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

March 8, 2026

Abstract

Can public R&D in developing countries raise productivity? We study how Brazil's Embrapa—a public research corporation founded in 1973 to develop locally-relevant agricultural technologies—shaped agricultural development. Researcher-level microdata reveal that Embrapa shifted innovation toward staple crops and Brazil-specific ecology, with no decline in researcher productivity. Exploiting Embrapa's staggered expansion alongside municipality-level variation in the ecological suitability of Embrapa's innovation, we find significant increases in agricultural productivity, concentrated in targeted staple crops. Our estimates imply that Embrapa raised aggregate agricultural productivity by 110% with a benefit–cost ratio of 17, driven by the applicability of Embrapa's innovation across Brazil.

*Inter-American Development Bank (IADB). Email: arielk@iadb.org.
†Massachusetts Institute of Technology (MIT). Email: moscona@mit.edu.
‡University of Notre Dame. Email: hpelleg3@nd.edu.
§Princeton University. Email: ksastry@princeton.edu.
Acknowledgments: We thank Daron Acemoglu, Chris Barrett (Discussant), Francisco Costa, Bruno Ferman, Martin Fiszbein, Thomas Fujiwara, Jon Gruber, Steven Helfand, Shawn Kantor (Discussant), David Lagakos, Bob Margo, Mushfiq Mobarak (Discussant), Ezra Oberfield, Richard Rogerson, Florian Trouvain, Chris Udry, and Leonard Wantchekon, as well as seminar participants at Berkeley ARE, Cornell, EESP-FGV, INSEAD, Northwestern, Oxford, Princeton, Sciences Po, Stockholm University, the TSE, Notre Dame, the University of Connecticut, the University of Copenhagen, Wharton, DePaul, the University of Oklahoma, LSE, IFPRI, IADB, the NBER Development Economics Workshop, the NBER Place-Based Policies and Entrepreneurship Workshop, the RIDGE Forum, the LACEA Annual Meeting, and the SED for helpful comments. We are also grateful to several researchers from Embrapa, including Mateus Batistella, Luiz Antonio de Araujo Cruz, Ana Beatriz Juca de Queiroz Fiuza, Geraldo Martha, and Ramon Bicudo, for many helpful discussions. *Disclaimer:* The views expressed in this article do not necessarily represent those of the IADB.

1 Introduction

Can public R&D investment in developing countries drive productivity growth? One common perspective is that innovation takes place exclusively in a small set of frontier nations and countries outside of this set grow by adopting foreign technology (Acemoglu et al., 2006). Underlying this view is a presumption that most countries have a low return on R&D investment but nonetheless a high return to technology upgrading (Parente and Prescott, 1994; Barro and Sala-i-Martin, 1997). By implication, investing in local R&D may be an ineffective tool to encourage growth, especially compared to removing barriers to technology adoption (e.g., Parente and Prescott, 2002). This perspective motivates the vast majority of research and policy intervention in developing countries (see discussion by Suri and Udry, 2022; Verhoogen, 2023; Acemoglu et al., 2026).

Opportunities for growth through adoption alone, however, may be more limited in practice. Foreign technology may not raise productivity if it is mismatched with the local context (Basu and Weil, 1998; Acemoglu and Zilibotti, 2001). This phenomenon is particularly stark in agriculture (Griliches, 1957; Evenson and Gollin, 2003; Moscona and Sastry, 2025), where environmental differences can lead to acute mismatch between the inventions of the "frontier" and local conditions.[1] From this perspective, public R&D investment in locally appropriate technology may be an important policy tool to spur development, especially given the weak intellectual property rights, long time horizons, and coordination challenges that deter private investment.[2] Yet, there is little empirical evidence about the impact or cost-effectiveness of public R&D as a policy tool to drive productivity growth.

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 "carry out research activities with the objective of producing knowledge and technology for the agricultural development of the Country" (Brazil, 1972). Original research was Embrapa's explicit goal. 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).

[1] That said, technology mismatch is also prevalent in medicine (Kremer and Glennerster, 2004), manufacturing (Basu and Weil, 1998; Acemoglu and Zilibotti, 2001), and modern services (Lerner et al., 2024). We discuss the potential implications of our findings outside of agriculture in the last section of the paper.

[2] R&D investment is characterized by large spillovers, especially in the form of nonrival and nonexcludable knowledge; long time horizons with uncertain payoffs (Jones, 2021); limited profit opportunities under weak intellectual property protection, as is common in low- and middle-income countries (Diwan and Rodrik, 1991); and high potential returns to scale and coordination (Jones, 2021; Gross and Roche, 2024).

Embrapa opened research centers throughout the country, including in areas with limited pre-existing R&D or agricultural production, and trained scientists to populate these regional hubs (Correa and Schmidt, 2014). Researchers at Embrapa's labs released a range of new technologies, from soil modification techniques adapted to the acidic soils of central Brazil to hundreds of new crop varieties, including the first that allowed for the production of soy in tropical latitudes (Monteiro et al., 2012). 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. But was public R&D investment truly responsible for this shift in the focus of innovation and rapid growth in agricultural productivity? And, if so, did these benefits justify the large investment costs?

We study the effect of public R&D investment on both agricultural innovation and productivity, highlighting the central role of appropriate technology for the returns to R&D. To do this, we obtain information on the size and geographic expansion of Embrapa, construct a novel dataset of the careers of all of Brazilian agricultural scientists, and draw on eight 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 almost entirely driven by the spread of research centers across the country, which unlocked the development of new technology suited to Brazil's diverse ecological zones. These findings suggest that targeted public R&D can help developing countries 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. 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 or recognition. From these data, we measure individual scientists' employment histories 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 to 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 differences between municipalities to proxy for agricultural technology mismatch.

Results: Agricultural Research. In the first part of our analysis, we investigate 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 had any effect on the aggregate direction of innovation. 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 on those topics, with no evidence of anticipation effects. This includes research by scientists not employed by Embrapa. Thus, Embrapa did not "crowd out" other research, for example in the private sector or universities—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 areas with low pre-existing human capital, research capacity, and researcher agglomeration. We 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 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 regions, even when we restrict attention to articles in high-impact, internationally recognized journals.

Results: Agricultural Productivity. Embrapa's effect on the rate and direction of *research* does not guarantee that it meaningfully affected agricultural *productivity*, the ultimate object of interest. For example, new research may or may not translate into the development of productive technologies, and once developed, new technologies may or may not be widely adopted. For these reasons, we next investigate Embrapa's effect on agricultural productivity and its underlying mechanisms. 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 sources 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*: Embrapa labs open over time as it expands to new regions. 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 Embrapa expanded.

We then study the effect of Embrapa exposure on local agricultural productivity, measured in the Census of Agriculture (1960-2017). Our empirical specification includes place and time fixed effects that absorb all fixed cross-sectional differences (e.g., direct effects of geography) and aggregate time trends that may spuriously coincide with Embrapa's expansion or other national policies. The identification assumption is that the founding of each new lab was orthogonal to agricultural productivity trends in ecologically similar compared to ecologically distant municipalities. Our approach does *not* require that the location or timing of lab openings is unrelated to local economic conditions or policies.

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 dropping municipalities that are close to any center, thereby exploiting differences in *ecological* similarity to Embrapa's labs among municipalities that are all *physically distant* from any lab. 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 measure of total factor productivity that also takes into account the use of labor and intermediate inputs.

To bolster our identification approach, we exploit additional variation across crops coming from the fact that Embrapa's R&D was focused on certain crops and not others. We find that exposure to Embrapa has a large positive effect on the crops that were the focus of its R&D and zero effect on other crops. We use this specification to estimate a more demanding "triple difference" model with location-by-time fixed effects, which absorb any location-specific trends that might bias our baseline estimates, and we recover a large and quantitatively similar effect of Embrapa's R&D on the productivity of targeted crops.

We also conduct falsification tests with placebo treatments constructed from alternative plausible expansion paths of Embrapa that meet the institution's selection criteria. If our results were spuriously picking up productivity growth in remote areas of Brazil that were inevitably exposed to the expansion of this national program, for example, we would expect to find large effects even of placebo treatments. Instead, our estimates based on the true expansion path are consistently in the far right (<1%) tail of the placebo estimates.

Turning to dynamics, we find no evidence of anticipation (i.e., no pre-trends). Instead, productivity effects emerge in the same decade as a change in Embrapa exposure and accumulate over time: the long-run effect on local productivity is 30-40% larger than the within-decade effect. Consistent with time lags in technology diffusion, the productivity effect of Embrapa is limited during its first decade but picks up and grows thereafter.

Finally, we investigate how these effects on agricultural production were mediated by input adoption, technology diffusion, and land use. First, we show that Embrapa exposure leads to higher expenditure on intermediate inputs like seeds, fertilizers, and chemical defenses, the main focus of Embrapa's technology development. Cropland expands, but at a slower rate than production, while pasture land declines, consistent with new technology inducing reallocation from lower-value pasture to crops. Second, we provide direct evidence of the adoption of Embrapa's novel seed varieties, documenting that ecological similarity to Embrapa's labs is a strong predictor of seed variety diffusion. Third, we find no evidence of changes in farm size or greater inequality, indicating that land consolidation was not a driving mechanism or side effect of productivity growth.

The Returns to R&D. In a final section, we 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 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 Brazilian public sector research is held at its initial 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 Embrapa's cost was high—about 1% of Brazil's total agricultural GDP at its peak, comparable to that of the USDA (Correa and Schmidt, 2014)—its benefits were considerably larger. One feature of our approach is that, unlike most existing studies of agricultural R&D, our analysis captures all of Embrapa's activity, including investments in failed projects. Thus, our estimates capture the returns to this large-scale program as a whole, rather than the effect of individual technological successes evaluated *ex post*.

We close our analysis by studying the extent to which Embrapa's returns were shaped by its scale versus its geographic scope. 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. If Embrapa had invested its full budget only in Rio de Janeiro, the epicenter of agricultural research prior to the 1970s, the benefits would have declined by over two thirds, implying a benefit-cost ratio of under five that is statistically indistinguishable from zero. Rio is not unique—the average return from investing the full budget into each of Embrapa's centers is almost as low. These findings highlight the importance of Embrapa's ecological scope and the resulting decline in technology mismatch across Brazil.

Related Literature. This paper relates to several strands of existing literature. First, we contribute to the literature on the global distribution and direction of innovation. Existing work on technology mismatch (or "appropriate technology") has argued that innovation is geographically concentrated, endogenously directed toward the conditions of high-income innovation hubs, and as a result, can have limited benefits elsewhere in the world (Stewart, 1978; Basu and Weil, 1998; Acemoglu and Zilibotti, 2001; Lerner et al., 2024; Moscona and Sastry, 2025). Policy analysis in this literature has focused on proposals to expand intellectual property protection in developing countries (e.g., Diwan and Rodrik, 1991) or to design new instruments to channel capital from rich-country donors toward global priorities (e.g., Kremer and Glennerster, 2004; Kremer et al., 2020). By contrast, we study a very different policy—public R&D that aims to generate home-grown innovation—and empirically evaluate its impact and cost-effectiveness at scale. We find that an important mechanism for the macroeconomic effect of Embrapa is the redirection of innovation toward local needs. Our results are, to our knowledge, the first direct empirical evidence of how a specific policy change—here, a "big push" in public R&D—can

allow a developing country to escape the "technology mismatch trap."[3]

Second, we build on existing work on the consequences of public R&D (see Jones, 2021, for a review). These studies have focused on episodes of large-scale US R&D investments, including the Space Race and Second World War (Kantor and Whalley, 2023; Gross and Sampat, 2023, 2025), and on the effects of US government grants on private-sector innovation (e.g., Azoulay et al., 2019). Most related, Kantor and Whalley (2019) study how the effects of US agricultural experiment stations on yields decline with distance. We study the effect of public R&D in a developing country, where both costs and benefits could be very different. Moreover, we estimate the *returns* to public R&D—an object that is crucial for policy but notoriously difficult to estimate in any context (Jones and Summers, 2022)—using a new strategy that leverages panel data on productivity and heterogeneous exposure to R&D investment. Compared to existing approaches, our strategy circumvents identification challenges associated with interpreting time-series trends in R&D and productivity (e.g., Fieldhouse and Mertens, 2023) and does not require enumerating the set of relevant "induced technologies" or their social values (e.g., Griliches, 1958).

Third, we contribute to the literature on determinants of global agricultural productivity gaps (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 importance 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., Foster and Rosenzweig, 1996; Evenson and Gollin, 2003; 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).

Fourth, we relate to a rich body of work studying Embrapa itself. Most of these studies are qualitative in nature, while some focus on quantitative analyses of specific technologies and assess their impact using data from experimental trials (e.g., Gasques et al., 2009; Pardey et al., 2006; de Andrade Alves et al., 2013; Correa and Schmidt, 2014; Klein and Luna, 2018). We move beyond case studies by combining new methods and data to evaluate the impact of Embrapa as a whole and study its impact on productivity growth in equilibrium. By studying the effect of all of Embrapa's activity—including investment in programs that fail or technologies that never make it to market—we avoid potential bias

[3]Thus, our analysis also relates to recent studies of sector-specific industrial policies in developing countries (e.g., Juhász et al., 2024). There has been less discussion of the potential of "big push" innovation policy.

from "picking technological winners" in our analysis of the returns to R&D. Moreover, our modeling framework makes it possible to study mechanisms and highlights how Embrapa's geographic design was central to its productivity effects.

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.

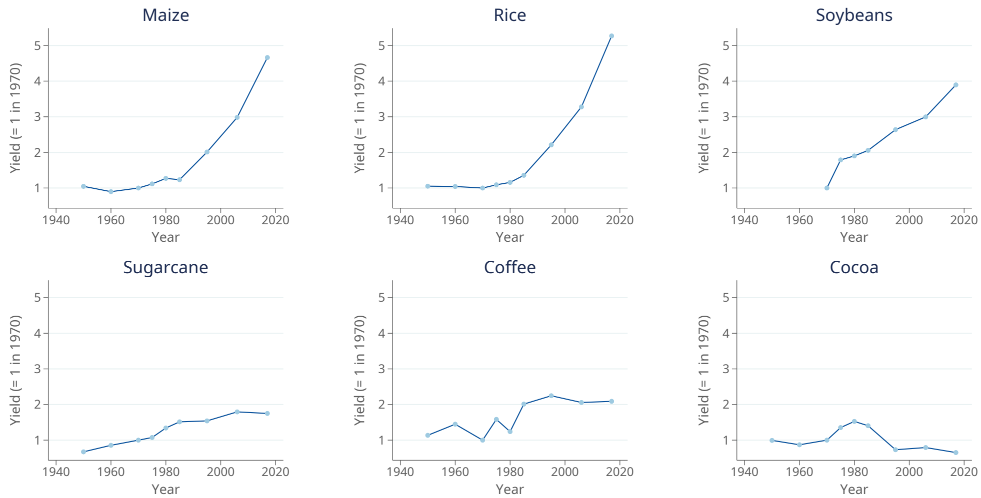
2.1 Brazilian Agriculture Before the 1970s

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) were 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 1960s, pressures on Brazil's food production sector intensified due to rapid urbanization and population growth. Already a net importer of food, Brazil became increasingly reliant on foreign donations and food aid (Martha Jr et al., 2012). There was a growing realization among policymakers that expanding domestic food production could help stave off food insecurity and political pressure from price-sensitive urban constituents.

