the limit could require double as much labour per acre. Column 3 shows that the share of workers employed in agriculture decreased in places where soy expanded and did not change significatively in places cultivated with maize. Share of workers employed in agriculture is defined as total number of workers in agriculture (from the Agricultural Census) divided by total number of workers (calculated as total number of workers in non-agricultural sectors from CEMPRE plus total number of workers in agriculture from the Agricultural Census). This change in agricultural employment is the shock to the labour market we will be exploiting in the rest of the paper.

Table 3 reports results from regression (2) using industrial labour market outcomes as dependent variables. We focus only on manufacturing plants (CNAE 2-digits codes 15 to 37) owned by firms that employ at least 30 employees: the sample for which the PIA Empresa survey contains the population of Brazilian firms. Industrial outcomes come from the yearly plant-level survey: we aggregate these data at AMC level and then collapse all years in 2 periods: one before 2003 and one after it, in order to maximize the number of observation used in the regressions. Total employment includes both production and non-production workers; plant size is calculated as total number of workers within an AMC divided by total number of plants; wage is aggregate wage bill within an AMC divided by total number of workers there.

The first column of table 3 shows that total employment in manufacturing significantly grew where soy cultivation expanded, but not where maize cultivation did. The second and third columns qualify this conclusion, and show that the increase came from both the intensive and the extensive margin: column 2 shows that places where soy expanded had on average more plants, while the third column shows that these plants were on average larger (plant size is calculated here as total employment divided by total number of plants). Finally, the last column of table 3 shows that wages did not move significatively were soy expanded while they grew were maize did.

Overall, the results on employment are consistent with a rightward shift of the labour supply schedule driven by soy expansion: although the positive sign on wages does not support this story, notice that this result might be driven by omitted variable bias, because the sign is reversed when the change in soy and maize area are inserted together in the regression.

---

5Brazilian manufacturing activity is very concentrated, and not all AMC municipalities have at least 1 plant from a large firm every year. Collapsing observations from more than 1 year allows us to use all municipalities for which we observe 1 plant at least once before and after the introduction of GE soy seeds.
3.2 Causality

In this section we provide direct empirical evidence on the effects of the widespread adoption of a new agricultural technology, GE soybean seeds, on industrial development in Brazil. The basic correlations in the data reported in the previous section show that areas where soy expanded experienced an increase in output per worker and a reduction in labor intensity in agriculture and an expansion in industrial employment. This could respond to the adoption of labor saving agricultural technologies reducing labor demand in the agricultural sector and thus inducing a reallocation of labor towards the industrial sector. Alternatively it could be due to other shocks to local labor markets. For example: an increase in labor demand in the industrial sectors could increase wages, inducing agricultural firms to switch to less labor intensive crops, like soy. To establish the direction of causality we exploit the timing of adoption and the differential impact of the new technology on potential yields across geographical areas. First, the new technology was commercially introduced in the U.S. in 1996, and legalized in Brazil in 2003. Thus, we use the periods before and after 2003 as our “pre and post-treatment” periods. Second, the new technology has a differential impact on potential yields depending on soil and weather characteristics. Thus, we exploit these exogenous differences on potential yields across geographical areas as our source of cross-sectional variation in the intensity of the treatment.

To implement this strategy, we need an exogenous measure of potential yields for soy and other crops, which we obtain from the FAO-GAEZ database. These potential yields are estimated using an agricultural model that predicts yields for each crop given climate and soil conditions. As potential yields are a function of weather and soil characteristics, not of actual yields in Brazil, they can be used as a source of exogenous variation in agricultural productivity across geographical areas. In addition, the database reports potential yields under different technologies or input combinations. Yields under low inputs are described as those obtained using traditional seeds and no use of chemicals, while yields under high inputs are obtained using high yielding varieties and optimum application of chemicals for weed control. Thus, the difference in yields between the high and low technology captures the effect of moving from traditional agriculture to a technology that uses optimum weed control, among other characteristics.  

6 FAO-GAEZ description of each technology is as follows:
Low-level inputs/traditional management

Under the low input, traditional management assumption, the farming system is largely subsistence based and not necessarily market oriented. Production is based on the use of traditional cultivars (if improved cultivars are used, they are treated in the same way as local cultivars), labor intensive techniques, and no application of nutrients, no use of chemicals for pest and disease control and minimum conservation measures.

