Public R&D Meets Economic Development: Embrapa and Brazil's Agricultural Revolution

Ariel Akerman* IADB
Jacob Moscona† MIT
Heitor S. Pellegrina‡ Notre Dame
Karthik A. Sastry§ Princeton

August 28, 2025

Abstract

Can public R&D investment in developing countries drive productivity growth? We study this question in the context of Brazilian agriculture and the Empresa Brasileira de Pesquisa Agropecuária (Embrapa), a public research corporation established in 1973 to develop locally suitable science and technology. First, we show that Embrapa redirected research toward prioritized staple crops and local ecological conditions, and increased research productivity, especially in remote and research-scarce regions. Second, exploiting the staggered rollout of research centers alongside heterogeneous local exposure to Embrapa's technology development, we find that Embrapa significantly increased agricultural output, driven both by higher productivity and expanded input use. Combined with a model, these estimates imply that public R&D investment increased national agricultural productivity by 110% with a benefit-cost ratio of 17. The decentralized structure, in which research labs were spread across many ecological zones instead of in a single hub, explains a large share of these benefits.

*Inter-American Development Bank, 1300 New York Ave NW, Washington, DC, USA, 20577. Email: arielk@iadb.org.

†Massachusetts Institute of Technology Department of Economics, 50 Memorial Drive, Cambridge, MA, USA, 02142. Email: moscona@mit.edu.

‡University of Notre Dame Department of Economics, 3060 Jenkins Nanovic Hall, South Bend, IN, USA, 46556. Email: hpelleg3@nd.edu.

§Princeton University Department of Economics, Julis Romo Rabinowitz Building, Princeton, NJ, USA, 08544. Email: ksastry@princeton.edu.

Acknowledgments: We thank Daron Acemoglu, Bruno Ferman, Martin Fiszbein, Thomas Fujiwara, David Lagakos, Bob Margo, Mushfiq Mobarak, Richard Rogerson, Chris Udry, and Leonard Wantchekon, as well as seminar participants at Princeton, Wharton, Northwestern, the University of Notre Dame, the University of Connecticut, the University of Copenhagen, the 2024 SED Winter Meeting, the RIDGE Forum, and EESP-FGV for helpful comments and conversations. We are also grateful for several insightful conversations with researchers from Embrapa, including Mateus Batistella, Luiz Antonio de Araujo Cruz, Ana Beatriz Juca de Queiroz Fiuza, Geraldo Martha, and Ramon Bicudo. *Disclaimer:* The views expressed in this article are those of the authors and do not necessarily represent those of the Inter-American Development Bank.

1 Introduction

Can public R&D investment in developing countries drive productivity growth? The conventional perspective in economics is that innovation takes place exclusively in a small set of frontier nations where the returns to R&D are large (Jones and Summers, 2020), and countries outside of this set grow by adopting foreign technology (Acemoglu et al., 2006). Underlying this view is a presumption that non-frontier countries have a low return on R&D investment due to a lack of human or physical capital, but nonetheless a high return to technology upgrading. By implication, policymakers outside the frontier should prioritize technology adoption over homegrown R&D (Parente and Prescott, 1994; Barro and Sala-i-Martin, 1997). This perspective is echoed in numerous microeconomic studies and development programs that aim to identify, and then overcome, the barriers to adopting foreign technology (Suri and Udry, 2022; Verhoogen, 2023).

Opportunities for growth without innovation, however, may be more limited in practice. There is mounting evidence that foreign technology adoption does not always raise productivity, especially if frontier innovation is mismatched with the local context (Basu and Weil, 1998; Acemoglu and Zilibotti, 2001). It is well understood that advanced technology in agriculture and medicine, for example, is tailored toward the environments of high-income countries and therefore less effective elsewhere (Kremer and Glennerster, 2004; Hotez, 2021; Moscona and Sastry, 2025). From this perspective, R&D investment in locally appropriate technology may be an important policy tool to spur growth in low and middle-income countries. Yet, there is little empirical evidence regarding the efficacy or cost-effectiveness of public R&D investment outside of frontier nations.

In this paper, we study one of the most prominent examples of public R&D investment in a developing country: the Brazilian Agricultural Research Corporation (Embrapa), established in 1973 to "promote, stimulate, coordinate and carry out research activities, with the objective of producing knowledge and technology for the agricultural development of the Country" (Brasil, 1972). Original scientific research was Embrapa's explicit goal from the outset. In the words of one of its founders, Eliseu Alves:

> [T]he major problem in Brazilian agriculture was not a lack of potential. The potential existed, but there was no science capable of generating technology suited to what we needed. To address this, we needed an institution capable of focusing high-level science on solving the concrete problems of Brazilian agriculture (Alves and Duarte, 2018, pp. 83-84).

Embrapa opened research centers throughout the country, including in areas with limited pre-existing R&D or agricultural production, and trained agricultural scientists to populate these regional hubs (Correa and Schmidt, 2014). Researchers at these centers developed soil modification tools adapted to the acidic and nutrient-poor soils of central Brazil, identified genetic traits that conferred resistance against Brazil's specific pest threats to crop production, and released hundreds of new crop varieties, including the first that allowed for the production of soy in tropical latitudes, among other advances (see Monteiro et al., 2012; Correa and Schmidt, 2014, and also Section 2). During the subsequent decades, Brazil transitioned from being a major food aid recipient to one of the world's largest agricultural exporters, a transformation that is particularly striking in light of the vast and persistent disparities in agricultural productivity across countries (Caselli, 2005). Qualitative accounts attribute much of this change to public R&D (Klein and Luna, 2018).

Our main contribution is to quantify the overall effect of public R&D investment on agricultural innovation and productivity growth in Brazil. To do this, we obtain information on the size and geographic expansion of Embrapa, construct a novel dataset of the research trajectories of all of Brazilian agricultural scientists, and draw on nine rounds of Brazil's Agricultural Census. Using our researcher-level data, we document that Embrapa re-directed research toward local ecological conditions and staple crops and overcame the obstacles to research productivity in places with limited pre-existing R&D capacity. Exploiting the staggered establishment of Embrapa's research centers, we find that R&D investments had large positive effects on agricultural productivity. Combined with a model, these estimates imply that Embrapa raised Brazilian agricultural productivity by 110% with a benefit-cost ratio of 17. This effect is largely driven by the spread of research centers across the country, which unlocked the development of appropriate technology for Brazil's diverse ecological zones. Our findings support the idea that targeted public R&D investments can allow countries to escape the "technology mismatch trap."

Measurement. To study how Embrapa affected the trajectory of scientific research and agricultural productivity, we assemble data from four types of sources. First, we construct a detailed history of Embrapa's expansion, including the location, founding year, and annual budget of each research center, as well as the name and position of all scientists.

Second, we construct a detailed record of agricultural science in Brazil from the near-universe of resumes of agricultural scientists in Brazil. Specifically, we compile data from Brazil's Lattes platform, a government-run database in which all researchers are required to have an up-to-date CV in order to apply for any form of public funding. From these data, we measure individual scientists' employment history and research output. Using keyword searches of all publication titles, we categorize research output by topics. Our final dataset covers over 35,000 researchers and more than 1.3 million research articles.

Third, to measure agricultural productivity we compile all rounds of the Agricultural

Census of Brazil from 1960 to present. The Census reports municipality-level information every five or ten years on revenues, yields, land use, land values, and input expenditure. We supplement the Census with additional data on crop-specific yields from the Municipal Agricultural Production (PAM) survey from 1974 the present.

Fourth, we compile municipality-level data on a broad range of geographic and ecological characteristics. These include the biome in which the municipality is located, the presence of specific crop pests and pathogens, and various measures of climate, topography, and soil characteristics. We use this to study the extent to which agricultural research focuses on characteristics of the immediate environment and to develop measures of ecological similarity between municipalities to proxy for agricultural technology mismatch.

Results: Agricultural Research. In the first part of our analysis, we study whether—and if so, how—Embrapa affected the trajectory of agricultural research in Brazil.

First, we study how Embrapa affected the focus of agricultural research across topics. Compared to other research in Brazil (e.g., at universities or private companies), articles written by Embrapa researchers are considerably more likely to mention a Brazilian biome, a major Brazilian pest or pathogen, or one of the staple crops singled out in Embrapa's founding to address the nation's food insecurity. The effects for ecological conditions are attenuated when we include location fixed effects to control for geographic differences in the composition of research. Consistent with the geography of research determining its ecological focus, we show that researchers are much more likely to study local ecological conditions. Our findings are consistent with historical accounts of Embrapa's aim to study all ecologies of Brazil by bringing research infrastructure to those locations (Alves, 1988).

Second, we study whether Embrapa's differential focus had any effect on the aggregate direction of innovation—that is, did Embrapa increase total research on specific ecological conditions and crops, or simply reallocate research to the public sector? Exploiting the opening of Embrapa research centers that have an explicit crop or biome focus, we find that these topic-specific center openings increase aggregate research activity on those topics, with no evidence of anticipation effects. This includes research by scientists who are not employed by Embrapa. Thus, Embrapa did not "crowd out" other research on its focus topics—if anything, there was "crowd in."

Third, we investigate the effect of Embrapa on researcher productivity. To cover Brazil's heterogeneous ecology, Embrapa established centers in potentially research-unproductive areas with low pre-existing human capital, research capacity, and researcher agglomeration. We can separately identify the effects of employer and place on research productivity using our unique researcher-level panel data, in which individuals move across both employers and locations. We find that employment at Embrapa substantially increases researcher productivity. Moreover, the effect of Embrapa is larger in more remote regions of the country and, quantitatively, more than compensates for the negative direct effect of working in these regions. These results suggest that Embrapa was able to overcome constraints to research in less developed parts of Brazil. The results are all similar if we restrict attention to articles published in high-impact, internationally-recognized journals.

Results: Agricultural Productivity. We next evaluate Embrapa's effect on agricultural productivity. Our approach is to use regional panel data and heterogeneous exposure to Embrapa across space and time for causal identification.

We measure local exposure to Embrapa by combining two pieces of variation. The first piece is *cross-sectional*: motivated by our results documenting the local ecological focus of agricultural research, we measure the potential suitability of research developed in each Embrapa center for all other municipalities using an environmental similarity index based on several characteristics of climate, topography, and soil conditions (as in Moscona and Sastry, 2025; Bazzi et al., 2016). The second piece is *time-varying*: we measure the staggered expansion of Embrapa across places. Combining these sources of variation, we construct *Embrapa Exposure* for each municipality and year as the maximum ecological similarity among centers that have been founded as of that year. This measures how Embrapa's research became more suited to different locations as the corporation expanded.

In our main empirical design, we study the effect of Embrapa Exposure on various local agricultural outcomes, as measured in the Census of Agriculture (1960-2017). We control for place and time fixed effects, absorbing fixed cross-sectional differences (e.g., direct effects of geography) as well as aggregate trends that may spuriously coincide with Embrapa's expansion. The central identification assumption is that the founding of new Embrapa centers was orthogonal to agricultural production trends in ecologically similar (compared to ecologically distant) Brazilian municipalities.

Our main finding is that exposure to Embrapa increases agricultural productivity. We show that this is not driven by physical distance to Embrapa centers by both controlling directly for physical distance to research centers and by dropping municipalities that are sufficiently close to the centers. Our findings are similar using several strategies to measure productivity, including the flow of production value per acre, the stock of land value per acre, or a broader measure of total factor productivity that also takes into account labor and intermediate inputs. Turning to dynamics, we find no evidence of anticipation (i.e., no pre-trends). Moreover, the effects accumulate over time: the long-run effect on local productivity is 30-40% larger than the within-decade effect. Finally, to support a causal interpretation of our findings, we conduct a falsification test in which we construct "placebo treatments" based on plausible alternative expansions of Embrapa across both time and space. Our estimates are in the right tail of the placebo estimates ($p < 0.01$).

We next investigate the mechanisms underlying these effects on productivity. First, we study the role of input intensification and land conversion. Embrapa exposure leads to higher expenditure on intermediate inputs like seeds, fertilizers, and chemical defenses, which were the main focus of Embrapa's technology development. Cropland also expands, but at a slower rate than production, while land devoted to pasture declines, consistent with reallocation driven by improved relative productivity of crops. Second, we isolate the effect of Embrapa on the staple crops that were the focus of its research efforts by exploiting variation in productivity trends across crops and within municipalities. We find that exposure to Embrapa has a large positive effect on the crops taht were the focus of its innovation, and little to no effect on other crops. These findings are consistent with directed innovation as the driving mechanism for our main findings.

The Returns to R&D. We finally combine our empirical strategy with a model to quantify the aggregate productivity consequences of Embrapa and the cost-effectiveness of its investments. The model captures not just technology mismatch between ecologically distinct places, the main focus of our reduced-form empirical analysis, but also scale effects and imperfect substitutability between research output from different centers. We estimate the model via a nonlinear least squares strategy that builds on and nests our reduced-form specification. Using the model, we compare observed agricultural productivity with a counterfactual in which Embrapa's research is held at its inception level.

We find that Embrapa increased aggregate agricultural productivity by 110%. This is 39 percent of the total agricultural productivity growth in Brazil between 1970 and the present as estimated by Fuglie (2015). Combining these estimates with the total expenditure of Embrapa, we calculate that the benefit-cost ratio of Embrapa's R&D was 17. Thus, while the cost of Embrapa was considerable—about 1% of Brazil's total agricultural GDP at its peak, comparable to the scale of investment in the US (Correa and Schmidt, 2014)—our analysis suggests that the benefits were considerably larger.

We close our analysis by studying how Embrapa's geographic scope shapes our benefit-cost analysis. To this end, we construct a series of counterfactuals, in which Embrapa pours its entire budget into a single large center in different locations of Brazil. Operating just in Brasília, where its largest headquarters is located, achieves less than two-thirds of the productivity gain of the full program. Moreover, the benefit-cost ratio is similarly lower, implying that expenditures in a single location are less effective than investments spread across many ecologically distinct locations. Yet, operating solely in Brasília is 50 percent better than the average return across different potential centers—indicating that the decision to place the headquarters there is consistent with maximization of the benefits.

Related Literature. This paper relates to several strands of existing literature. The first is a body of work studying the consequences of public R&D. Existing work focuses primarily on high-income countries (e.g., Howell, 2017; Azoulay et al., 2019; Kantor and Whalley, 2023; Gross and Sampat, 2023, 2025; Moretti et al., 2025). Most related, Kantor and Whalley (2019) study the historical effects of agricultural experiment stations in the US and how their effects decline with distance. Our contribution here is twofold. First, we investigate the effect of public R&D in a developing country, where both costs and benefits could be very different, and directly estimate how new technology affects productivity. Second, we show how public R&D can spur the development of locally appropriate technology, the absence of which has been shown to reduce productivity in low and middle income countries (Stewart, 1978; Lerner et al., 2024; Moscona and Sastry, 2025).

A small body of work estimates the returns to R&D in rich countries, generally following one of two strategies: linking R&D investments to productivity in the aggregate time series (Jones and Summers, 2020; Fieldhouse and Mertens, 2023) or separately estimating the value of the individual technologies that result from public R&D investments (Griliches, 1958; Azoulay et al., 2019). We develop a new approach that leverages heterogeneous exposure to research investment in panel data to causally identify the effect of R&D on productivity directly. This circumvents some of the identification challenges associated with time-series approaches without requiring the researcher to take a stand on the social value of individual technologies, a challenging object to estimate that can often vary substantially from private valuations (Griliches, 1979; Nordhaus, 2004).

Second, we contribute to the literature on the geography of innovation. Internationally, existing work on "inappropriate technology" has documented that the concentration of innovation in a few places can limit its benefits for the rest of the world (Basu and Weil, 1998; Acemoglu and Zilibotti, 2001). We show that public R&D can spur the development of locally-appropriate technology and raise productivity. Within the US, existing work has found evidence of external returns to scale (agglomeration externalities) in innovation and drawn implications for place-based policies that target "high tech clusters" (Moretti, 2021; Gruber and Johnson, 2019). In contrast to these studies, we explicitly model the geographic specificity of innovation and show that the scope of new technology investments contribute substantially to their productivity effects. In this context, spreading research activity out—and not concentrating it in a few clusters—has the highest return.

Third, we contribute to the literature on determinants of global productivity gaps in agriculture (e.g., Gollin et al., 2014; Adamopoulos and Restuccia, 2022), especially those investigating drivers of agricultural productivity growth in Brazil (Bustos et al., 2016; Pellegrina, 2022; Pellegrina and Sotelo, 2024). A dominant strand of this literature has emphasized frictions inhibiting technology adoption (e.g., Conley and Udry, 2010; Duflo et al., 2011; Suri and Udry, 2022). Our results, on the other hand, emphasize the important role of technology development. This is consistent with work emphasizing the role of philanthropic investments in tropical agricultural R&D during the Green Revolution (e.g., Evenson, 2001; Evenson and Gollin, 2003; Pingali, 2012; Moscona, 2019; Gollin et al., 2021) and, more generally, how ecological mismatch with centers of agricultural R&D can help explain global disparities in agricultural productivity (Moscona and Sastry, 2025). We show that public R&D investment in Brazil helped escape this "technology mismatch trap."

Fourth, we relate to an existing body of work studying Embrapa itself. Many of these studies are qualitative in nature and, to assess impact, focus on case studies of Embrapa's research and technology development (e.g., Gasques et al., 2012; Correa and Schmidt, 2014; Klein and Luna, 2018). Others have quantified the economic effects of specific crop varieties developed by Embrapa (e.g., Pardey et al., 2006; Gasques et al., 2009). We compile new data and develop new empirical strategies to systematically measure the effects of Embrapa on agricultural innovation and productivity growth in Brazil.

Outline. This paper is organized as follows. Section 2 provides background information about Embrapa and Brazilian agriculture. Section 3 describes our data and measurement strategies used in our empirical analysis. Section 4 presents our results on the impact of Embrapa on the rate and focus of agricultural innovation, and Section 5 presents our results on the impact of Embrapa on agricultural productivity growth. Section 7 concludes.

2 Background: Brazilian Agriculture and Embrapa

This section reviews the history of Brazilian agriculture, from the country's dependence on foreign food aid as recently as the 1960s to its ascendance as the world's third largest agricultural exporter. In this context, we introduce the institutional background of Embrapa and discuss its founding, development, and role in agricultural R&D.

2.1 Brazilian Agriculture Before the 1970s

In the mid 20th century, Brazil's agricultural productivity was relatively stagnant and low by global standards. Figure 1 plots yields over time for six major crops. Yields for major staples (maize, wheat, and rice) are close to flat between 1950 and 1970. Even historically important cash crops like sugarcane, which fueled Brazil's colonial economy, were under half as productive as the corresponding sectors in the United States (Klein and Luna, 2023).

During the late 1960s, pressures on Brazil's food-production sector intensified due to rapid urbanization and population growth. The country, already a net importer of food, became increasingly reliant on foreign donations and food aid (Vieira Filho and Fishlow, 2017; Martha Jr et al., 2012). There was a growing realization among policy makers that expanding domestic food production could be essential for staving off food insecurity and political pressure from urban constituents facing high food prices.

A key roadblock, according to agronomists, was the absence of available technology that was productive on Brazilian land (Alves and Duarte, 2018). While the Green Revolution had more immediate effects on developing countries that hosted major breeding centers (e.g., Mexico and the Philippines), it did not directly benefit Brazil. A potential explanation was that Brazil had a very different geography, ecology, and farming practice. For example, new crop varieties were ill-suited to Brazil's acidic soils (Vilaça de Vasconcelos et al., 2022). Moreover, many were intended for use alongside substantial irrigation, but Brazilian agriculture was largely rain-fed (Cabral et al., 2022).

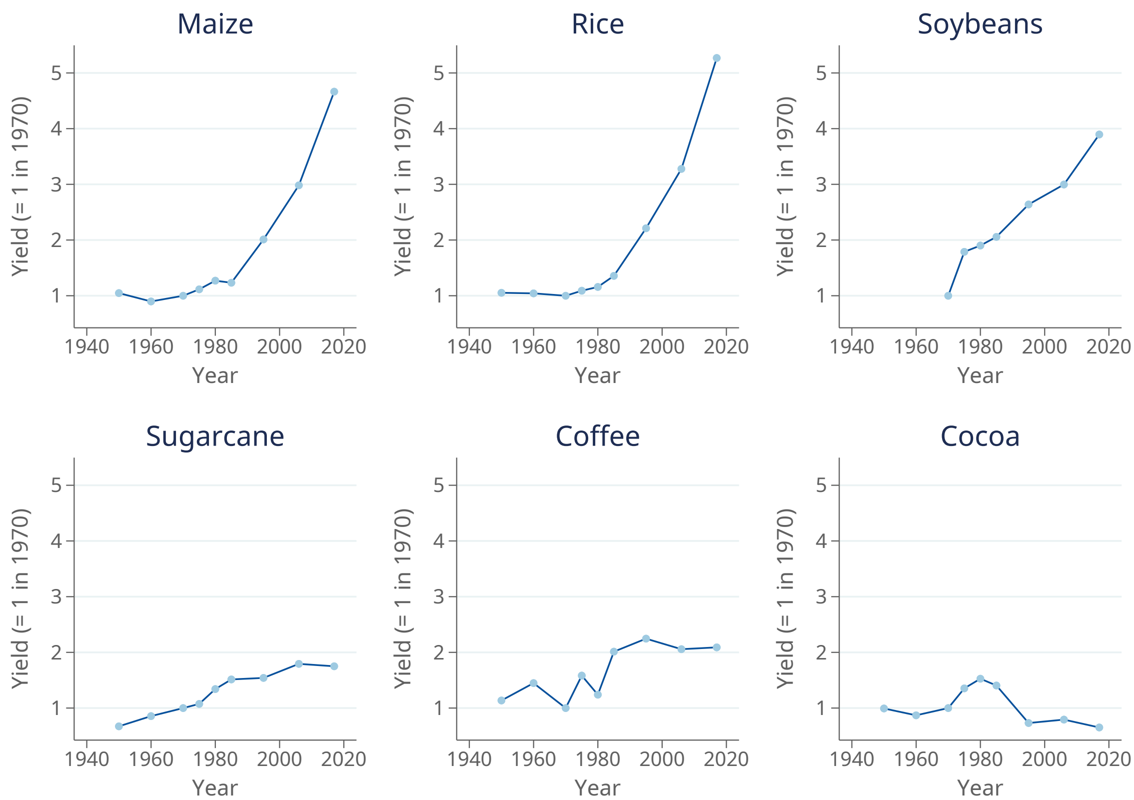
Brazil's own agricultural research was limited and concentrated on wealthier states and export-oriented cash crops (Embrapa, 2006; de Barros and de Barros, 2005). For example, by 1960 more than half of the non-fruit varieties developed by the Agronomic Institute of Campinas, one of Brazil's main research centers, were for coffee, sugarcane, and cotton. Much of the remaining investment was devoted to importing and testing foreign technology, based on the presumption that technology adoption could close the productivity gap with the rest of the world (Martha Jr et al., 2012).

