Technology Adoption and the Agricultural Supply Response Function

Guilherme DePaula

Working Paper 18-WP 582
July 2018

Center for Agricultural and Rural Development
Iowa State University
Ames, Iowa 50011-1070
www.card.iastate.edu

Guilherme DePaula is Assistant Professor, Department of Economics, Iowa State University, Ames, IA 50011, E-mail: gdepaula@iastate.edu.

For questions or comments about the contents of this paper, please contact Guilherme DePaula, gdepaula@iastate.edu

Technology Adoption and the Agricultural Supply Response Function

Guilherme DePaula*

June 11, 2018

Abstract

In this study, I exploit the recent technology-driven soy boom in Brazil to assess how the diffusion of different technologies, namely the genetically modified soy and biological nitrogen-fixing soy varieties adapted to the Brazilian savanna, change the agricultural supply response function. I use a novel panel dataset combining farm-level data for 1.5 million commercial farms from the 1996 and 2006 Brazilian agricultural census surveys to estimate the price effects on the expansion of the soy acreage. I find that the acreage response functions become increasingly elastic towards the agricultural frontier because of the existence of different technological diffusion processes. The large price effect on the adoption of nitrogen-fixing soy designed to convert marginal savanna pastureland into soy production explains most of the heterogeneity in the acreage supply function in Brazil. The estimated long-run price elasticity of soy acreage is 0.6 in the south and 1.8 in the savanna. On the agricultural frontier close to the savanna–Amazon border, the price elasticity of agricultural land is 0.13, implying that a 10% permanent increase in soy prices would result in the conversion of 1 million hectares of natural vegetation to farmland.

Key words: Acreage response, Technological change, Agricultural micro-data, Brazil, Soy

JEL codes: O33, Q12, Q16, Q55

---

* Assistant Professor, Department of Economics and Center for Agricultural and Rural Development, Iowa State University. 518 Farm House Lane, Ames, IA 50011. gdepaula@iastate.edu. This article was originally one section of my doctoral dissertation at Yale University. The first version of this analysis was a section of my job market paper titled "Technological Diffusion and Bundled Contracts: Soy Boom in Brazil". I thank Leandro Justino for excellent research assistance with the analysis of the confidential agricultural census datasets. I am grateful for the invaluable guidance and support of Robert Mendelsohn, Arnulf Grubler, and Kenneth Gillingham. I thank Xiaohong Chen, Ahmed Mushfiq Mobarak, Ary Fortes, Glaucia Ferreira, and the seminar participants at Yale, the Association of Environmental Resources Economics conference, and the University of Ottawa Environmental Economics Workshop for their comments and suggestions. I also thank the research staff from the Brazilian Institute of Geography and Statistics (IBGE) and Brazilian Agricultural Research Company (Embrapa), especially Carlos Lessa, Luis Carlos Pinto, Flavio Alves, Geraldo Souza, and Rosanna Guidicci for their support in accessing the micro census dataset. I gratefully acknowledge the financial support from the Council for Latin American and Iberian Studies at Yale, the Yale Institute for Biospheric Studies, and the Yale MacMillan Center. Any errors are my own.

How does technology adoption influence long-run agricultural supply response functions? There has recently been renewed focus on the identification of agricultural supply and demand elasticities to evaluate the effectiveness of biofuel and forest conservation policies (Hausman, 2012; Roberts and Schlenker, 2013). Such policies have a potential “leakage” effect through their influence on commodity prices. For instance, higher demand for corn for biofuel production in the United States could increase commodity prices and lead to agricultural expansion in Brazil. Land diversion policies designed to protect natural vegetation and conserve carbon sinks could also lead to land-use change through increases in commodity prices. In both cases, the degree of the “leakage” effect depends on the shape of the long-run agricultural supply response function.

Although technological innovation is the principal factor changing long-run agricultural supply, estimating the effects of technology adoption in long-run agricultural supply functions remains a challenge. Most empirical analyses of agricultural supply use country or county time-series data to estimate supply elasticities and model technological change with trend variables. The implicit assumption of these models is that technological change in the agriculture sector is a slow-moving process that can be captured by time effects. Although the development of agricultural technologies such as new seed varieties can take years or even decades, the technological adoption process as well as institutional changes that influence adoption do not necessarily follow a slow-moving or approximately linear pattern. In fact, some technological adoption cases such as the soybean expansion in the Brazilian savanna tend to follow a typical S-shaped pattern with a period of rapid technological adoption. Moreover, the adoption process itself could be responsive to changes in commodity prices, leading to a non-linear response in agricultural supply.