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 its major breeding centers, it did not lead to large benefits for Brazil. A potential explanation was that Brazil had a very different geography, ecology, and farming practices. For example, new crop varieties were ill-suited to Brazil's acidic soils (Vilaça de Vasconcelos et al., 2022) and intended for use alongside substantial irrigation, while agriculture in Brazil was largely rainfed (Cabral et al., 2022). Brazil's own agricultural research was limited and concentrated on wealthier states and a few export-oriented cash crops, including coffee, sugarcane, and cotton (Embrapa, 2006). A large share of investment was devoted to importing and testing foreign technology, often with limited success (Martha Jr et al., 2012).

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.

2.2 Embrapa: Origins and Design

In 1973, against this backdrop—and amid a commodity price shock that put its dependence on food imports into stark relief—Brazil's national government established the Empresa Brasileira de Pesquisa Agropecuária (Embrapa), a public corporation devoted to agricultural R&D.[4]

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.

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

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 as a share of agricultural GDP was comparable to that of all public agricultural R&D in many high-

[4]More precisely, the Law No. 5851 of December 7th of 1972 authorized the creation of a federal public company dedicated to agricultural research in Brazil. Based on this law, the government formally constituted the company a few months later in April 26 of 1973, which is recognized as the founding date.

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 $1.15 billion USD, roughly thirteen times the value of Brazilian public agricultural R&D investment at the time of its founding. For comparison, the USDA R&D budget in 2010 was $2.3 billion USD (Sargent et al., 2009). Embrapa is similarly large measured by its employment of agricultural researchers. By 2010, Embrapa employed 2,300 agricultural scientists, nearly half the total number in Brazil. This is again comparable to the US, where the Agricultural Research Service, the USDA's research arm, employs roughly two thousand scientists and post-doctoral researchers (USDA ARS, 2024).

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). Consistent with this approach, 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 (Correa and Schmidt, 2014; Embrapa, 2006). As one important example of these efforts, Embrapa directly participates 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 across Brazil in order to tailor its investments to their specific constraints and production.

Geographic Scope and Specialization. Finally, Embrapa was designed from the start to study Brazil's diverse ecological conditions by opening research labs in all regions of the country (Alves, 1988). The geographic expansion of Embrapa over time was driven by several guiding principles and feasibility constraints.

First, a central goal was to expand research centers over time to 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 and a lack of science and technology suited to Brazil's diverse geographic features. Figure 2 highlights this expansion over time and across space.

Second, beyond the general goal of geographic expansion, Embrapa identified a series of key priorities for agricultural research and established new centers to achieve these goals. One priority was increasing crop productivity in the Cerrado (see Figure 2), a two-million square kilometer tropical savanna in central Brazil whose highly acidic soils made farming with existing technology unproductive (Embrapa, 2006). 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 Cerrados was established in 1975 as one of the organization's first centers. A similar impulse motivated the early establishment of Embrapa Rondônia in the heart of the Amazon, more than a thousand kilometers from any pre-existing agricultural research station (Embrapa, 2025).

Third, the choice of new locations was constrained by where the national leadership thought it was possible to open a new research lab. Some states were deemed to not have sufficient capacity for the establishment of a research center and, within states, municipalities with some existing agricultural research and transportation infrastructure were chosen based on sets of "feasibility studies" conducted by the central administration (Embrapa, 1974). In a handful of cases, new centers subsumed smaller, regional agricultural experiment stations; in most cases, however, no such institution existed and centers were established based on assessments of nearby infrastructure and state capacity.

Our analysis of Embrapa's effect on agricultural productivity exploits the establishment of new centers over time (Section 5). However, as we noted in the Introduction and describe in greater detail in Section 5.1, we never exploit differences across municipalities with and without an Embrapa research center. Instead, we exploit differences across municipalities in ecological similarity to municipalities with research labs, which shifts the potential usefulness of their innovation. Nevertheless, we use the information outlined above to construct a series of falsification exercises based on alternative plausible expansion paths for Embrapa to complement our main empirical design (see Section 5.3.2).

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

Embrapa became the main developer of agricultural technology in Brazil. Its researchers have developed more than 350 crop varieties and have 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. At the same time, not all of Embrapa's investments have been met with success and, as is the case with many government-led R&D programs, there is ongoing debate about the extent to which these failures imply that Embrapa's overall approach to R&D investment has limited returns. Below, we briefly highlight some of the most salient examples.

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 with toxic aluminum. Norman Borlaug, the father of the Green Revolution, claimed that "nobody thought these soils were ever going to be productive" during the 1970s (Rohter, 2007).

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

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 (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 received the World Food Prize for research in this area. Norman Borlaug himself averred: in the Cerrado, "Embrapa was able to put all the pieces together" (Rohter, 2007).

The Growth of Soy and Other Successes. One important engine of Brazil's agricultural growth has been the expansion of cultivation of soybeans, a crop for which Brazil now 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 and soybean varieties that were adapted to tropical latitudes.[5] 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 adapted to tropical latitudes and, in subsequent decades developed about 200 more. These were critical to the "first phase" of Brazil's soy revolution (Monteiro et al., 2012).

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 late 1990s and early 2000s (Bustos et al., 2016). Both of these phases are visible in Figure 1, which shows soybean yields increasing substantially in each decade starting from 1970.

There are many additional examples of successful new technologies developed by Embrapa. These include high-yielding rice varieties that have led to yield increases of nearly one percent per year (Magalhães Jr. and Stone, 2018); new "tropicalized" carrot varieties

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

that are widely adopted (Nehring, 2024); cotton varieties adapted to Mato Grosso, thousands of kilometers from the previous centers of production on the coast (Klein and Luna, 2023); and, more recently, transgenic crop traits, culminating in the release of a genetically-modified maize variety that is resistant to the fall armyworm, one of the most damaging pests in Brazil (Embrapa and Helix, 2022). These case studies, however, may not reflect the overall effect of Embrapa or imply high overall returns on investment.

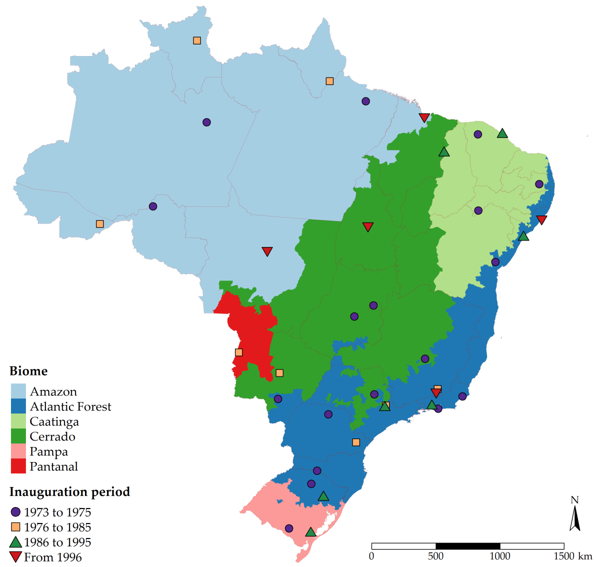
Embrapa's Failures and the Road Ahead. In addition to the marquee cases of Embrapa's success, there are many examples of failed projects and abandoned research directions. For example, Avila et al. (2005) report comparatively low returns for new eucalyptus varieties, as well as limited benefits to new techniques developed for pepper cultivation. Embrapa's long-standing program to develop improved cassava varieties has lagged due to a range of technological hurdles. None of these varieties have been commercialized to date (Embrapa, 2018b). There are several examples of improved grass varieties developed for the Cerrado becoming invasive and harming crop production in nearby areas (Damasceno et al., 2018). There are even continued challenges associated with maintaining resistance against key soybean diseases, especially soybean rust (Embrapa, 2023). Within resistant cultivars, epidemics have been widespread (Godoy, 2019). Evaluating any specific investment or technology in isolation may fail to capture the overall effect of this decades-long, large-scale experiment in publicly funded R&D.

Our goal in the next section is to move beyond case studies and investigate whether government investment in R&D as a whole had a systematic effect on innovation and productivity growth in Brazil. Like other national research programs that span many decades and research projects—including the US's NIH or USDA—it is possible to list a large number of both *ex post* failures and *ex post* successes. This makes it challenging to learn about the average effect of Embrapa from case studies without taking into account the effect of the overall portfolio. Did Embrapa "lead" technological progress and productivity growth, or merely follow pre-existing trends? Did new technology development systematically 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 new 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

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 are sourced from Embrapa and IBGE.

ecology. We summarize these data sources below, relegating details to Appendix A.

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. Embrapa expanded over time, 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, 2022). Thus, these cost data provide a unique opportunity to perform a benefit-cost analysis, which we turn to in 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, collaboration, and 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.[6] This includes each individual's educational history (i.e., degrees, graduation years, and institutions); full employment history (i.e., institutions, employment years, and position titles); and full publication history (i.e., titles, journals, and publication years).[7] 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 ever employed by Embrapa—and approximately 1.3 million publications. We also geo-locate all employers and institutions by municipality using ChatGPT (GPT-4o). In addition, we use keyword searches in the title of each article to identify all articles related to Brazil's major biomes, crop pests, and crops (see Appendix A.2).

Agricultural Production. Our main source of data on agricultural production is the Agricultural Census of Brazil. While the Census has been conducted every five or ten years since 1920, digital versions are available only for the 1995-6, 2006, and 2017 rounds. We therefore digitize all rounds since 1960 (1960, 1970, 1975, 1980, 1985) from scanned files published by the Brazilian Institute of Geography and Statistics (IBGE). Using these data, we construct a municipality-level panel dataset with information on agricultural output, land use, land values, input use, labor use, and farm size (see Appendix Table C.1).

We supplement the agricultural censuses with data from the Municipal Agricultural Production (PAM) survey from 1974 to the 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 broader set of crops than the census.

Finally, since Brazil's municipal borders shifted during the sample period, to make

[6] 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.

[7] 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).

geographic units consistent over time we follow Brazil's statistical agencies and link all data to minimal consistent border units (*Áreas Mínimas Comparáveis*, AMCs) (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 classification from Brazil's main statistical agency (IBGE, 2024).[8] Figure 2 maps these biomes, with Embrapa's research centers superimposed. This variation in ecology, much of which is unique to Brazil, was a driving motivation behind the design of 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, these data allow us study how the expansion of Embrapa to new ecological zones shifted research focus.

To quantify ecological differences across Brazilian municipalities in greater detail, we use geo-spatial data on the distribution of nine agro-climatic characteristics that are critical to agricultural production, including temperature, precipitation, elevation, ruggedness, growing season length, soil acidity, soil clay content, soil silt content, and soil coarse fragment content (see Appendix Table A.2).[9] We use these values to construct a measure of ecological similarity between municipalities and the locations of Embrapa's research labs, which we treat as a shifter of the local suitability of each lab's R&D (see Section 5).

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 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-classified database of all agricultural

[8]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.

[9]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.

research publications. Our main regression equation is:

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

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. If employment by Embrapa made agricultural researchers more likely to study Brazil's ecology, as represented by a focus on local biomes or local pest and pathogen threats, we would expect that $\beta > 0$.

One reason that researchers at Embrapa may be more likely to study Brazil's ecological conditions is that Embrapa set up new centers in environments that are characteristic of Brazil (e.g., the Amazon rainforest or the Cerrado). 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 tailored to the local environment (see Section 2.2). Breeding and experimentation with new varieties that are productive in a particular ecosystem, for example, is often only possible in that ecosystem itself. 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 with location fixed effects.

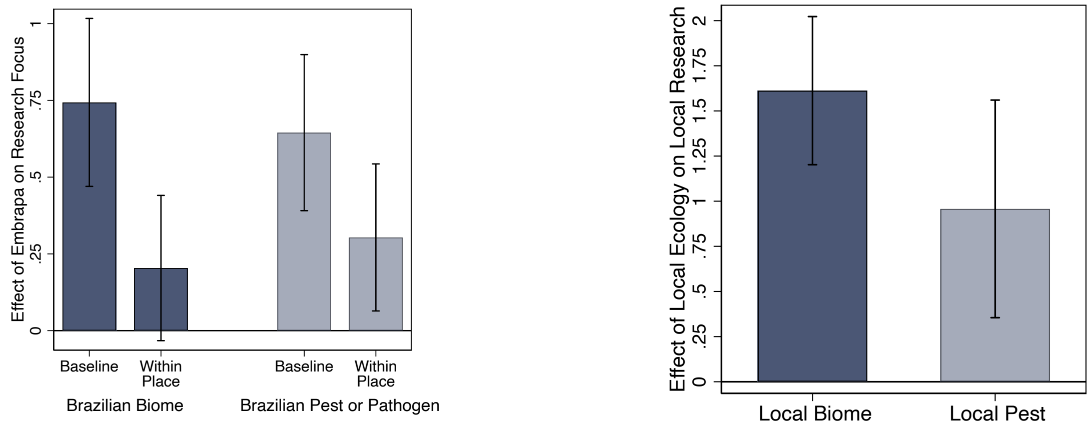
In Figure 3a, we report our estimates of β normalized by the mean of the outcome variable. Embrapa researchers write significantly more articles about Brazilian biomes and crop-affecting pests. The effects are equal to 74% and 64% of the sample mean, respectively (first and third bars). Moreover, the 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. These findings are larger if we weight research articles by the impact factor of their journal, suggesting that the findings are driven by high-quality research (see Figure C.2).

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

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

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

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.

major pests and pathogens. For biomes, $\mathbb{I}\{\text{Local}\}_{ik}$ is an indicator that equals one if municipality i is located within biome k. For 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 to find that $\xi > 0$.

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.

4.2 The Direction of Research: Crops

In addition to building knowledge about under-studied ecological zones, Embrapa also prioritized staple crops that were most relevant for food security (Martha Jr et al., 2012). This was a departure from previous agricultural research in Brazil, much of which had historically focused on exported cash crops like coffee and sugarcane.

We therefore study the extent to which Embrapa 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). We find that researchers employed by Embrapa are substantially more likely to study these staple crops (Figure C.4). Unlike the equivalent for ecological conditions, this effect is not driven by location, since estimates are similar after including municipality fixed effects (second column). This indicates that Embrapa's focus on staple crops was national, not confined to specific regions. 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 will motivate a refinement of our main identification strategy in Section 5. To more precisely isolate the effect of Embrapa's R&D on agricultural productivity, we will exploit variation in exposure to Embrapa not only across locations, based on their differential ecological similarity to Embrapa's research centers, but also across crops, which received different levels of attention from Embrapa's researchers in all parts of the country.