2.2 Embrapa: Origins and Design

Against this backdrop—and a 1973 price shock that put Brazil's dependence on food imports into starker relief—Brazil's central administration embarked on large-scale investments to modernize Brazilian agriculture. The centerpiece of this project was the Empresa Brasileira de Pesquisa Agropecuária (Embrapa), a public corporation devoted to agricultural research and development.

The working group tasked with designing Embrapa, consisting mainly of economists and agricultural scientists, identified two main challenges blocking Brazilian agricultural productivity growth. The first was the geographic centralization of the research structure, which limited agricultural science to a small fraction of the country. The second was the lack of trained and specialized personnel and of attractive career paths in agricultural research (Cabral, 2005, 38-50). The final report of this group, widely known as "The Black Book" (*Livro Preto*), served as the guiding document in Embrapa's formation in 1973.

Figure 1: Brazilian Crop Yields Over Time



Notes: This graph shows the evolution of physical yields (tons of output per hectare) for six crops in Brazil, using data from the Brazilian Census of Agriculture from 1950 to 2017. Each dot indicates a separate observation. In each panel, we normalize the yield in 1970 to 1. Data for soybeans are not available before 1970.

Three key organizational principles for Embrapa emerged from this plan: the organization's scale, geographic scope, and structure as a public corporation.

Scale. The investment in Embrapa was large relative both to past efforts in Brazil and contemporary efforts elsewhere in the world. By the 2000s, Embrapa's spending on agricultural R&D as a share of agricultural GDP was comparable to that of all public agricultural R&D in many high-income countries, including the US (Correa and Schmidt, 2014), and roughly triple that of public agricultural R&D in India and China (OECD, 2022). In 2010, Embrapa's budget reached 1.9 billion reais in current values, or about 1.15 billion dollars, roughly thirteen times the value of Brazilian public agricultural R&D at the time of its founding. For comparison, the US Department of Agriculture (USDA) R&D budget in 2010 was 2.3 billion dollars (Sargent et al., 2009).

Embrapa was a similarly large program measured by its employment, especially of agricultural researchers. By 2010, Embrapa's employed about 2,300 agricultural scientists in Brazil, or 40% of the total number estimated to be in the country by the Agricultural Science and Technology Indicator (ASTI). This number is again comparable to the US, where the Agricultural Research Service (ARS), the USDA's research arm, employs roughly two-thousand scientists and post-doctoral researchers (USDA ARS, 2024).

Geographic Scope and Specialization. Embrapa's central goal was to study all regions and ecosystems of Brazil (Alves, 1988). This was a response to the founders' diagnoses for the earlier stagnation of Brazil's agriculture and agricultural research: over-concentration in a few areas, an excessive focus on existing cash crops (which were grown in the same areas), and a lack of science and technology suited to Brazil's diverse geographic features. At the time of Embrapa's founding, the priority regions for expanding research and agricultural production were the Northeast, the Cerrado, and the Amazon (Embrapa, 2006).

Embrapa's efforts in the Cerrado typify the organization's approach to incorporating new geographies in agricultural research. The Cerrado is a two-million square kilometer tropical savanna (see also Section 3 and Figure 2). A defining feature that made farming with existing technology unproductive is the Cerrado's highly acidic soils. In the words of Embrapa's founder, agricultural expansion into the Cerrado required "a better understanding of the climate, soils, water availability, flora, land, and, ultimately, an entire ecosystem" that had been neglected by existing science (Cabral, 2005). Embrapa Cerrado was established in 1975 as one of the organization's first centers.[1]

Structure. Embrapa was structured as a national public company operating under an autonomous legal framework. In principle, this structure enabled flexible interactions with the private sector, public universities, and other organizations (Cabral, 2005; Embrapa, 2006; Martha Jr et al., 2012).

Relatedly, as we will further describe below, much of Embrapa's efforts centered on developing technologies that could be widely marketed to farmers. Embrapa explicitly encouraged a "problem-based" approach among its researchers and discouraged "curiosity-driven" research, insofar as the latter distracted from the goal of developing Brazilian agriculture at scale (Correa and Schmidt, 2014; Embrapa, 2006). As one important example of these efforts, Embrapa directly participated in the commercialization of seed varieties, either by itself or in cooperation with external partners (Correa and Schmidt, 2014). It also conducts extensive interviews with farmers from all regions of Brazil in order to tailor its research investments to their specific constraints and production threats (Cruz, 2025).

[1] A similar impulse motivated the 1975 establishment of Embrapa Rondônia in the heart of the Amazon, more than a thousand kilometers from any pre-existing agricultural research station. This unit's task was to develop new biotechnology to improve agriculture in the Amazon region (Embrapa, 2025b).

2.3 Embrapa's R&D and the Growth of Brazilian Agriculture

Embrapa is the main developer of agricultural technology in Brazil. Its researchers have developed more than 350 crop varieties and 200 international patents to their credit (Correa and Schmidt, 2014). Case-study evidence suggests that Embrapa technology was central to several developments in Brazilian agriculture over the last 50 years. Below, we briefly highlight some of the most salient developments.

Expansion into the Cerrado. An early priority of Embrapa was to establish a presence in the Cerrado. As of the 1970s, the Cerrado was a region of "low-productivity activity, such as extensive cattle ranching" (Correa and Schmidt, 2014, p. 3). The constraints to agricultural production in the Cerrado are myriad, including high temperatures, lengthy dry spells, very low soil pH, significant nitrogen deficiency, and high saturation in toxic aluminum (Scheid Lopes et al., 2012). Norman Borlaug, the father of the Green Revolution, averred that "nobody thought these soils were ever going to be productive" in the global research community during the 1960s and 1970s (Rohter, 2007).

Nonetheless, regions within the Cerrado have become central to Brazil's modern agricultural economy, developing into a highly diversified agricultural hub. According to the 2006 agricultural census, the Cerrado accounted for 20 percent of national maize production, 42 percent of soy, 7 percent of vegetables, 65 percent of cotton, and 10 percent of sugarcane. In contrast, during the 1970s the region accounted for no more than 5 percent of national production of any of these crops.

Prior studies suggest that Embrapa and its research into soil chemistry were central to widespread use of agricultural liming, which allowed farmers to neutralize the Cerrado's acidic soils (Vieira Filho and Fishlow, 2017; Correa and Schmidt, 2014). Embrapa scientists were also active in researching techniques for nitrogen fixation that could overcome the soil's nutrient deficiency. Two Embrapa researchers have, in subsequent decades, received the World Food Prize for research related to each of these strands: Edson Lobato, for studying the soil chemistry of the Cerrado, and Mariangela Hungria, for developing techniques for bacterial nitrogen fixation. Borlaug himself, in an interview, summarized thusly: in the Cerrado, "Embrapa was able to put all the pieces together" (Rohter, 2007).

The Growth of Soy. An important engine of Brazil's agricultural growth has been the enormous expansion of cultivation of soybeans, a crop for which Brazil today ranks as the world's second largest producer and exporter. Soybeans are, by nature, a temperate crop, originating from the northeast of China. The expansion of soybeans into Brazil's tropical latitudes necessitated new technology on at least two fronts: techniques to tame the soil of the tropical savanna, described above, and soybean varieties that were adapted to tropical latitudes.[2] In 1975, Embrapa Soja was established in the state of Paraná with the goal of "tropicalizing" the soybean. In 1980, Embrapa created a first soybean variety adopted to tropical latitudes and, in subsequent decades developed about 200 total soy varieties. Monteiro et al. (2012) argue that Embrapa's investments were critical to the "first phase" of expansion of the Brazilian soy industry.

The 1970s and 1980s saw Brazil become the world's second largest soy producer, with a more than five-fold increase in global market share relative to 1970 (Monteiro et al., 2012). This phase predates subsequent developments, including the introduction of genetically modified soy and entry of multinational processing companies, which began during the mid-1990s (see Bustos et al., 2016). Both of these phases are visible in Figure 1, which shows soybean yields increasing dramatically in each decade starting from 1970.

The Broader Scope. There are many additional examples of successful new technologies developed by Embrapa. This includes the high-yielding rice varieties that have led to yield increases of nearly one percent per year (Magalhães Jr. and Stone, 2018); cotton varieties adapted to Mato Grosso, thousands of kilometers away from the previous centers of cotton production in São Paulo and Paraíba (Klein and Luna, 2023); and, more recently, transgenic crop traits, culminating in the release of a genetically-modified maize variety that is resistant to local pests and pathogens, including the fall armyworm, one of the most damaging pests in Brazil (Embrapa and Helix, 2022).

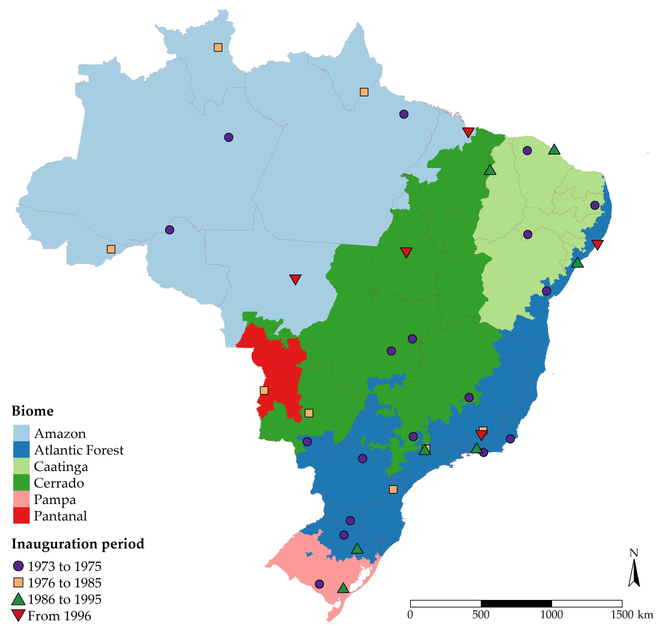
Thus, the historical record establishes Embrapa's size, scope, and footprint on several classes of agricultural technology. Our goal is to move beyond a handful of case studies and investigate whether government investment in R&D had a systematic and causal effect on innovation and productivity growth in Brazil. Did Embrapa "lead" technological progress and productivity growth, or merely follow pre-existing trends? Did new technology development translate into meaningful changes in agricultural productivity? And if so, were the productivity benefits sufficient to justify the scale of research investment? Answering these questions requires turning to data, which we describe in the next section.

3 Data and Measurement

Our empirical analysis combines information on the organization and geographic expansion of Embrapa with detailed data on Brazilian agricultural research, production, and ecology. We summarize these data sources below, relegating details to Appendix A.

[2] See Pellegrina (2022) for more details about the expansion of soybeans to tropical regions.

Figure 2: Embrapa's Research Centers



Notes: Locations of Embrapa research centers by year of creation, overlaid with Brazil's major biomes and state boundaries. Data sourced from Embrapa and IBGE.

Embrapa's Organization and Budget. We compile comprehensive information about Embrapa's organizational structure through a government transparency request (Embrapa, 2022a). First, we collect information on the founding year and address of each unit. Figure 2 shows the distribution of Embrapa research centers across Brazil, categorized by founding year. While Embrapa was founded in 1973, it continued to expand over the subsequent decades, opening new centers well into the 2000s. The centers are spread across all of Brazil's major ecological zones and regions.

Second, we obtain detailed data on Embrapa's budget through a second transparency request (Embrapa, 2022b). For every center and year since 1974, we compile information on all personnel, operational, and capital expenses. Comprehensive data on the costs of R&D programs are rare—especially over extended periods of time and for large-scale investments—which is a key barrier to generating credible estimates of the returns to R&D investment (Jones and Summers, 2020). Thus, these cost data provide a unique opportunity to perform a benefit-cost analysis (see Section 6).

Agricultural Research. To measure agricultural research across topics, institutions, and time, we construct a novel database of the career trajectories and research output of agricultural researchers in Brazil. This allows us to study how the growth of Embrapa—as well as employment by one of its centers—shifted the rate and direction of research.

The key source of information used to construct this database is Brazil's Lattes platform, an integrated information system managed by the Brazilian government that stores researchers' CVs, including their publication, employment, and educational histories. Each researcher is responsible for updating their own information and an up-to-date Lattes profile is required for all government funding, support, and collaboration or recognition. As a result, there are strong incentives for all researchers to maintain a complete and current profile. Moreover, the Lattes platform is publicly available, maintained by the National Council for Scientific And Technological Development of Brazil (CNPq), making it possible to freely access the detailed history of all researchers in the database.

We collect the full Lattes profiles of all individuals with any listed expertise in the agricultural sciences.[3] This includes each individual's educational history (i.e., degrees, graduation years, institutions, and dissertation titles); full employment history (i.e., institutions, employment years, and position titles); and full publication history (i.e., publications, journals, publication years, and titles).[4] We validate the coverage of this database by comparing it with the official list of all Embrapa-affiliated researchers, and find that that 94% of all Embrapa researchers appear in our Lattes-derived data. Our final database contains 35,602 unique researchers—6,259 of whom were employed by at least one Embrapa research center during their career—and approximately 1.3 million research articles.

To study how Embrapa affected the location and direction of research, we geo-located all employers and institutions by municipality using ChatGPT (GPT-4o). In addition, we used keyword searches in the title of each article to determine its topic (see Appendix A.2). We identify all articles related to Brazil's major biomes, crop-affecting pests, and crops.

Agricultural Production. Our main source of data on agricultural production is the Agricultural Census of Brazil. While the Census has been conducted since 1920, digital versions are available only for the 1995-6, 2006, and 2017 rounds. We therefore digitize all rounds since 1960 from scanned *pdf* files published online by the Brazilian Institute of Geography and Statistics (IBGE). Using these data, we construct a municipality panel data with information on agricultural output, land use (i.e., devoted to crops versus pasture), land values, technological input use (e.g., use of tractors, chemicals, fertilizers), labor input use, and farm size (see Appendix Table C.1).

We supplement the agricultural censuses with data from the Municipal Agricultural

[3]Namely, those who listed *Agronomia*, *Ciência e Tecnologia de Alimentos*, *Engenharia Agrícola*, *Recursos Florestais e Engenharia Florestal*, *Recursos Pesqueiros e Engenharia de Pesca*, and *Zootecnia*, all subfields of the "Agricultural Sciences" field, following the CNPq classification.

[4]We merged all publications in the database to the metadata of the associated journal, including journal publishing institution, country, and impact factor (SciELO, 2025; SCImago, 2025; Elsevier, 2025).

Production (PAM) survey from 1974 to present, which collects annual information on the output and land devoted to sixty-four crops and products across all municipalities (IBGE, 2023). PAM has the advantage of covering a much broader set of crops than the census.

Finally, since Brazil's municipal borders shifted during the sample period, to make geographic units consistent over time we follow Brazil's statistical agencies and link all data to minimal consistent border units (*Área Mínima Comparável*, AMC) (IBGE, 2011).

Geo-spatial and Ecological Data. Finally, we compile data on Brazil's ecological characteristics. First, we categorize regions of Brazil into its major biomes, using the biome classification from Brazil's main statistical agency (IBGE, 2024).[5] Figure 2 displays the distribution of these biomes, with Embrapa's research centers superimposed. This variation in ecology, much of which is unique to Brazil and not present in parts of the world where most agricultural innovation takes place, was a driving motivation behind Embrapa. Second, we compile data on the geographic distribution of Brazil's most damaging crop-affecting pests and pathogens (see Appendix Table A.1) using data from the Center for Agricultural Biosciences International Crop Protection Compendium. We identify the set of Brazilian states in which each pest and pathogen is known to be present. Together with the biome classifications—and our data on research by topic—these data allow us study how the expansion of Embrapa to new ecological zones shifted research focus.

Last, to quantify ecological differences across Brazilian municipalities, we use geo-spatial data on the distribution of nine agro-climatic characteristics that are critical to agricultural production, including temperature, precipitation, elevation, ruggedness, the length of the growing season, soil acidity, soil clay content, soil silt content, and soil coarse fragment content (see Appendix Table A.2).[6] We normalize each characteristic across municipalities and use these values to construct a measure of ecological distance between municipalities and Embrapa research centers, which is a central element in our evaluation of the impacts of Embrapa's expansion on agricultural productivity in Section 5.1.

4 Results: Embrapa and Agricultural Research

In this section, we investigate the effect of Embrapa on agricultural research in Brazil. First, we analyze how Embrapa shifted the direction of innovation. Second, we investigate

[5]The six Brazilian biomes are (a) the Amazon (49% of Brazil), a dense tropical rainforest; (b) the Cerrado (24% of Brazil), a tropical savanna; (c) the Atlantic Forest (13% of Brazil), a forested region along the Eastern seaboard; (d) the Caatinga (10% of Brazil), a semi-arid biome unique to northeastern Brazil; (e) the Pampa (2% of Brazil), temperate grasslands; and (f) the Pantanal (2% of Brazil), the world's largest tropical wetland.

[6]Our focus on this set of characteristics builds on Bazzi et al. (2016) and Moscona and Sastry (2025), who confirm that these characteristics are key determinants of crop-specific technology and knowledge transfer.

how Embrapa affected the productivity of innovation, particularly as it expanded to more remote parts of the country where existing research infrastructure was more limited.

4.1 The Direction of Research: Ecological Conditions

We first investigate whether employment at Embrapa shifted researchers' focus toward Brazilian ecological conditions, exploiting our topic-labeled database of all agricultural research publications. We define articles related to Brazilian ecological conditions as those that mention one of Brazil's major biomes or those that mention one of Brazil's major crop-affecting pests and pathogens. Our main regression equation is:

$$\mathbb{I}\{\text{Article } p \text{ mentions topic } k\}_{prit} = \beta \cdot \mathbb{I}\{\text{Embrapa}\}_{pit} + \alpha_t + \delta_i + \epsilon_{prit}, \tag{1}$$

where r indexes researchers, i indexes municipalities, and t indexes years. The unit of observation is an article p and $\mathbb{I}\{\text{Embrapa}\}_{pit}$ is an indicator that equals one if the article was written by an individual employed by Embrapa in year t. The outcome is an indicator that equals one if the article mentions (i) one of Brazil's biomes or (ii) one of Brazil's major crop-affecting pests or pathogens.

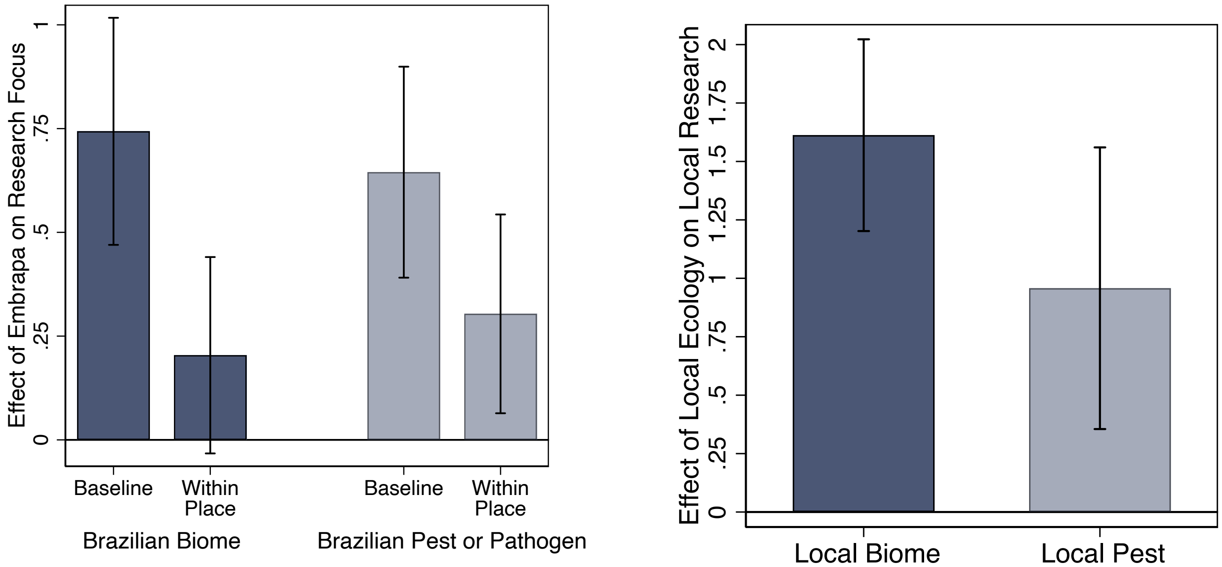
One reason researchers at Embrapa may have been disproportionately likely to study Brazil's ecological conditions is that Embrapa set up centers in parts of the country where existing research was limited. This difference in *where* research takes place could go a long way in determining the areas of focus since crop breeding and agricultural technology development often have to be finely tailored to the local environment (see Section 2.2). As a preliminary test of this mechanism, we include municipality fixed effects δ_i in estimates of equation (1); if the location of research drives the differential focus of Embrapa research, we would expect the main effect to be attenuated when location fixed effects are included.

In Figure 3a, we report our estimates of β normalized by the mean of the outcome variable. Embrapa researchers write significantly more articles about Brazilian biomes and crop-affecting pests. The effects are equal to 74% and 64% of the sample mean, respectively (first and third bars). Moreover, the estimates are substantially attenuated after including municipality fixed effects (second and fourth bars) suggesting that the location of Embrapa's research can explain a large share of the differential research focus.

We next provide direct evidence that research is disproportionately focused on local ecological conditions. We collapse the article-level database to the municipalities-topic-year level and estimate the following regression model by Poisson maximum likelihood:

$$\text{Articles}_{ikt} = \exp\{\xi \cdot \mathbb{I}\{\text{Local}\}_{ik} + \alpha_{kt} + \delta_{it} + \epsilon_{kit}\} \tag{2}$$

Figure 3: The Direction of Research Across Ecological Conditions



(a) Embrapa's Effect on Topic Focus ($p \times i \times t$)

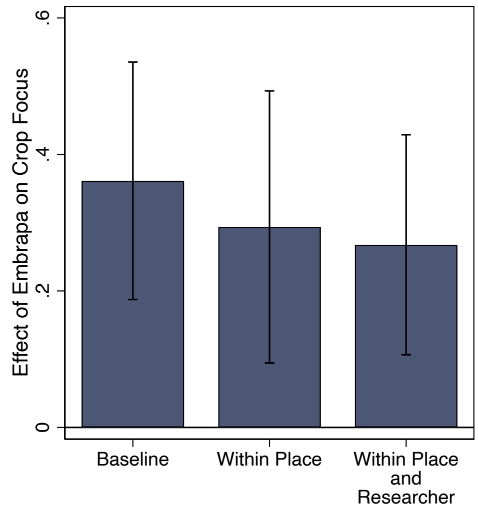
(b) Local Research Focus ($i \times k$)

Notes: In Panel A, the unit of observation is an article, and each bar represents a coefficient estimate from equation (1). We report β normalized by the mean of the outcome variable. In the first two bars, the outcome is an indicator that equals one if the article mentions a Brazilian biome and in the second two bars, the outcome is an indicator that equals one if the article mentions a Brazilian pest or pathogen. The second and fourth bars include municipality fixed effects as controls. In Panel B, the unit of observation is a municipality-topic pair, and each bar represents a coefficient estimate from equation (2). In both panels, standard errors are clustered by municipality and 95% confidence intervals are reported.

where the outcome variable is the total number of articles written about topic k in municipalities i and year t. Again, we define two sets of topics k: Brazil's biomes and Brazil's major pests and pathogens. When we focus on biomes, $\mathbb{I}\{\text{Local}\}_{ik}$ is an indicator that equals one if municipality i is located within biome k (see Figure 2). When we focus on pests, $\mathbb{I}\{\text{Local}\}_{ik}$ is an indicator that equals one if the municipality i is located in a state where pest k is present. If research is directed toward local ecological conditions, we would expect that $\xi > 0$ in both cases; that is, research focuses on the local biome or on pests and pathogens that are locally present.