The main contribution of this study is to model the technological adoption process to estimate agricultural supply response functions. I contrast the agricultural supply effect of the adoption of two agricultural technologies in Brazil: genetically modified soy (GE soy) in the south of Brazil and biological nitrogen-fixing soy varieties adapted to the Brazilian savanna (NF soy). Both technologies diffused rapidly in Brazil after 1996, but had different effects on the soy supply function. The pesticide-resistant GE soy variety was designed for traditional soy-producing regions such as the south of Brazil, and this rapidly replaced traditional soy varieties. By contrast, the NF soy technology was developed to convert marginal pastureland in the savanna into large soy plantations, and this diffused rapidly through partnerships between farmers and trading companies.

The Brazilian case of the soy expansion provides a unique experiment to assess how the diffusion of different technologies changes the agricultural supply response. Figure 1 illustrates the elusive effect of technological change in the analysis of country time series. The historical yield time series for soy in Brazil shows a clear linear trend. However, the soy acreage time series reflects the non-linear technological diffusion processes. The apparent slow-moving technological change reflected in the yield time series conceals the historical technological change process that altered the soy agricultural supply function in Brazil. The booming soy expansion after 2006 happened through the conversion of marginal land in the savanna, which was originally considered to be unsuitable for agricultural production. In this study, I estimate the acreage component of agricultural supply separately for the south and savanna regions of Brazil and use measurements for the return to the adoption of GE and NF soy to assess the differentiated effect of these technologies on the soy supply function and implications for land-use change in Brazil.

I use the most comprehensive agricultural dataset in Brazil to estimate the regional acreage response functions for soy. I combine for the first time confidential farm-level data for 1.5 million commercial farms from the 1996 and 2006 Brazilian agricultural census surveys (IBGE, 1996, 2006), capturing the fastest period of soy expansion in Brazil. The agricultural census dataset is the only source of agricultural data covering the entire country that contains farm-level information on farmers' land-use and technology choices as well as on their characteristics and agricultural outcomes such as measures of productivity and forestland. The agricultural census is also the only official dataset reporting pastureland at the farm level for Brazil. The aggregated time-series dataset in Brazil does not track pastureland, and it is not possible to model land-use change in Brazil without pastureland since it represents about 50% of all farmland and 73% of agricultural land. Therefore, I link the census farm-level data at the census block level and then integrate the census panel with a panel dataset with spot price data at the municipality level (IBGE PAM, 2015)[1].

I estimate the soy acreage function by using a long first differences model, which replicates the long-run response to a permanent price increase. The unit of observation is the rural census block. I explore the permanent increase resulting from two macroeconomic policy changes in Brazil after 1996: the Kandir law, which eliminated state taxes for soy exports, and the 1999 currency devaluation. These two policies combined resulted in a permanent increase in soy prices of over 50% despite the relatively stable global soy prices during this period. I show that these policy changes in combination with the large variation in distance from producers' locations to markets and ports in Brazil generated spatial variation in price changes. I then explore this variation in price changes to identify price effects conditional on the census block fixed effects, the state and mesoregion fixed effects, and a rich set of baseline characteristics to capture local trends in agricultural expansion. I also instrument price changes with the straight-line distance to market measurements, conditioned on the baseline characteristics.

I use two methods to incorporate the technological change effect into the acreage model. I first adapt the approach of Foster and Rosenzweig (1996) and Bustos, Caprettini, and Ponticelli (2016) by using the variation in the potential yield measures in a first differences model to identify the technological change effects. Second, I explicitly model the technological choice in a selection model for the acreage response function to separate the effect of prices on the soy acreage expansion through technological adoption. The identification of the technological change effects relies on the combination of exogenous potential yield measures (GAEZ FAO/IIASA, 2017) and the timing of the introduction of GE soy and new farmer–trader NF soy contracts.