4.3 Aggregate Effects and Crowd-Out

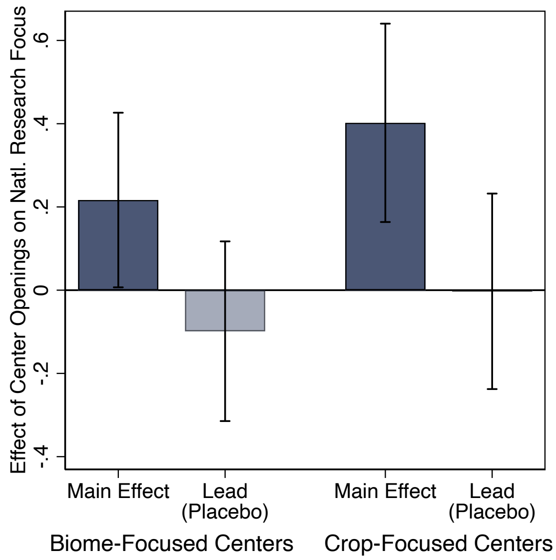
So far, we have focused on how Embrapa shaped the direction of research at the individual level. However, the fact that Embrapa shifted individuals' research focus does not necessarily imply that it shifted the aggregate direction of innovation across topics. One possibility, for example, is that researchers 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. When k indexes biomes, Centers_{kt} is the number of Embrapa centers as of year t that are located in biome k. Finding that $\gamma > 0$ would imply that Embrapa shifted the overall direction of innovation.

Figure 4: 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. In both panels, standard errors are heteroskedasticity robust and 95% confidence intervals are reported.

Figure 4 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). 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).[10] The direction of research shifts only after new Embrapa centers open. We finally note that this aggregate effect is driven by an increase both in research conducted by Embrapa affiliates themselves *and* research conducted by other, non-Embrapa scientists (see Appendix Figure C.3).[11] There is no evidence of crowd out.

Together, these results indicate that if anything, Embrapa-sponsored research had positive spillovers on other researchers, for example in the private sector or in universities (i.e., "crowd in"). The finding of potential crowd in of public R&D is consistent with the findings of Azoulay et al. (2019) regarding biomedical research in the US and Moretti et al. (2025) regarding defense R&D in OECD countries.

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

[11] 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.$= 0.0172$).

4.4 Research 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 inefficiencies in large-scale government programs like Embrapa, and because, as part of its strategy to cover topics that had not been the focus of prior innovation, Embrapa expanded to regions with limited pre-existing research infrastructure and therefore potentially lower research productivity.

To study the impact of Embrapa on researcher productivity, we use a movers-based empirical design. Our baseline specification is:

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

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 $\gamma_{a(r,t)}$ is a set of tenure fixed effects.[12]

We then 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 a "hub" (i.e., among the top ten in terms of either total agricultural research output or its share of college graduates) or not.[13] 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 (Correa and Schmidt, 2014). 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 this effect varies across regions (columns 3-6). Intuitively, we find that moving away from a research hub is associated with a decline in research productivity, captured by the coefficient on the "Non Hub" indicator. However, working

[12] We calculate researcher tenure as the current year minus the first recorded year of any employment spell.
[13] The results are very similar if we choose alternative cut-offs (e.g., define the hubs as the top twenty).

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 Non Hub			0.203***	0.189***	0.121***	0.119***
			(0.072)	(0.067)	(0.028)	(0.026)
Embrapa x Hub			-0.019	0.048	0.043*	0.061**
			(0.063)	(0.069)	(0.024)	(0.026)
Non Hub			-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 whether the municipality is or is not a "Hub." In particular, using either previous agricultural research (columns 3-4) or the share of college graduates (columns 5-6), we define "Hubs" as the top 10 municipalities and "Non Hubs" as the complement. 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.

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 earlier 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 agricultural research (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 enabled 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 or use alternative parameterizations of the dependent variable (Appendix Table C.3). Thus, the findings are not driven by insubstantial publications or by the details of our regression specification.

Taking Stock. Together, these estimates suggest that Embrapa shifted the focus and geography of agricultural research without sacrificing research productivity. The fact that Embrapa shifted the rate and direction of *research* output, however, does not guarantee that it meaningfully affected *agricultural productivity*, either because new research output did not translate into the development of new technologies, or because new technologies were developed but were not widely adopted by farmers. For these reasons, the next section turns to directly measuring the effect of Embrapa on downstream agricultural outcomes.

5 Results: Embrapa and Agricultural Productivity

We next study Embrapa's effects on agricultural productivity and technology adoption. First, we describe our empirical design that exploits the staggered expansion of Embrapa alongside regional heterogeneity in the suitability of its innovation. Second, we present our main result that Embrapa significantly increased agricultural productivity, and describe a range of additional tests to validate our estimates. Third, we show direct evidence of the adoption of improved inputs and the diffusion of Embrapa's new crop varieties.

5.1 Empirical Strategy

Measuring the returns to a large-scale R&D investment on productivity is challenging due to fundamental difficulties in 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 (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 then develop tailored models to estimate their effects (e.g., Pardey et al., 2006). This approach has three independently challenging steps. The first is identifying the relevant set of technologies, which can lead to "picking winners" and thereby overstating the effects of R&D (Jones and Summers, 2022).[14] The second is estimating the *private* value of these technologies, which requires myriad intermediate assumptions (see Azoulay et al., 2019, for a discussion). The third

[14]In the case of Embrapa, previous evaluations have largely focused on the returns to specific crop varieties in experimental settings (e.g., Pardey et al., 2006, on upland rice, edible beans, and soybeans). This analysis faces all three challenges outlined above, in addition to the facts that a substantial share of Embrapa's research (67%, in our data) is not directly related to the development of new seeds at all.

is estimating the social value, which may differ substantially from the private value due to knowledge spillovers, crowd-out or crowd-in of private investment, and the split of surplus between innovators and technology users (Griliches, 1979; Nordhaus, 2004).

In light of these challenges, we develop a different approach to estimate the effect of Embrapa on productivity at the *regional* level by measuring each municipality's changing exposure to Embrapa's research laboratories. Using regional data allows us to account for indirect effects while also sweeping out other confounding forces and policy changes. 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.

Our approach is motivated by three observations. First, agricultural research in Brazil is heavily focused on targeting local ecosystem characteristics (see Section 4). Second, narrative accounts of Embrapa put special emphasis on the development and diffusion of technologies specific to particular ecological regions of Brazil (Correa and Schmidt, 2014). Third, existing work has documented that ecological similarity between locations of invention and of production is a strong predictor for the diffusion and productivity effects of agricultural technology (e.g., Griliches, 1957; Moscona and Sastry, 2025).

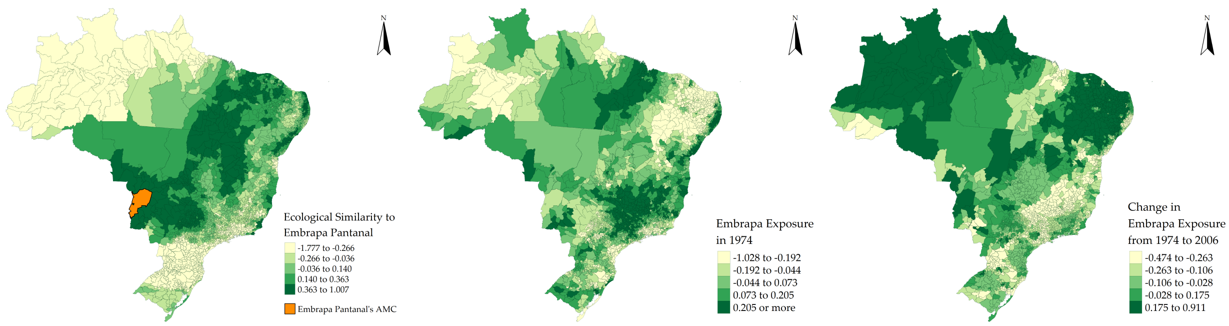
Our specific measure of ecological similarity, based on prior work by Bazzi et al. (2016) and Moscona and Sastry (2025), is an index that aggregates similarity in climate, topography, and soil characteristics. For each pair of municipalities i and j, we compute:

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

the sum of absolute deviations across nine ecological characteristics 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. Below, we will show that this

Figure 5: Embrapa Exposure Across Space and Time

(a) Similarity to Pantanal Lab **(b)** Embrapa Exposure, 1974 **(c)** Δ Embrapa Exposure, 74-06



Notes: Panel (a) displays ecological similarity, defined in equation (5), with respect to Embrapa Pantanal, whose location is colored in orange. Panel (b) shows Embrapa Exposure in 1974, defined in equation (6). Panel (c) shows the change in Embrapa Exposure from 1974 to 2006. In all panels, we linearly project out physical distance to the nearest Embrapa center.

measure of ecological similarity strongly predicts crop variety diffusion across municipalities, including municipalities that are geographically far from one another (Section 5.5.2).

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.[15] Cross-sectional variation in Embrapa Exposure$_{it}$ comes from the network of pairwise ecological similarity across municipalities, as captured by variation in Ecological Similarity$_{ij}$. Time-series variation comes from the fact that new centers are founded over time, shifting each municipality's most similar Embrapa center.

To illustrate the variation underlying this measure, Figure 5a plots each municipality's ecological similarity to the Embrapa center in the Pantanal, a biome in Brazil's southwest. The municipalities most ecologically similar to Embrapa Pantanal are in central Brazil, including some that are geographically close to the lab and others that are very far away.

[15]We explore the robustness of our main findings to alternative parameterizations of Embrapa Exposure, including measures that include the top two or three most ecologically similar centers (rather than only the maximum) as well as measures that weight these centers by their operating budgets (see Table C.8). In our quantification and cost-benefit analysis, we introduce and estimate a more general form for exposure to Embrapa's research that incorporates potential benefits from all centers (see Section 6).

More generally, as new centers opened over time, the parts of the country that were positioned to benefit from Embrapa research shifted dramatically, since each center was ecologically similar to different parts of the country. Figure 5b plots Embrapa Exposure$_{it}$ as of the end of 1974, when only the first handful of research centers had been opened, while Figure 5c 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 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 or clustering 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, crop yields, and measures of total factor productivity—and decompose the effect on productivity into changes in input use, land use, and productivity.

The central identification assumption is that when a new Embrapa center is opened, ecologically similar municipalities are on similar trends to ecologically distant ones. Our identification approach allows for the presence of simultaneous policy changes or other government investments (e.g., the construction of roads), so long as trends in those investments are not systematically associated with trends in ecological similarity to Embrapa's labs. Moreover, policy changes at the national (e.g., concurrent trade policy changes) or state level (e.g., agricultural support programs) can be absorbed by time and state-by-time fixed effects, respectively. 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 (see Section 2.2) and because physical proximity to a research center could have benefits beyond those of new innovation (e.g., extension services). We introduce several strategies to address this, before turning to an additional identification approach that exploits variation in Embrapa's innovation across crops and within municipalities (Section 5.3.1) which, along with a set of falsification exercises (Section 5.3.2), makes it possible to account for any omitted municipality-level trends.

Table 2: Embrapa Exposure Increases Agricultural Productivity

	(1)	(2)	(3)	(4)	(5)
			A. Baseline results		
Embrapa Exposure	0.730***	0.825***	0.985***	0.799***	0.599***
	(0.080)	(0.084)	(0.120)	(0.082)	(0.086)
Observations	18386	18109	11821	18101	18109
R^2	0.954	0.954	0.945	0.955	0.976
			B. Weighted by 1970 Agricultural Area		
Embrapa Exposure	0.758**	1.266***	1.141***	1.174***	0.824***
	(0.302)	(0.244)	(0.196)	(0.227)	(0.221)
Observations	18372	18101	11818	18101	18101
R^2	0.964	0.965	0.961	0.966	0.980
Municipality FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
log(Distance to Embrapa) x Round FE	-	Y	Y	Y	Y
Drop if $< 100km$ from Embrapa	-	-	Y	-	-
Drop if neighbor to Embrapa	-	-	Y	-	-
log(Initial Prod.) x Round FE	-	-	-	Y	-
log(Initial Pop.) x Round FE	-	-	-	Y	-
log(Initial Roads) 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: log of distance 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; proximity to roads 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 Main Estimates: 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. One potential concern is that the specific municipalities where Embrapa chooses to open its labs are on different trends. For example, while not part of its explicit design (see Section 2.2), Embrapa may open new labs in municipalities that are politically connected to the central government, and these political connections could have independent effects on productivity trends. To the extent that our measure of Embrapa exposure is correlated with physical distance to one of these labs, the endogeneity of lab construction decisions could bias our results. We investigate this possibility directly by exploiting the fact that there are municipalities with both high and low levels of exposure to each center in geographically distant parts of the county (see Figure 5).

Our results are very similar after controlling for (the logarithm of) Euclidian distance to Embrapa centers interacted with census-round fixed effects, thus allowing for a flexibly time-varying effect of proximity (column 2). The results are also stable under a more conservative strategy of additionally restricting the sample to municipalities that neither border a municipality with an Embrapa research center nor are within 100 kilometers of one (column 3), thereby exploiting variation in ecological similarity only among municipalities that are physically distant from any Embrapa lab. Beyond its benefit as a validation of our identification approach, this finding demonstrates that our estimates are 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 infrastructure that might accompany local investments in Embrapa. Instead, the estimates are consistent with our interpretation that the results capture variation in the suitability of new technology.

The results are also similar controlling flexibly for municipality-level characteristics that might affect local productivity trends, including initial agricultural productivity, population, and proximity to major roads (column 4). These estimates suggest that the main result is not biased by mean reversion or other trends in local characteristics, including public investment. Finally, to further zoom in on the precise geographic variation spanned by our exposure variable, we find that effect of Embrapa is also similar after including state-by-year fixed effects (column 5), which absorb any differential trends in state-level research (e.g., at state universities), agricultural support mechanisms, regional policy variation, or local state capacity that might have affected lab siting decisions (see Section 2.2).

Finally, the results are not driven by our exact parameterization of Embrapa exposure. In particular, our findings are similar if we drop any component of the ecological similarity index in equation (5) (Appendix Figure C.5). The findings are also similar using alternative parameterizations of Embrapa exposure, including measures that incorporate potential technological benefits from multiple centers and/or scale these effects by each center's budget to model an "intensive margin" of investment (Appendix Table C.8).

Alternative Productivity Measures and TFP. We next show 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 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 technology development was capitalized into local land values.

The third and fourth are measures of local agricultural total factor productivity (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.[16] 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 our baseline outcome, the value of production per unit of land. That is, intensification 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 5.5.1.

5.3 Additional Identification Approaches

The previous section documented that exposure to Embrapa increased local agricultural productivity and presented a range of evidence consistent with a causal interpretation of the results. Next, we present two additional approaches to further rule out the possibility that our estimates capture municipality-level trends unrelated to Embrapa's innovation.

5.3.1 Exploiting Variation Across Crops

If Embrapa's technology development is the mechanism driving our main findings, then Embrapa exposure should have the largest effect on the crops that were the focus of its

[16] 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 $\alpha_N = 0.434, \alpha_L = 0.342, \alpha_K = 0.167$, and $\alpha_M = 0.057$, and the "2010s" calibration sets $\alpha_N = 0.083, \alpha_L = 0.373, \alpha_K = 0.214$, and $\alpha_M = 0.331$.

R&D and little or no effect on crops that were not. Our earlier finding (Section 4.2), consistent with narrative accounts (Martha Jr et al., 2012), was that Embrapa focused its research on specific staple crops: beans, cassava, maize, rice, soy, and wheat. It comparatively ignored traditional cash crops like sugarcane, coffee, and cocoa. We exploit this fact by studying the differential effect of Embrapa exposure on the production of different crops.