High-level inputs/advanced management

Under the high input, advanced management assumption, the farming system is mainly market oriented. Commercial production is a management objective. Production is based on improved high yielding varieties, is fully
good predictor of the profitability of adopting herbicide resistant GE soy seeds. Thus, we can then exploit the predicted differential impact of the high technology on yields across geographical areas in Brazil as our source of cross-sectional variation in agricultural productivity. This research design allows us to investigate whether exogenous shocks to local agricultural productivity lead to changes in the size and efficiency of the local industrial sectors. Note that this identification strategy relies on the assumption that although Brazil is an open economy, the existence of transport costs implies that local markets are important sources of labor and demand for industrial firms.

More formally, our basic empirical strategy is to estimate an equation of the following form:

$$y_{jt} = \alpha_j + \alpha_t + \beta A_{jt}^{soy} + \varepsilon_{jt}$$

where $y_{jt}$ is an outcome that varies across municipalities and time, $j$ indexes municipalities, $t$ indexes time, $\alpha_j$ are municipality fixed effects, $\alpha_t$ are time fixed effects and $A_{jt}^{soy}$ = potential yield of soy under high (low) inputs if $t \geq 2003$ ($t < 2003$).

Note that a potential problem with this identification strategy is that the productivity of land is positively correlated across crops, thus we could be capturing the effect of overall technical change instead of the labor saving technical change associated to GE soy. For example, during this period there were also increases in yields of maize, a labor intensive crop. Thus, we need to control in the above regressions for the changes in yields of maize when switching from the low to the high technology. We then include the following variable as a control: $A_{jt}^{maize} = potential yield of maize under high (low) inputs if $t \geq 2003$ ($t < 2003$). In addition, we want to control for changes in the prices of crops, that can also have an influence on the expansion of soy relative to other agricultural activities. Note that the overall effect of price changes would

$$y_{jt} = \alpha_j + \alpha_t + \beta A_{jt}^{soy} + \gamma A_{jt}^{maize} + \sum_z \theta_z p_z^j A_{j0}^z + \varepsilon_{jt}$$

where $z = soy, maize and sugar$, $p_z^j$ is the international price of crop $z$, $A_{j0}^z$ is potential yield of crop $z$ under low inputs for maize, medium for sugar.

First we estimate the effects of the agricultural technology shock $A^{soy}$ on the share of agricultural land devoted to soy. As the increase in yields resulting from adoption of the high technology ($A_{high}^{soy} - A_{low}^{soy}$) is expected to be correlated with the profitability of technology adoption, we can use changes in $A^{soy}$ as an instrument for changes share of agricultural land devoted to soy. Then, we can perform an instrumental variables estimation of equation (1) where the first stage is given
by equation (3). Alternatively, we can estimate reduced form equations of the form of (3) where we study the direct effect of the change in potential yields driven by technology adoption \((A_{soy}^{high} - A_{soy}^{low})\) on the set of outcomes we are interested in, namely: value of output per worker and labor intensity of agricultural production and industrial employment. In the following subsections, we report our first stage, reduced form and instrumental variable estimates.

### 3.3 First Stage

We document here the relation between the change in agricultural area cultivated with soy and maize on our technological shock described above. The purpose of these regressions is twofold. On the one hand we see them as a "sanity check" on the information content of our technological shock; on the other hand these regressions represent the first stage used to address the endogeneity of the OLS regressions reported above.