Estimates of equation (2) are reported in Figure 3b. Researchers are substantially more likely to publish articles related to their local environment. In addition to highlighting the mechanism underlying Embrapa's re-direction of research, this finding directly motivates our empirical strategy in Section 5 that uses ecological similarity to Embrapa's research centers as a shifter of the ecological suitability of new technology development.

Figure 4: The Direction of Research Across Crops



Notes: The unit of observation is an article and each bar represents a coefficient estimate from equation (1). The outcome is an indicator that equals one if the article mentions one of Embrapa's focus crops (beans, cassava, maize, rice, soy, and wheat). In the second column, municipality fixed effects are included as controls, and in the third column, both municipality and researcher fixed effects are included as controls. Standard errors clustered by municipality and 95% confidence intervals are reported.

4.2 The Direction of Research: Crops

In addition to focusing on building knowledge about under-studied ecological zones, Embrapa also set its focus on a handful of staple crops that were most relevant for food consumption (Martha Jr et al., 2012). This was a departure from pre-existing agricultural research in Brazil, much of which had concentrated on historically high-value cash crops for export, such as coffee and sugarcane.

In this section, we study the extent to which Embrapa has redirected research focus across different crops. To do so, we estimate a version of equation (1) in which the outcome is an indicator that equals one if the article mentions one of Embrapa's focus crops (beans, cassava, maize, rice, soy, and wheat). Figure 4 presents our results. Researchers employed by Embrapa are substantially more likely to study these staple crops. Unlike ecological conditions, this effect is not driven by geographic location, since estimates are similar after including municipality fixed effects (second column). Moreover, the estimate is similar after including researcher fixed effects, suggesting that moving to Embrapa leads researchers to shift their focus toward Embrapa's focus crops (third column).

These results motivate a refinement of our main identification strategy in Section 5. To ensure that our estimated effect of Embrapa on productivity is driven by new technology development, we exploit variation in exposure to Embrapa not only across locations (due to their differential ecological similarity to Embrapa's centers) but also across crops (due to the fact that certain crops were the focus of Embrapa's research while others were not).

4.3 Aggregate Effects and Crowd-Out

So far, we have focused on how Embrapa shaped the direction of research at the individual level. However, the fact that Embrapa shifted individuals' research focus does not necessarily imply that it shifted the aggregate direction of innovation across topics. One possibility, for example, is that researchers employed by Embrapa crowded out research that would have taken place anyway. If this is the case, Embrapa could have had no impact on the overall amount of research conducted on its priority areas.

To investigate how the expansion of Embrapa shifted the aggregate focus of research, we exploit the opening of new Embrapa research centers over time and estimate the effect on the *national* distribution of research across topics. Our regression model is:

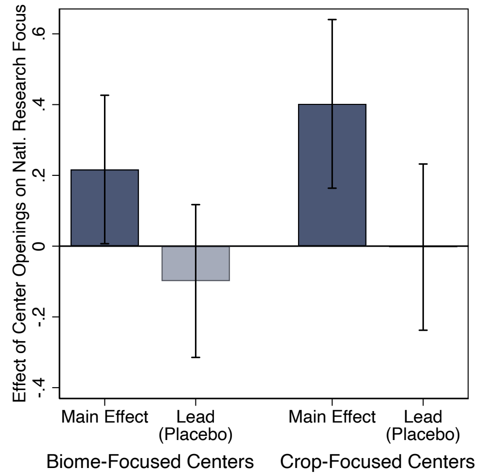
$$\text{Articles}_{kt} = \exp\{\gamma \cdot \text{Centers}_{kt} + \alpha_k + \delta_t + \epsilon_{kt}\} \tag{3}$$

where Articles_{kt} is the total number of articles written in Brazil about topic k in year t and Centers_{kt} is the number of Embrapa centers focusing on topic k. We exploit the fact that certain Embrapa centers had an explicit focus on certain crops (e.g., Embrapa Rice and Beans located in Santo Antônio de Goiás) or were located in specific biomes to learn about local ecology. Using this information, we estimate versions of equation (3) across crops and biomes. When k indexes crops, Centers_{kt} is the number of centers as of year t with an explicit focus on crop k, and when k indexes biomes, Centers_{kt} is the number of Embrapa centers as of year t that are located in biome k. $\gamma > 0$ would imply that Embrapa shifted the overall direction of innovation.

Figure 5 reports our estimates of equation (3). Across both margins, the expansion of Embrapa significantly shifted the aggregate focus of agricultural research ($\gamma > 0$; first and third bars). While no Embrapa centers had an explicit focus on individual pests, we estimate a version of equation (3) in which k indexes pests and Centers_{kt} is the number of centers located in states in which the pest is present; we estimate a smaller but still positive effect ($\gamma = 0.0502$, S.E.$= 0.0172$). One potential concern is that Embrapa centers were opened in response to aggregate research trends. However, we find no evidence of pre-existing trends: future changes in Embrapa center openings are not correlated with the current direction of research (second and fourth bars). The direction of research shifts only after new Embrapa centers open.[7] This aggregate effect is driven by an increase both in research

[7]Specifically, we estimate equation (3) after adding $\text{Centers}_{k(t+1)}$ in addition to Centers_{kt}.

Figure 5: Embrapa and the National Direction of Research



Notes: Each pair of bars corresponds to an estimate of equation (3) that includes both the contemporaneous and one-year leading (future) value of Centers$_{kt}$, in addition to topic and time fixed effects. The unit of observation is a biome-year in the first two columns and a crop-year in the second two columns. Heteroskedasticity robust standard errors are constructed and 95% confidence intervals are reported.

conducted by Embrapa affiliates *and* research conducted by other, non-Embrapa scientists (see Appendix Figure C.2). This result indicates that if anything, Embrapa-sponsored research had positive spillovers on other researchers (i.e., "crowd in").

4.4 Research Productivity

We now investigate how Embrapa affected researcher productivity. Embrapa's incentives to carry agricultural research to new parts of the country and new topics may have come at the expense of research productivity. This is both because there could be major inefficiencies in large-scale government programs like Embrapa, and because, as part of its strategy to cover topics and ecosystems that had not been the focus of prior innovation, Embrapa expanded to regions with limited pre-existing research infrastructure and potentially lower overall research productivity.

To study the impact of Embrapa on researcher productivity—both in research hubs and in more remote parts of the country—we use a movers-based design to test how employment at Embrapa affects research output. Our baseline specification is:

$$y_{rit} = \beta \cdot \mathbb{I}\{\text{Embrapa}\}_{rit} + \alpha_r + \xi_i + \delta_t + \gamma_{a(i,t)} + \epsilon_{rit}, \tag{4}$$

where y_{rit} is a monotone transformation of the number of papers published by researcher r in year t and municipality i, $\mathbb{I}\{\text{Embrapa}\}_{rit}$ is an indicator that equals one if researcher

i is employed by Embrapa at time t, and α_r is a researcher fixed effect. δ_t is a year fixed effect, ξ_i is a municipality fixed effect, and $\gamma_{a(i,t)}$ is a set of tenure fixed effects.[8]

We also estimate the effect of Embrapa separately in traditional research hubs and in more remote areas where it hoped to expand agricultural research. To do this, we include interaction terms between $\mathbb{I}\{\text{Embrapa}\}_{rit}$ and indicators that equal one if the municipality is in the top ten in terms of either total agricultural research output or human capital (as proxied by the share of college graduates).[9] A key question is whether Embrapa—by linking researchers across centers and connecting all affiliated researchers to its national research network—was able to overcome the research productivity disadvantages that may have existed in more remote areas where it hoped to spur new innovation.

Table 1 reports estimates of equation (4). Working for Embrapa is associated with higher research output, even conditional on individual fixed effects that absorb any differences in ability (column 1). This is despite the fact that, if anything, Embrapa-affiliated researchers have weaker publication incentives than researchers at other institutions given the emphasis on immediate translation and application (Martha Jr., 2025). The results are similar after absorbing municipality-by-year fixed effects that control for any changes in local policy, local funding, or other trends (column 2).

Next, we investigate how the effect of working at Embrapa varies across regions (columns 3-6). Intuitively, we find that moving outside one of the top twenty research hubs is associated with a decline in research productivity, captured by the coefficient on the "Low Research" indicator. However, working for Embrapa fully reverses the productivity disadvantage of these more remote regions (column 3). The positive association between employment at Embrapa and research productivity estimated on the full sample is driven almost entirely by the effect of Embrapa outside of traditional research hubs. The findings are similar defining research hubs based on the total output of agricultural innovation (columns 3-4) or based on the share of the population with a college degree (columns 5-6), and the point-estimates are very similar after fully absorbing all municipality-specific trends (columns 4 and 6). These results indicate that Embrapa's organizational structure enables high research productivity even in remote regions.

Turning to dynamics, we show that there is no evidence of pre-existing trends: movers' research productivity rises only *after* moving to an Embrapa center, and this effect remains concentrated in centers outside of traditional research hubs (Appendix Table C.2). Moreover, results are similar if we quality adjust the outcome by weighting each publication by

[8] We calculate researcher tenure as the current year minus the first recorded year of any employment spell.

[9] The results are very similar if we choose alternative cut-offs (e.g., define the hubs as the top fifteen or twenty municipalities by research output or college graduates).

Table 1: Effects of Embrapa Affiliation on Researcher Productivity

	ihs(Number of Papers)					
	(1)	(2)	(3)	(4)	(5)	(6)
Embrapa	0.081***	0.087***				
	(0.027)	(0.029)				
Embrapa x Low Research			0.203***	0.189***	0.121***	0.119***
			(0.072)	(0.067)	(0.028)	(0.026)
Embrapa x High Research			-0.019	0.048	0.043*	0.061**
			(0.063)	(0.069)	(0.024)	(0.026)
Low Research			-0.077***		-0.016**	
			(0.020)		(0.007)	
Adj. R2	0.428	0.454	0.430	0.455	0.428	0.454
Observations	530672	519562	530672	519562	530377	519291
Heterogeneity	-		Previous Research		College Degree	
Year FE	Y	Y	Y	Y	Y	Y
Municipality × Year FE	–	Y	–	Y	–	Y
Researcher FE	Y	Y	Y	Y	Y	Y
Tenure FE	Y	Y	Y	Y	Y	Y

Notes: Estimates of equation (4), with the inverse hyperbolical sine of publications as the dependent variable are reported. Columns 3-6 interact the Embrapa indicator with an indicator for low or high research capacity in the municipality, defined using either previous agricultural research (columns 3-4) or the share of college graduates (columns 5-6). All regressions include year, researcher, and job tenure fixed effects. Columns 2, 4, and 6 include municipality-year fixed effects. Standard errors clustered by municipality.

its journal impact factor and if we use alternative parameterizations of the dependent variable (Appendix Table C.3). Thus, the findings are not driven by insubstantial publications or by the functional form of our regression specification.

Together, these estimates suggest that Embrapa shifted the focus and geography of agricultural research without sacrificing research productivity. While researchers are generally less productive outside of Brazil's main research hubs, Embrapa researchers are not, due to the organization's larger productivity advantage in remote areas.

5 Results: Embrapa and Agricultural Productivity

The previous section has explored the effects of Embrapa on the direction and productivity of agricultural research. We now study the effects of Embrapa on agricultural productivity. We develop an empirical strategy that exploits time-series variation from the staggered introduction of Embrapa's research centers and cross-regional variation in the suitability of Embrapa's technology for different ecological conditions. After establishing that Embrapa had a large, positive impact on agricultural productivity, we decompose this effect across different margins of outputs and inputs and explore whether the effects were more pronounced for staple crops targeted by Embrapa's technology development.

5.1 Empirical Strategy

Measuring the returns to a large-scale R&D investment on economic productivity is challenging due to a fundamental trade-off between ruling out confounding effects and capturing indirect effects. One strategy in existing work is to focus on aggregate productivity outcomes, like the yields for major crops visualized in Figure 1. While the trend break in productivity for staple crops in the 1970s is suggestive of a large effect of policy changes at this time (Klein and Luna, 2023), it is difficult to isolate the role of a single policy of interest (e.g., the founding and expansion of Embrapa) from other factors and time-series trends taking place over the same period.

Another strategy is to break down large-scale investments in R&D into specific technologies that resulted from these investments, and develop tailored models to estimate their productivity effects and social returns (e.g., Pardey et al., 2006). This approach is challenging because it requires first identifying all relevant technologies and then accounting for the difference between their social and private returns, which involves measuring for knowledge spillovers, crowd-out of private investment, and rent-sharing between innovators and technology users. The private returns to R&D can differ substantially from overall benefits, often representing a tiny fraction of the overall benefits (Griliches, 1979; Nordhaus, 2004). Moreover, even accounting for only the private value of new technology is a challenging task, requiring myriad intermediate assumptions (see Azoulay et al., 2019, for a discussion). A further concern is whether case studies of particular technologies might "pick winners" and thereby overstate impacts (Jones and Summers, 2020).

In light of these challenges, we develop a different approach to estimate the effect of Embrapa on productivity at the *regional* level. Using regional data allows us to account for indirect effects while also sweeping out the major confounding forces and policy changes that operate at the national level. In what follows, we use this strategy to measure the effects of Embrapa on agricultural productivity as well as other agricultural outcomes. In conjunction with an economic model and detailed data on investment costs, we then use the same approach to estimate the aggregate returns to public R&D (Section 6).

5.1.1 Measuring Embrapa Exposure

Our empirical strategy is based on a measure of exposure to Embrapa's research that varies across both time and space. The time-series variation comes from the staggered introduction of Embrapa's research centers. The cross-sectional variation comes from the bilateral *ecological similarity* between Brazilian municipalities, which we treat as a shifter of the differential suitability of Embrapa's technologies across municipalities.

Conceptually, our approach is motivated by three observations. First, as shown in Section 4, agricultural research in Brazil is heavily focused on targeting local ecosystem characteristics. Second, narrative accounts of Embrapa's history put special emphasis on the development and diffusion of technologies specific to particular ecological regions of Brazil, such as agricultural liming techniques for the central Cerrado and soybean varieties bred to survive in tropical soils (Correa and Schmidt, 2014). Third, existing work has documented that ecological similarity between locations of invention and locations of use is a strong predictor for the diffusion and eventual productivity effects of agricultural technology (e.g., Griliches, 1957; Moscona and Sastry, 2025).

Our specific measure of ecological similarity, based on prior work by Bazzi et al. (2016) and Moscona and Sastry (2025), is an index that aggregates similarity in climate, topography, and soil characteristics. For each pair of municipalities i and j, we measure the sum of absolute deviations

$$\text{Ecological Similarity}_{ij} = -\sum_{x} |x_i - x_j|, \tag{5}$$

across the nine ecological characteristics that we collect x. We express all of these characteristics in z-score units, normalized by their mean and standard deviation across all municipalities. A higher value of Ecological Similarity$_{ij}$ means that i and j are more ecologically similar, implying that agricultural technology designed in or for municipality i is more likely to be suitable in municipality j.

Combining our time-series and cross-sectional variation, we measure each municipality's changing exposure to Embrapa's research as:

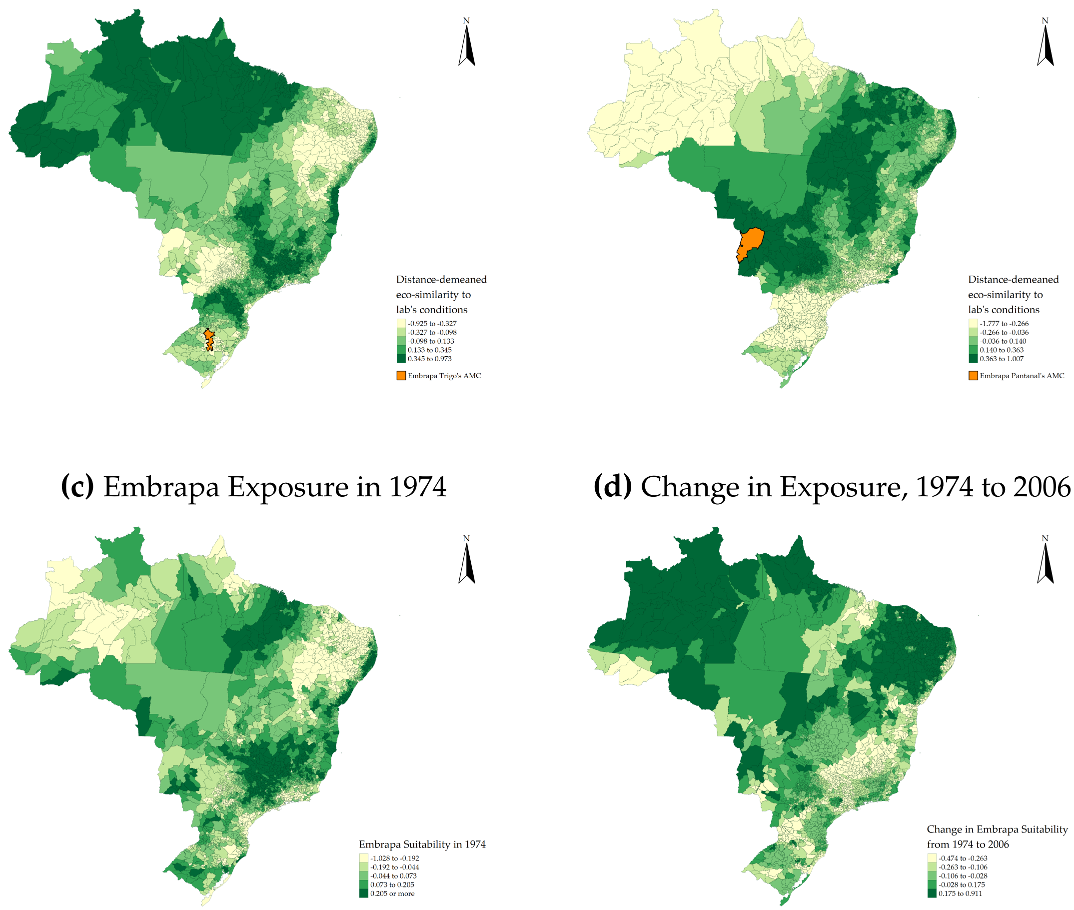
$$\text{Embrapa Exposure}_{it} = \max_{j \in \mathcal{R}_t} \text{ Ecological Similarity}_{ij} \tag{6}$$

where $\mathcal{R}_t$ is the set of centers that exist by time t. In words, this measure captures each municipality's ecological similarity to the most ecologically similar Embrapa center that exists as of time t. Cross-sectional variation in Embrapa Exposure$_{it}$ comes from the network of pairwise ecological similarity across municipalities, as captured by variation in

Figure 6: Embrapa Exposure Across Space and Time

(a) Similarity to Embrapa Trigo **(b)** Similarity to Embrapa Pantanal

(c) Embrapa Exposure in 1974 **(d)** Change in Exposure, 1974 to 2006



Notes: Panels (a) and (b) display ecological similarity, defined in equation (5), with respect to Embrapa Trigo (Wheat) and Pantanal, whose locations are colored in orange. Panel (c) shows Embrapa Exposure in 1974, defined in equation (6). Panel (d) shows the change in Embrapa Exposure from 1974 to 2006. In all panels, we linearly project out physical distance to the nearest Embrapa center.

Ecological Similarity$_{ij}$. Time-series variation comes from the fact that new centers are founded over time, shifting each municipality's most appropriate Embrapa center.

To illustrate the variation underlying this measure, Figure 6a plots each municipality's ecological similarity to the Embrapa center for Wheat and Figure 6b plots the same for the Embrapa Pantanal. The technology developed in these two research centers would have served very different parts of the country. The municipalities most ecologically similar to Embrapa Wheat are in the South East and Northern Brazil, while the municipalities most ecologically similar to Embrapa Pantanal are in central Brazil. More generally, as new centers opened over time, the parts of the country that were positioned to benefit from

Embrapa research shifted dramatically. Figure 6c plots Embrapa Exposure$_{it}$ as of the end of 1974, when only the first handful of research centers had been opened, while Figure 6d plots the change in Embrapa Exposure$_{it}$ over the course of the sample period.

5.1.2 Estimating Equation and Identification

Our baseline specification to estimate the effect of Embrapa on agricultural outcomes is:

$$y_{it} = \beta \cdot \text{Embrapa Exposure}_{it} + \chi_i + \chi_t + \gamma' X_{it} + \varepsilon_{it}, \tag{7}$$

where y_{it} is the outcome of interest, χ_i and χ_t are municipality and census wave fixed effects, and X_{it} is a vector of time-varying controls. The coefficient β captures the extent to which exposure to Embrapa's research affected agricultural outcomes. Standard errors are clustered by municipality in our baseline analysis, but the precision of our estimates is very similar using Conley (1999) standard errors to adjust for spatial correlation or clustering standard errors by state (see Appendix Table C.7).

Our main outcome variable y_{it} is agricultural productivity, measured as the logarithm of total agricultural production value divided by total farm area. We also estimate the effect of Embrapa exposure on a number of other alternative measures, including land values and crop yields, and decompose the effect on productivity into changes in input, technology, land use, and total factor productivity.

The central identification assumption is that the founding of new Embrapa centers was orthogonal to agricultural production trends in ecologically similar (compared to ecologically distant) municipalities. That is, we assume that when a new Embrapa center is opened, ecologically similar municipalities are on similar trends to ecologically distant ones. One potential concern is that ecological similarity to an Embrapa center is correlated with physical distance, both because precise research center locations were chosen under a range of constraints and because physical proximity to a research center could have benefits (e.g., extension services) beyond the impact of new innovation. To address this, we show that controlling flexibly for geographic distance to research centers, or fully excluding municipalities that are close to any center, have little impact on our estimates. If anything, the coefficient magnitude increases after making these restrictions. We also present pre-trend analysis and a host of falsification exercises to further support a causal interpretation of our findings.