To do this, we compile data on the output of each crop in each municipality and year from the Municipal Agricultural Production (PAM) survey. We then study the effect of Embrapa exposure on crop-specific output, separately for crops that were the focus of Embrapa's R&D and for crops that were not by estimating

$$y_{ikt} = \beta_1 \cdot \text{Embrapa Exposure}_{it} \cdot \text{EC}_k + \beta_2 \cdot \text{Embrapa Exposure}_{it} \cdot \text{NEC}_k + \chi_{ik} + \chi_{tk} + \varepsilon_{ikt} \quad (8)$$

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

We find that β_1 is large and highly significant, while β_2 is close to zero and statistically insignificant (column 1 of Table 3). That is, exposure to Embrapa's new technology raises output for the crops Embrapa directed research *toward*, and not for the crops that Embrapa comparatively ignored. The fact that β_2 is not negative furthermore indicates that Embrapa did not crowd out productivity growth in areas that were not the focus of its technology development. Thus, the effect of Embrapa exposure on productivity is concentrated *only* in the crops that were the focus of its R&D investment.

We can further augment equation (8) by adding municipality-year fixed effects (χ_{it}):

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

This strategy allows us to estimate the disproportionate effect of Embrapa Exposure on the productivity of targeted crops, while fully absorbing *all* other location-level changes over time, such as local infrastructure investments or local development policies that could be spuriously correlated with our exposure measure. We find a very similar point estimate for β_1 in this more demanding regression specification (column 2).

Finally, using crop-specific outcomes, we can investigate whether our results are driven by soybeans, the crop that is perhaps most emblematic of Brazil's agricultural expan-

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

	Outcome is log Crop-Specific Output			
	(1)	(2)	(3)	(4)
Embrapa Exposure × EC	1.426***	1.410***	1.276***	1.279***
	(0.235)	(0.348)	(0.226)	(0.340)
Embrapa Exposure × NEC	0.059		0.059	
	(0.283)		(0.283)	
Observations	188619	188455	180797	180620
R^2	0.832	0.876	0.829	0.875
Crop-Year FE	Y	Y	Y	Y
Municipality-Crop FE	Y	Y	Y	Y
Municipality-Year FE	-	Y	-	Y
Drop Soybeans	-	-	Y	Y

Notes: Each column reports an estimate of equation (8) 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 drop soy from the sample. The outcome variable is the logarithm of crop-specific output. Standard errors are clustered by municipality and reported in parentheses.

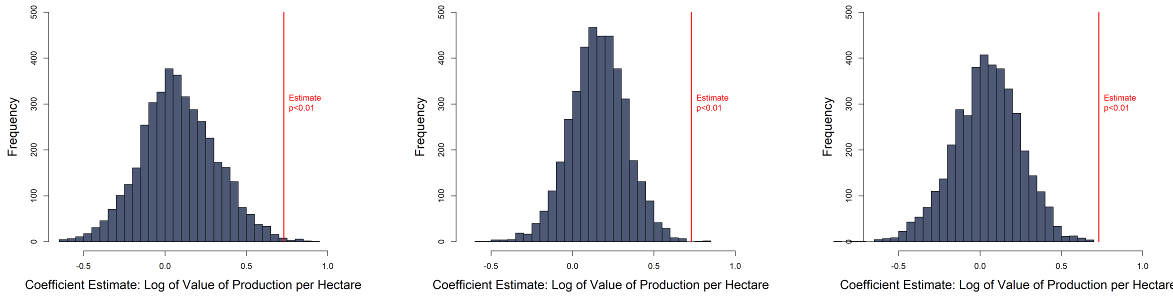
sion. Reassuringly, we find similar effects in a specification that fully excludes soybeans (columns 3 and 4), suggesting that the effect for soy is comparable to the effect for all other crops that were also the focus of R&D investment.

5.3.2 Falsification Tests

Next, we conduct a series of falsification tests to further support a causal interpretation of the findings. Our results never directly exploit variation in whether a municipality hosts an Embrapa lab or not; instead, we exploit only variation in ecological exposure to new technology induced by these siting decisions. That said, given the endogeneity of lab siting decisions, one may still be concerned that these induced changes could spuriously correlate with other trends. For example, if there happened to be productivity growth in parts of Brazil that are ecologically more remote, this may be picked up by our treatment variable given Embrapa's goal of expanding research to all of Brazil's ecological zones.

To address these issues, we conduct a series of falsification tests in which we randomize the timing and geography of the expansion of Embrapa. We then investigate whether exposure to these counterfactual expansion patterns has a similar effect on changes in productivity. If our main results truly capture the effect of exposure to Embrapa's technology development, we would expect our actual estimate to be in the far right tail of the effect of these placebo measures of Embrapa exposure, which are constructed from data that

Figure 6: Falsification Tests



(a) Random Opening Years **(b)** All Biomes **(c)** All Biomes and PS Sampling

Notes: Each sub-figure reports placebo draws from separate falsification exercises. In Panel (a), we randomize only the timing of center openings. In Panel (b), we simulate alternative Embrapa expansion patterns, randomizing both the locations of the 40 centers and their opening year, and requiring that there is at least one center in each biome for every draw. We further 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 (c), we repeat the exercise from Panel (b) and further draw counterfactual center locations with probability proportional to municipalities' estimated propensity to receive an Embrapa center, predicted using the (log of) population, (log of) distance to the nearest highway, and (log of) distance to the nearest state capital. For each counterfactual, we 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.

resemble but do not exactly replicate Embrapa's actual expansion pattern.

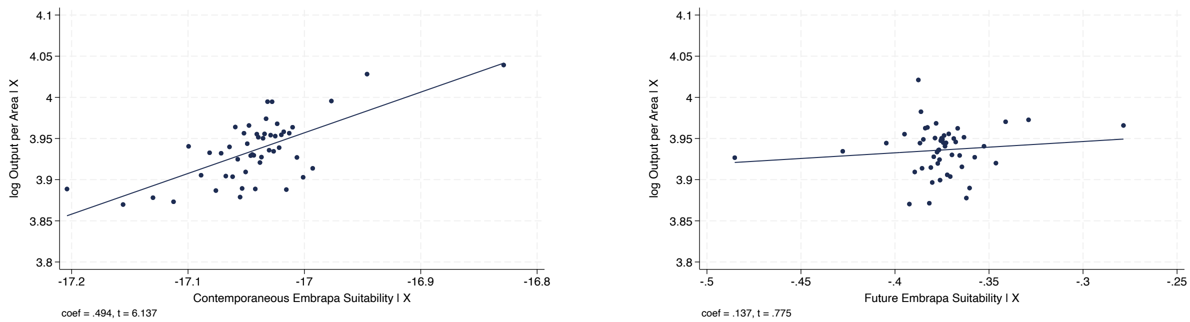
We construct three sets of placebo Embrapa expansion patterns. First, most conservatively, we keep the location of all Embrapa centers fixed and simply randomize the opening year across centers (i.e, only the temporal component of Embrapa exposure). Our main estimate is in the far right tail of the placebo coefficient distribution, which is centered close to zero (Figure 6a).[17] Second, we randomize across both opening years and municipalities, but force all placebo expansion patterns to include at least one center in each biome, consistent with the goals of Embrapa's leadership (see Section 2.2). Again, our actual estimate is in the right tail of the distribution (Figure 6b). Third, we repeat the same exercise but further weight municipalities by their estimated propensity to receive an Embrapa center, where the propensity score is predicted using population, road density, and proximity to the state capital. This allows us to identify plausible counterfactual expansion patterns taking into account local constraints listed by Embrapa's leadership. Once again, our actual estimate is in the far right tail of the placebo distribution (Figure 6c). If the results were driven by spurious trends, we would have likely found similar effects regardless of the exact timing of lab openings or specific lab locations within the feasible set.

[17] In Appendix Table C.5, we control for the expected value of the treatment variable from the simulations in Panel (a) (see Borusyak and Hull, 2023). The estimates are quantitatively similar to our baseline.

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.

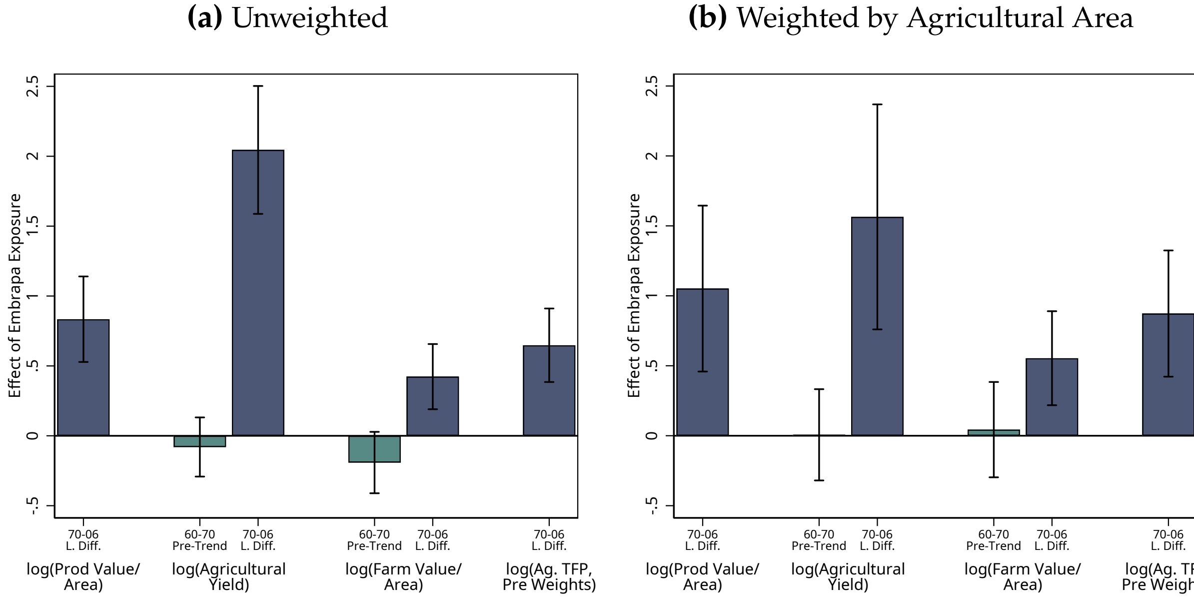
5.4 Dynamics

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, we investigate how long the effect of Embrapa on productivity takes to materialize. We separately estimate the relationship between the municipality-level change in Embrapa exposure over the full period on agricultural productivity in each Census round (see Appendix Figure C.6). While we find little impact during the 1970s, consistent with time lags in technology development and diffusion, a positive effect emerges during the 1980s and continues to grow thereafter. This timing is consistent with historical accounts of Embrapa's impact in the 1980s and 1990s (Monteiro et al., 2012). It is inconsistent with our result picking up the expansion of genetically modified soybean varieties driven by the arrival of multinational crop breeders in the late 1990s and early 2000s (Bustos et al., 2016), which itself could have been an outcome of new innovation seeded by Embrapa.

Figure 8: Long Difference Estimates and Pre-Existing Trends



Notes: The regression model is equation (10), and we control for the log 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: 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 (Fuglie, 2015). For blue bars, the outcome is the difference in each variable between 1970 and 2006; for green bars, the outcome is the difference in each variable between 1960 and 1970 (pre-trends). Standard errors are heteroskedasticity robust and 95% confidence intervals are reported.

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

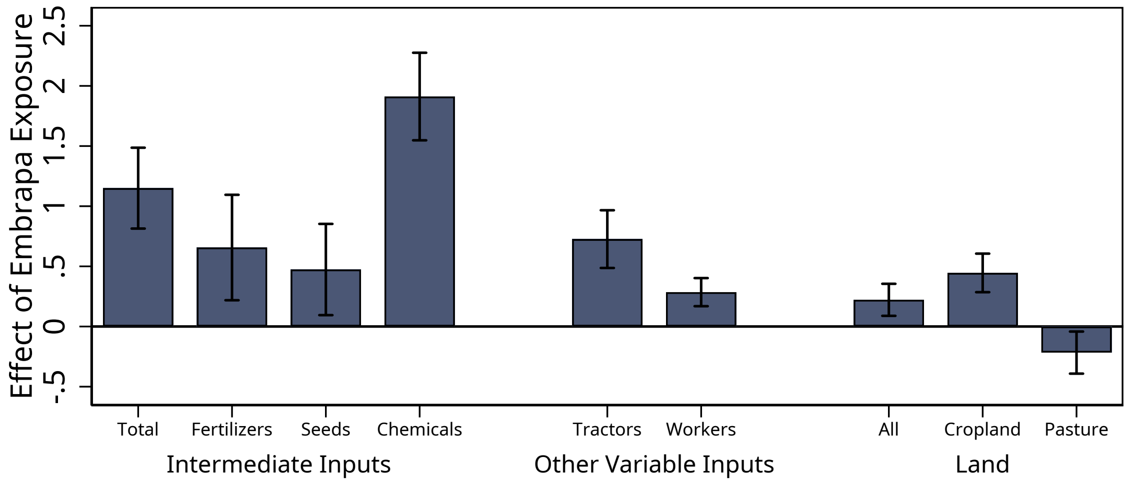
$$\Delta y_i = \beta \cdot \Delta \text{Embrapa Exposure}_i + \gamma' X_i + \epsilon_i \tag{10}$$

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 (a) and farmland area weighted (b) regression specifications. These long-difference effects are 30-40% larger than our baseline estimates, 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

Figure 9: 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 log of distance 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.

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.

5.5 Inputs, Technology Diffusion, and Land Use

Having documented that exposure to Embrapa had a positive effect on agricultural production, we next investigate its effect on input and land use. First, we show that Embrapa increased the use of inputs—especially intermediates that were the focus of its technology development—and land devoted to crops. Second, using a new database of variety use by municipality, we show how the diffusion of Embrapa's novel varieties was strongly predicted by ecological similarity to Embrapa's research labs. Third, we explore the effect of Embrapa on the *distribution* of agricultural land holdings and find no evidence that Embrapa increased inequality, indicating that consolidation is not a key mechanism.

5.5.1 Input Intensification and Land Conversion

We explore the impact of Embrapa on input and land use by estimating our main model, equation (7), with these outcomes. First, we show that Embrapa had a positive effect on intermediate input use, including fertilizers, seed, and chemicals (Figure 9, column 1). These are all areas that were the focus of Embrapa's research. The positive effect on total intermediate input use is driven by independent positive effects on fertilizers (column 2), seeds (column 3), and chemicals (column 4).

Second, we find that Embrapa exposure is associated with increased mechanical input use (column 5) and 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 a small expansion of agricultural land (column 7). This is driven by a large increase in cropland (column 8) and decline in pastureland that can explain roughly half of the cropland expansion (column 9). Thus, Embrapa's focus on crop technology both led crop cultivation to replace pasture land and opened new land for crop production.[18] A further implication is that our main strategy of focusing on total agricultural output properly accounts for this reallocation from pasture to cropland, whereas measuring gains only for crop agriculture might overstate Embrapa's net effects.

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 index of inputs (Appendix Table C.4, Panels C-D).