I find significant heterogeneity in the acreage response functions across Brazil, consistent with the results presented by Hausman (2012) for the soy expansion and Nagaravapu (2010) for the sugarcane expansion. In the south of the country, the estimated long-run price elasticity of the soy expansion is 0.25 at the region level and 0.62 for locations with a higher return to GE adoption compared with a price elasticity of the GE soy diffusion of 1.11. These results are consistent with the rapid replacement of traditional soy varieties with GE soy. The price effect in the soy acreage expansion through the technological adoption channel is small in the south and my estimates are comparable with the price elasticities for the south region estimated by Hausman (2012), using time-series data.

In the savanna, the soy acreage function is more elastic. The region-level price elasticity is 1.15, which increases to 1.8 in locations with a high return to NF soy adoption. On the agricultural frontier of the savanna, on the border with the Amazon biome, the estimated long-run price elasticity is 3.13. These estimates are statistically significant and higher than the short-run price elasticities of Hausman (2012) for the savanna[2]. In the savanna, the large acreage response to price changes is explained by the indirect effect of prices through the technological adoption channel. The permanent price increase enabled NF soy production on new land, resulting in an elastic supply function. This result suggests that the standard time-series model for acreage response functions underestimates supply elasticities on the agricultural frontier where technological change is designed for the conversion of land into agricultural production.

I also estimate a first differences model for the expansion of all agricultural land, including pastureland, to assess the environmental impact of a permanent change in soy prices. The estimated price elasticity of agricultural land is 0.025 for Brazil and 0.101 for the savanna. The elasticity for the south region is not statistically significant. The elasticity for Brazil is comparable to the agricultural land elasticities estimated by Barr et al. (2010) for the period between 2004 and 2006, and is lower than the cropland elasticities estimated by Hausman (2012) and Roberts and Schlenker (2013), as expected, since these studies do not account for pastureland. These elasticities imply that a 10% increase in soy prices, holding technology and policy fixed, would result in a deforestation of 936 thousand hectares in Brazil if we use the country elasticities and approximately 953 thousand hectares only in the Midwest savanna using the savanna elasticity. Although the simulated land-use change is relatively small compared with historical deforestation in Brazil, these results suggest that simulated land-use change based on country-level models can significantly underestimate the environmental impact of price increases on the agricultural frontier.

This study is the first acreage supply analysis of soy in Brazil using micro census data and, to the best of my knowledge, the first econometrics study that models the technological change effect to estimate the acreage supply function. This analysis thus builds on a rich literature that measures agricultural supply functions (Nerlove, 1956; Cochrane, 1955; Tomek and Robinson, 1981; Just, 1993, 2000) and complements the series of recent econometrics studies that estimate supply response functions to simulate the impact of environmental policies on land-use change (Barr et al., 2010; Hausman, 2012; Roberts and Schlenker, 2013; Nagavarapu, 2010). My empirical approach combines the model for estimating the technological change effects used by Bustos, Caprettini, and Ponticelli (2016) with the econometric estimation of the price elasticity of the soy expansion used by Hausman (2012). Bustos, Caprettini, and Ponticelli (2016) use aggregated data to estimate the effects of the soy technological change in agricultural and development outcomes in Brazil, but they do not account for price changes. Hausman uses a panel of aggregated data for prices and acreage to estimate the long-run price elasticities of the soy expansion in Brazil, but models technological change by using different functional forms for time trends. My empirical analysis therefore builds on these two studies by modeling both technological changes and price changes using micro-level data.

Institutional Context: Technological Change and Price Increase

The technological change that transformed the Brazilian agricultural sector was the adaptation of soy for production in the large savanna biome. Since the 1960s, the Brazilian government has sponsored a plant-breeding program that combines tropicalized soy seeds with nitrogen-fixing bacteria strains. Brazilian scientists designed new seed–bacteria combinations that allowed production in savanna soils without the application of nitrogen fertilizer to enable profitable soy growth in the undeveloped regions of Brazil. The adoption of NF soy is expensive because in addition to buying these new seed–bacteria combinations, farmers must clean the land and correct the chemical deficiencies of the soil through the widespread application of lime and fertilizers. This high upfront investment initially slowed the diffusion of NF soy, which accelerated only in the mid-1990s after the government implemented a series of market reforms that opened the Brazilian market to foreign investment and changed soy prices. For a detailed description of NF soy technology, see Hungria, Campos, and Mendes (2001), and Alves, Boddey, and Urquiaga, (2002). DePaula (2017), Junior (2011), and Silva (2012) explain the technological diffusion mechanisms for NF soy.