Table 2: Embrapa Exposure Increases Agricultural Productivity

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Baseline Results					
Embrapa Exposure	0.730***	0.825***	0.985***	0.844***	0.819***	0.599***
	(0.080)	(0.084)	(0.120)	(0.084)	(0.084)	(0.086)
Observations	18386	18109	11821	18101	18101	18109
R^2	0.954	0.954	0.945	0.955	0.955	0.976
	Panel B: Weighted by 1970 Agricultural Area					
Embrapa Exposure	0.758**	1.266***	1.141***	1.277***	1.272***	0.824***
	(0.302)	(0.244)	(0.196)	(0.242)	(0.243)	(0.221)
Observations	18372	18101	11818	18101	18101	18101
R^2	0.964	0.965	0.961	0.965	0.966	0.980
Municipality FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
log(Distance to Embrapa) x Round FE	-	Y	Y	Y	Y	Y
Drop if $< 100km$ from Embrapa	-	-	Y	-	-	-
Drop if neighbor to Embrapa	-	-	Y	-	-	-
log(Initial Prod.) x Round FE	-	-	-	Y	Y	-
log(Initial Pop.) x Round FE	-	-	-	-	Y	-
State x Round FE	-	-	-	-	-	Y

Notes: The unit of observation is a municipality-census-round pair, where municipalities are harmonized to minimal consistent border units (IBGE, 2011). The regression model is equation (7). The outcome variable is the log of production value per farm area. In Panel B, estimates are weighted by each municipality's agricultural area in 1970. The control variables included are: distance (in km.) to the nearest Embrapa center times census-round fixed effects; log of production value per farm area in 1970 interacted with census-round fixed effects; log of population in 1970 interacted with census-round fixed effects; and state by census-round fixed effects. In column 3, we drop municipalities that are ever less than 100 km from an Embrapa center or neighbor a municipality with an Embrapa center. Standard errors are clustered at the municipality level.

5.2 Embrapa Increases Agricultural Productivity

Table 2 presents our main estimates of equation (7) using log of agricultural production value per area as the outcome. Panel A reports our baseline estimates and Panel B reports estimates in which each observation is weighted by municipality agricultural area in 1970. We find that β is positive and significant ($p < 0.01$): exposure to Embrapa increases agricultural productivity. The magnitude of the effect in Column 1 implies that an increase in Embrapa Exposure equal to one *cross-sectional* standard deviation increases agricultural productivity by 12%. Later, in Section 6, we revisit the calculation of *aggregate, time-series gains* by integrating our estimates with a more structured economic model.

We next show that the baseline result is not driven by geographic proximity to research centers. The estimate is quantitatively similar after controlling for (the logarithm of) distance to Embrapa centers interacted with census-round fixed effects, thus allowing for a flexibly time-varying effect of proximity (column 2). The result is also stable under a more conservative strategy of additionally restricting the sample to municipalities that neither border an Embrapa research center nor are within 100 km of one (column 3). Our main estimates are therefore not merely picking up the effects of proximity to a research center, which could proxy for extension services, access to information about Embrapa, or other stimulative effects of Embrapa on the local economy. Instead, the estimated effect on geographically distant but ecologically similar locations is consistent with our interpretation that the results capture variation in the ecological suitability of new technology.

The results are also similar controlling flexibly for pre-existing trends related to initial agricultural productivity (column 4) and population (column 5). These estimates suggest that the main result is not biased by mean reversion or other trends in local economic outcomes. Finally, to further zoom in on the precise geographic variation spanned by our exposure variable, we find that effect of Embrapa is also virtually unchanged after including state-by-year fixed effects (column 6), which fully absorb any different trends in state-level research investments (e.g., at state universities) or agricultural support mechanisms.

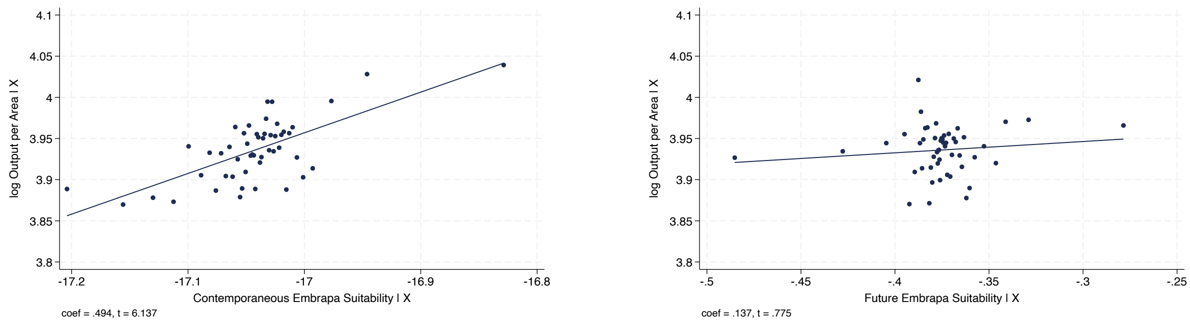
In additional results, we find that Embrapa Exposure has a positive effect on other measures of agricultural productivity (Appendix Table C.4). The first is the value of crop output per area, which we compute by summing the output of each major crop listed in the census weighted by its national price in 1970 (Panel A). Compared to our baseline, this measure focuses entirely on crop (rather than livestock) output and, by construction, does not take into account local variation in output prices. Potentially for both reasons, we find larger coefficient estimates compared to our baseline. The second productivity measure is local farm value per acre, which captures the effect of Embrapa on the net present value of future agricultural profits under a hedonic interpretation (Panel B). These findings suggest that new technology development was capitalized into local land values.

The third and fourth are measures of local agricultural TFP, measured as total production value relative to input use. We use a four-factor production function (land, labor, capital, and intermediates) calibrated to the estimates of Fuglie (2015) for Brazil.[10] Because Fuglie's (2015) estimates suggest a significant decline in the importance of labor and land and increase in the importance of intermediates of capital, we use two different calibrations: one designed to match Brazilian agriculture in the 1970s (Panel C) and another designed to match Brazilian agriculture in the 2010s (Panel D). The effects on both measures are positive, but smaller than the effects on output over land value. That is, intensification

[10]We define TFP $= Y/(N^{\alpha_N} L^{\alpha_L} M^{\alpha_M} K^{\alpha_K})$, where we measure output Y as the value of agricultural production, labor N as the total number of workers, intermediates M as the sum of expenditure on fertilizers, seeds, and chemical defenses (e.g., insecticides, herbicides, and fungicides), and capital K as the number of tractors. The "1970s" calibration sets $\alpha_N = 0.342, \alpha_L = 0.342, \alpha_K = 0.167$, and $\alpha_M = 0.057$, and the "2010s" calibration sets $\alpha_N = 0.083, \alpha_L = 0.373, \alpha_K = 0.214$, and $\alpha_M = 0.331$.

Figure 7: Embrapa Affects Productivity Contemporaneously, with no Anticipation

(a) Contemporaneous Exposure **(b)** Future Exposure



Notes: The regression model is equation (7), augmented with the leading value Embrapa Exposure$_{i,t+1}$, and controlling for distance to the nearest Embrapa center times round fixed effects and state times round fixed effects. The left and right panel respectively show the binned scatterplot of the outcome, log of production value per farm area, against the contemporaneous value Embrapa Exposure$_{i,t}$ and the lead value Embrapa Exposure$_{i,t+1}$. In each case, the binned scatterplot partials out other variables and the included fixed effects. The printed t-statistics are based on standard errors clustered by municipality.

of variable inputs explains *some* but not *all* of the increase in output per land area, leaving a sizable remainder to TFP growth. We further explore this breakdown in Section 5.3.1.

We finally observe that our findings are not driven by any single component of the ecological similarity index (equation (5)). Appendix Figure C.3 reports estimates after dropping each component from the measure. The results are very similar across all components and outcome variables. This suggests that our findings are not unduly sensitive to the precise method for measuring ecological similarity in our treatment variable.

5.2.1 Dynamics

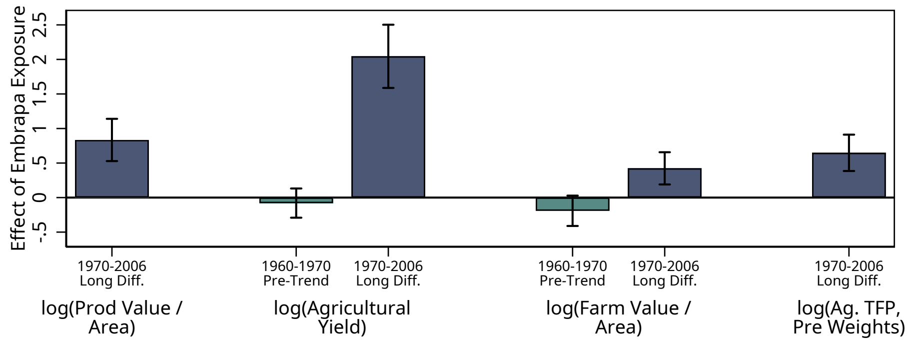
So far, we have studied the effect of exposure to Embrapa on productivity in same decade or five-year period (i.e., within census round). We next investigate dynamics.

First, we show that there is no relationship between changes in Embrapa exposure and *pre-existing* changes in productivity since the prior census round. To do this, we estimate an augmented version of equation (7) that also includes the leading value of Embrapa Exposure$_{it}$ as a regressor. The coefficient on the contemporaneous value remains positive and significant ($p < 0.01$) while the coefficient on the leading value is small in magnitude and indistinguishable from zero (Figure 7). Thus, the main results do not seem to be driven by pre-existing trends, which are flat and insignificant.

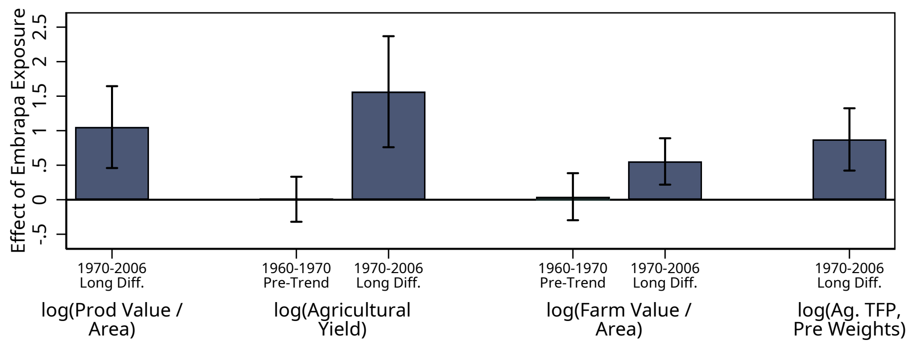
Next, to study the long-run effects of Embrapa, we estimate long-difference regressions that capture how changes in exposure to Embrapa affected changes in productivity

Figure 8: Long Difference Estimates and Pre-Existing Trends

(a) Unweighted



(b) Weighted by Agricultural Area



Notes: The regression model is equation (8), and we control for the logarithm of distance to the nearest Embrapa center and state fixed effects. The top panel reports unweighted estimates, and the bottom panel reports estimates weighted by farmland in 1970. The outcome variables are listed at the bottom of each panel: log of total production value per area, log of crop yields, log of total farm value per area, and log of agricultural TFP, based on weights corresponding to Brazilian agriculture in the 1960s (see Fuglie, 2015). For blue bars, the outcome is the difference in each variable between 1970 and 2006 and for green bars, the outcome is the difference in each variable between 1960 and 1970. 95% confidence intervals are reported.

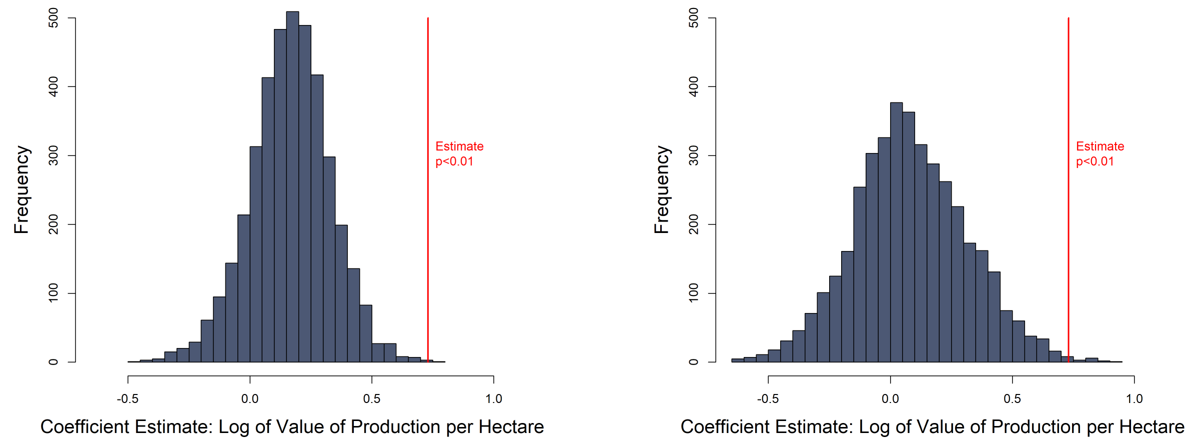
over the full sample period. Accounting for these longer-run effects could be important to the extent that research investment had dynamic knowledge spillovers or new technology took time to diffuse and generate returns. The estimating equation is:

$$\Delta y_i = \beta \cdot \Delta \text{Embrapa Exposure}_i + \gamma' X_i + \epsilon_i \tag{8}$$

where ΔEmbrapa Exposure$_i$ is the change in Embrapa Exposure$_{it}$ from 1970 to present and X_i includes geographic distance to the nearest Embrapa center and state fixed effects (i.e., the equivalent specification to column 6 from Table 2).

Figure 8 (blue bars) presents long-difference estimates for our four main outcome variables, using both unweighted (Panel A) and farmland area weighted (Panel B) regression specifications. These long-difference effects are 30-40% larger than our baseline estimates,

Figure 9: Falsification Tests



(a) Randomizing Across Geography and Time

(b) Randomizing Across Time Only

Notes: Each sub-figure reports placebo draws from separate falsification exercises. In Panel (a), we simulate 4,000 alternative Embrapa expansion patterns, randomizing both the locations of the 40 centers and the opening year. We restrict the set of municipalities with a research center to those above 90% of the minimum municipality population across all actual centers in the opening year, and we use the empirical distribution of years with a center opening. In Panel (b) we randomize only the timing of center openings. For each counterfactual, we repeat estimate equation (7) and report a histogram of the resulting coefficient estimates. The coefficient estimate using Embrapa's actual expansion pattern is displayed with a vertical red line.

suggesting that some of the effect takes longer than a decade to materialize.

Finally, we compare these long-difference estimates to changes in productivity prior to the expansion of Embrapa as a further test for pre-existing local productivity trends. Due to differences in data collection during census rounds before 1970, this is only possible for crop yields and agricultural land values. We find no evidence that Embrapa exposure is positively correlated with changes in these outcomes prior to the establishment of Embrapa (Figure 8, green bars). Like our in-sample pre-trend analysis, these findings suggest that exposure to Embrapa was not related to pre-existing productivity dynamics.

5.2.2 Falsification Tests

To further support a causal interpretation of the findings, we conduct a falsification test in which we randomize the geography and timing of the expansion of Embrapa and investigate whether exposure to these counterfactual expansion patterns has a similar effect on changes in productivity. If, for example, there were pre-existing positive trends in parts of Brazil that are ecologically remote, then even placebo expansion patterns for Embrapa would spuriously correlate with these pre-existing trends, and there would be no additional effect of true exposure to Embrapa. If, on the other hand, our main results capture the causal effect of exposure to Embrapa, then our estimates should be in the far right tail of the counterfactual coefficient distribution. Reassuringly, whether we construct placebo Embrapa expansion paths by randomizing both the location and timing of Embrapa center openings (Figure 9a) or by randomizing *just* the timing, thereby exploiting only the temporal component of our identification strategy (Figure 9b), our main estimate is in the far right tail of the placebo coefficient distribution. These results rule out the possibility that our baseline findings are driven by spurious trends, which would likely lead to similar estimates regardless of the exact timing of Embrapa's center opening.[11]

5.2.3 Heterogeneous Effects and Inequality

Our main analysis has focused on the effect on overall agricultural productivity in municipalities. However, in principle, Embrapa may have also affected inequality both across and within regions. Across regions, we test for heterogeneous effects of Embrapa Exposure on agricultural productivity in regions that differ in their baseline productivity and farm size. We find no statistically significant differences and, if anything, slightly larger effects for less-productive regions (Appendix Figure C.4). Within regions, we test the effect of Embrapa exposure on average size and the Gini coefficient of the farm size distribution. We find a null effect on average farm size and a negative effect on farm size inequality (Appendix Table C.6). Together, these estimates suggest that the expansion of Embrapa led to an increase in productivity without a corresponding increase in inequality.

5.3 Mechanisms

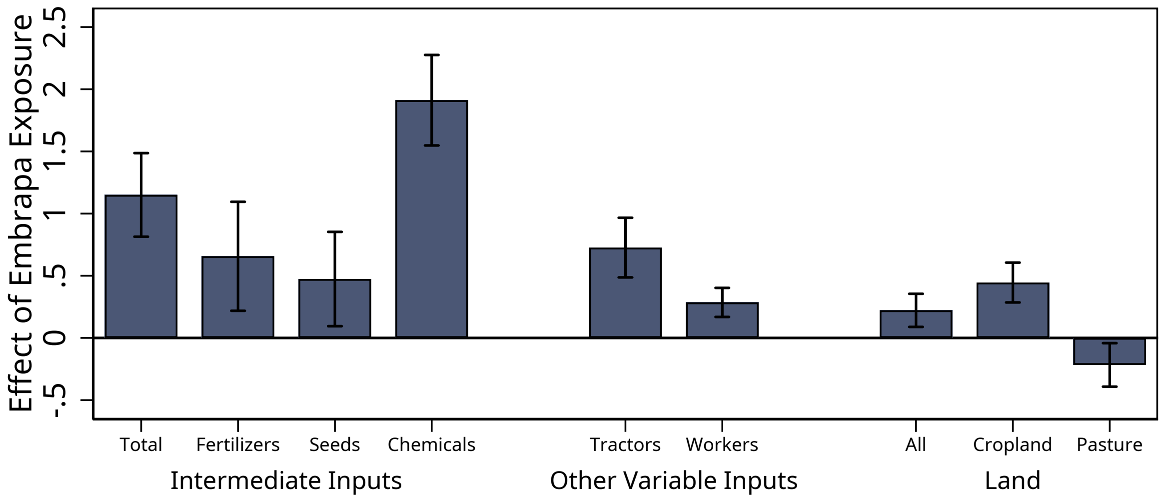
Having documented that exposure to Embrapa had a large positive effect on agricultural productivity, we next investigate the mechanisms underlying this main result. First, we study the effect of Embrapa on technology, input, and land use. Second, using annual data on crop-level output in each municipality, we show that the effects of Embrapa on productivity are concentrated only in the crops that were an explicit focus of Embrapa innovation. This is further evidence that innovation is the key driving mechanism.

5.3.1 Input Intensification and Land Conversion

Figure 10 reports estimates of equation (7) that describe the impact of Embrapa on input and land use. First, we show that Embrapa had a large positive effect on intermediate input and technology use, including fertilizers, seed, and chemicals (column 1). These

[11] In Appendix Table C.5, we control for the expected value of the treatment variable from simulations that vary the timing of center openings, following the logic of Borusyak and Hull (2023). The estimates are quantitatively similar to those in the baseline table (Table 2).

Figure 10: Embrapa Exposure and Agricultural Inputs



Notes: The unit of observation is a municipality-census-round pair, where municipalities are harmonized to minimal consistent border units (IBGE, 2011). The regression model is equation (7) and, in each specification, we control for the logarithm of distance (in km.) to the nearest Embrapa center times census-round fixed effects and state by census-round fixed effects. The outcome variables, denoted below each bar, are (all in logarithms): the total expenditure on fertilizers, seeds, and chemicals; fertilizers; seeds; chemical defenses (e.g., insecticides, herbicides, and fungicides); number of tractors; number of workers; all agricultural land; agricultural land for crops; and agricultural land for pasture. Standard errors are clustered at the municipality level and error bars are 95% confidence intervals.

are all areas that were the focus of Embrapa's research; thus, this finding indicates that research led to meaningful changes in farmers' technology and input use. The positive effect on total intermediate input use is driven by independent positive effects on fertilizer use (column 2), seed use (column 3), and especially chemical use (column 4).

Second, we investigate the effect on other variable inputs. Embrapa exposure is associated with increased mechanical input use (column 5) and also increased labor use (column 6). However, the effect on labor is substantially smaller than the effect on other inputs and less than one third the magnitude of the effect on intermediate inputs, implying that agricultural production became less labor intensive.

Third, we find that Embrapa led to an expansion of cropland (column 7), driven by an increase in land devoted to crop production (column 8), over half of which came from land that was converted from pastureland (column 9). The positive effect on cropland is larger in absolute value than the negative effect on pasture land, driven by the fact that technology development also opened new land for agricultural production. The negative effect on pasture land, likely due to Embrapa's research focus on crop technology rather than livestock, also recalls the earlier finding that Embrapa exposure increased crop-specific production in excess of overall agricultural output (Appendix Table C.4, Panel A). A further implication is that our strategy of focusing on total agricultural output, which we will also pursue in our counterfactual analysis in Section 6, properly accounts for this reallocation from pasture to cropland, whereas a strategy that measured gains only for crop agriculture might over-state the net economic effects of Embrapa.

Finally, it is worth noting that the effect on overall productivity is substantially larger than the average effect on input use. This underlies our earlier finding that exposure to Embrapa had a positive effect on measured TFP, computed as production value relative to an aggregate of variable and non-variable input use (Appendix Table C.4, Panels C-D).

5.3.2 Directed Innovation Across Crops

If innovation is the mechanism driving our main findings, then Embrapa exposure should have the largest positive effect on the productivity of crops that were the focus of Embrapa's technology development and little or no effect on crops that were not. We investigate this prediction, exploiting the fact that Embrapa focused its attention on a specific set of crops that it saw as the priority for staving off food shortages—in particular, beans, cassava, maize, rice, soy, and wheat (Martha Jr et al., 2012). Consistent with these historical accounts, we showed direct evidence earlier that Embrapa-affiliated researchers focused disproportionately on these crops in their publication output (see Section 4.2).

We compile data on the output of each crop in each municipality and year from the Municipal Agricultural Production (PAM) survey. We then investigate the effect of Embrapa exposure on crop-specific output, separately for crops that were the focus of Embrapa's innovation and for crops that were not. In particular, we estimate:

$$y_{ikt} = \beta_1 \cdot \text{Embrapa Exposure}_{it} \cdot \text{EC}_k + \beta_2 \cdot \text{Embrapa Exposure}_{it} \cdot \text{NEC}_k + \chi_{ik} + \chi_{tk} + \varepsilon_{ikt} \quad (9)$$

where k indexes crops, y_{ikt} is the (log of) either output or yield for crop k in municipality i and year t, EC_k is an indicator that equals one if crop k was one of Embrapa's focus crops, and NEC_k is an indicator that equals one if crop k was not one of Embrapa's focus crops. χ_{ik} and χ_{tk} are two-way fixed effects at the municipality-year and crop-year, respectively. If Embrapa's technology development drives the positive effect of Embrapa exposure on productivity, we would expect that $\beta_1 > 0$ and that $\beta_1 > \beta_2$.