5.5.2 Crop Variety Diffusion

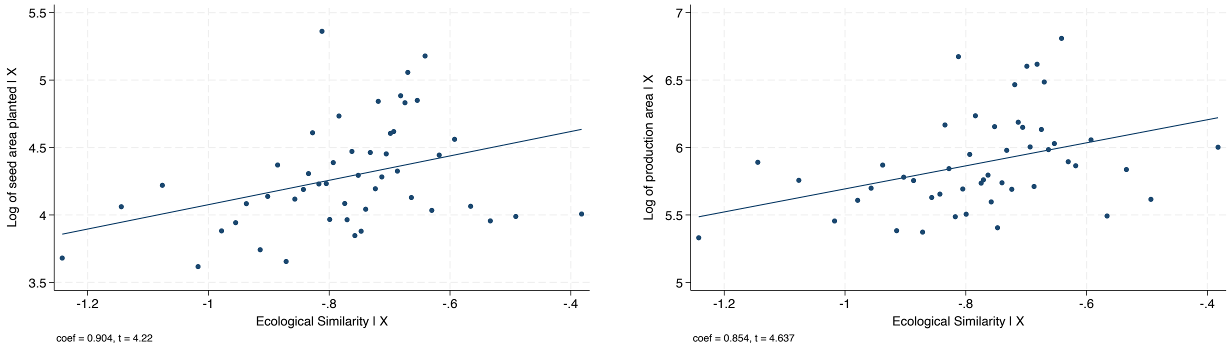
Next, we provide direct evidence of the diffusion of Embrapa's technology. Brazil's Ministry of Agriculture records the production area in each municipality for a large number of seed and crop varieties (see Appendix Section A.3). Using the National Register of Cultivars, we link each variety to the Embrapa research center that developed it. We then directly test whether individual plant varieties diffused disproportionately to municipalities that are ecologically similar to the Embrapa labs that developed them.

In particular, we estimate the following regression specification:

$$\log \text{Production Area}_{v(j),i} = \xi \cdot \text{Ecological Similarity}_{ij} + \alpha_v + F(\text{Distance}_{ij}) + \epsilon_{vij} \tag{11}$$

[18]The negative effect on pasture land recalls the earlier finding that Embrapa exposure increased crop-specific production in excess of overall agricultural output (Appendix Table C.4, Panel A).

Figure 10: Ecological Similarity and Embrapa Variety Adoption



(a) Seed Area Planted

(b) Production Area

Notes: This figure reports estimates of equation (11), based on data from SIGEF on the production of seeds in Brazil. Panel (a) shows the relationship between the area produced of a given seed variety (i.e. cultivar) in a region and the ecological similarity to the origin of that seed variety, measured by the Embrapa Center that developed that seed, controlling for origin fixed effects, seed-type fixed effects, and a cubic polynomial of the log of geographic distance between the location of seed production and the origin location of that seed. Panel (b) shows the same relationship using the area estimated for the production of the final crop that uses a given seed. The final data include 158 plant varieties and 2,242 municipality-variety pairs.

where i indexes municipalities, j indexes Embrapa labs, and v indexes crop varieties. Production Area$_{v(j),i}$ is the total production area of variety v (developed in lab j) in municipality i, and Ecological Similarity$_{ji}$ is defined in equation (5). Variety fixed effects absorb all differences (including differences in quality) across seed varieties and the associated research lab, isolating only ecological differences across "destination" municipalities holding the variety fixed. If the ecological appropriateness of Embrapa's technology determined its subsequent diffusion patterns, we would expect that $\xi > 0$.

Panel (a) of Figure 10 shows a positive and statistically significant estimate of ξ, even after controlling flexibly for geographic distance between i and j. A one-standard deviation increase in ecological similarity is associated with a 50% higher variety-specific production area. The results are similar if we use total crop production using the variety as the dependent variable (Panel (b)). Moreover, the patterns of variety diffusion highlight that there was substantial adoption in municipalities that were geographically very far from the Embrapa lab that developed the variety, but nevertheless ecologically similar (see Appendix Figure C.8). These results provide direct evidence that exposure to Embrapa (as we measure it) is associated with technology diffusion.

5.5.3 Land Inequality

Our main analysis has focused on the effect on overall agricultural productivity in municipalities. These aggregate effects may mask unequal benefits across farmers, affecting inequality both across and within regions. However, we find no statistically significant differences and slightly larger point-estimates for regions with lower baseline productivity or average farm size (Appendix Figure C.7). We also find no evidence that Embrapa increased inequality within regions, proxied using the average farm size and farmland Gini index (Appendix Table C.6). These estimates therefore suggest that the expansion of Embrapa increased productivity without a corresponding increase in inequality.

6 The Returns to Public R&D in Agriculture

This section estimates the returns to the public R&D investment associated with Embrapa. While the previous section established that exposure to Embrapa increased local productivity, it is not yet clear that Embrapa's aggregate benefits justify its costs. 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 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 similarity g_{ij}. 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

$$A_{it} = \bar{A}_{it} \left(\sum_{j \in \mathcal{R}_t} [\exp(\beta g_{ij})(N_{jt})^{\gamma}]^{\theta} \right)^{\frac{1}{\theta}} \tag{12}$$

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

$$\log A_{it} = \beta g_{i\bar{\jmath}} + \gamma \log N_{\bar{\jmath}} + \frac{1}{\theta} \log \left(\sum_{j \in \mathcal{R}_t} \left[\frac{\exp(\beta g_{ij})(N_{jt})^{\gamma}}{\exp(\beta g_{i\bar{\jmath}})(N_{\bar{\jmath}})^{\gamma}} \right]^{\theta} \right) + \log \bar{A}_{it} \tag{13}$$

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 similarity. 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 (13) 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 (13) 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 (13) 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 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 similarity between each pair of municipalities (see Section 5.1). For estimation, we transform equation (13) 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 (13) 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 0.09 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).

The structure we impose allows us to recover not only the differential gains from research in more exposed regions relative to less exposed ones—captured by the reduced-form estimates in Section 5.1—but also the aggregate gains generated by Embrapa, reflected in the third term of equation (13). With estimated parameters β, γ, and θ, together with data on A_{iT}, N_{it}, and g_{ij}, we can recover the implied values of the productivity shifters $\bar{A}_{it}$, net of Embrapa's aggregate contribution to agricultural productivity. This framework thus enables us to quantify the aggregate gains in counterfactual scenarios in which we vary the research investments across labs, N_{it}, in equation (13).

Our key assumptions about the "missing intercept" are embodied in our strategy to identify each $\bar{A}_{it}$ as well as the assumption that these terms are held fixed in counterfactuals. For estimating the productivity gains of Embrapa, our model is conservative insofar as it ignores benefits from general-purpose technologies whose usefulness does not depend on ecological similarity. Our structure also does not explicitly consider the effects of Embrapa on private-sector research at the national level. In light of our earlier finding that the establishment of Embrapa centers crowded *in* others' research (Section 4.3), this omission also likely biases our estimates downward. The model accounts for *local* general equilibrium effects in output markets, input markets, and innovation, because our empirical specification measures the effects on aggregate productivity (in value terms) within municipality. For example, if new technology from Embrapa leads to lower output prices, higher input prices, and/or transformation from livestock to crop agriculture, these effects

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
		$[0.058; 0.123]$	$[-]$
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 (13). 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.

are incorporated into our estimates. However, the model does not explicitly consider other *national* general equilibrium effects that are not mediated by R&D. We favor our approach due to its simplicity and transparency, allowing us to focus on our question of interest. We leave additional analysis of national general equilibrium effects to future research.

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.[19] 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. Relative to this benchmark, our estimates imply 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, 2022) and even scanter evidence in low- and middle-

[19] 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.

income countries.[20] 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 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.[21] To compute the costs, we begin with administrative data on Embrapa's expenditures that were obtained via transparency request. These include capital expenses, such as the costs of centers, their land, and their research equipment, and personnel and organizational expenditures, such as salaries and human capital investments.[22] We 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).[23] This is somewhat larger than prior estimates for advanced economies: for example, Jones and Summers (2022) 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 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 some caution. For example, the potential environmental effects of Embrapa are beyond the scope of our analysis. The sign of these effects is ambiguous, since it depends on whether agricultural intensification increased or decreased local environmental harm (both of which are possi-

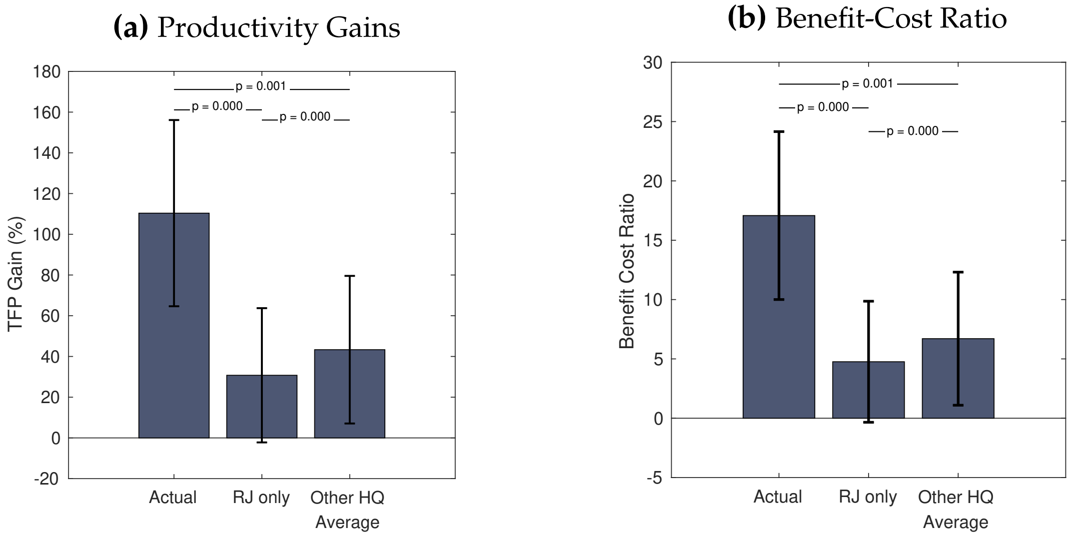
[20]While Embrapa generates revenues through royalties and the sale of various products, it operates as a non-profit company. In a typical year, revenues fall below its operational costs. The bulk of the returns that we estimate in this section are therefore not privately appropriated by the company.

[21]These assumptions are consistent with Embrapa's own reports, which use discount rates in the 4–7.7 percent range (Embrapa, 2018a), as well as World Bank estimates for Latin America (Lopez, 2008).

[22]Human capital investments were a notable portion of Embrapa's budget: Correa and Schmidt (2014) estimate that, in Embrapa's first decade, one fifth of its budget was devoted to training employees.

[23]These estimates also imply an internal rate of return of 25% (see Appendix B.4).

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



Notes: This figure shows the aggregate effect of Embrapa on aggregate agriculture productivity (Panel (a)) and the overall benefit-cost ratio of the program (Panel (b)). In each panel, we show three scenarios: the "Actual" scenario, comparing observed Embrapa with a counterfactual of pre-period agricultural R&D; the "HQ only" scenario, comparing observed Embrapa with only one center in Brasília (using the full budget); and the "Other HQ Average," comparing observed Embrapa with the average return from 38 individual scenarios in which we take each observed Embrapa location as the sole research center (using the full budget). Error bars are 95 percent confidence intervals computed based on the delta method. P-values are based on the null hypothesis of no difference between columns.

ble), and on how the rise of Brazilian exports affected land use elsewhere in the world.[24] That said, we believe that there are several reasons why our approach would lead us (if anything) 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 assume that existing research spending needs to be fully maintained in order to sustain existing gains. Third, our empirical strategy ignores other potential benefits of Embrapa including the effect of new technology on weather resilience, food security, and productivity outside of Brazil (see Lachaud and Bravo-Ureta, 2022). Moreover, even our lower-bound benefit-cost ratio of ten would be considered high among development interventions (see Copenhagen Consensus Center, 2015).[25]

[24] Existing evidence, including from Brazil (Hsiao et al., 2025), suggests that agricultural intensification leads to *less* local deforestation and environmental degradation (e.g., Barrett, 1999; Carreira et al., 2024), and that the rise of Brazilian agriculture also likely reduced global deforestation (Farrokhi et al., 2025).

[25] In Appendix Figure C.9, 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.

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 scale and scope, we simulate the effect of Embrapa under different institutional structures. First, we consider a scenario in which Embrapa doubled down on the pre-existing epicenter of agricultural research in Rio de Janeiro. This holds the scale of investment constant but reduces its geographic scope. This scenario reduces the productivity gains from Embrapa to 31% and the benefit-cost ratio to 4.7 (Figures 11a and 11b, second bars). These estimates substantially lower than the effect of Embrapa's actual design, and our lower bound for the benefit-cost ratio is now below zero.

Second, we consider a collection of alternative scenarios, in which Embrapa invested all of its resources in a single hub in the location of one of the research centers that it ended up constructing. One possibility is that Rio de Janeiro is a uniquely bad location to build a research hub since it is ecologically distinct from much of the rest of Brazil. However, while larger than the effect of scaling up the center in Rio, the average of both the benefits and benefit-cost ratio of scaling up across all other possible centers is still well below our baseline counterfactual estimates (Figures 11a and 11b, third bars). In particular, this design achieves less than half of the observed benefits and has a lower-bound benefit-cost ratio of just 1.1.

These findings suggest that its geographic and ecological scope were central to Embrapa's cost-effectiveness. By contrast, scaling up agricultural R&D in just one location would have much more limited benefits.

6.4 Implications for Policy Design

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

The (Potential) 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). However, these lessons may not apply in all cases.

We show that the productivity effects of innovation are shaped not only by scale effects but also by mismatch between the location for which technology is developed and the location in which it is applied ($\beta > 0$ in the model). Prior work has documented that this effect is present not only for agriculture (e.g., Griliches, 1957; Moscona and Sastry, 2025), but also for medical technology that addresses different diseases (Kremer and Glennerster, 2004), high-tech start-ups that cater to different markets (Lerner et al., 2024), and perhaps even green technology that is adapted to different environments, market structures, and capital stocks (Dugoua and Moscona, 2025). A 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.

Thus, while mismatch effects may be more stark in agriculture due to the ecological specificity of technology than they are in other settings (see, e.g., De Souza, 2024), other differences across markets—including factor endowments, culture, tastes, or complementary tools and infrastructure—could lead to variation in the applicability of non-agricultural technology as well. Measuring the extent of these mismatch effects in fields such as computer science and or climate mitigation technology, which are the focus of considerable policy attention, could be important for designing R&D policy.

A second finding that is crucial to our conclusion that Embrapa benefited from geographic scope is that Embrapa was able to overcome the productivity disadvantage of conducting research outside of R&D "clusters" (see Table 1). For this reason, we abstracted from *external* agglomeration effects—spillovers from non-Embrapa researchers on Embrapa's research productivity—in our structural analysis.[26] In principle, strong agglomeration effects for other (non-agricultural) technologies (see, e.g., Moretti, 2021) could tip the balance toward concentrating research in a few hubs. Understanding the relative importance of concentration (agglomeration) versus expansion (reducing mismatch) in different sectors could be an interesting area for future work.