Figure 2 shows the six biomes of Brazil and soy expansion into the savanna. Soy production started in the temperate south region of Brazil, mostly in the Atlantic Forest biome, where the soil and climate characteristic are similar to other major soy-producing regions in the United States and Argentina. In 2011, soy production intensified in the south, also extending into the Pampa biome, and expanded throughout the savanna biome, reaching the frontier of the Amazon. In 2001, soy production reached as far north as 7 degrees latitude south of the equator line, in the state of Maranhao.

NF soy technology was specifically designed for the conversion of marginal land in the Brazilian savanna. By contrast, the GE soy developed by Monsanto was created for traditional soy-producing regions and was first commercialized in the United States and Argentina in 1996. The new GE seed was resistant to glyphosate pesticides, reducing the necessary number of pesticide applications. GE soy technology also reduced labor costs in the farm and facilitated no-tillage planting, improving soil conservation practices. GE soy was officially approved for commercial use in Brazil in 2001 and diffused rapidly, mostly in the south of Brazil. For more information on the characteristics and adoption process of GE soy technology, see Bustos, Caprettini, and Ponticelli (2016), Qaim and Zilberman (2003) and Ainsworth et al. (2012).

Change in Soy Prices in Brazil from 1996 to 2006

After years of hyperinflation, the Brazilian government implemented a series of macroeconomic policies in the mid-1990s to stabilize the country’s economy. Two policy changes raised the price farmers received for exporting soy by over 50% between 1996 and 2006, despite relatively constant international prices.

The first policy was the Kandir Law approved in 1996 (Kume and Piani, 1997; Soares, 2012). The Kandir Law eliminated the state tax on the circulation of products and transport services (ICMS) related to the export of primary goods, an effective tax cut of about 20%. ICMS is the main state tax in Brazil that varies across states and transportation routes. Figure 3A shows the average ICMS rate applied to soy exports over time. The Kandir Law also changed the relative prices of raw and processed soy, favoring exports of raw soy and reducing investment in new soy-processing capacity in Brazil (Junior, 2011).

The second policy intervention was the devaluation of the Real currency during the financial crisis of 1999. The Kandir Law, which had the effect of a fiscal devaluation, was created in part to compensate for the appreciation of the Real after the implementation of Plano Real (Kume and Piani, 1997). Figure 3B shows the evolution of the effective real exchange rate of the Real. In 1994, the Brazilian government pegged the Real to the Dollar to control inflation; however, the peg system proved too expensive and had to be removed in January 1999, resulting in a 66% currency devaluation. Finally, the Brazilian government reduced tariffs on imports to spur competition and increase productivity. Between 1989 and 1994, the index of agricultural input prices, which includes purchases of fertilizers and new farming equipment, decreased by about 25% in real terms (De Melo, 1999). The change in import tariffs compensated for the negative effect of the currency devaluation on the import of agricultural inputs and machinery.

The combination of the currency devaluation and tax cuts led to a permanent increase in the price of soy for the export market. Furthermore, the changes in prices and profit margins varied significantly across Brazil because of the large variation in the quality of the national transportation infrastructure. Most of the soy produced in Brazil and exported is transported to ports by trucks, in many cases on unpaved roads. This variation in transport cost induces differentiated changes in farm gate prices and farm profits across Brazil.

Data

I combine three datasets for my empirical analysis: the Brazilian Agricultural Census produced by the Brazilian Institute of Geography and Statistics (IBGE, 1996, 2006), the IIASA/FAO Global Agro-Ecological database (GAEZ) (FAO/ IIASA, 2017), and the Municipal Agricultural Production Survey, also maintained by IBGE (2015). I summarize in this section the key features of the dataset. For more information see DePaula (2017). Table 1 presents the summary statistics for the main variables used in the empirical analysis. The panel dataset has 34,115 common census blocks, accounting for 1.59 million and 1.5 million commercial farms in 1996 and 2006, respectively. Appendix A describes the variables used in the analysis and IBGE (1998, 2012, 2016) documents the IBGE agricultural census and IBGE PAM survey.