We find that β_1 is large and highly significant, while β_2 is close to zero and statistically insignificant (Table 3). The estimates are qualitatively similar when crop-specific output is measured in terms of total output (columns 1-4) or yield (columns 5-8). The coefficient estimates are larger for total output, suggesting that land devoted to Embrapa's focus crops also increased but not by enough to offset the positive effect on average productiv-

Table 3: Embrapa Increases Output and Yield for Innovation-Focus Crops

	Outcome Variable is:							
	log Crop-Specific Output				*log Crop-Specific Yield*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Embrapa Exposure × EC	1.426***	1.410***	1.276***	1.279***	0.140**	0.368***	0.121*	0.356***
	(0.235)	(0.348)	(0.226)	(0.340)	(0.066)	(0.109)	(0.068)	(0.110)
Embrapa Exposure × NEC	0.059		0.059		-0.102		-0.102	
	(0.283)		(0.283)		(0.095)		(0.095)	
Observations	188619	188455	180797	180620	188619	188455	180797	180620
R^2	0.832	0.876	0.829	0.875	0.937	0.956	0.937	0.957
Crop-Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Municipality-Crop FE	Y	Y	Y	Y	Y	Y	Y	Y
Municipality-Year FE	-	Y	-	Y	-	Y	-	Y
Drop Soybean	-	-	Y	Y	-	-	Y	Y

Notes: Each column reports an estimate of equation (9) in which the unit of observation is a municipality-crop-year triplet. Crop-by-year and municipality-by-crop fixed effects are included in all specifications. Even numbered columns also include municipality-by-year fixed effect, thus fully absorbing Exposure × Non-Embrapa Crop. Columns 3-4 and 7-8 drop soy from the sample. The outcome variable is the logarithm of crop-specific output, measured either in physical units (columns 1-4) or by production value (columns 5-8). Standard errors are clustered by municipality and reported in parentheses.

ity. The estimates are similar if soybeans are fully excluded from the analysis (columns 3-4 and 7-8), suggesting that the findings are not driven exclusively by the crop that would become Brazil's top export commodity. Finally, the differential effect of Embrapa exposure on crops that were the explicit focus of Embrapa's innovation remains if we further include municipality-by-year fixed effects in the regression, thereby fully absorbing any municipality-level trends (columns 2, 4, 6, and 8). Together, these findings build further confidence that the effect of Embrapa exposure on agricultural productivity is not capturing any spurious municipality-level trend and is driven by directed technological progress.

6 The Returns to Public R&D in Agriculture

This section examines the returns to public R&D investment associated with Embrapa. To this end, we develop and estimate the minimal theoretical framework necessary to assess the macroeconomic implications of Embrapa, while maintaining consistency with the reduced-form results presented in Section 5. Our baseline estimates suggest that Embrapa increased Brazilian agricultural productivity by 110%, implying a benefit-cost ratio of 17. The majority of these returns come from the geographic structure of Embrapa and its spread across ecological conditions, rather than the overall scale of investment. We focus on the essential components of the estimation, leaving details to Appendix B.

6.1 Model and Estimation

Set-Up. We begin by describing a model of regional agricultural productivity and its relationship with agricultural research. Let $i \in \mathcal{I}$ index the regions of Brazil. Each pair of regions $i, j \in \mathcal{I}$ has a primitive ecological dissimilarity $g_{ij} > 0$. A subset of the regions, $\mathcal{R}_t \subseteq \mathcal{I}$ has an Embrapa research center at time t, and the center in each region $j \in \mathcal{R}_t$ employs N_{jt} scientists at time t. The agricultural productivity of a region i in period t is given by

$$A_{it} = \bar{A}_{it} \left(\sum_{j \in \mathcal{R}_t} [\exp(-\beta g_{ij})(N_{jt})^{\gamma}]^{\theta} \right)^{\frac{1}{\theta}} \tag{10}$$

where β measures how the effectiveness of research output scales with ecological similarity, γ is the elasticity of research output to labor inputs, θ determines the elasticity of substitution between research centers, and $\bar{A}_{it}$ is an exogenous shifter capturing all other determinants of productivity, such as land quality and access to non-Embrapa research.

We can transform equation (10) into the following estimating equation, which is a generalization of the main empirical specification from Section 5:

$$\log A_{it} = -\beta g_{i\bar{\jmath}} + \gamma \log N_{\bar{\jmath}} + \frac{1}{\theta} \log \left(\sum_{j \in \mathcal{R}_t} \left[\frac{\exp(-\beta g_{ij})(N_{jt})^{\gamma}}{\exp(-\beta g_{i\bar{\jmath}})(N_{\bar{\jmath}})^{\gamma}} \right]^{\theta} \right) + \log \bar{A}_{it} \tag{11}$$

where $\bar{\jmath}$ indexes the municipality with the closest Embrapa center from i in period t.

This model for local agricultural productivity incorporates three key economic effects of R&D, each of which is summarized by a different parameter.

First, the parameter β governs how quickly the effect of research output on productivity decays with ecological dissimilarity. This captures the inappropriateness or ecological mismatch of agricultural technology. Unlike the other two parameters, β is not standard in existing models of innovation and productivity, but it was the key focus of Embrapa and of our analysis in Section 5. The first term of equation (11) captures the effect of ecological similarity of the most ecologically proximate Embrapa center.

Second, the parameter γ governs how agricultural research output relates to the number of researchers: an x% increase in researchers in all locations raises productivity everywhere by $\gamma \times x$%. This captures "scale effects" in R&D (e.g., Jones, 1995), which could be important for determining the overall impact of Embrapa's investment. Estimating γ will also make it possible to determine the extent to which the impact of Embrapa was due to greater overall R&D investment versus the specific geographic structure that spread R&D investments across space. The second term of equation (11) captures the effect of the scale of the most ecologically proximate research center.

Third, the parameter θ governs the degree of substitutability across research from different Embrapa centers. While our reduced form analysis only took into account the effect of the most ecologically proximate research center, in practice there could be imperfect substitutability between the agricultural products or techniques that different centers develop. The third term of equation (11) captures the effect of all other Embrapa activities taking place outside the most ecologically proximate center on local productivity.

Our estimating equation from Section 5—equation (7)—is obtained in the limit where $\gamma = 0$ and $\theta \to \infty$. This case shuts down scale effects, eliminating the effect of the nearest center's scale, and makes centers' research outputs perfect substitutes. In this case, our earlier estimates recover the parameter β.

Estimation. We estimate the model's three parameters (β, γ and θ) via nonlinear least squares. Mapping from theory to the data, we measure A_{it} as the value of agricultural production per hectare, N_{it} as expenditure on labor in each Embrapa center, and g_{ij} as the ecological dissimilarity between each pair of municipalities (see Section 5.1). For estimation, we transform Equation (11) into a long difference between a reference year $t_1 = 2006$, the end of a large period of expansion from Embrapa, and a pre-Embrapa period. This mirrors our earlier long difference analysis in Section 5.2. Appendix B presents details of our estimation procedure and robustness exercises.

Table 4 reports estimates of equation (11) using different methods. Column 1 presents the OLS estimate of β, assuming $\gamma = 1/\theta = 0$, consistent with our strategy in Section 5. Column 2 shows the results from our non-linear least squares estimation. The coefficient on β slightly *increases*, confirming the importance of ecological mismatch in mediating the benefits from research even in this richer framework. Our estimate of $\gamma = 0.090$ implies mild scale effects: a 1 percent increase in the total number of Embrapa researchers raises agricultural productivity by 0.07 percent. To better assess the magnitude of our estimate for γ, column 3 reports results from a constrained estimation where γ is set high enough for Embrapa's research alone to account for the entirety of Brazil's agricultural productivity growth between 1970 and 2010, estimated at 280% (Fuglie, 2015).[12] This yields an estimate of $\gamma = 0.14$; nonetheless, even with scale effects at this natural upper bound, there remains a large and quantitatively stable effect of ecological mismatch.

[12] As a reference point, Jones (2002) calibrates a scale effect of $\gamma = 1/3$ in the US to reconcile aggregate TFP growth with the observed growth of the R&D workforce.

Table 4: Estimates of the Agricultural Productivity Function

	Model Specification		
	(1)	(2)	(3)
Parameter	OLS	NLLS	NLLS
β	0.820	0.934	0.941
	$[0.611; 1.030]$	$[0.691; 1.177]$	$[0.686; 1.196]$
γ		0.090	0.143
		$[-0.014; 0.056]$	$[-]$
θ		7.494	6.212
		$[3.210; 11.778]$	$[2.979; 9.444]$
p-value (H_0: equal to Col. (1))		0.000	0.000
p-value (H_0: equal to Col. (2))			0.107

Notes: This table shows estimates of equation (11). 95% confidence intervals are reported in square brackets. P-values are generated using the log-likelihood ratio tests. Column 3 restricts γ to be large enough so that Embrapa's research account for all the productivity growth in Brazil between 1975 and 2010.

6.2 Embrapa's Productivity Effects and the Returns to Public R&D

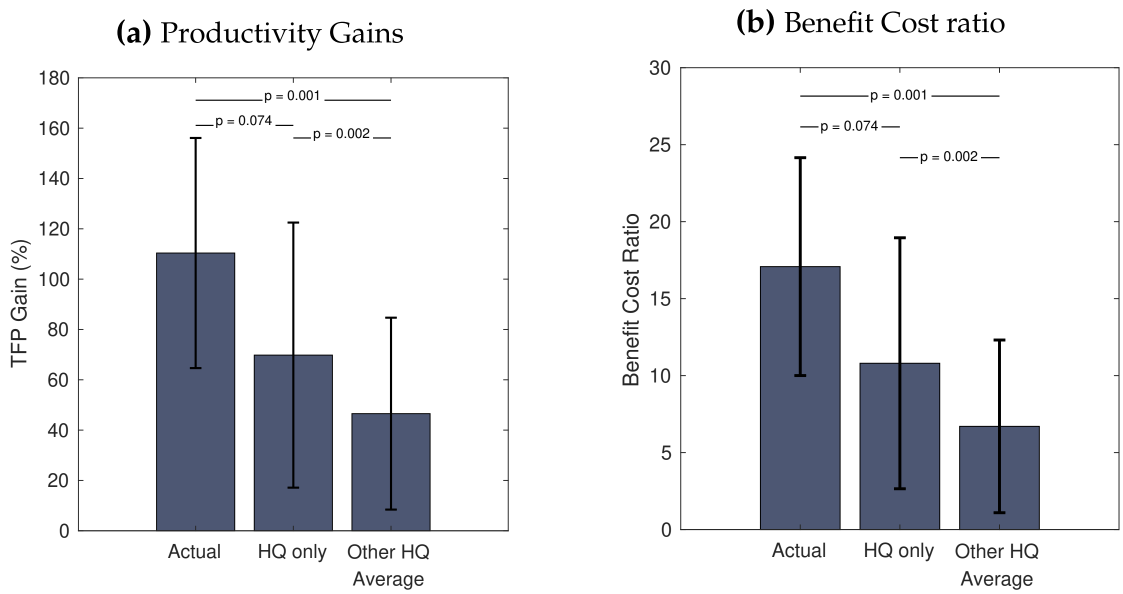
We first use the model to evaluate the aggregate agricultural productivity gains from Embrapa. Specifically, we compare productivity in 2006 to a counterfactual in which public agricultural R&D was held fixed at pre-Embrapa levels.[13] We find that Embrapa induced a 110% gain in average productivity (Figure 11a, first bar). To put this number in perspective, Fuglie (2015) estimates that Brazil's aggregate agricultural productivity rose by 280% between 1970 and 2010, implying that Embrapa accounts for 39 percent of the total productivity gains over this period.

A notable feature of our setting is we can benchmark our estimates of the value of public R&D against data on its total cost. That is, we can compute a social return on public R&D investment, an object on which there is relatively scant evidence in advanced economies (e.g., Jones and Summers, 2020) and even scanter evidence in low- and middle-income countries. Understanding the cost-effectiveness of R&D investments is essential to determine whether policies like Embrapa are legitimate strategies to foster growth.

To do this, we construct annual series of costs and benefits, which we discount to their present value as of 2006. To compute benefits, we convert the productivity gains attributed to Embrapa into value-added gains in 2006. For years prior to 2006, we assume a phased-in

[13]We discipline this counterfactual using historical data on the structure of agricultural research under the pre-existing DNPEA (Departamento Nacional de Pesquisa e Experimentação Agropecuária). These research centers were in Belém (Pará), Cruz das Almas (Bahia), Sete Lagoas (Minas Gerais), Pelotas (Rio Grande do Sul), and Manaus (Amazonas). We rescale our measurements for the size of these research centers, taken between 1971 and 1973, to match the initial scale of Embrapa.

Figure 11: Productivity Gains and Benefit-Cost Analysis of Embrapa



Notes: Panel (a) effect the impact of Embrapa's research on aggregate agricultural productivity. Error bars are 95 percent confidence intervals computed based on the delta method. Panel (b) shows the benefit-cost ratio implied by the productivity gains presented in Panel (a) and data on the costs of research. P-values are based on the null hypothesis of no difference between columns.

benefit structure: a linear increase in gains from 1975 until 2000 and a constant annual gain thereafter, equal to the value-added gain in 2006. All pre-2006 benefits are discounted to present value using a 7 percent social discount rate and post-2006 benefits are valued using a perpetuity formula using a 5 percent rate.[14] To compute the costs, we first discount all expenditures prior to 2006 to their present value using the same 7 percent interest rate. We then assume that maintaining Embrapa's research at its 2006 level is necessary to sustain the gains and apply the perpetuity formula to calculate the present value of all future costs using a 5 percent discount rate. Finally, we compute the ratio of the present value of benefits to the present value of costs.

Our baseline estimate for the benefit-to-cost ratio of Embrapa is 17 (Figure 11b, bar 1).[15] This is somewhat larger than prior estimates for advanced economies: for example, Jones and Summers (2020) report a "conservative estimate" of 5 for overall public R&D in the US economy, though they note that there are many reasons that the true number could be much larger. Griliches (1958) computes a benefit-cost ratio of 7 for research on hybrid corn. One possibility is that returns to well-structured R&D in developing countries are

[14]These assumptions are consistent with Embrapa's own reports, which use discount rates in the 4–7.7 percent range (Embrapa, 2018, 2020), as well as World Bank estimates for Latin America (Lopez, 2008).
[15]These estimates also imply an internal rate of return of 25 (see Appendix B.4.)

particularly high because of the relative absence of locally appropriate technology and the high returns to developing it. Indeed, a handful of studies estimating the returns to R&D in tropical agriculture obtain figures that are even larger than ours (Rosegrant et al., 2023).

There are several reasons that our estimates should be interpreted with caution; however, we believe that, if anything, our approach would lead us to understate the true returns. First, we impose a slow benefit ramp-up, which is conservative relative to other work arguing that benefits peak ten years after investment (Rosegrant et al., 2023). Second, we make the conservative assumption that existing research spending needs to be held constant in order to maintain existing gains. Third, while we measure all of Embrapa's *costs*, there are likely several components of Embrapa's *benefits* that are not captured by our empirical strategy, including the effect of new technology on weather resilience (i.e., lower production variance), food security, and agricultural productivity outside of Brazil (see Lachaud and Bravo-Ureta, 2022, on international productivity spillovers from agricultural R&D). Moreover, even our lower-bound benefit-cost ratio of ten would be considered high among development interventions (see, e.g., Copenhagen Consensus Center, 2004).[16]

6.3 Mechanisms: Research Scale vs. Geographic Scope

We next investigate the mechanisms underlying Embrapa's returns. Embrapa increased both the scale of agricultural R&D, by substantially increasing overall investment, and its geographic scope, by establishing research centers in many different regions and ecological zones. How much of the overall impact on agricultural productivity was due to greater overall R&D investment versus the re-direction of R&D toward the development of technology suited to Brazil's varied ecological conditions?

To separately account for the roles of scale and scope, we simulate the effect of Embrapa under different institutional structures. First, we consider a scenario in which Embrapa opened only its headquarters in Brasília and invested all of its resources there. This holds the scale of investment constant but reduces its geographic scope. This scenario reduces the productivity gains from Embrapa to 70% and the benefit-cost ratio to 11, 37% lower than our estimate under Embrapa's true scope (Figures 11a and 11b, second bar). Moreover, the lower-bound benefit-cost ratio of 2.7 is no longer large and similar to existing estimates of the multiplier effect from simple cash transfers (Egger et al., 2022). Thus, there were large and cost-effective gains to spreading investment across many regions.

Brasília may be a particularly favorable place to scale up agricultural R&D because of its ecological "centrality," leading us to over-state the returns to scaling up research more

[16] In Appendix Figure C.6, we report alternative benefit-cost calculations under other plausible assumptions for the ramp-up period of benefits, the maintenance cost after 2006, and the cost of capital.

generally. To address this, we instead consider a scenario in which Embrapa invested all of its resources in a single hub in the location of one of its other existing research centers. The average of both the benefits and benefit-cost ratio across these scenarios are considerably lower than those for the actual design and for the Brasília-only counterfactual (Figures 11a and 11b, third bar). These findings are consistent with large returns to Embrapa's geographic scope and limitations to scaling up agricultural R&D in any one location.

6.4 Implications for Policy Design

We conclude by summarizing two lessons for designing R&D programs that emerge from our analysis.

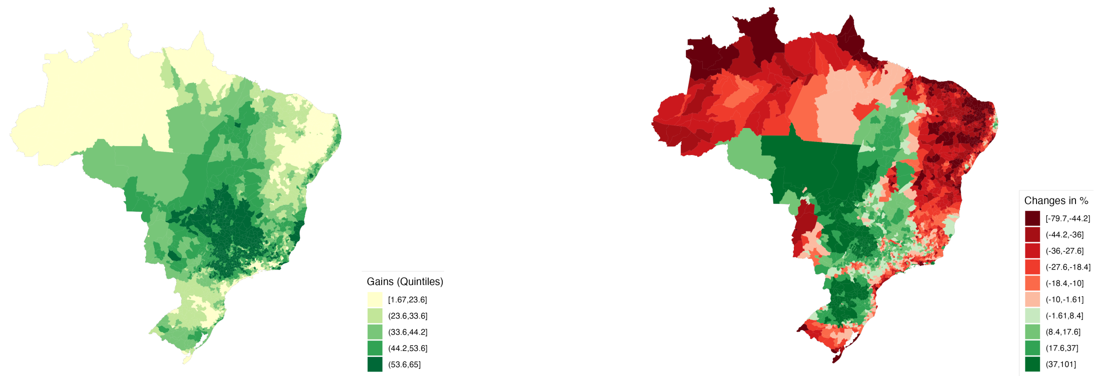
The Benefits of Spreading Out. Our counterfactual analysis suggests that much of the economic gains from public agricultural R&D in Brazil arose from its diffuse geographic structure, spread across the regions and biomes of the country. This broad conclusion contrasts with that of a separate literature that has studied the concentration of innovative activities in "high-tech clusters" of the United States. For example, Moretti (2021) documents that local concentration of inventors in computer science, semiconductors, and biology increases the productivity of marginal inventors, and suggests that further agglomeration would increase total innovative output (see also Gruber and Johnson, 2019).

One key point of our study is that we document that the productivity effects of innovation is shaped by mismatch between the location for which technology is developed and the location in which it is applied—in the language of the structural model, $\beta > 0$. This force is present in previous work on agricultural technology (e.g., Griliches, 1957; Moscona and Sastry, 2025). Other work suggests that similar forces are at play for medical technology that addresses different diseases (Kremer and Glennerster, 2004; Hotez, 2021) and high-tech start-ups that cater to different markets (Lerner et al., 2024). A conceptual take-away from our analysis is that gauging the extent of "mismatch effects" may be crucial not only for assessments of whether R&D investments are worthwhile but also for determining their ideal structure and design. We also show that Embrapa was able to overcome the productivity disadvantage of conducting research outside of R&D "clusters" (see Table 1)—understanding how this was accomplished in greater detail could be an important input into the implementation of future R&D programs.

Heterogeneous Returns and Targeting. A related conclusion is that the returns to R&D investments vary considerably across space. To illustrate this in our setting, Figure 12a displays the productivity gains from constructing a single large research center each municipality. Brasília—Embrapa's true headquarters in Brazil's central region—is located in

Figure 12: Geographic Distribution of the Productivity Gains of an Embrapa Center

(a) Initial Aggregate Productivity Gain **(b)** Change in Aggregate Productivity Gain



Notes: Panel (a) shows the gains in aggregate agricultural productivity of opening one large Embrapa center across all municipalities compared to a baseline with no centers. Panel (b) shows the difference in the relative gains from opening one additional center when no centers exist versus all centers from 2006 exist. Reg regions would generate a lower aggregate productivity gain relative to the baseline scenario and green regions would generate a higher aggregate productivity gain relative to the baseline scenario.

the region that would maximize the gains from a single, large center (darkest green in the map). This is driven by the fact that this region is ecologically close to many areas of the country, and our analysis reveals that ecological mismatch mediates the effects of R&D investments on productivity. Placing a single, large center in other parts of the country, by contrast, could yield returns as low as 1-2 percent.

Moreover, since research centers are substitutes in our model ($\theta > 1$), the best places to target change over time. Figure 12b maps the change in aggregate productivity gains between 1970 (when no centers exist) and 2006 from creating a single new center in each municipality. As Embrapa spread to more remote parts of the country, there was a clear decline over time in the gains from establishing a new center in municipalities far from the center of Brazil (red and dark red in the map). Once a center is established in a particular ecological zone, the additional returns to a center in an ecologically similar area becomes much lower. These changing returns over time can be used to target new investments. Indeed, one of the last new Embrapa centers was established in 2012 in Sinop, a municipality in Western Brazil where returns to new research are highest (darkest green on the map).[17]

[17] An additional implication is that there has been a decline in over time potential productivity gains from adding additional centers (see Appendix Figure C.7). Before Embrapa existed, adding a new center could raise aggregate productivity by up to 60 percent; in contrast, a new center of the same size could only increase aggregate productivity by 10 today. This could explain Embrapa's greater focus in recent years on alternative goals such as climate-damage mitigation and production resilience (Embrapa, 2025a).

7 Conclusion

Global R&D investment is concentrated in a handful of high-income countries. Existing work documents that frontier R&D is developed to match the specific needs and demands of these high-income countries, limiting the productivity benefits of this technology elsewhere (Kremer and Glennerster, 2004; Moscona and Sastry, 2025). Can targeted public R&D in a developing country allow it to escape this technology mismatch trap?

To answer this question, we study Brazil's Embrapa (the Empresa Brasileira de Pesquisa Agropecuária), perhaps the most prominent example of a publicly-funded R&D program in a developing country, which was established in 1973 to spur the development of locally-suitable agricultural science and technology. Combining detailed data on Embrapa's structure and research costs with a novel dataset of the research and career trajectories of all Brazilian agricultural scientists and nine rounds of Brazil's census of agriculture, we investigate the impact of public R&D on research output and agricultural productivity growth.