Heterogeneous Returns and Targeting. A related conclusion is that the returns to R&D investments vary considerably across space and time. To illustrate this in our setting, Figure C.11a displays the productivity gains from constructing a single large research center

[26] The parameter γ allows for *within*-Embrapa agglomeration effects. By also abstracting from positive agglomeration effects of Embrapa research on non-Embrapa researchers—which were suggested by our finding of "crowd-in" in Section 4.3 and Figure C.3—we may be underestimating the overall effect of Embrapa.

in each municipality. Brasília—Embrapa's true headquarters in Brazil's central region—is located in 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, 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 C.11b 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.[27]

7 Conclusion

Global R&D investment is concentrated in a few 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 its benefits in developing countries. 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, a large-scale, half-century long public R&D program that was established in 1973 to spur the development of locally suitable agricultural science and technology. Combining detailed data on Embrapa's structure, the research and career trajectories of all Brazilian agricultural scientists, and municipality-level agricultural outcomes, we investigate the impact of public R&D on research output and agricultural productivity growth.

We present three main sets of findings. First, using 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, we find that Embrapa substantially increased agricultural productivity. Fi-

[27] For the same reason, the potential productivity gains from adding additional centers has declined over time (see Appendix Figure C.10). Adding a new center could raise aggregate productivity by up to 60 percent before Embrapa existed, versus 10 percent today.

nally, 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.

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., Akcigit, U., and Johnson, S. (2026). Technology and economic development. Forthcoming in Handbook of Development Economics.

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.

Avila, A. F. D., Castelo, M. M., Vedovoto, G. L., Irias, L. J. M., and Rodrigues, G. S. (2005). Impactos econômicos, sociais e ambientais dos investimentos na embrapa. *Revista de Política Agrícola*, 14(4):86–101.

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.

Barrett, C. B. (1999). Stochastic food prices and slash-and-burn agriculture. *Environment and Development Economics*, 4(2):161–176.

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.

Brazil (1972). Lei nº 5.851, december 1972. https://www.planalto.gov.br/ccivil_03/LEIS/1970-1979/L5851.htm.

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.

Carreira, I., Costa, F., and Pessoa, J. P. (2024). The deforestation effects of trade and agricultural productivity in brazil. *Journal of Development Economics*, 167:103217.

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 (2015). Post-2015 consensus: Cost-benefit analyses for the sustainable development goals.

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).

Damasceno, G., Souza, L., Pivello, V. R., Gorgone-Barbosa, E., Giroldo, P. Z., and Fidelis, A. (2018). Impact of invasive grasses on Cerrado under natural regeneration. *Biological Invasions*, 20(12):3621–3629.

de Andrade Alves, E. R., e Souza, G. d. S., and Gomes, E. G. (2013). *Contribuição da Embrapa para o desenvolvimento da agricultura no Brasil*. Embrapa.

De Souza, G. (2024). The labor market consequences of appropriate technology. Working Paper 2022-53, Federal Reserve Bank of Chicago.

Diwan, I. and Rodrik, D. (1991). Patents, appropriate technology, and north-south trade. *Journal of International Economics*, 30(1-2):27–47.

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.

Dugoua, E. and Moscona, J. (2025). The economics of climate innovation: technology, climate policy, and the clean energy transition. Working Paper 34601, NBER.

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

Embrapa (1974). Agriculture research in brazil: The need for research. Institutional report, Embrapa – Brazilian Agricultural Research Corporation, Brasília, Brazil.

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

Embrapa (2018b). Brazil in the race to develop waxy cassava starch. https://www.embrapa.br/en/busca-de-noticias/-/noticia/31435206/brazil-in-the-race-to-develop-waxy-cassava-starch. Accessed: 2026-02-09.

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 (2023). Asian soybean rust fungus genome reveals clues to help control. https://www.embrapa.br/en/busca-de-noticias/-/noticia/81067648/asian-soybean-rust-fungus-genome-reveals-clues-to-help-control. Accessed: 2026-02-09.

Embrapa (2025). 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. and Gollin, D. (2003). Assessing the impact of the Green Revolution, 1960 to 2000. *Science*, 300(5620):758–762.

Farrokhi, F., Kang, E., Pellegrina, H. S., and Sotelo, S. (2025). Deforestation: A global and dynamic perspective. Technical report, NBER.

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

Foster, A. D. and Rosenzweig, M. R. (1996). Technical change and human-capital returns and investments: evidence from the green revolution. *The American economic review*, pages 931–953.

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 crescimento da agricultura. *Nota Técnica. Brasília: Ministério da Agricultura, Pecuária e Abastecimento, Assessoria de Gestão Estra tégica*.

Godoy, C. V. (2019). Invasion of soybean rust and its management, from brazilian experiences. In *JIRCAS International Symposium Proceedings*, pages 53–58, Tsukuba, Japan. Japan International Research Center for Agricultural Sciences. Accessed: 2026-02-09.

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 Roche, M. P. (2024). Coordinated R&D programs and the creation of new industries. *Harvard Business School Strategy Unit Working Paper*, (24-027):24–027.

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, NBER.

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.

Hsiao, A., Moscona, J., Olken, B. A., and Sastry, K. A. (2025). Climate change, deforestation, and the expansion of the global agricultural frontier. Working Paper.

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

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.

Jones, B. F. (2021). Science and innovation: The under-fueled engine of prosperity. *The Hamilton Project and Aspen Economic Strategy Group, Policy Proposal 2021*. Available at: https://www.economicstrategygroup.org/wp-content/uploads/2021/11/8-Jones.pdf.

Jones, B. F. and Summers, L. H. (2022). A calculation of the social returns to innovation. In *Innovation and Public Policy*, pages 13–60. University of Chicago Press.

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

Juhász, R., Lane, N., and Rodrik, D. (2024). The new economics of industrial policy. *Annual Review of Economics*, 16(1):213–242.

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, NBER.

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.

Kremer, M., Levin, J., and Snyder, C. M. (2020). Advance market commitments: insights from theory and experience. In *AEA Papers and Proceedings*, volume 110, pages 269–273. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

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, NBER.

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.

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, NBER.

Nehring, R. (2024). The technopolitics of agronomic knowledge and tropical (izing) vegetables in brazil. *Environmental Science & Policy*, 162:103911.

Nordhaus, W. D. (2004). Schumpeterian profits in the American economy: Theory and measurement. Working Paper 10433, NBER.

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.

Parente, S. L. and Prescott, E. C. (2002). *Barriers to riches*. MIT Press.

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.

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., Esworthy, R., Matthews, C. M., Morgan, D., Moteff, J. D., Schacht, W. H., Smith, P. W., and Upton, H. F. (2009). Federal research and development funding: Fy2010. Technical Report R40710, CRS, Washington, D.C.

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.

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). We assembled data for the rounds of 1950, 1960, 1970, 1975, 1980, 1985, 1996, 2006 and 2017.

The censuses rounds of 1996, 2006 and 2017 are available online from the SIDRA system (Sistema IBGE de Recuperação Automática). 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 almost 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.

A.3 Seed Production

To provide direct evidence on the role of ecological similarity in shaping technology adoption from Embrapa, we collected data from SIGEF (Sistema de Gestão e Fiscalização) on the production and control of seed production. These data provide information on all the seed production by municipality and by plant variety (i.e., cultivar). To measure the origin of each seed, we bring in data from Registro Nacional de Cultivares (National Registry of Cultivar), which gives information on the institution or company that registered each plant variety. We select the subset of plant varieties owned by Embrapa's research centers. We then measure the ecological similarity between the research center that registered that plant variety and the actual location of production of that seed. Our final data contain 158 plant varieties, including for example several soybean cultivars, and a total of 2,242 production-variety observations.

A.4 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 $\text{Ecological Similarity}_{ij}$, defined for pairs of municipalities $i, j \in \mathcal{I}$ as

$$\text{Ecological Similarity}_{ij} = -\sum_{x} |x_i - x_j|, \tag{14}$$

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:

$$x_i = \frac{\tilde{x}_i - \text{mean}(\tilde{x}_i)}{\text{sd}(\tilde{x}_i)} \tag{15}$$

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

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.4). 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).

attribute to construct ecological similarity (equation (14)).

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.5 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 (12)

$$A_{it} = \bar{A}_{it} \underbrace{\left(\sum_{j \in \mathcal{R}_t} \left[\exp(\beta g_{ij}) (N_{jt})^{\gamma} \right]^{\theta} \right)^{\frac{1}{\theta}}}_{\equiv EE_{it}} . \tag{16}$$

For exposition, it is useful to define EE_{it}, or "Embrapa Exposure," as the second term. Taking logarithms, and re-arranging terms, we obtain equation (13) from the main body of the paper:

$$\log A_{it} = \beta g_{i\bar{j}} + \gamma \log N_{\bar{j}} + \frac{1}{\theta} \log \left(\sum_{j \in \mathcal{R}_t} \left[\frac{\exp(\beta g_{ij})(N_{jt})^{\gamma}}{\exp(\beta g_{i\bar{j}})(N_{\bar{j}})^{\gamma}} \right]^{\theta} \right) + \log \bar{A}_{it}. \tag{17}$$

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

$$\log A_{iT} - \log A_{i0} = \beta g_{i\bar{j}} + \gamma \log N_{\bar{j}} + \frac{1}{\theta} \log \left(\sum_{j \in \mathcal{R}_T} \left[\frac{\exp(\beta g_{ij})(N_{jT})^{\gamma}}{\exp(\beta g_{i\bar{j}})(N_{\bar{j}})^{\gamma}} \right]^{\theta} \right) + \epsilon_i. \tag{18}$$

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

$$\widehat{A}_{0Tc} = \frac{\sum_i \bar{A}_{iT} EE_{ic}}{\sum_i \bar{A}_{iT} EE_{i0}}. \tag{19}$$

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

$$GV_{0Tc} = \widehat{A}_{0Tc} - 1 \tag{20}$$

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.

$$GV_{0tc} = \begin{cases} \frac{(t-1975)}{(2000-1975)} GV_{0Tc} & \text{if } t \geq 1975 \text{ and } t < 2000 \\ GV_{0Tc} & \text{if } t \geq 2000 \end{cases} \tag{21}$$

We then compute the present value of all benefits using

$$PVB_T = \sum_{t=1975}^{T} \frac{V_{0T} \times GV_{0tc}}{(1+0.07)^{T-t}} + \frac{V_{0T} \times GV_{0Tc}}{0.05} \tag{22}$$

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

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

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

The benefit-cost ratio is then

$$BC_T = \frac{PVB_T}{PVC_T}. \tag{24}$$

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 (2006) value using the same interest rates applied in the benefit cost ratios.

$$PVC_0 = \sum_{t=1975}^{T} \frac{RC_t}{(1+0.07)^t} + \frac{RC_T}{0.05 \times (1+0.07)^T} \tag{25}$$

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:

$$NPV_0 = \sum_{t=0}^{\infty} \frac{V_{0T} \times GV_{0tc}}{(1+r)^t} - PVC_0. \tag{26}$$

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

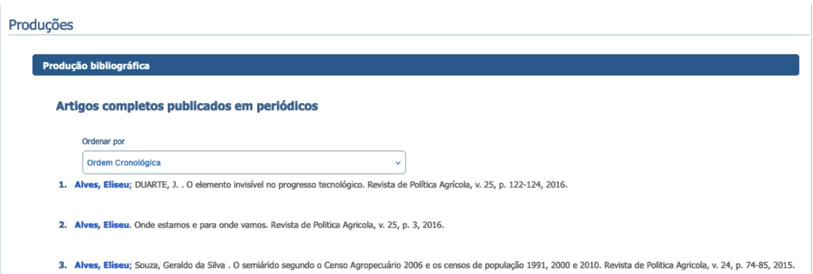
Endereço Profissional Empresa Brasileira de Pesquisa Agropecuária.
Parque Estação Biológica pqeb ed. Sede av.w3 norte(final)
Asa Norte
70770901 - Brasília, DF - Brasil
Telefone: (61) 34484202
Ramal: 4202

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

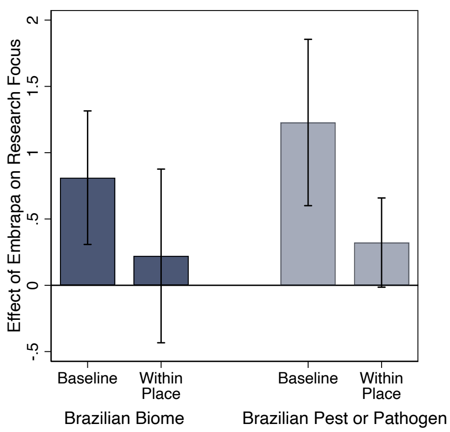
1. **Alves, Eliseu**; DUARTE, J. . O elemento invisível no progresso tecnológico. Revista de Política Agrícola, v. 25, p. 122-124, 2016.

2. **Alves, Eliseu**. Onde estamos e para onde vamos. Revista de Politica Agricola, v. 25, p. 3, 2016.

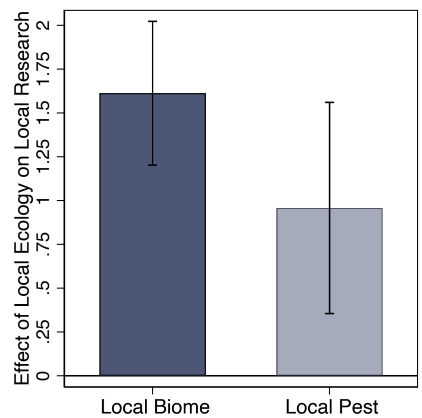
3. **Alves, Eliseu**; Souza, Geraldo da Silva . O semiárido segundo o Censo Agropecuário 2006 e os censos de população 1991, 2000 e 2010. Revista de Politica Agricola, v. 24, p. 74-85, 2015.

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: The Direction of Research Across Ecological Conditions: Quality Adjusted



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

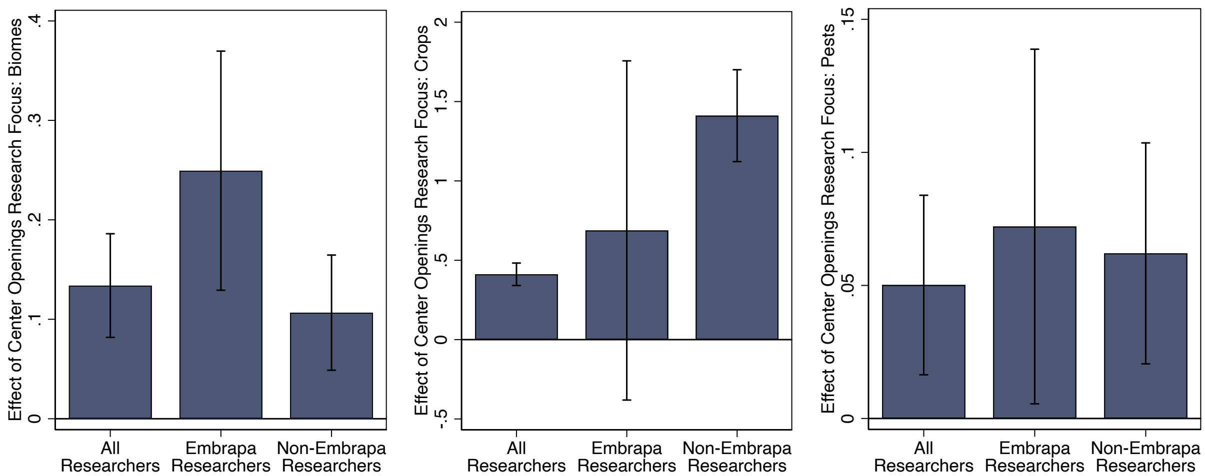


(b) Local Research Focus ($i \times k$)

Notes: This Figure replicates the analysis of Figure 3, but with the research outcome variables weighted by the impact factor of the publication outlet. 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 the journal impact factor of each article mentioning a Brazilian biome and in the second two bars, it is the same for each article mentioning 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). The outcome variable is the number of journal impact factor weighted articles. In both panels, standard errors are clustered by municipality and 95% confidence intervals are reported.