The novel dataset is a two-period panel linking the 1996 and 2006 Brazilian agricultural census surveys at the census block level. Due to confidentiality restrictions, it is not possible to link the census data at the farm level. I thus aggregated the farm-level variables at the census block level for approximately 1.5 million commercial farmers and then connected the census blocks for the census surveys of 1996 and 2006. I followed the commercial farm definition based on farm production value used by the Brazilian Agricultural Research Agency (Alves et al., 2012). Commercial farming accounts for more than 80% of agricultural production in Brazil. Although most census blocks appear in both census surveys, some were combined or divided between 1996 and 2006. I thus created the common census block unit as the smallest set of census blocks identical in both surveys. The common census block is the unit of observation of the panel dataset. One census block in Brazil has on average 40 farms. Appendix C contains a map of the rural census blocks in Brazil.

The census panel dataset contains the acreage harvested with soy in each farm by type of seed (traditional, certified, genetically modified) as well as the characteristics of the farm and production system. The dependent variables are the change in soy harvested area, change in the share of farmland allocated to soy in a census block, and change in GE soy area harvested. Over three-quarters (76%) of the 5.4 million hectares of the soy expansion occurred in the savanna region, while 3.2 million hectares were harvested with GE soy in 2006, 80% in the south. The average change in soy harvested area was 338 hectares.

The GAEZ dataset (FAO/ IIASA, 2017) has estimates of the potential yield of soybean and other crops under three levels of input intensity: low, medium, and high. I adapt the approach of Bustos, Caprettini, and Ponticelli (2016) to explore the cross-sectional variation in potential yield in a first differences model to identify the technological change effects. In particular, I use the differences in potential yield under the different levels of input use to capture the marginal value of agricultural intensification. The potential yield measures used are for rain-fed agriculture and represent 30-year averages defined in 0.5° by 0.5° grid cells (FAO/ IIASA, 2017). I integrate the potential yield measures with the census panel by using a geographical information system to determine the potential yield at the census block level. The GAEZ dataset of potential yields has also been used in studies of agricultural development (Nunn and Qian, 2011) and climate change impacts in agriculture (Costinot, Donaldson, and Smith, 2016). For additional information about the GAEZ dataset, see the model documentation (Fischer et al., 2012).

In addition to the agricultural census, IBGE conducts annual surveys of producers of major crops to track the quantity produced, area planted and harvested, and average producer price (hereafter referred to as spot prices) at the municipality level (IBGE, 2002; Hausman, 2012). Data collectors from IBGE are allocated to different municipalities to track the output and price variables monthly and compute weighted average estimates for the annual productivity and spot prices in each municipality, generating a publicly available panel dataset that spans 1990 to 2014 (IBGE PAM, 2015). Hausman (2012) also used the IBGE PAM agricultural dataset to estimate sugarcane and soy price elasticities in Brazil. I use the PAM panel of spot prices to compute changes in expected prices before and after the macroeconomic reforms.

My preferred measure of the expected soy price is the six-year average spot price before the census survey because the goal is to assess the long-run effect of permanent price changes. This long average minimizes annual variation, especially during the mid-1990s when the Brazilian economy was unstable. Figure 4 shows the distribution of changes in expected prices for the two major soy-producing regions in Brazil. The large changes in soy spot prices are the direct result of the introduction of the Kandir law in 1996 and change in exchange rate policy in 1999. There is a large variation in price changes, measured as differences in the log of expected prices, in both the savanna and the south regions. However, as expected, the price changes are significantly higher in the savanna given its high transportation costs. The estimated price effects are robust to alternative measures of the expected soy price and to the addition of controls for soy price volatility, measured in terms of weighted deviations from average prices (Hausman, 2012).