We have three main sets of findings. First, using granular data on Brazilian agricultural research, we find that Embrapa shifted the focus of agricultural research toward Brazil's ecological conditions and main staple crops; moreover, Embrapa increased researchers' productivity, especially in remote and resource scarce regions. Second, exploiting the staggered expansion of Embrapa's research centers and heterogeneous ecological similarity to new centers across municipalities, we find that Embrapa substantially increased agricultural productivity. These effects are driven by particularly pronounced yield increases for the staple crops which were the focus of Embrapa's research. Finally, combining these estimates with a model and data on Embrapa's cost structure, we find that Embrapa increased Brazilian agricultural productivity by 110% with a benefit-cost ratio of 17. Counterfactual analyses suggest that the geographic scope of Embrapa's research efforts and development of appropriate technology for Brazil's varied ecological conditions was an important mechanism. Together, these results suggest that investment in public R&D can be an important component of development policy and catalyst for productivity growth.

References

Acemoglu, D., Aghion, P., and Zilibotti, F. (2006). Distance to frontier, selection, and economic growth. *Journal of the European Economic Association*, 4(1):37–74.

Acemoglu, D. and Zilibotti, F. (2001). Productivity differences. *The Quarterly Journal of Economics*, 116(2):563–606.

Adamopoulos, T. and Restuccia, D. (2022). Geography and agricultural productivity: Cross-country evidence from micro plot-level data. *The Review of Economic Studies*, 89(4):1629–1653.

Alves, E. and Duarte, J. (2018). *Prosa com Eliseu*. Embrapa, Brasília, DF. Interview with Jorge Duarte.

Alves, M. H. M. (1988). Dilemmas of the Consolidation of Democracy from the top in Brazil: A Political Analysis. *Latin American Perspectives*, pages 47–63.

Azoulay, P., Graff Zivin, J. S., Li, D., and Sampat, B. N. (2019). Public R&D investments and private-sector patenting: evidence from NIH funding rules. *The Review of Economic Studies*, 86(1):117–152.

Barro, R. and Sala-i-Martin, X. (1997). Technological diffusion, convergence, and growth. *Journal of Economic Growth*, 2(1):1–26.

Basu, S. and Weil, D. N. (1998). Appropriate technology and growth. *The Quarterly Journal of Economics*, 113(4):1025–1054.

Bazzi, S., Gaduh, A., Rothenberg, A. D., and Wong, M. (2016). Skill transferability, migration, and development: Evidence from population resettlement in Indonesia. *American Economic Review*, 106(9):2658–2698.

Borusyak, K. and Hull, P. (2023). Nonrandom exposure to exogenous shocks. *Econometrica*, 91(6):2155–2185.

Brasil (1972). Lei nº 5.851, de 7 de dezembro de 1972. https://www.planalto.gov.br/ccivil_03/LEIS/1970-1979/L5851.htm. Dispõe sobre a criação da Empresa Brasileira de Pesquisa Agropecuária – Embrapa.

Bustos, P., Caprettini, B., and Ponticelli, J. (2016). Agricultural Productivity and Structural Transformation: Evidence from Brazil. *American Economic Review*, 106(6):1320–1365.

Cabral, J. I. (2005). *Sol da Manhã: Memória da Embrapa*. UNESCO, Representação no Brasil.

Cabral, L., Pandey, P., and Xu, X. (2022). Epic narratives of the Green Revolution in Brazil, China, and India. *Agriculture and Human Values*, 39(1):249–267.

Caselli, F. (2005). Accounting for cross-country income differences. *Handbook of Economic Growth*, 1:679–741.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1):1–45.

Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in Ghana. *American Economic Review*, 100(1):35–69.

Copenhagen Consensus Center (2004). Copenhagen consensus 2004: Results. https://www.copenhagenconsensus.com/copenhagen-consensus-2004.

Correa, P. and Schmidt, C. (2014). Public research organizations and agricultural development in Brazil: how did Embrapa get it right? Economic Premise Report 88490, World Bank Poverty Reduction and Economic Management Network (PREM).

Cruz, L. A. d. A. (2025). Interview with Luiz Antônio de Araújo Cruz. Interview conducted in person. Embrapa researcher interviewed for research purposes.

de Barros, J. R. M. and de Barros, A. L. M. (2005). A geração de conhecimento e o sucesso do agronegócio brasileiro. *XXXVIII Encontro Nacional da ANPAD 2014*, (2003):1–16.

Duflo, E., Kremer, M., and Robinson, J. (2011). Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya. *American Economic Review*, 101(6):2350–2390.

Egger, D., Haushofer, J., Miguel, E., Niehaus, P., and Walker, M. (2022). General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya. *Econometrica*, 90(6):2603–2641.

Elsevier (2025). Scopus Database. https://www.scopus.com. Accessed: January 2025.

Embrapa (2006). *Sugestões para a Formulação de um Sistema Nacional de Pesquisa Agropecuária*. Embrapa Informação Tecnológica, Brasília, DF, Facsimile of the June 1972 edition.

Embrapa (2018). Avaliação econômica de sistemas de Integração Lavoura-Pecuária-Floresta. Technical report, Embrapa.

Embrapa (2020). Avaliação de impactos do sistema de análise temporal da vegetação – satveg. Technical report, Embrapa Informática Agropecuária, Campinas.

Embrapa (2022a). Response to Information Request via Lei de Acesso à Informação (LAI). Obtained through the Lei de Acesso à Informação, Request NUP 21212000600202276, Accessed: February 23, 2025.

Embrapa (2022b). Response to Information Request via Lei de Acesso à Informação (LAI). Obtained through the Lei de Acesso à Informação, Request NUP 21212.000633/2022-16, Accessed: February 23, 2025.

Embrapa (2025a). Climate Change Portfolio. https://www.embrapa.br/en/portfolio/mudanca-climatica. Accessed: 2025-08-21.

Embrapa (2025b). História - Embrapa Rondônia. https://www.embrapa.br/en/rondonia/historia. [Accessed: February 23, 2025].

Embrapa and Helix (2022). Embrapa and Helix present a transgenic maize fully developed in Brazil. https://www.embrapa.br/en/busca-de-noticias/-/noticia/76710465/embrapa-and-helix-present-a-transgenic-maize-fully-developed-in-brazil. Accessed: 2025-06-28.

Evenson, R. E. (2001). Economic impacts of agricultural research and extension. *Handbook of Agricultural Economics*, 1:573–628.

Evenson, R. E. and Gollin, D. (2003). Assessing the impact of the Green Revolution, 1960 to 2000. *Science*, 300(5620):758–762.

Fieldhouse, A. and Mertens, K. (2023). The returns to government R&D: Evidence from US appropriations shocks. Working paper, Federal Reserve Bank of Dallas.

Fuglie, K. (2015). Accounting for growth in global agriculture. *Bio-based and Applied Economics*, 4(3):201–234.

Gasques, J. G., Bastos, E. T., and Bacchi, M. R. (2009). Produtividade e fontes de cres cimento da agricul tura. *Nota Técnica. Brasília: Ministério da Agricultura, Pecuária e Abastecimento, Assessoria de Gestão Estra tégica*.

Gasques, J. G., Bastos, E. T., Valdes, C., and Bacchi, M. R. P. (2012). Total factor productivity in Brazilian agriculture. In *Productivity growth in agriculture: an international perspective*, pages 145–161. CABI Wallingford UK.

Gollin, D., Hansen, C. W., and Wingender, A. M. (2021). Two blades of grass: The impact of the green revolution. *Journal of Political Economy*, 129(8):2344–2384.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014). Agricultural productivity differences across countries. *American Economic Review*, 104(5):165–170.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, pages 501–522.

Griliches, Z. (1958). Research costs and social returns: Hybrid corn and related innovations. *Journal of Political Economy*, 66(5):419–431.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics*, 10(1):92–116.

Gross, D. P. and Sampat, B. N. (2023). America, jump-started: World War II R&D and the takeoff of the US innovation system. *American Economic Review*, 113(12):3323–3356.

Gross, D. P. and Sampat, B. N. (2025). The therapeutic consequences of the war: World war ii and the 20th-century expansion of biomedicine. Working Paper 33457, National Bureau of Economic Research.

Gruber, J. and Johnson, S. (2019). *Jump-starting America: How breakthrough science can revive economic growth and the American dream*. PublicAffairs (Hachette Book Group), New York.

Hotez, P. J. (2021). *Forgotten people, forgotten diseases: the neglected tropical diseases and their impact on global health and development*. John Wiley & Sons.

Howell, S. T. (2017). Financing innovation: Evidence from R&D grants. *American Economic Review*, 107(4):1136–1164.

IBGE (2023). Municipal Agricultural Survey (PAM): Agricultural Production. https://sidra.ibge.gov.br. Accessed: 2025-06-21.

IBGE (2024). Biomas, 1:5000mil, shapefile. https://www.ibge.gov.br/geociencias/cartas-e-mapas/informacoes-ambientais/15842-biomas.html.

IBGE, D. d. G. (2011). Evolução da divisão territorial do Brasil 1872-2010. *Rio de Janeiro: IBGE*.

Jones, B. F. and Summers, L. H. (2020). A calculation of the social returns to innovation. Working Paper 27863, National Bureau of Economic Research.

Jones, C. I. (1995). R&D-based models of economic growth. *Journal of Political Economy*, 103(4):759–784.

Jones, C. I. (2002). Sources of us economic growth in a world of ideas. *American Economic Review*, 92(1):220–239.

Kantor, S. and Whalley, A. (2019). Research proximity and productivity: long-term evidence from agriculture. *Journal of Political Economy*, 127(2):819–854.

Kantor, S. and Whalley, A. T. (2023). Moonshot: Public R&D and growth. Working Paper 31471, National Bureau of Economic Research.

Klein, H. S. and Luna, F. V. (2018). *Feeding the World*. Cambridge University Press.

Klein, H. S. and Luna, F. V. (2023). *Brazilian crops in the global market: the emergence of Brazil as a world agribusiness exporter since 1950*. Springer.

Kremer, M. and Glennerster, R. (2004). *Strong medicine: creating incentives for pharmaceutical research on neglected diseases*. Princeton University Press.

Lachaud, M. A. and Bravo-Ureta, B. E. (2022). A Bayesian statistical analysis of return to agricultural R&D investment in Latin America: Implications for food security. *Technology in Society*, 70:102054.

Lerner, J., Liu, J., Moscona, J., and Yang, D. Y. (2024). Appropriate entrepreneurship? The rise of China and the developing world. Working Paper 32193, National Bureau of Economic Research.

Lopez, H. (2008). The Social Discount Rate: Estimates for Nine Latin American Countries. Policy Research Working Paper 4639, World Bank, Washington, DC.

Magalhães Jr., A. M. d. and Stone, L. (2018). Breeding better Brazilian rice. *Phys.org*. Accessed June 27, 2025.

Martha Jr, B. G., Contini, E., and Alves, E. (2012). 12. Embrapa: its origins and changes. *The regional impact of national policies: the case of Brazil*, 204.

Martha Jr., G. (2025). Personal interview. Interview conducted by the authors on March 7, 2025.

Monteiro, J., da Rocha, A., Kury, B., and Darzé, A. (2012). Soybeans in the Savannahs of Brazil. In Sabel, C. F., Fernández-Arias, E., Hausmann, R., Rodríguez-Clare, A., and Stein, E., editors, *Export Pioneers in Latin America*. Inter-American Development Bank and David Rockefeller Center for Latin American Studies, Harvard University, Washington, DC and Cambridge, MA.

Moretti, E. (2021). The effect of high-tech clusters on the productivity of top inventors. *American Economic Review*, 111(10):3328–3375.

Moretti, E., Steinwender, C., and Van Reenen, J. (2025). The intellectual spoils of war? Defense R&D, productivity, and international spillovers. *Review of Economics and Statistics*, 107(1):14–27.

Moscona, J. (2019). Agricultural development and structural change within and across countries. *Working Paper*.

Moscona, J. and Sastry, K. A. (2025). Inappropriate technology: Evidence from global agriculture. Working Paper 33500, National Bureau of Economic Research.

Nordhaus, W. D. (2004). Schumpeterian profits in the American economy: Theory and measurement. Working Paper 10433, National Bureau of Economic Research.

OECD (2022). *Agricultural Policy Monitoring and Evaluation 2022: Reforming Agricultural Policies for Climate Change Mitigation*. Agricultural Policy Monitoring and Evaluation. OECD Publishing, Paris.

Pardey, P. G., Alston, J. M., Chan-Kang, C., Magalhães, E. C., and Vosti, S. A. (2006). International and institutional r&d spillovers: Attribution of benefits among sources for brazil's new crop varieties. *American Journal of Agricultural Economics*, 88(1):104–123.

Parente, S. L. and Prescott, E. C. (1994). Barriers to technology adoption and development. *Journal of Political Economy*, 102(2):298–321.

Pellegrina, H. S. (2022). Trade, productivity, and the spatial organization of agriculture: Evidence from brazil. *Journal of Development Economics*, 156:102816.

Pellegrina, H. S. and Sotelo, S. (2024). Migration, Specialization, and Trade: Evidence from Brazil's March to the West. Technical report, Forthcoming, Journal of Political Economy.

Pingali, P. L. (2012). Green revolution: Impacts, limits, andthe path ahead. *Proceedings of the National Academy of Sciences of the United States of America*, 109(31):12302–12308.

Rohter, L. (2007). Scientists Are Making Brazil's Savannah Bloom. *The New York Times*. Accessed: 2025-08-12.

Rosegrant, M. W., Wong, B., Sulser, T. B., Dubosse, N., and Lybbert, T. J. (2023). Benefit–cost analysis of increased funding for agricultural research and development in the global south. *Journal of Benefit-Cost Analysis*, 14(S1):181–205.

Sargent, John F., J., Matthews, C. M., Morgan, D., Moteff, J. D., Sargent, J. F., Schacht, W. H., Smith, P. W., and Upton, H. F. (2009). Federal research and development funding: Fy2010. CRS Report for Congress R40710, Congressional Research Service, Washington, D.C.

Scheid Lopes, A., Guilherme, L. R. G., and Ramos, S. J. (2012). The Saga of the Agricultural Development of the Brazilian Cerrado. e-ifc: Research Findings 32, International Potash Institute.

SciELO (2025). Scientific Electronic Library Online (SciELO) Database. https://www.scielo.org. Accessed: January 2025.

SCImago (2025). SCImago Journal & Country Rank Database. https://www.scimagojr.com. Accessed: January 2025.

Stewart, F. (1978). *Technology and underdevelopment*. Springer.

Suri, T. and Udry, C. (2022). Agricultural technology in Africa. *Journal of Economic Perspectives*, 36(1):33–56.

USDA ARS (2024). About ARS. https://www.ars.usda.gov/about-ars/. Accessed: 2025-08-20.

Verhoogen, E. (2023). Firm-level upgrading in developing countries. *Journal of Economic Literature*, 61(4):1410–1464.

Vieira Filho, J. E. R. and Fishlow, A. (2017). *Agricultura e indústria no Brasil: inovação e competitividade*. Instituto de Pesquisa Econômica Aplicada (IPEA).

Vilaça de Vasconcelos, M. J., Figueiredo, J. E. F., de Oliveira, M. F., Schaffert, R. E., and Raghothama, K. G. (2022). Plant phosphorus use efficiency in acid tropical soil. *Revista Brasileira de Milho e Sorgo*, 21.

Appendix

A Data

In this appendix, we summarize additional details about measurement.

A.1 Agricultural Censuses

The Census interviews every farm in every municipality in intervals of five or ten years, achieving near universal coverage (Klein and Luna, 2018) and making it an invaluable source of information on agricultural output, land allocation, and input use. We assembled data for the rounds of 1950, 1960, 1970, 1975, 1980, 1985, 1996, 2006 and 2017.

The censuses rounds of 1996, 2006 and 2018 are available online from the SIDRA system. All other rounds were made available as scanned *pdf* files available in the digital library of the Brazilian Institute of Geography and Statistics (IBGE). We digitized virtually all the information from these earlier censuses.

While some prior work had relied on the extraction of select variables from a subset of the census rounds, we are unaware of an existing complete digitization and harmonization of this database. We exclude the 1920 and 1940 rounds because they were conducted long before the founding of Embrapa and because there are more substantial issues with data completion (i.e., an absence of information on land use) and methodological changes that make it challenging to compare data across rounds.

We summarize the available information in each round of the agricultural census in Appendix Table C.1.

A.2 Article Topic Classification

In our analysis of the direction of research (Sections 4.1, 4.2, and 4.3), we rely on a keyword classification of article topics. To do this, we first process the titles of articles into lowercase strings with no leading and trailing spaces and no accented characters. We then produce keyword dictionaries for each topic. These are printed in full in Table A.1. We build the dictionaries by first enumerating the English, Portuguese, and Spanish translation of the term. We then enumerate common synonyms. For crops and pests, we also include scientific names.

Table A.1: Keywords for Topic Classification

Category	Topic	Keywords
Biomes	Cerrado	cerrado, tropical savanna
	Pantanal	pantanal, tropical wetland
	Amazonia	amazon
	Caatinga	caatinga
	Pampa	pampa
	Mata Atlantica	atlantic forest, atlantic rainforest, mata atlantica, foret atlantique, selva misionera, bosque atlantica, floresta atlantica
Crops	Wheat	trigo, triticum aestivum, t. aestivum, triticum, wheat
	Soy	soja, glycine max, g. max, glycine, soy
	Rice	arroz, rice, oryza sativa, o. sativa, oryza
	Beans	feijao, feijoes, feijoeiro, phaseolus vulgaris, p. vulgaris, phaseolus, common bean
	Corn	milho, maiz, maize, corn, zea mays, z. mays
Pests	Whiteflies	whitefl, mosca branca, mosca blanca, mosca-branca, mosca-blanca, moscas blancas, moscas, brancas, bemisia, b. tabaci
	Fusarium	fusarium ear blight, fusarium head blight, fusariose do trigo, fusarium graminearum, f. graminearum
	Boll Weevil	boll weevil, anthonomus grandis, a. grandis, bicudo-do-algodoeiro, bicudo do algodoeiro, grillo de la capsula del algodonero
	Wheat Rust	wheat rust, cereal rust, stem rust, ferrugen do colmo, ferrugem do trigo, ferrugem-do-trigo, ferrugem do colmo, ferrugem-do-colmo, polville de la cana, roya del tallo, roya del trigo, roya negra, (Wheat & (rust, ferrugem, ferrugen))
	Witches' Broom	moniliophthora perniciosa, m. perniciosa, crinipellis perniciosa, witches' broom, witches broom, escoba de bruja, vassoura de bruxa, vassoura-de-bruxa
	Coffee Berry Borer	hypothenemus hampei, h. hampei, coffee berry borer, barrenador del cafe, broca del cafe, broca del fruto del cafe, totaladro de las cerezas del cafeto
	Coffee Leaf Rust	hemileia vastatrix, h. vastatrix, coffee leaf rust, ferrugem do cafeeiro, ferrugem do cafe, roya del caf, coffee rust, (Coffee & (rust, ferrugem, ferrugen))
	Fall Armyworm	spodoptera frugiperda, s. frugiperda, fall armyworm, lagarta do cartucho, lagarta militar
	Corn Earworm	helicoverpa zea, h. zea, corn earworm, bollworm, lagarta da espiga, broca grande do fruto, gusano bellotero del algodon
	Soybean Rust	phakopsora, p. pachyrhizi, p. meibomiae, soybean rust, soyabean rust, roya de la soya, roya de la soja, ferrugem da soja, ferrugem asiatica, (Soybean & (rust, ferrugem, ferrugen))
	Soybean Cyst Nematode	soybean cyst nematode, heterodera de la soja, nematodo de la soya, nematoide de cisto da soja, cisto da soja

Notes: This table prints our keywords for classifying the topics of research articles in the analysis of Section 4. The second column prints the topics, arranged in three categories (biomes, crops, and pests/pathogens). Before searching the keywords, we transform the titles to be lowercase strings with no accented characters. In the keyword lists, commas denote an "or" condition. For three pests (wheat rust, coffee leaf rust, and soybean rust), we include one compound rule that identifies articles that match the respective crop and any of three synonyms for "rust."

Table A.2: Components of Ecological Similarity Index

Measure	Original Unit	Notes	Source
Temperature	°C	Annual mean from 1981 to 2010	Willmott and Matsuura
Precipitation	mm	Annual mean from 1981 to 2010	Willmott and Matsuura
Growing season	days	Sufficiently warm and moist days	FAO GAEZ
Elevation	m	Distance above sea level	GTOPO30 Digital Elevation Model
Ruggedness	m^2	Relative elevation to neighboring grid cells	Riley et al. (1999) and Nunn and Puga (2012)
Soil acidity	pH	in water to 250m	SoilGrids and WoSIS
Clay content	% mass	to 250m	SoilGrids and WoSIS
Silt content	% mass	to 250m	SoilGrids and WoSIS
Coarse fragments	% volume	to 250m	SoilGrids and WoSIS

Notes: This table describes the 9 geographic attributes used in the construction of ecological similarity (see Section 5.1 and Appendix A.3). In our analysis, we summarize all of these measures at the level of municipalities and convert them from their original units (identified in column 2) to z-score units across municipalities. The "growing season" is defined as days in which temperature exceeds 5 °C and the sum of precipitation and soil moisture exceeds 0.5 times potential evapotranspiration. The soil categories are defined by particle sizes: "clay" is from 0 to 2 μm, "silt" is from 2 to 50 μm, and coarse fragments are over 2mm. The "Willmott and Matsuura" data correspond to Matsuura and National Center for Atmospheric Research Staff (2023). The WoSIS (World Soil Information Service) data are described in Batjes et al. (2017).

A.3 Geospatial Data and Ecological Similarity Index

In Section 5.1, we use an index of geospatial attributes to construct a measure of ecological similarity between pairs of locations. We refer to this variable as Ecological Similarity$_{ij}$, defined for pairs of municipalities $i, j \in \mathcal{I}$ as

$$\text{Ecological Similarity}_{ij} = -\sum_{x} |x_i - x_j|, \tag{12}$$

where each x is a separate geographic attribute. Here, we describe the construction of this index in more detail.

We construct the index using nine geospatial data sources, which are described in Appendix Table A.2. Three describe the climate: temperature, precipitation, and growing season. Two describe topography: elevation and ruggedness. Four describe soil characteristics: acidity and relative content of clay, silt, and coarse fragments.

To construct the index, we first measure each of the nine characteristics by taking spatial averages over each Brazilian municipality (i.e., *Área Mínima Comparável*, or AMC). We let $\tilde{x}_i$ denote the value of a given characteristic in municipality i, expressed in its original units. We next transform each characteristic into a z-score:

$$x_i = \frac{\tilde{x}_i - \text{mean}(\tilde{x}_i)}{\text{sd}(\tilde{x}_i)} \tag{13}$$

where we take the mean and standard deviation across the municipalities. This expresses each characteristic in a common unit. Finally, we sum the absolute differences of each attribute to construct ecological similarity (equation (12)).