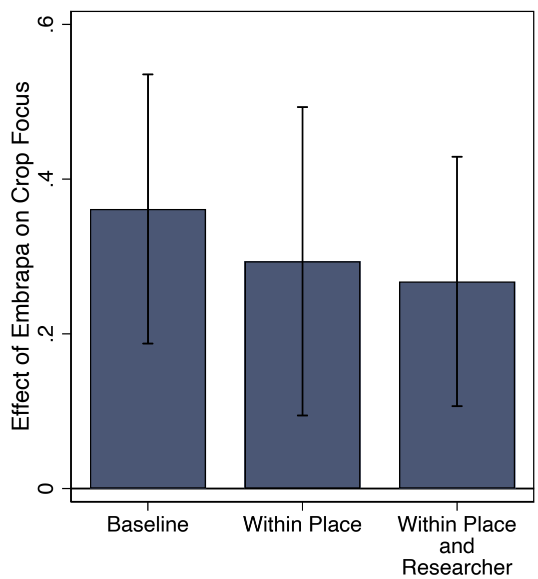
Figure C.3: Effect of Embrapa on Research Direction: Embrapa vs. non-Embrapa Research

(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 (a) biomes, (b) crops, and (c) pests and pathogens. 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. Standard errors are heteroskedasticity robust and 95% confidence intervals are reported.

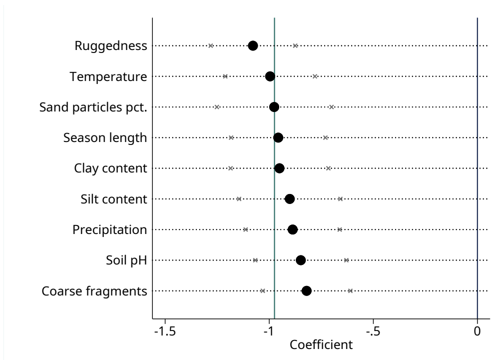
Figure C.4: The Direction of Research Across Crops



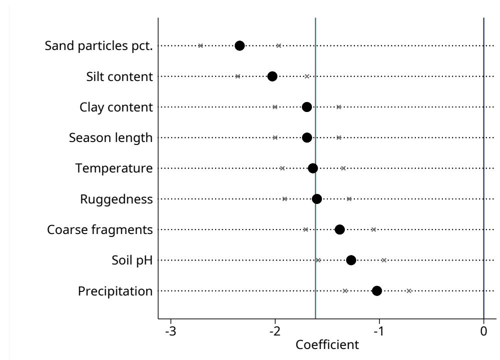
Notes: The unit of observation is an article and each bar represents a coefficient estimate from equation (1). The outcome is an indicator that equals one if the article mentions one of Embrapa's focus crops (beans, cassava, maize, rice, soy, and wheat). In the second column, municipality fixed effects 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.

Figure C.5: Effect of Embrapa Exposure on Productivity Dropping Index Components

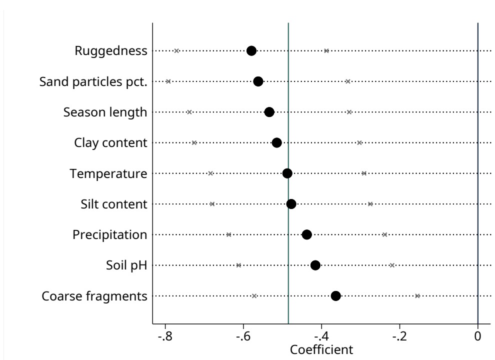
(a) log(Production Value / Area)



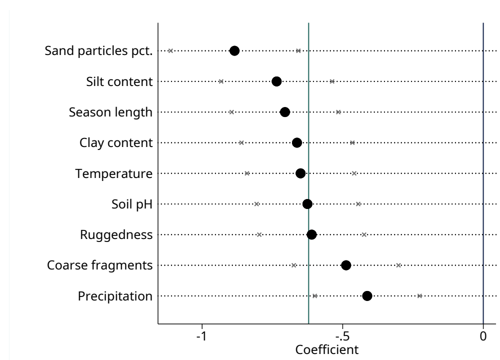
(b) log(Agricultural Yield)



(c) log(Farm Value / Area)



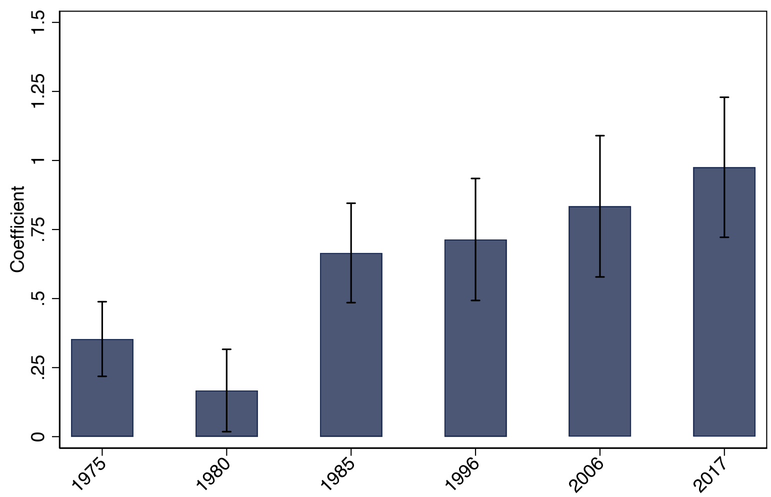
(d) log(Agricultural TFP, Pre Weights)



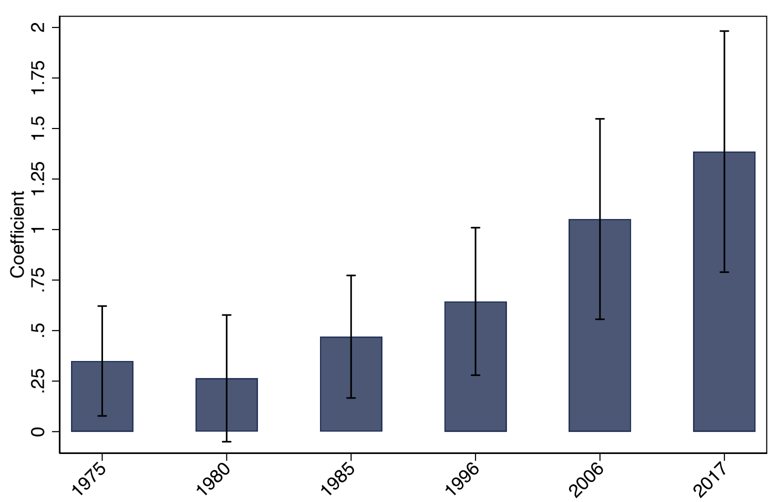
Notes: Each subfigure shows the robustness of our main estimates of equation (7) under variant constructions of Embrapa Exposure that exclude each indicated component of agro-climatic similarity (see Table A.2). Each subfigure corresponds to a different productivity outcome variable. The dots and error 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.6: Dynamic Effects of Embrapa Exposure

(a) Unweighted

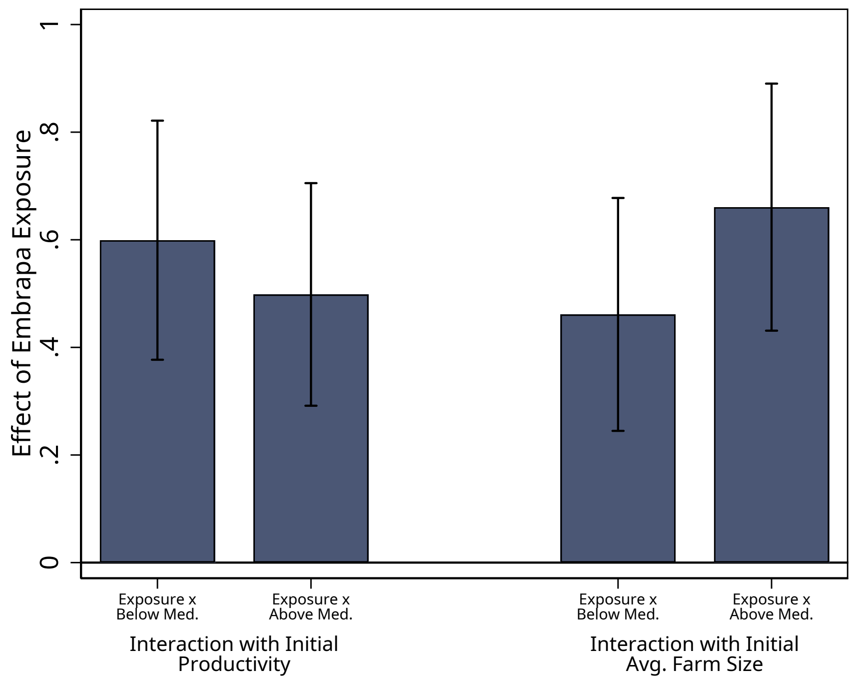


(b) Weighted



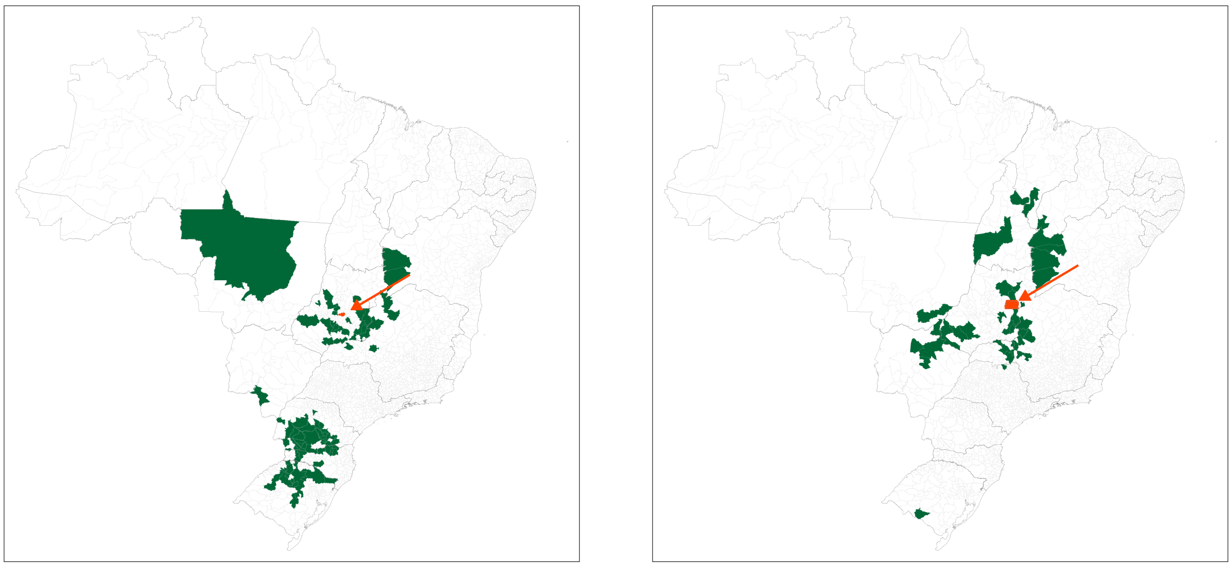
Notes: This figure presents the point-estimates of separate long-difference regressions in which we substitute the dependent variable in equation (10) by different time lags, as specified in the figure. Panel (a) presents the unweighted results and Panel (b) the weighted ones by agricultural land area in 1970.

Figure C.7: Heterogeneous Effects of Embrapa Exposure: 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 (bars 1-2) and indicators for above versus below baseline average farm size (bars 3-4). All specifications include municipality and year fixed effects, in addition to the logarithm of distance 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.8: Municipalities Using Embrapa's Varieties: Two Examples