The equations we estimate are:

\[
\Delta \left( \frac{\text{Crop Area}}{\text{Agricultural Area}} \right)_j = \Delta \alpha + \beta \Delta A_{soy} + \Delta \varepsilon_j
\]

and

\[
\Delta \left( \frac{\text{Crop Area}}{\text{Agricultural Area}} \right)_j = \Delta \alpha + \beta \Delta A_{soy} + \gamma \Delta A_{maize} + A_{sugar} + \Delta \varepsilon_j
\]  

(5)

where \(\Delta \left( \frac{\text{Crop Area}}{\text{Agricultural Area}} \right)\) is the change in share of farm land reaped with either soy or maize between 1996 and 2006 and it is defined as above. \(A_{soy}^{high}, A_{maize}^{high}\) and \(A_{sugar}\) are potential yields in AMC \(j\) for soy, maize and sugar while the technological shocks are defined as explained in the previous section: \(\Delta A^z_j = A^z_{j;\text{HIGH inputs}} - A^z_{j;\text{LOW inputs}}\), with \(z = \text{soy, maize}\). The soy shock intends to capture the legalization of GE soy in 2003, while the maize shock is meant to capture technological progress in the cultivation of this crop, as highlighted in graph 2 above. We control for general suitability to sugar production with intermediate inputs \(A_{sugar}\) because sugar is the other major seasonal crop in Brazil. However, we do not use its change in potential yield because sugar production does not seem to be more productive in 2006 relative to 1996 (see graph 3 above).

Column 1 and 3 in table 4 show that the soy and maize shock correctly predict soy and maize expansion over the period: when inserted alone, an increase in soy (maize) potential yield is associated with significantly greater farm land reaped with soy (maize). Column 2 and 4 strengthen this
results: they show that controlling for both shocks and for the potential yield of sugar, increase
the effect of both shock: on soy, the effect more than doubles, while on maize the effect goes up
by one third. Also, the fact that the maize shock has a significantly negative effect on the share
of farm land cultivated with soy and that the potential yield of sugar has a significantly negative
effect on both shares, is also consistent with optimal behavior of farmers, who choose which crop to
cultivate based on the specific suitability of their plots. Note, that all of these suitability measures
tend to be positively correlated (this is especially true for maize and soy), so the result that the
shocks correctly predict the expansion or retrenchment of specific crops means that they capture
the dimensions that are more relevant for the decisions of farmers.

All in all, table 4 support the choice of our instrument. The effect of the shocks are extremely
significant (the F-tests for the joint significance of the regressors range from 141.8 to 15.51) and
also economically relevant. The estimated coefficient on soy implies that municipalities with a one
standard deviation above the mean increase in potential soy yields increased the share of soy in
planted land area by 36% of a standard deviation.

3.4 Reduced Form

Once we established the relevance of our instruments, we turn to the study of the effect of techno-
logical change in agriculture on production and employment in both agriculture and manufacturing:
our reduced form. In this section we show that the change in potential soy yield is associated with
a reduction in the use of labour in agriculture and with a rightward shift of the labour supply
schedule in manufacturing.

We start with agricultural production: table 5 reports the results of running an equation similar
to (5) on the agricultural outcomes used in table 2:

$$\Delta y_j = \Delta \alpha + \beta \Delta A_j^{soy} + \gamma \Delta A_j^{maize} + A_j^{sugar} + \Delta \varepsilon_j$$

where $y_j$ is value produced per worker in seasonal crops, labor intensity or share of workers employed
in agriculture all defined as in table 2 above.

Table 5 reinforce the results on agriculture highlighted in section 3.1: productivity increased
significatively in places where potential soy yield increased relatively more, while in the same places
labor intensity in agriculture dropped and agriculture shrunk in terms of total employment. At the same time, in places where potential maize yield increased relatively more, the number of workers per acre increased and the share of population employed in agriculture grew. Also, the maize shock seems to be negatively associated with the seasonal crop value per worker, but this effect is not significantly different from 0. It is interesting to note that these results do not hold when we run these regressions only including AMC in the North-East of the country, the region where sugar has been cultivated historically, and neither soy nor maize are widespread (results available upon request).

Overall, table 5 is consistent with our interpretation of the technological change brought about by GE soy seed. The new technology seems to require much less labour per acre, and this in turn seems to have driven labour out of the agricultural sector.