The form of our index, including the choice of the attributes and the choice of the ℓ_1 distance, is based on related prior work. Bazzi et al. (2016) use a related index of agro-climatic similarity between agricultural regions of Indonesia to proxy for the transferability of agricultural workers' skills. Moscona and Sastry (2025) use a related index to study the appropriateness of internationally transferred agricultural technology.

While our main analysis uses the composite ecological similarity index, we also investigate the sensitivity of our findings to individual components of the index. Appendix Figure C.3 replicates our baseline estimates of equation (7) after dropping individual components of the index. Our findings indicate that our results are not unduly quantitatively sensitive to any specific component.

B Structural Model

In Section 6, we briefly presented the structural model and its estimation. In this appendix, we provide a thorough description of the estimation procedure. To keep the section self-contained, we repeat some of the equations from the main text.

B.1 Estimation of Agricultural Productivity Function

Our estimating equation is derived from equation (10)

$$A_{it} = \bar{A}_{it} \underbrace{\left(\sum_{j \in \mathcal{R}_t} \left[\exp(-\beta g_{ij})(N_{jt})^{\gamma} \right]^{\theta} \right)^{\frac{1}{\theta}}}_{\equiv EE_{it}} \tag{14}$$

where we define EE_{it}, or "Embrapa Exposure," as the second term. Taking logarithms, and re-arranging terms, we obtain equation (11) from the main body of the paper:

$$\log A_{it} = -\beta g_{i\bar{\jmath}} + \gamma \log N_{\bar{\jmath}} + \frac{1}{\theta} \log \left(\sum_{j \in \mathcal{R}_t} \left[\frac{\exp(-\beta g_{ij})(N_{jt})^{\gamma}}{\exp(-\beta g_{i\bar{\jmath}})(N_{\bar{\jmath}})^{\gamma}} \right]^{\theta} \right) + \log \bar{A}_{it}. \tag{15}$$

We write the equation above in differences, between reference period T and initial period 0, which gives

$$\log A_{iT} - \log A_{i0} = -\beta g_{i\bar{\jmath}} + \gamma \log N_{\bar{\jmath}} + \frac{1}{\theta} \log \left(\sum_{j \in \mathcal{R}_T} \left[\frac{\exp(-\beta g_{ij})(N_{jT})^{\gamma}}{\exp(-\beta g_{i\bar{\jmath}})(N_{\bar{\jmath}})^{\gamma}} \right]^{\theta} \right) + \epsilon_i. \tag{16}$$

Here, the residual term is $\epsilon_i \equiv \log(\bar{A}_{iT}) - \log(\bar{A}_{i0}) - \log(EE_{i0})$, and captures both the initial exposure to Embrapa and changes in the exogenous productivity shifter. We estimate this equation using non-linear least squares, including a constant term. We note that, by taking the first difference, the specification absorbs any municipality-specific factors that are constant over time—such as natural agroclimatic conditions—as well as period-specific effects that are constant across municipalities, such as changes in national-level prices.

B.2 Aggregate Productivity Gain

Using data on cost of researchers N_{iT} for a reference period T, together with our estimates of β, γ, and θ and our measurements of ecological similarity $[g_{ij}]_{i,j \in \mathcal{I}}$, we construct Embrapa Exposure in period T:

$$EE_{iT} = \left(\sum_{j \in \mathcal{R}_T} [\exp(-\beta g_{ij})(N_{jT})^{\gamma}]^{\theta} \right)^{\frac{1}{\theta}},$$

we then recover the productivity shifter $\bar{A}_{iT}$ by combining data on A_{iT} with EE_{iT}

$$\bar{A}_{iT} = \frac{A_{iT}}{EE_{iT}}$$

We create our baseline productivity level for each municipality i, from which we evaluate our counterfactuals, using data on the initial level of research activity N_{i0}, productivity in baseline 0 given a reference year T, based on the following expression

$$A_{i0T} = \bar{A}_{iT} \left(\sum_{j \in \mathcal{R}_0} [\exp(-\beta g_{ij})(N_{j0})^{\gamma}]^{\theta}. \right)^{\frac{1}{\theta}}$$

In counterfactuals, we specify a counterfactual level of research cost N_{ic}, compute the resulting exposure to Embrapa EE_{ic}, and evaluate the aggregate agricultural productivity gains from baseline year 0 to the counterfactual c, using a reference year T, based on

$$\widehat{A}_{0Tc} = \frac{\sum_i \bar{A}_{iT} EE_{ic}}{\sum_i \bar{A}_{iT} EE_{i0}}. \tag{17}$$

All reported aggregate productivity gains are based on this equation, using 2006 as the reference year and 1970 as our baseline.

B.3 Benefit-Cost Analysis

To compute the benefit, we assume that changes in aggregate value-added are proportional to changes in aggregate productivity. We therefore compute baseline value-added as:

$$V_{0T} = \overline{V}_T \frac{\sum_{i \in \mathcal{I}} \bar{A}_{iT} EE_{i0}}{\sum_{i \in \mathcal{I}} \bar{A}_{iT} EE_{iT}}$$

where $\overline{V}_T$ is national agricultural value-added, which we observe (in US dollars) using data from the UN Food and Agriculture Organization (FAO).

We then compute the gains in US dollars associated with $\widehat{A}_{0Tc}$ as

$$GV_{0Tc} = \widehat{A}_{0Tc} - 1 \tag{18}$$

Using the gains in value added for the reference year, we assume a simple phase-in structure. Specifically, we assume that, between 1975 and 2000, the gains increase linearly, reaching the full value by 2000. From 2000 onward, the gains remain constant at the level observed in the reference year.

$$GV_{0tc} = \begin{cases} \dfrac{GV_{0Tc}}{(2000-1975)} & \text{if } t \geq 1975 \text{ and } t \leq 2000 \\ GV_{0Tc} & \text{if } t \geq 2000 \end{cases} \tag{19}$$

We then compute the present value of all benefits using

$$PVB_T = \sum_{t=1975}^{T} \frac{V_{0T} \times GV_{0Tc}}{(1+0.07)^{T-t}} + \frac{V_{0T} \times GV_{0Tc}}{0.05} \tag{20}$$

where the first term brings past benefits to present value of reference year T at an interest rate of 7 percent and discount future values after T at a discount rate of 5 percent. Note that we assume that the benefits are repeated throughout the future.

The present value of the cost is simpler to compute. We have data on the costs of research in local currency, adjusted by inflation. We convert these values into US dollars given the exchange rate in the year of reference. We then compute

$$PVC_T = \sum_{t=1975}^{T} \frac{RC_t}{(1+0.07)^{T-t}} + \frac{RC_T}{0.05} \tag{21}$$

where RC_t is the total research cost of Embrapa in year t, including personnel and capital.

The benefit-cost ratio is then

$$BC_T = \frac{PVB_T}{PVC_T}. \tag{22}$$

B.4 Internal Rate of Return

We compute the internal rate of return as of 1974, at the beginning of the project. To do so, we bring all the costs to present value using the same interest rates applied in the benefit cost ratios.

$$PVC_0 = \sum_{t=1975}^{T} \frac{RC_t}{(1+0.07)^{T-t}} + \frac{RC_T}{0.05 \times (1+0.07)^{T}} \tag{23}$$

We then compute the interest rate that would make the net present value of Embrapa equal to zero. The net present value is given by:

$$NPV_0 = \sum_{t=0}^{\infty} \frac{GV_{0Tc}}{(1+r)^t} - PVC_0. \tag{24}$$

And the IRR is the value of r that makes $NPV_0 = 0$.

C Additional Tables and Figures

Figure C.1: Example of a Lattes Profile

(a) Professional summary

Eliseu Roberto de Andrade Alves

Endereço para acessar este CV: http://lattes.cnpq.br/2601943746052858

ID Lattes: **2601943746052858**

Última atualização do currículo em 21/03/2017

Ph. D em Economia Agrícola pela Purdue University Indiana (1972), mestre em Agricultural Economics - Purdue University Indiana (1968). Graduado em engenharia agrônoma pela Univrsidade Fedeal de Viçosa (1954) .É servidor público há 58 anos, sem interrupção: na EMATER-MG e EMBRAPA. Atuou como chefe do departamento de planejamento e avaliação da ACAR MG., hoje, EMATER MG. de 1968 a 1973; diretor da EMBRAPA de 1973 a 1979; presidente da EMBRAPA de 1979 a 1985; presidente da CODEVASF de 1985 a 1989; secretário nacional de irrigação 1989; Assessor da Presidência da Embrapa de 1990-2010.Professor de estatística, microeconomia e política agrícola do curso de posgraduação em Economia Regional, CEDEPLAR,da Universidade Federal de Minas Gerais, Professor de economia rural da curso de posgraduação da Fundação Getúlio Vargas, professor de economia rural do curso de posgraduação da Faculdade de Economia e Administração da Universidade de São Paulo.Tem orientado teses em nível de mestrado e doutorado, na no nível de mestrado e doutorado: CEDEPLAR, UFV e USP. Durante 17 anos contribuiu para consolidar o serviço de extensão rural, de 1955 a 1973. Junto com um grupo de cientistas sociais, ajudou conceber e formular o modelo no qual aEMBRAPA se baseou e ajudou a defender o modelo no governo. Período 1971-73.Como diretor, concebeu e executou o programa de posgraduação. Como presidente da EMBRAPA, fez deste programa a primeira prioridade, atingindo a marca de 1600 posgraduados, em 12 anos. Ainda, ajudou consolidar a EMBRAPA e torná-la uma referência nacional e internacional. Como presidente da CODEVASF, criou o conceito do distrito de irrigação, pelo qual os projetos públicos passaram a ser administrados pelos irrigantes. Emancipou todos os projetos da CODEVASF. Concebeu e implantou o programa de exportação de frutas. Negociou empréstimos com o Banco Mundial, BID e Governo Japonês que estão permitindo uma expansão de mais de cem mil hectares de área irrigada. Ajudou conceber e executar o programa de 1 milhão de hectares irrigados do Presidente Sarney. Este programa foi inspirado num trabalho feito por técnicos da EMBRAPA, do qual sou o primeiro autor. Iniciou o processo de mudança da lei de irrigação, ainda em curso hoje. Prestamos, presentemente, assessoria ao grupo em carregado de reformulá-la. Na FGV ajudou formular e implementar uma pesquisa por amostragem que serviu para definir melhor as contas do setor agrícola. De 1990 em diante, totalmente dedicado à pesquisa em política agrícola e em ciência e tecnologia. Nesta condição, tem dado assessoria a governos, na condição de consultor do BID, Banco Mundial e FAO. Destacam-se Venezuela, Paraguai e Equador. Tem atendido a convites de universidades, cooperativas, organizações de produtores e científicas para fazer palestras sobre a dinâmica da agricultura brasileira, tema objeto de minhas pesquisas. **(Texto informado pelo autor)**

Identificação

Nome	Eliseu Roberto de Andrade Alves
Nome em citações bibliográficas	ALVES, E. R. A.;Alves, Eliseu
Lattes iD	http://lattes.cnpq.br/2601943746052858
País de Nacionalidade	Brasil

Endereço

Endereço Profissional	Empresa Brasileira de Pesquisa Agropecuária. Parque Estação Biológica pqeb ed. Sede av.w3 norte(final) Asa Norte 70770901 - Brasília, DF - Brasil Telefone: (61) 34484202 Ramal: 4202

Formação acadêmica/titulação

1968 - 1972	Doutorado em Agricultural Economics. Indiana University-Purdue University Indianapolis, IUPUI, Estados Unidos. Título: An Econometric Study of the Agricultural Labor Market in Brazil:, Ano de obtenção: 1972. Orientador: G.E Schuh. Grande área: Ciências Agrárias Grande Área: Ciências Agrárias / Área: Agronomia. Setores de atividade: Educação; Agricultura, Pecuária e Serviços Relacionados.
1965 - 1968	Mestrado em Economia Agrícola. Indiana University-Purdue University Indianapolis, IUPUI, Estados Unidos. Título: An Econometric Evaluation of the Impact of an Extension Program: Minas Gerais,, Ano de Obtenção: 1968. Orientador: G.E Schuh.

(b) Articles

Produções

Produção bibliográfica

Artigos completos publicados em periódicos

Ordenar por

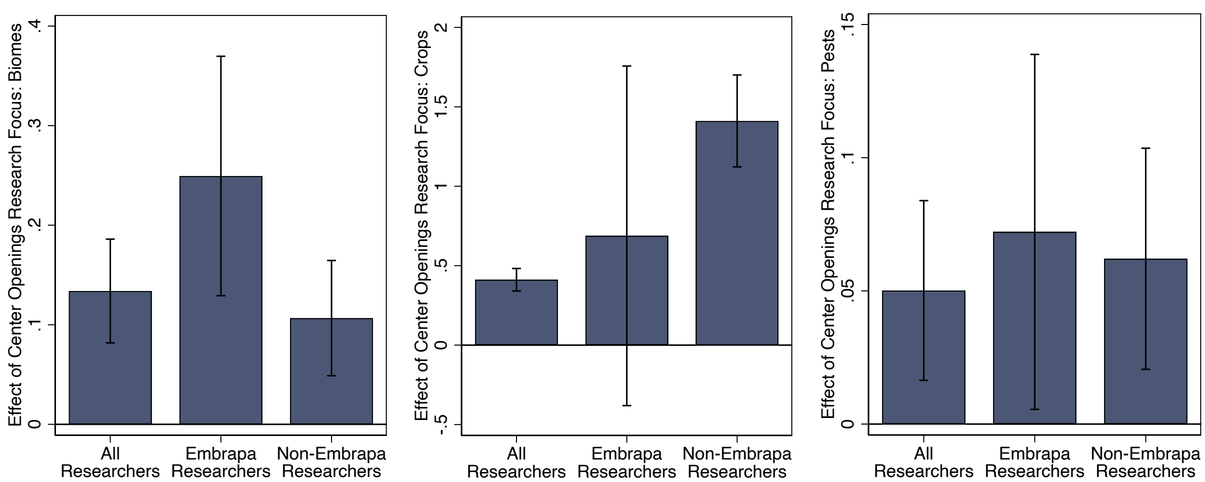
Ordem Cronológica

1. **Alves, Eliseu**; DUARTE, J. . O elemento invisível no progresso tecnológico. Revista de Política Agrícola, v. 25, p. 122-124, 2016.
2. **Alves, Eliseu**. Onde estamos e para onde vamos. Revista de Politica Agricola, v. 25, p. 3, 2016.
3. **Alves, Eliseu**; Souza, Geraldo da Silva . O semiárido segundo o Censo Agropecuário 2006 e os censos de população 1991, 2000 e 2010. Revista de Politica Agricola, v. 24, p. 74-85, 2015.

Notes: This figure shows an example of an individual researcher's CV on Lattes. Panel (a) shows the professional summary, which identifies the individual's education (degrees, institutions, and thesis titles) and employment spells, identified by years and job titles. Panel (b) shows the beginning of the profile's listing of publications, from which we observe the author(s), title, forum of publication, and year.

Figure C.2: Effect of Embrapa on Research Direction: Embrapa vs. non-Embrapa Research

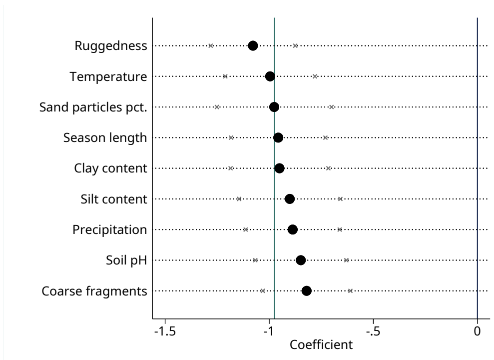
(a) Biome Focus **(b)** Crop Focus **(c)** Pest and Pathogen Focus



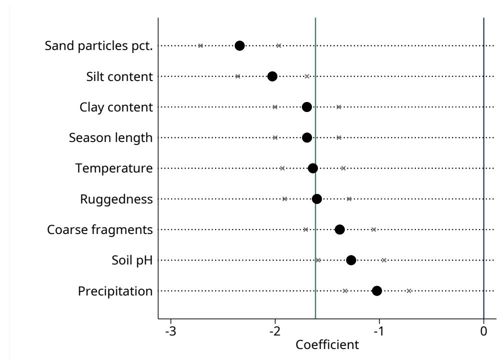
Notes: Each subfigure reports estimates of equation (3), showing the effect of the opening of new Embrapa centers on the re-direction of research across biomes (a), crops (b), and pests and pathogens (c). In each subfigure, the first bar reports the effect on total topic-specific research, the second reports the effect on topic-specific research by Embrapa affiliates, and the third reports the effect on topic-specific research by researchers unaffiliated with Embrapa. 95% confidence intervals are reported.

Figure C.3: Effect of Embrapa Exposure on Productivity Dropping Index Components

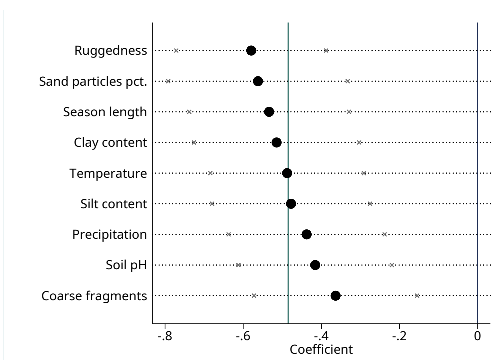
(a) log(Production Value / Area)



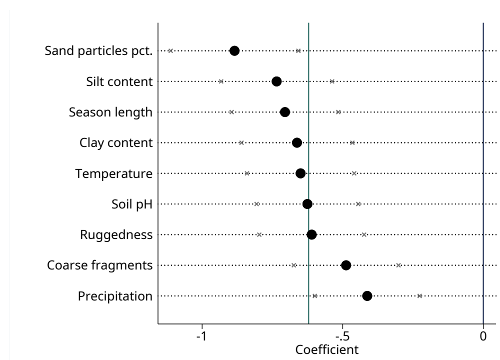
(b) log(Agricultural Yield)



(c) log(Farm Value / Area)

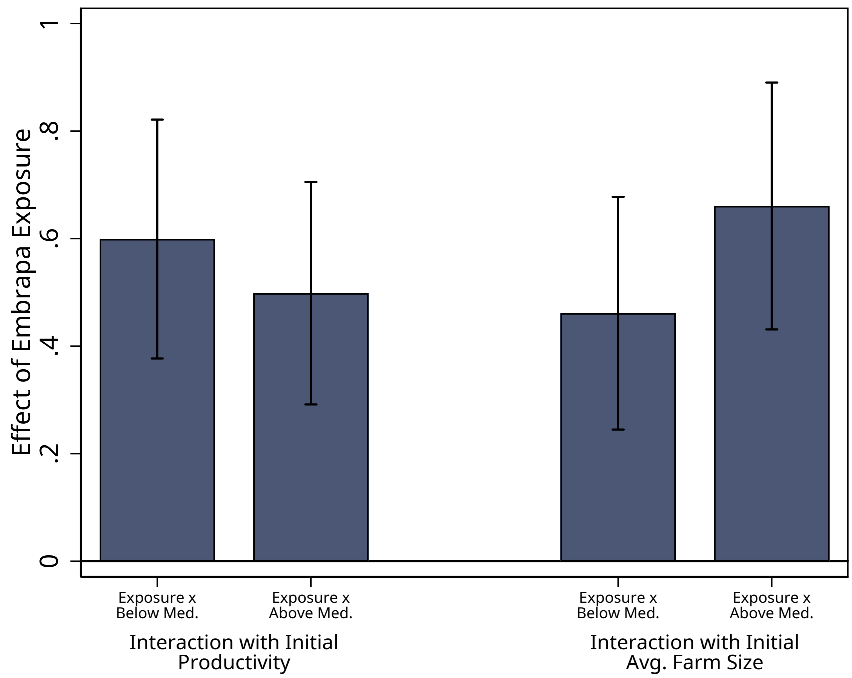


(d) log(Agricultural TFP, Pre Weights)



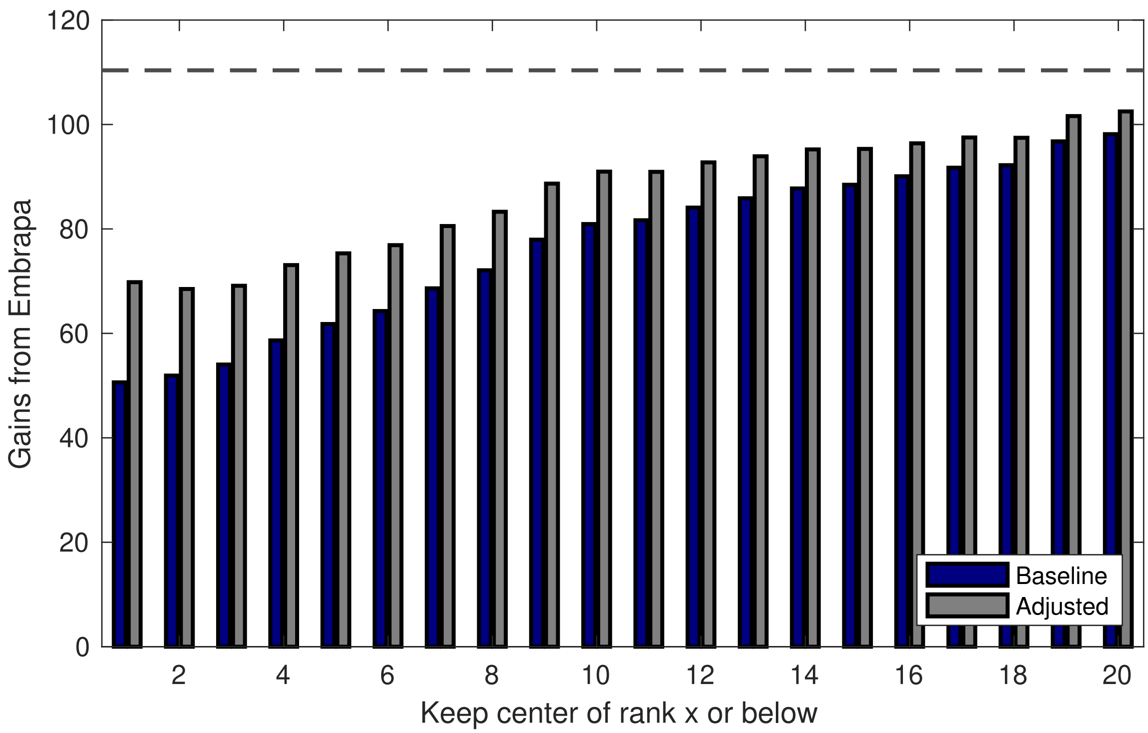
Notes: Each subfigure shows the robustness of our main estimates of equation (7) under variant constructions of Embrapa Exposure that exclude each indicated component of agro-climatic similarity (see Table A.2). Each subfigure corresponds to a different productivity outcome variable. The dots and error bars correspond to estimates and 95% confidence intervals for each variant model, and the vertical green line corresponds to the point estimate using the main index.