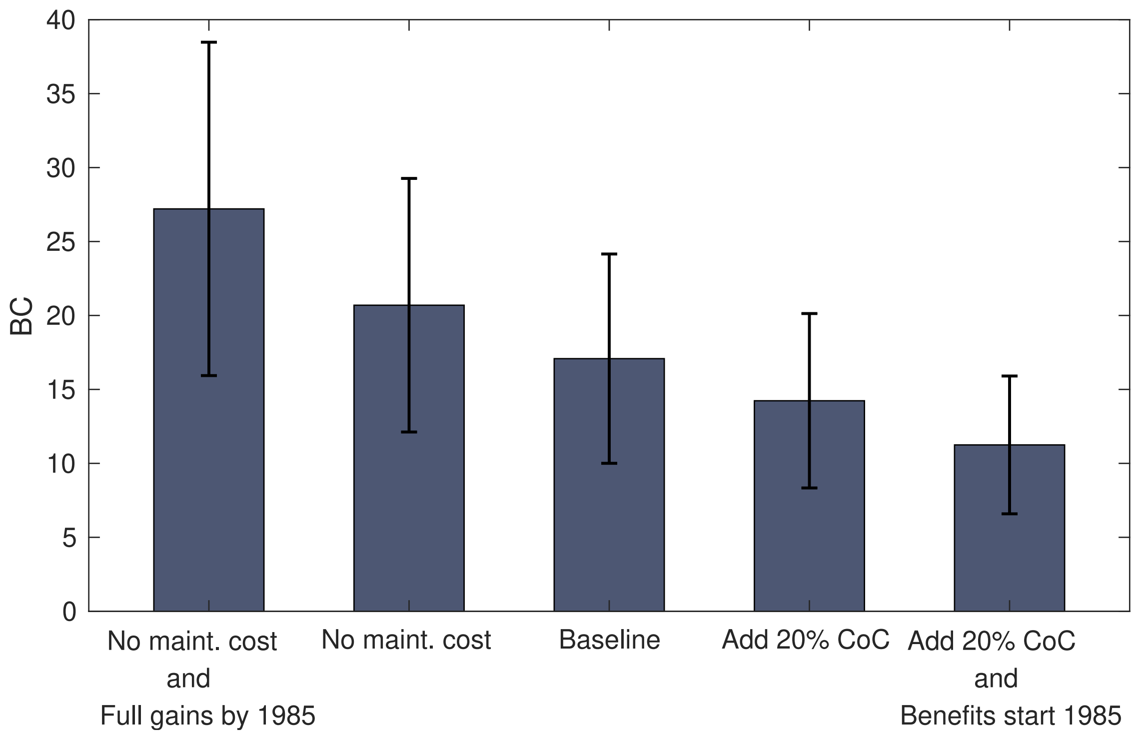


(a) *BRS Esteio* black beans

(b) *BRS 7380RR* soybean

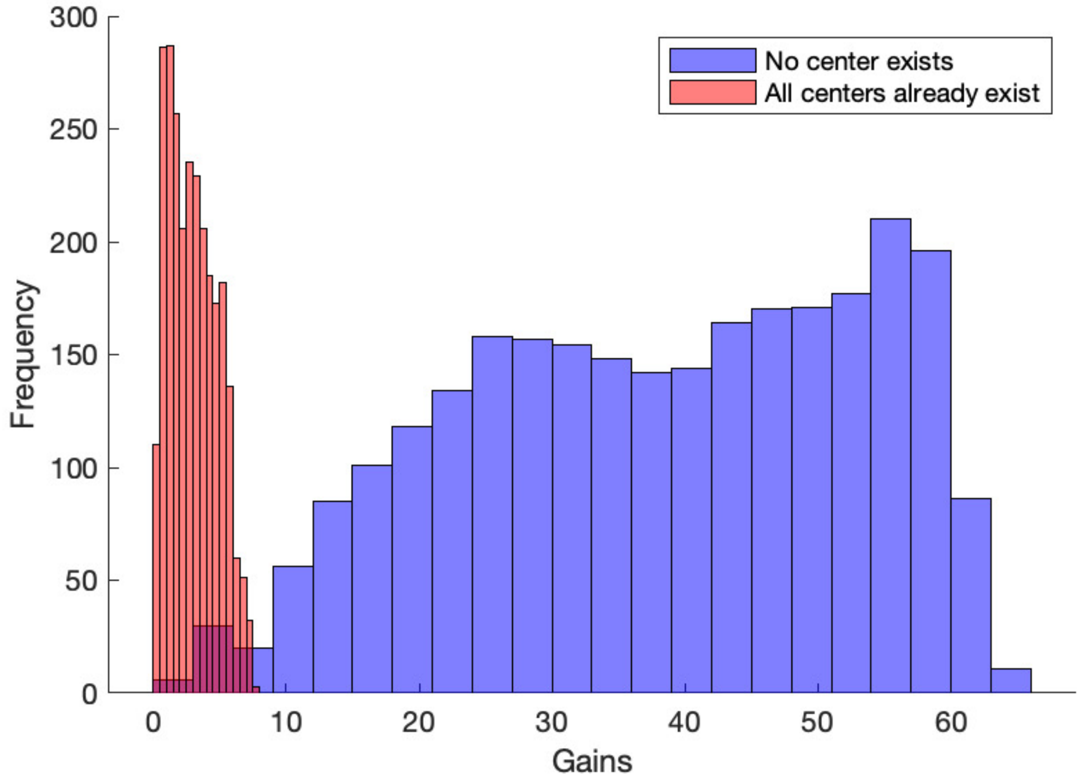
Notes: Each subfigure displays the municipalities in which seed production for the respective variety has been recorded, based on farmer self-reported data from the Sistema de Gestão e Fiscalização (SIGEF) of the Ministry of Agriculture and Livestock (MAPA). Panel C.8a reports the areas where the black bean variety *BRS Esteio* has been produced. *BRS Esteio* was released in 2014 by Embrapa Rice and Beans (located in Santo Antônio de Goiás, GO, highlighted and pointed in orange). The use of *BRS Esteio* quickly spread across the federation, potentially due to its resistance to prevalent pests (Pereira et al., 2013). Panel C.8b shows where *BRS 7380RR* has been produced. This variety was produced by Embrapa Cerrados, in Planaltina, DF. Particularly resistant to nematodes, this soybean proved well-suited for for the Cerrado soils of Goiás, Distrito Federal, Western Bahia, Northwest Minas Gerais, and Mato Grosso (Embrapa, 2017, 2015). Figure 10 shows that, across all Embrapa varieties, these diffusion patterns are strongly predicted by ecological similarity with the location where the variety was developed, even after controlling flexibly for geographic distance.

Figure C.9: 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.10: Productivity Gains of New Research Centers Before and After Embrapa

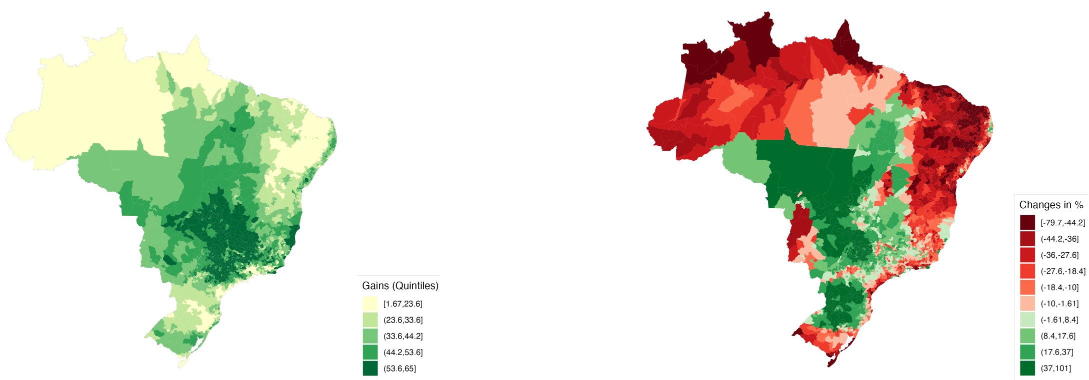


Notes: This figure displays the aggregate productivity gains from adding a large Embrapa center under two scenarios: first, when no centers exist; and second, when all centers established by 2006 are already in place. In both cases, the size of the new center is held constant.

Figure C.11: 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. Red 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.

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 Non Hub, t+2	0.034	0.017	0.034	0.006
	(0.039)	(0.100)	(0.038)	(0.094)
Embrapa x Non Hub, t+1	0.022	0.020	0.027	0.029
	(0.043)	(0.111)	(0.040)	(0.102)
Embrapa x Non Hub, t	-0.005	-0.017	-0.002	-0.003
	(0.046)	(0.141)	(0.042)	(0.132)
Embrapa x Non Hub, t–1	0.040	0.081	0.044	0.099
	(0.037)	(0.104)	(0.035)	(0.097)
Embrapa x Non Hub, t–2	0.051*	0.102	0.051*	0.102
	(0.029)	(0.082)	(0.029)	(0.084)
Embrapa x Non Hub, t–3	0.073***	0.185**	0.059**	0.143**
	(0.027)	(0.074)	(0.026)	(0.070)
Embrapa x Hub, t+2	0.028	-0.051	0.022	-0.068
	(0.029)	(0.086)	(0.030)	(0.087)
Embrapa x Hub, t+1	0.012	-0.031	0.004	-0.049
	(0.040)	(0.113)	(0.042)	(0.122)
Embrapa x Hub, t	-0.022	-0.040	-0.021	-0.045
	(0.042)	(0.119)	(0.046)	(0.137)
Embrapa x Hub, t–1	0.016	0.035	0.008	0.012
	(0.036)	(0.108)	(0.040)	(0.117)
Embrapa x Hub, t–2	0.024	0.013	0.026	0.012
	(0.031)	(0.087)	(0.031)	(0.088)
Embrapa x Hub, 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. “Hubs” are defined as the top ten municipalities in terms of total agricultural research (columns 1-2) or the college graduate share (columns 3-4), and “Non Hubs” are the complement. Standard errors clustered at the AMC level and reported in parentheses.

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

(a) Outcome is (asinh) Citation-Weighted Articles

	(1)	(2)	(3)	(4)	(5)	(6)
Embrapa	0.178***	0.173***				
	(0.054)	(0.058)				
Embrapa x Non Hub			0.269***	0.257***	0.249***	0.242***
			(0.054)	(0.049)	(0.053)	(0.048)
Embrapa x Hub			0.097**	0.122**	0.109**	0.120**
			(0.042)	(0.048)	(0.043)	(0.047)
Non Hub			-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

	(1)	(2)	(3)	(4)	(5)	(6)
Embrapa	0.080	0.101				
	(0.071)	(0.075)				
Embrapa x Non Hub			0.129***	0.123***	0.195***	0.189***
			(0.027)	(0.027)	(0.071)	(0.066)
Embrapa x Hub			0.039	0.065**	-0.009	0.031
			(0.025)	(0.026)	(0.062)	(0.066)
Non Hub			-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

	(1)	(2)	(3)	(4)	(5)	(6)
Embrapa	0.068***	0.068***				
	(0.011)	(0.012)				
Embrapa x Non Hub			0.085***	0.081***	0.078***	0.078***
			(0.012)	(0.012)	(0.012)	(0.012)
Embrapa x Hub			0.049***	0.060***	0.052***	0.061***
			(0.013)	(0.011)	(0.012)	(0.011)
Non Hub			-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: (a) the inverse hyperbolic sine of citation-weighted articles, (b) the count of articles winsorized at the 99th percentile of the researcher-by-year data, and (c) an indicator (0/1) for whether a researcher published any article. The regression model is equation (4). In columns 1 and 2, the main regressor indicates whether the researcher works for Embrapa. Columns 3 to 6 interact the Embrapa indicator with municipality research characteristics. "Hubs" are defined as the top ten municipalities in terms of total agricultural research (columns 3-4) or the college graduate share (columns 5-6), and "Non Hubs" are the complement. 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)
Panel A: Outcome is log(Crop Yield)					
Embrapa Exposure	1.373***	1.595***	1.519***	1.309***	1.262***
	(0.119)	(0.130)	(0.172)	(0.128)	(0.140)
Observations	13093	12895	8432	12892	12895
R^2	0.614	0.615	0.635	0.640	0.743
Panel B: Outcome is log(Farm Value / Farm Area)					
Embrapa Exposure	0.383***	0.480***	0.594***	0.395***	0.341***
	(0.072)	(0.077)	(0.110)	(0.073)	(0.083)
Observations	13134	12936	8444	12930	12936
R^2	0.964	0.964	0.964	0.965	0.968
Panel C: Outcome is log(Agricultural TFP, Pre Weights)					
Embrapa Exposure	0.457***	0.544***	0.592***	0.474***	0.476***
	(0.072)	(0.075)	(0.099)	(0.074)	(0.079)
Observations	12237	12044	7888	12041	12044
R^2	0.984	0.984	0.985	0.985	0.986
Panel D: Outcome is log(Agricultural TFP, Post Weights)					
Embrapa Exposure	0.268***	0.205**	0.125	0.187**	0.260***
	(0.077)	(0.081)	(0.108)	(0.080)	(0.082)
Observations	12237	12044	7888	12041	12044
R^2	0.984	0.984	0.985	0.985	0.986
Municipality FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
log(Distance to Embrapa) x Round FE	-	Y	Y	Y	Y
Drop if $< 100km$ from Embrapa	-	-	Y	-	-
Drop if neighbor to Embrapa	-	-	Y	-	-
log(Initial Prod.) x Round FE	-	-	-	Y	-
log(Initial Pop.) x Round FE	-	-	-	Y	-
log(Initial Roads) 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 1970s (Panel C) and 2010s (Panel D). The control variables included are: log of distance 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; proximity to roads 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: Effects of Embrapa Exposure: Controlling for Expected Treatment

	(1)	(2)	(3)	(4)	(5)
	Log of production value per total farm area				
Embrapa Exposure	0.508***	0.609***	0.820***	0.606***	0.394***
	(0.098)	(0.109)	(0.150)	(0.107)	(0.115)
Observations	18386	18109	11821	18101	18109
R^2	0.954	0.954	0.945	0.955	0.976
Municipality FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
log(Distance to Embrapa) x Round FE	-	Y	Y	Y	Y
Drop if $< 100km$ from Embrapa	-	-	Y	-	-
Drop if neighbor to Embrapa	-	-	Y	-	-
log(Initial Prod.) x Round FE	-	-	-	Y	-
log(Initial Pop.) x Round FE	-	-	-	Y	-
log(Initial Roads) 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). 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: log of distance 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; proximity to roads 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: Effects of Embrapa Exposure 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: Effects of Embrapa Exposure: Alternative Standard Errors

	(1)	(2)	(3)	(4)	(5)
	Panel A: Conley Standard Errors				
Embrapa Exposure	0.730***	0.825***	0.985***	0.799***	0.599***
	(0.186)	(0.199)	(0.278)	(0.179)	(0.128)
Observations	18386	18109	11821	18101	18109
	Panel B: State-Level Clustered Standard Errors				
Embrapa Exposure	0.730***	0.825***	0.985***	0.799***	0.599***
	(0.195)	(0.205)	(0.305)	(0.186)	(0.135)
Observations	18386	18109	11821	18101	18109
Municipality FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
log(Distance to Embrapa) x Round FE	-	Y	Y	Y	Y
Drop if $< 100km$ from Embrapa	-	-	Y	-	-
Drop if neighbor to Embrapa	-	-	Y	-	-
log(Initial Prod.) x Round FE	-	-	-	Y	-
log(Initial Pop.) x Round FE	-	-	-	Y	-
log(Initial Roads) 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: log of distance 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.

Table C.8: Effects of Embrapa Exposure: Alternative Parameterization of Exposure

	Parameterization of Exposure Treatment:				
	Top	Top Two	Top Two Budget Weighted	Top Three	Top Three Budget Weighted
	(1)	(2)	(3)	(4)	(5)
	Panel A: Baseline Results				
Embrapa Exposure	0.730***	0.864***	0.716***	0.918***	0.749***
	(0.080)	(0.085)	(0.078)	(0.087)	(0.082)
Observations	18386	18386	18306	18386	18386
R^2	0.954	0.954	0.954	0.954	0.954
	Panel B: Weighted by 1970 Agricultural Area				
Embrapa Exposure	0.758**	1.057***	0.775**	1.189***	0.784*
	(0.302)	(0.354)	(0.343)	(0.330)	(0.411)
Observations	18372	18372	18292	18372	18372
R^2	0.964	0.965	0.964	0.965	0.964
Municipality FE	Y	Y	Y	Y	Y
Year FE	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). 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 Embrapa exposure measure if computed differently across columns, either as ecological similarity to the most ecologically similar center (column 1); ecological similarity to the two most ecologically similar centers, where both get equal weights (column 2); ecological similarity to the two most ecologically similar centers, where the centers are weighted by their budgets (column 3); ecological similarity to the three most ecologically similar centers, where both get equal weights (column 4); and cological similarity to the three most ecologically similar centers, where the centers are weighted by their budgets (column 5). All columns include municipality and year fixed effects. Standard errors are clustered at the municipality level.

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.

Embrapa (2015). Soja BRS 7380RR. https://www.embrapa.br/busca-de-solucoes-tecnologicas/-/produto-servico/4459/soja-brs-7380rr. Access date: October 23, 2025.

Embrapa (2017). Cultivar BRS 7380 RR aumenta produtividade dos sojicultores do Cerrado. https://www.embrapa.br/busca-de-noticias/-/noticia/21357046/cultivar-brs-7380-rr-aumenta-produtividade-dos-sojicultores-do-cerrado. Access date: October 23, 2025.

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

IBGE (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. 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, NBER.

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

Pereira, H. S., Melo, L. C., Faria, L. C. d., Wendland, A., Del Peloso, M. J., Costa, J. G. C. d., Nascente, A. S., Díaz, J. L. C., Carvalho, H. W. L. d., Almeida, V. M. d., et al. (2013). Brs esteio: common bean cultivar with black grain, high yield potential and moderate resistance to anthracnose. *Crop Breeding and Applied Biotechnology*, 13:373–376.

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