We now turn to the regressions on manufacturing outcomes. Here we are able to exploit the full panel structure of our data: the equation we estimate with these data have the form:

\[
y_{jt} = \alpha_j + \alpha_t + \beta A_{jt}^{sog} + \gamma A_{jt}^{maize} + \sum_z \theta_z P_t^z A_{j0}^z + \varepsilon_{jt}
\]

where \( y_{jt} \) are industrial outcomes of interest; \( A_{jt}^{sog} (A_{jt}^{maize}) \) are potential yield of soy (maize) under low inputs for all years before 2003 and under high input for all years from 2003 on. We observe all years from 1996 to 2007 and control for the real price of soy, maize and sugar times potential yield for these crops in 1996: \( A_{j0}^z \) is potential yield under low inputs when \( z = \text{soy or maize} \) and it is potential yield under medium inputs when \( z = \text{sugar} \). These controls are intended to make sure that changes are truly driven by technological change rather than the evolution of commodity prices. Although international commodity prices will affect all Brazilian AMC in the same way, they might still have heterogeneous effect in places that were more suitable to the cultivation of some particular crop at the beginning. Controlling for the interaction of these prices with potential yield in 1996 makes sure that our results are not driven by commodity prices. In all specification we control for both AMC and year fixed effects (\( \alpha_j \) and \( \alpha_t \)) and cluster standard errors at AMC level to avoid that serial correlation in our shock makes the precision of our estimates artificially high (Bertrand, Duflo and Mullainathan, 2004).

Table 6 shows reduced form results with industrial variables. All variables are defined as above and are built as averages at AMC level starting from plant level data. Again we focus on man-
ufacturing plants owned by firms with at least 30 employees (for which we observe the whole population).

The first column of table 6 reproduces closely the patterns highlighted in table 3. In particular, it confirms that industrial employment (both production and non-production workers) grew significantly in places where soy potential yield increased more. The effect of maize potential yield is negative (as in the OLS) but not significantly different from 0. Contrary to the OLS results however, the effect on employment seems to come exclusively from the growth of existing plants: in places where the change in potential soy yield increased more, average plant size grew significantly, while the total number of plant did not change. Average wage significantly fell in places where the potential yield of soy increased more, and significantly grew in places where the change in potential yield of maize was greater.

Overall, the results shown in table 6 are consistent with rightward shift of the labour supply curve in areas where the change in potential yield of soy was greatest, and a (less clear) leftward shift of the labour supply schedule in the places where the change in potential yield of maize was bigger. Together with table 5, these results support our story of a labour-saving technological change in soy production promoting a growth of industrial employment by liberating labour in agriculture.

4 Final Remarks

To be Added

References


5 Figures and Tables

Figure 1: distribution of actual soy yields across Brazilian municipalities in 1996 and 2006.
Figure 2: distribution of actual maize yields across Brazilian municipalities in 1996 and 2006.
Figure 3: distribution of actual sugar yields across Brazilian municipalities in 1996 and 2006.
Figure 4: potential soy yield under low technology.
Figure 5: potential soy yields under intermediate inputs.
Figure 6: potential soy yields under high technology.
Figure 7: technological change in soy: potential yield under high technology minus potential yield under low technology.
<table>
<thead>
<tr>
<th>Area Reaped (million ha)</th>
<th>1996</th>
<th>2006</th>
<th>change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soy</td>
<td>9.2</td>
<td>15.7</td>
<td>6.4</td>
</tr>
<tr>
<td>Maize</td>
<td>10.5</td>
<td>11.7</td>
<td>1.3</td>
</tr>
<tr>
<td>Sugar</td>
<td>4.2</td>
<td>5.6</td>
<td>1.4</td>
</tr>
<tr>
<td>All seasonal crops</td>
<td>36.8</td>
<td>48.2</td>
<td>11.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Employment (million workers)</th>
<th>1996</th>
<th>2006</th>
<th>change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seasonal crops</td>
<td>6.8</td>
<td>6.4</td>
<td>-0.4</td>
</tr>
<tr>
<td>Agriculture</td>
<td>17.9</td>
<td>16.6</td>
<td>-1.3</td>
</tr>
</tbody>
</table>