Theory: Heterogeneity in the Long-run Acreage Response Function

A permanent increase in commodity prices can affect the farmed acreage through (i) a direct channel, namely the reallocation of land from less profitable activities, and (ii) a technology adoption channel, namely the conversion of marginal land through the adoption of new production systems. Both channels can increase heterogeneity in the acreage response function. The farm acreage response function is $A_{ijt} = A(P_{ijt}, T_{ijt}(P_{ijt}, U_{ij}))$, where i, j, and t indexes the farm, crop, and year. A_{ijt} is the area harvested, P_{ijt} is the farm-gate spot price, T_{ijt} is the production technology, and U_{ijt} is a measure of land quality for the production of crop j. I omit the crop subscript j to simplify the notation, as we focus on soy in this article. The price effects are

$$(1) \quad \frac{\partial A_{it}}{\partial P_{it}} = \underbrace{\frac{\partial A_{it}}{\partial P_{it}}}_{Direct\ Price\ Effect} + \underbrace{\frac{\partial A_{it}}{\partial T_{it}} \frac{\partial T_{it}}{\partial P_{it}}}_{\substack{Technology\ Adoption \\ Price\ Effect}}$$

Both terms on the right-hand side of equation (1) are functions of land quality U_{ij}. The first term captures the standard land-use substitution effect. The second term represents the indirect price effect through technological adoption, and this will vary with the availability of new production technologies and the return to adoption for land of quality U_{ij}. Figure 5 illustrates the heterogeneity of the soy acreage response function in the two regions of Brazil. For simplicity, the acreage response function was initially identical in both regions. The direct price response is represented by the expansion $\Delta A = A_1 - A_0$ given the price increase $\Delta P = P_1 - P_0$. In the south, the price change incentivizes the adoption of GE soy, thereby shifting the acreage function to the right. The final expansion is then $\Delta A = A_2 - A_0$ and the acreage response function A_{South} is more elastic than the original acreage function.

In the savanna, the increase in prices stimulates the adoption of capital-intensive NF soy, designed for the conversion of marginal land for soy production. The indirect effect is larger because the adoption of NF soy enables the conversion of large parcels of land. The soy expansion in the savanna is $\Delta A = A_3 - A_0$, and the savanna acreage response function, $A_{Savanna}$ is more elastic. Hence, the indirect technological adoption channel explains the heterogeneity in the acreage response functions given the variation in technologies.

Long-run Acreage Response Identification Problem

How does heterogeneity in the technological adoption process affect the identification of the long-run acreage functions? In the long run, technological change is the main driver of agricultural supply and both the profitability of technological change and the set of technologies available to farmers vary spatially. To examine this identification problem, I assume linear approximations for the acreage response and technology adoption equations. Let the acreage response function be represented as $A_{it} = \delta + \beta P_{it} + \theta T_{it} + U_i + \epsilon_{it}$ and the technology equation as $T_{it} = \alpha(U_i) + \gamma(U_i)P_{it} + v_{it}$. Technology T_{it} indexes the soy production system in a simplified linear probability model with random coefficients. Land quality U_i is observed by the farmer but not by the econometrician. The farmer will choose the most profitable technology for his/her farm based on U_i. α, β, γ, and δ are parameters and ϵ and v error terms. By plugging the technology equation into the acreage equation and rearranging the terms, we obtain the reduced-form acreage response function typically used in the empirical analysis of agricultural supply:

(2) $$A_{it} = \delta + [\beta + \theta\gamma(U_i)]P_{it} + [\theta\alpha(U_i) + U_i] + \theta v_{it} + \epsilon_i$$

The heterogeneous effect of prices on technological adoption due to the variation in land quality is $\tilde{\beta} = \beta + \theta\gamma(U_i)$. The price elasticity of acreage increases with the effect of a permanent price increase on technology adoption, γ, and the effect of the new production system on the relative profitability of soy production, θ. For example, in the Brazilian savanna, the adoption of NF soy increases the relative profitability of soy compared with extensive grazing, a large θ, resulting in a large price effect, $\tilde{\beta} \gg \beta$. Equation (2) is typically estimated by using panel datasets such that the time-invariant term $\theta\alpha(U_i) + U_i$ is differenced out of the model. However, if U_i is unobserved, the error term in equation (2) is $\theta\gamma(U_i)P_{it} + \theta v_{it} + \epsilon_i$, which is a function of prices. In this case, using instrumental variables for the price such as exogenous demand shocks would not identify the price effects, except when prices do not affect technology adoption, $\theta\gamma(U_i) = 0$. I propose estimating the long-run price effect by explicitly modeling the indirect price effect $\theta\gamma(U_i)$, using the exogenous spatial variation in potential yields as a proxy for $\gamma(U_i)$.