Figure C.4: Heterogeneous Effects of Embrapa Exposure on Productivity by Baseline Productivity and Farm Size



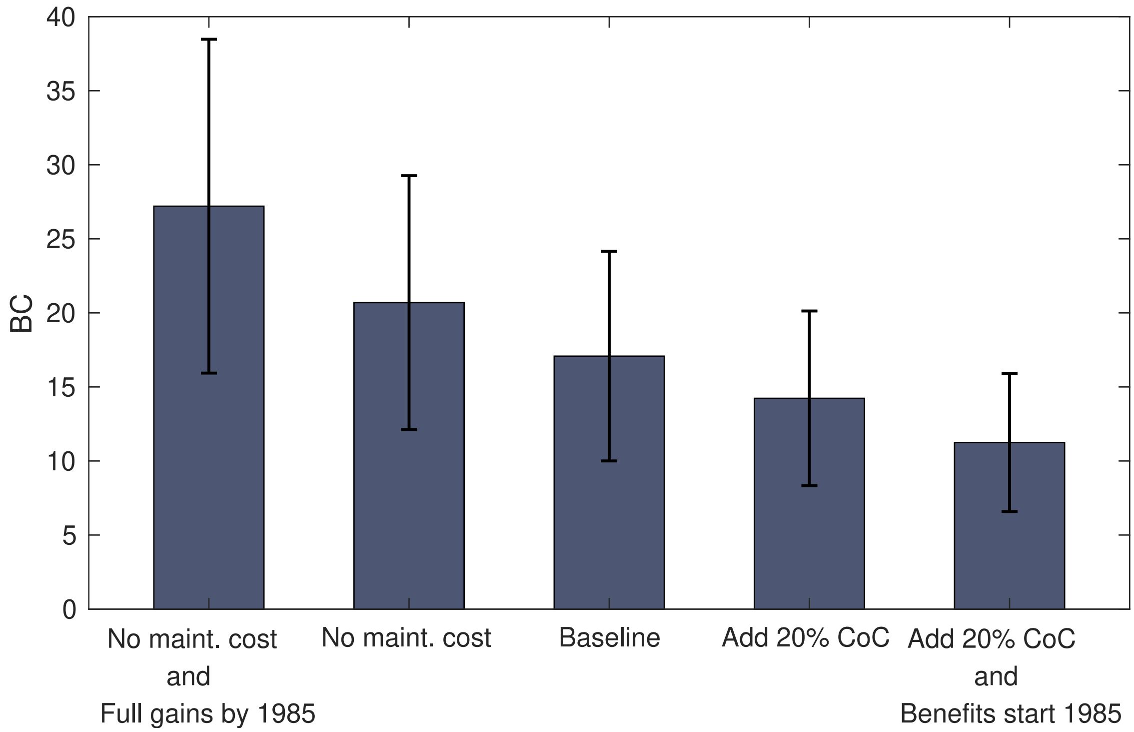
Notes: This figure presents regression estimates of equation (7), augmented to include interaction terms between Embrapa Exposure and indicators for above versus below baseline agricultural productivity (columns 1-2) and indicators for above versus below baseline average farm size (columns 3-4). All specifications include municipality and year fixed effects, in addition to the logarithm of distance (in km.) to the nearest Embrapa center times census-round fixed effects and state-by-census-round fixed effects. Standard errors are clustered by municipality and 95% confidence intervals are reported.

Figure C.5: Productivity Gains from Sequentially Opening Research Centers



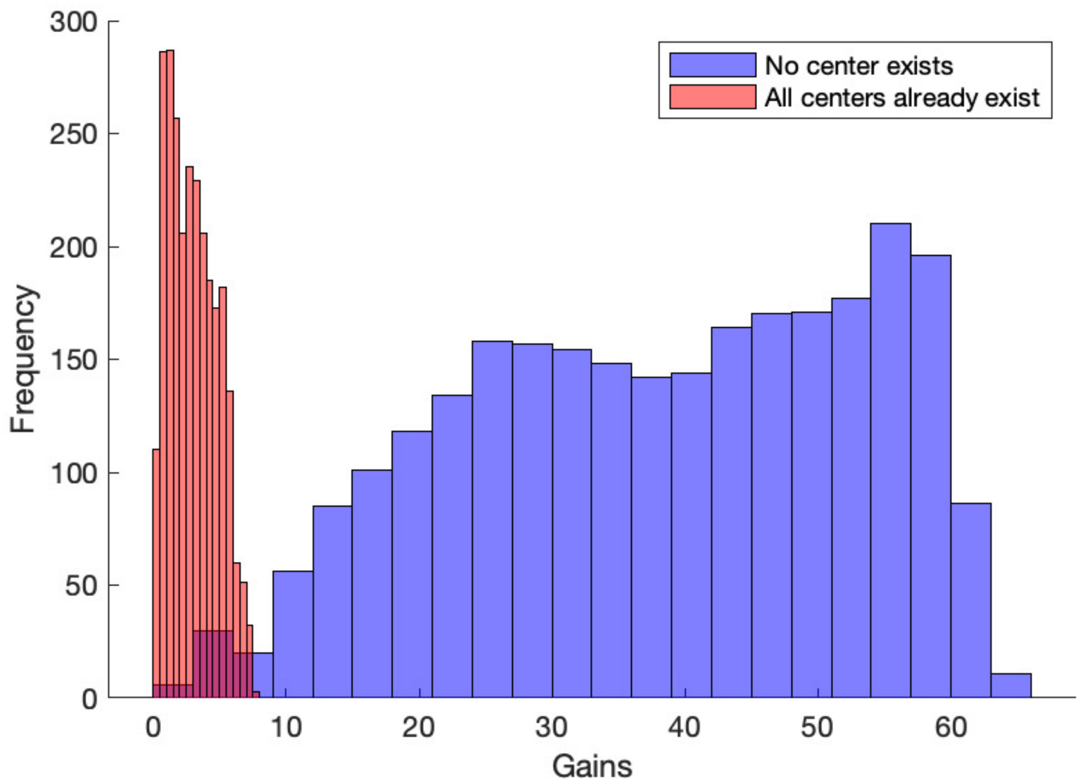
Notes: This figure illustrates the productivity gains from sequentially adding research centers, starting with the largest and progressing to the smallest. The dotted gray line indicates the total gains from Embrapa under the observed allocation. The blue bar represents the gains if remove any other center, the gray bar the gains if all researchers from Embrapa are reallocated to the active centers, keep the relative size of the active centers fixed.

Figure C.6: Benefit-Cost Calculation Under Different Assumptions



Notes: This figure reports the benefit–cost ratio under alternative assumptions about costs and benefits. The first bar excludes Embrapa's maintenance expenditures by setting all research costs to zero after the 2006 reference year and a shorter phase in period in which the full benefits are realized in 1985. The second bar only excludes Embrapa's maintenance expenditures. The third bar corresponds to our preferred specification. The fourth bar raises the Cost of Capital (CoC) by 20 percent, so that each dollar invested in Embrapa requires 1.20 in funding. The fifth bar assumes that the benefits of Embrapa's research phase are null until 1985 and phase in linearly after then until 2000.

Figure C.7: Productivity Gains of New Research Centers Before and After Embrapa



Notes: This figure displays the productivity gains from adding a large Embrapa center under two scenarios: first, when no centers exist; and second, when all centers established by 2006 are already in place. In both cases, the size of the new center is held constant.

Table C.1: Digitization of the Agricultural Census of Brazil

	Historical censuses (*newly digitized*)						Modern censuses		
	1950	1960	1970	1975	1980	1985	1996	2006	2017
Farm value	✓	✓	✓	✓	✓	✓	✓	✓	✓
Land use	✓	✓	✓	✓	✓	✓	✓	✓	✓
Farming output per crop	✓	✓	✓	M	✓	✓	✓	✓	✓
Use of land per crop	✓	✓	✓	✓	✓	✓	✓	✓	✓
Farm size distribution	✓	✓	✓	✓	✓	✓	✓	✓	✓
Livestock	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tractors	✓	✓	✓	✓	✓	✓	ø	✓	✓
Agricultural workers	✓	✓	✓	M	✓	✓	✓	✓	✓

Notes: This table describes the data we collect from the Agricultural Census of Brazil. ✓: all reported tables have been completely digitized; M: all available tables have been digitized, but there were missing pages in the original documents; ø: not collected by the census or never made available.

Table C.2: Dynamic Effects of Embrapa Affiliation on Researcher Productivity

Defn. of High v. Low Research:	Agricultural Research		College Graduate Share	
	ihs(Papers)	Norm. Count	ihs(Papers)	Norm. Count
Embrapa x Low Research, t+2	0.034	0.017	0.034	0.006
	(0.039)	(0.100)	(0.038)	(0.094)
Embrapa x Low Research, t+1	0.022	0.020	0.027	0.029
	(0.043)	(0.111)	(0.040)	(0.102)
Embrapa x Low Research, t	-0.005	-0.017	-0.002	-0.003
	(0.046)	(0.141)	(0.042)	(0.132)
Embrapa x Low Research, t–1	0.040	0.081	0.044	0.099
	(0.037)	(0.104)	(0.035)	(0.097)
Embrapa x Low Research, t–2	0.051*	0.102	0.051*	0.102
	(0.029)	(0.082)	(0.029)	(0.084)
Embrapa x Low Research, t–3	0.073***	0.185**	0.059**	0.143**
	(0.027)	(0.074)	(0.026)	(0.070)
Embrapa x High Research, t+2	0.028	-0.051	0.022	-0.068
	(0.029)	(0.086)	(0.030)	(0.087)
Embrapa x High Research, t+1	0.012	-0.031	0.004	-0.049
	(0.040)	(0.113)	(0.042)	(0.122)
Embrapa x High Research, t	-0.022	-0.040	-0.021	-0.045
	(0.042)	(0.119)	(0.046)	(0.137)
Embrapa x High Research, t–1	0.016	0.035	0.008	0.012
	(0.036)	(0.108)	(0.040)	(0.117)
Embrapa x High Research, t–2	0.024	0.013	0.026	0.012
	(0.031)	(0.087)	(0.031)	(0.088)
Embrapa x High Research, t–3	0.031	0.051	0.037	0.074
	(0.026)	(0.070)	(0.029)	(0.081)
Adj. R2	0.500	0.508	0.500	0.508
Observations	351089	351089	350672	350672
Year FE	Y	Y	Y	Y
AMC × Year FE	Y	Y	Y	Y
Researcher FE	Y	Y	Y	Y

Notes: This table reports estimates of regression 4, with the inverse hyperbolical sine of published papers (columns 1 and 3) and normalized amount of papers (columns 2 and 4) as the dependent variable, as well as two leads and three lags of the main independent variables. "Low Research" is defined as below the top ten AMCs in terms of total agricultural research (columns 1-2) and the college graduate share (columns 3-4). Standard errors clustered at the AMC level and reported in parentheses.

Table C.3: Effects of Embrapa on Alternative Measures of Researcher Productivity

	(1)	(2)	(3)	(4)	(5)	(6)
(a) Outcome is (asinh) Citation-Weighted Articles						
Embrapa	0.178***	0.173***				
	(0.054)	(0.058)				
Embrapa x Low Research			0.269***	0.257***	0.249***	0.242***
			(0.054)	(0.049)	(0.053)	(0.048)
Embrapa x High Research			0.097**	0.122**	0.109**	0.120**
			(0.042)	(0.048)	(0.043)	(0.047)
Low Research			-0.075***		-0.033***	
			(0.011)		(0.009)	
Adj. R2	0.355	0.379	0.355	0.379	0.355	0.379
(b) Outcome is the Normalized Article Count						
Embrapa	0.080	0.101				
	(0.071)	(0.075)				
Embrapa x Low Research			0.129***	0.123***	0.195***	0.189***
			(0.027)	(0.027)	(0.071)	(0.066)
Embrapa x High Research			0.039	0.065**	-0.009	0.031
			(0.025)	(0.026)	(0.062)	(0.066)
Low Research			-0.038***		-0.020	
			(0.010)		(0.016)	
Adj. R2	0.430	0.455	0.428	0.454	0.430	0.455
(c) Outcome is an Article Indicator						
Embrapa	0.068***	0.068***				
	(0.011)	(0.012)				
Embrapa x Low Research			0.085***	0.081***	0.078***	0.078***
			(0.012)	(0.012)	(0.012)	(0.012)
Embrapa x High Research			0.049***	0.060***	0.052***	0.061***
			(0.013)	(0.011)	(0.012)	(0.011)
Low Research			-0.023***		-0.013***	
			(0.006)		(0.003)	
Adj. R2	0.335	0.362	0.336	0.362	0.335	0.362
Observations	530672	519562	530672	519562	530377	519291
Heterogeneity	-		Previous Research		College Degree	
Year FE	Y	Y	Y	Y	Y	Y
Municipality × Year FE	–	Y	–	Y	–	Y
Researcher FE	Y	Y	Y	Y	Y	Y
Tenure FE	Y	Y	Y	Y	Y	Y

Notes: This table replicates the analysis of Table 1 using alternative outcome measures: the inverse hyperbolic sine of citation-weighted articles (a), the count of articles winsorized at the 99th percentile of the researcher-by-year data (b), and an indicator (0/1) for whether a researcher published any article. The regression model is equation (4). Columns 1 and 2 main regressor indicates whether the researcher works for Embrapa. Columns 3 to 6 interact the Embrapa indicator with municipality research characteristics. Columns 3 and 4 indicates whether the municipality had a low or high previous agricultural research production. Columns 5 and 6 indicate whether the municipality had a low or high share of college graduates. All regressions include Year, Researcher, and job tenure fixed effects. Columns 2, 4, and 6 include municipality-year fixed effects. Standard errors clustered at the municipality level.

Table C.4: Effects of Embrapa Exposure on Other Measures of Productivity

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Outcome is log(Crop Yield)					
Embrapa Exposure	1.373***	1.595***	1.519***	1.502***	1.366***	1.262***
	(0.119)	(0.130)	(0.172)	(0.130)	(0.132)	(0.140)
	13093	12895	8432	12892	12892	12895
	0.614	0.615	0.635	0.623	0.629	0.743
	Panel B: Outcome is log(Farm Value / Farm Area)					
Embrapa Exposure	0.383***	0.480***	0.594***	0.470***	0.410***	0.341***
	(0.072)	(0.077)	(0.110)	(0.075)	(0.073)	(0.083)
	13134	12936	8444	12930	12930	12936
	0.964	0.964	0.964	0.965	0.965	0.968
	Panel C: Outcome is log(Ag. TFP, Pre Weights)					
Embrapa Exposure	0.457***	0.544***	0.592***	0.552***	0.489***	0.476***
	(0.072)	(0.075)	(0.099)	(0.074)	(0.074)	(0.079)
	12237	12044	7888	12041	12041	12044
	0.984	0.984	0.985	0.984	0.985	0.986
	Panel D: Outcome is log(Ag. TFP, Post Weights)					
Embrapa Exposure	0.268***	0.205**	0.125	0.218***	0.183**	0.260***
	(0.077)	(0.081)	(0.108)	(0.079)	(0.080)	(0.082)
	12237	12044	7888	12041	12041	12044
	0.984	0.984	0.985	0.984	0.985	0.986
Municipality FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
log(Distance to Embrapa) x Round FE	-	Y	Y	Y	Y	Y
Drop if $< 100km$ from Embrapa	-	-	Y	-	-	-
Drop if neighbor to Embrapa	-	-	Y	-	-	-
log(Initial Prod.) x Round FE	-	-	-	Y	Y	-
log(Initial Pop) x Round FE	-	-	-	-	Y	-
State x Round FE	-	-	-	-	-	Y

Notes: The unit of observation is a municipality-census-round pair, where municipalities are harmonized to minimal consistent border units (IBGE, 2011). The regression model is equation (7). The outcome variables are: log of average yields of major crops weighted by 1970 prices (Panel A; see main text for details); log of agricultural land value per farm area (Panel B); and log of total factor productivity, based on farm area, labor use, intermediates use, and mechanical inputs (Panels C and D). For calculating agricultural TFP, we use a constant-returns-to-scale, Cobb-Douglas production function with weights reported in Fuglie (2015) (Table A.2), based on calculations in the Brazilian agricultural census corresponding to the 1960s (Panel C) and 2010s (Panel D). The control variables included are: logarithm of distance (in km.) to the nearest Embrapa center times census-round fixed effects; log of production value per farm area in 1970 interacted with census-round fixed effects; log of population in 1970 interacted with census-round fixed effects; and state by census-round fixed effects. In column 3, we drop municipalities that are ever less than 100 km from an Embrapa center or neighbor a municipality with an Embrapa center. Standard errors are clustered at the municipality level.

Table C.5: Embrapa Exposure Increases Productivity Controlling for Expected Value of Exposure Treatment

	(1)	(2)	(3)	(4)	(5)	(6)
Embrapa Exposure	0.585***	0.683***	1.063***	0.726***	0.713***	0.483***
	(0.152)	(0.165)	(0.246)	(0.166)	(0.166)	(0.163)
	18386	18109	11821	18101	18101	18109
	0.954	0.954	0.945	0.955	0.955	0.976
Municipality FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
log(Distance to Embrapa) x Round FE	-	Y	Y	Y	Y	Y
Drop if $< 100km$ from Embrapa	-	-	Y	-	-	-
Drop if neighbor to Embrapa	-	-	Y	-	-	-
log(Initial Prod.) x Round FE	-	-	-	Y	Y	-
log(Initial Pop.) x Round FE	-	-	-	-	Y	-
State x Round FE	-	-	-	-	-	Y
Expected Treatment	Y	Y	Y	Y	Y	Y

Notes: The unit of observation is a municipality-census-round pair, where municipalities are harmonized to minimal consistent border units (IBGE, 2011). The regression model is equation (7). All specifications include a control for the expected value of the treatment variable across simulations that fix the locations of Embrapa's centers but vary their timing, following the logic of Borusyak and Hull (2023). The outcome variable is the log of production value per farm area. The control variables included are: distance (in km.) to the nearest Embrapa center times census-round fixed effects; log of production value per farm area in 1970 interacted with census-round fixed effects; log of population in 1970 interacted with census-round fixed effects; and state by census-round fixed effects. In column 3, we drop municipalities that are ever less than 100 km from an Embrapa center or neighbor a municipality with an Embrapa center. Standard errors are clustered at the municipality level.

Table C.6: The Effects of Embrapa on Farm Size Inequality

	Outcome is:			
	Log of Avg. Farm Size		Farmland Gini Index	
	(1)	(2)	(3)	(4)
Embrapa Exposure	-0.090	0.038	-0.050***	-0.041***
	(0.059)	(0.067)	(0.010)	(0.011)
Observations	18399	18121	17565	17293
R^2	0.894	0.907	0.761	0.783
Municipality FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
log(Distance to Embrapa) x Round FE	-	Y	-	Y
State x Round FE	-	Y	-	Y

Notes: The unit of observation is a municipality-census-round pair, where municipalities are harmonized to minimal consistent border units (IBGE, 2011). The regression model is equation (7). The outcome variables are the logarithm of average farm size and the Gini index of the farm size distribution. Standard errors are clustered at the municipality level.

Table C.7: Embrapa Exposure Increases Agricultural Productivity

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Conley Standard Errors					
Embrapa Exposure	0.730***	0.825***	0.985***	0.844***	0.819***	0.599***
	(0.186)	(0.199)	(0.278)	(0.194)	(0.183)	(0.128)
Observations	18385	18108	11821	18101	18101	18108
	Panel B: State-Level Clustered Standard Errors					
Embrapa Exposure	0.730***	0.825***	0.985***	0.844***	0.819***	0.599***
	(0.195)	(0.205)	(0.305)	(0.197)	(0.196)	(0.135)
Observations	18386	18109	11821	18101	18101	18109
Municipality FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
log(Distance to Embrapa) x Round FE	-	Y	Y	Y	Y	Y
Drop if $< 100km$ from Embrapa	-	-	Y	-	-	-
Drop if neighbor to Embrapa	-	-	Y	-	-	-
log(Initial Prod.) x Round FE	-	-	-	Y	Y	-
log(Initial Pop.) x Round FE	-	-	-	-	Y	-
State x Round FE	-	-	-	-	-	Y

Notes: The unit of observation is a municipality-census-round pair, where municipalities are harmonized to minimal consistent border units (IBGE, 2011). The regression model is equation (7). The outcome variable is the log of production value per farm area. In Panel A, standard errors are computed using the spatial HAC estimator of Conley (1999), allowing for spatial correlation within 250 km and serial correlation over two time periods. In Panel B, the standard errors are clustered at the state level. The control variables included are: distance (in km.) to the nearest Embrapa center times census-round fixed effects; log of production value per farm area in 1970 interacted with census-round fixed effects; log of population in 1970 interacted with census-round fixed effects; and state by census-round fixed effects. In column 3, we drop municipalities that are ever less than 100 km from an Embrapa center or neighbor a municipality with an Embrapa center.

References

Batjes, N. H., Carvalho Ribeiro, E. D., van Oostrum, A. J. M., Leenaars, J. G. B., Hengl, T., and Mendes de Jesus, J. S. (2017). WoSIS: providing standardised soil profile data for the world. *Earth System Science Data*, 9(1):1–14.

Bazzi, S., Gaduh, A., Rothenberg, A. D., and Wong, M. (2016). Skill transferability, migration, and development: Evidence from population resettlement in Indonesia. *American Economic Review*, 106(9):2658–2698.

Borusyak, K. and Hull, P. (2023). Nonrandom exposure to exogenous shocks. *Econometrica*, 91(6):2155–2185.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1):1–45.

Fuglie, K. (2015). Accounting for growth in global agriculture. *Bio-based and Applied Economics*, 4(3):201–234.

IBGE, D. d. G. (2011). Evolução da divisão territorial do Brasil 1872-2010. *Rio de Janeiro: IBGE*.

Klein, H. S. and Luna, F. V. (2018). *Feeding the World*. Cambridge University Press.

Matsuura, K. and National Center for Atmospheric Research Staff (2023). The Climate Data Guide: Global (land) precipitation and temperature: Willmott & Matsuura, University of Delaware. https://climatedataguide.ucar.edu/climate-data/global-land-precipitation-and-temperature-willmott-matsuura-university-delaware. Last modified 2023-11-27. Retrieved on 2025-08-26.

Moscona, J. and Sastry, K. A. (2025). Inappropriate technology: Evidence from global agriculture. Working Paper 33500, National Bureau of Economic Research.

Nunn, N. and Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. *Review of Economics and Statistics*, 94(1):20–36.

Riley, S. J., DeGloria, S. D., and Elliot, R. (1999). Index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences*, 5(1-4):23–27.