Table 1: Summary Statistics.
<table>
<thead>
<tr>
<th></th>
<th>Δ Value per Worker</th>
<th>Δ Labor Intensity</th>
<th>Δ % Agri Workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ % Soy Area</td>
<td>3.303***</td>
<td>-0.630***</td>
<td>-0.0734**</td>
</tr>
<tr>
<td></td>
<td>(0.281)</td>
<td>(0.210)</td>
<td>(0.0358)</td>
</tr>
<tr>
<td>N</td>
<td>3,841</td>
<td>3,838</td>
<td>3,921</td>
</tr>
<tr>
<td>Panel B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ % Maize Area</td>
<td>2.907***</td>
<td>0.679***</td>
<td>0.0204</td>
</tr>
<tr>
<td></td>
<td>(0.209)</td>
<td>(0.160)</td>
<td>(0.0252)</td>
</tr>
<tr>
<td>N</td>
<td>4,062</td>
<td>4,053</td>
<td>4,112</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 2. OLS regressions: changes in agricultural production on changes of area reaped with soy and maize.
<table>
<thead>
<tr>
<th></th>
<th>Δ Total Employment</th>
<th>Δ Number of Plants</th>
<th>Δ Plant Size</th>
<th>Δ Wage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ % Soy Area</td>
<td>1.087***</td>
<td>0.547**</td>
<td>0.662**</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td>(0.379)</td>
<td>(0.268)</td>
<td>(0.334)</td>
<td>(0.155)</td>
</tr>
<tr>
<td>N</td>
<td>2,048</td>
<td>2,063</td>
<td>2,048</td>
<td>2,048</td>
</tr>
<tr>
<td><strong>Panel B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ % Maize Area</td>
<td>0.117</td>
<td>0.132</td>
<td>0.046</td>
<td>0.250**</td>
</tr>
<tr>
<td></td>
<td>(0.260)</td>
<td>(0.175)</td>
<td>(0.188)</td>
<td>(0.122)</td>
</tr>
<tr>
<td>N</td>
<td>2,172</td>
<td>2,187</td>
<td>2,172</td>
<td>2,172</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3. OLS regressions: changes in the industrial labour market on changes of area reaped with soy and maize.
<table>
<thead>
<tr>
<th></th>
<th>Δ % Soy Area</th>
<th>Δ % Maize Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta A^{soy} )</td>
<td>0.012***</td>
<td>0.025***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>( \Delta A^{maize} )</td>
<td>-0.003***</td>
<td>0.003***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>( \Delta A^{sugar} )</td>
<td>-0.007***</td>
<td>-0.006***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>N</td>
<td>3,921</td>
<td>3,921</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.054</td>
<td>0.074</td>
</tr>
<tr>
<td>F-test for joint significance</td>
<td>141.28</td>
<td>52.77</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4. First stage: changes in area reaped with soy and maize on potential yield shocks.
<table>
<thead>
<tr>
<th></th>
<th>$\Delta$ Value per Worker</th>
<th>$\Delta$ Labor Intensity</th>
<th>$\Delta$ % Agri Workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta A^{soy}$</td>
<td>0.143***</td>
<td>-0.088**</td>
<td>-0.027***</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.035)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>$\Delta A^{maize}$</td>
<td>-0.025</td>
<td>0.049***</td>
<td>0.010***</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.013)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$A^{sugar}$</td>
<td>-0.036*</td>
<td>-0.027</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.017)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>N</td>
<td>4,150</td>
<td>4,146</td>
<td>4,254</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.003</td>
<td>0.007</td>
<td>0.013</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Reduced form: changes of agricultural production on potential yield shocks.
<table>
<thead>
<tr>
<th></th>
<th>Total Employment</th>
<th>Number of Plants</th>
<th>Plant Size</th>
<th>Wage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A^{soy}$</td>
<td>0.087***</td>
<td>-0.006</td>
<td>0.094***</td>
<td>-0.046***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.016)</td>
<td>(0.025)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>$A^{maize}$</td>
<td>-0.021</td>
<td>-0.001</td>
<td>-0.020</td>
<td>0.019***</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.008)</td>
<td>(0.013)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>$P^2 A^2$ controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AMC &amp; year FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>25,258</td>
<td>25,517</td>
<td>25,258</td>
<td>25,235</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.922</td>
<td>0.948</td>
<td>0.809</td>
<td>0.777</td>
</tr>
</tbody>
</table>

Standard errors clustered at AMC level in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Reduced form results: changes in the industrial labor market on potential yield shocks.