Differential Price Effects with the Diffusion of GE and NF Soy Technologies in Brazil

I rewrite equation (1) by using the first-order condition for the farmer's land allocation problem to derive empirical implications for the soy expansion with technological adoption. The price response of the soy expansion can be represented by

$$(3) \quad \frac{\partial A^*}{\partial P} = \Psi(A^*)\left[\underbrace{\frac{\partial\, \Pi^*(A^*, T)}{\partial P}}_{Direct\ Price\ Effect} + \underbrace{\frac{\partial\, \Pi^*(A^*, T)}{\partial T}\frac{\partial\, T}{\partial P}}_{\substack{Technology\ Adoption\\ Price\ Effect}}\right]$$

I derive Equation (3) by applying the implicit function theorem to the optimal condition of land-use conversion, $\Pi^*(\, A^*, T\, , P) - OC\, (\, A^*, T\,) = 0$. The farmer converts land to soy production to equate the marginal benefits and costs of conversion. Π^* is the optimized farmer profit function, OC is the opportunity cost of the land, A^* is the optimal acreage converted into soy production and $\Psi(A^*) = \frac{1}{\frac{\partial\, \Pi^*}{\partial A} - \frac{\partial\, OC}{\partial A}}$ captures the variation in the productivity of land at the optimal acreage A^*. I use the three components on the right-hand side of equation (3) to explain the different price responses in Brazil.

Implication 1 (NF Soy): *The combination of a large endowment of land in the savanna and the diffusion of capital-intensive NF soy technology generates a large acreage* *response to price increases.* In the savanna, $\Psi(S^*)$ is large. Land productivity decreases slowly as marginal land is converted to soy production through the diffusion of NF soy. The indirect price effect is large for the adoption of NF soy because the technology requires a large capital investment in soil correction. The direct price effect is small because the savanna soil is unsuitable for soy production without the NF technology.

Implication 2 (GE Soy): *Labor-saving GE soy will mostly replace traditional soy plantations on productive land in the south of Brazil and will generate a small acreage response to price increases.* Land in the south is more suitable for agricultural production than that in the savanna and the variation in soil productivity is larger, implying a small $\Psi(A^*)$. Both the direct and the indirect price effects in the case of GE soy are small because the technology does not require capital investment and is cost-saving. GE soy would thus diffuse rapidly – even without an increase in soy prices. At the same time, the savings on labor expenditure are unlikely to be sufficiently large to convert the land used for other crops to soy production. An exception could be the substitution of grazing land for soy production in the southern frontier of Brazil.

Empirical Framework: Long First Differences Model

To estimate the long-run effect of a permanent price shock on the soybean expansion, we would ideally track identical farms subject to different permanent price shocks over a period sufficiently long for farmers to change their production technologies; we would then contrast farmers' land-use decisions. In this analysis, I use a long first differences model that approximates this ideal experiment. I explore a permanent price shock resulting from macroeconomic reforms in a large country in which farmers have different technological options. I then observe a farmer's choices before and after the price changes by using a long panel dataset with two periods 10 years apart. Furthermore, I observe the diffusion of two technologies across the country to quantify the price effect through the technological change channel. To estimate the price effects, I explore the spatial variation in price shocks due to large differences in transportation costs.

The level equation for the first differences model is

$$(4)\quad A_{it} = \delta_0 + \delta_1 t + \delta_2 t d_{UF} + \beta_1 t PY_i + \beta_2 P_{MU,t} + \beta_3 PY_i P_{MU,t} + \beta_4 X_{it} + \beta_5 X_i + v_i + \epsilon_{it}$$

where $t = 0, 1$ indexes the agricultural census years of 1996 and 2006, i indexes the census block, and MU indexes the municipality. A_{it} is the area harvested with soybeans in census block i in year t and P_{MUt} is the expected price in municipality MU. d_{UF} are dummy variables for states (referred to as UFs in Brazil). PY_i represents potential soybean yields under the maximum and medium levels of input use. These potential yield measures are time-invariant. The time-invariant observed and unobserved heterogeneity, X_i and v_i, are differenced out. Examples of time-invariant characteristics are soil and climate attributes, geographical features, historical land use, historical drivers of agricultural productivity, and unobserved land quality at each census block. X_{it} is a vector of the time-variant characteristics.

The estimated baseline model is the first differences equation:

$$(5)\quad \Delta A_i = \delta_1 + \delta_2 d_{UF} + \beta_1 PY_i + \beta_2 \Delta P_{MU} + \beta_3 PY_i \Delta P_{MU} + \beta_4 \Delta X_i + \Delta\epsilon_i$$

The parameter of interest is the average partial effect of prices on the soy harvested acreage: $\mu(PY_i) = \frac{\partial \Delta A_i}{\partial \Delta P_{MU}} = \frac{\partial A_{it}}{\partial P_{MUt}} = \beta_2 + \beta_3 PY_i$. In a first differences model, the cross-sectional variation in potential yield captures the effect of technological change, as the return from technology adoption, the $\frac{\partial\, \Pi^*(A^*,T)}{\partial T}$ term in equation (3), varies with the potential yield of the land (Foster and Rosenzweig, 1996; Bustos, Caprettini, and Ponticelli, 2016). The intuition is that the available set of production technologies expanded and that land with higher returns to intensification has a comparative advantage for the adoption of new technologies. The coefficient of the interaction term between the price and potential yield, β_3, captures the differential price response due to a technological diffusion process. I estimate equation (5) separately for the south and savanna regions, using measures of potential yield associated with the profitability of the NF and GE soy technologies to approximate the indirect price effect through the technological adoption channel. I use the estimated price effect to compute the long-run price elasticities of the soy expansion. The coefficients δ_1 and δ_2 capture the trends in the soy expansion at the country and state levels. I also model trends by the baseline characteristics such as historical land productivity and profitability to capture the differentiated trends correlated with the opportunity cost of the land. The coefficient β_1 captures the technological change effect similar to the parameters estimated by Bustos, Caprettini, and Ponticelli (2016).

The key identifying assumption for the first differences estimator with two time periods is

(6) $E(\Delta\epsilon_i \mid d_{UF}, PY_i, \Delta P_{MU}, \Delta X_i) = 0$

The identification assumption in the long first differences acreage model is different from the assumption for the fixed-effects model estimated with a panel of annual agricultural output. The endogeneity problem in the fixed-effects model is the incorporation of anticipated yield shocks in expected prices (Roberts and Schlenker, 2013; Hendricks, Janzen, and Smith, 2014). Roberts and Schlenker (2013) propose an instrumental variable approach based on storage shocks to estimate a global agricultural supply function. Hendricks, Janzen, and Smith (2014) show that controlling for current yield shocks or estimating an acreage response function mitigates the endogeneity problem. In a 10-year first differences model, it is unlikely that the expected prices in 1996 incorporate information about yield shocks 10 years later.

The long first differences model uses the longitudinal and cross-sectional variation to identify the price effects. In this case, unobserved changes in local policies, local infrastructure, or local market structures could be correlated with permanent changes in local prices. I expect the baseline characteristics and the state fixed effects to capture most of the variation in the local trends for two reasons. The first is because agricultural policies and public infrastructure projects in Brazil are designed at the federal or state level. The second is because the expansion of the soy market in Brazil from 1996 to 2006 was driven by the export market and not by changes in local markets. For example, Brazilian soy exports almost tripled between 1996 and 2006. Another potential source of endogeneity would be concentration of investors such as trading companies in particular locations generates spatial variation in demand for soy. In this case, we could have reverse causality. This is unlikely in the Brazilian case for two reasons. First, we know the specific policy changes that affected soy prices and how the interaction of the policies with the distance to market generated a large spatial variation in price shocks. Second, it is clear that the profitability of soy trading companies that invested in Brazil during this period depended on the scale of trading, as profit margins were small. Trading companies operated in multiple locations, and there is survey evidence of competition among trading companies in local markets of the agricultural frontier (Rezende, 2008).

Finally, part of the soy expansion could be a result of immigrant farmers purchasing cheap land in the agricultural frontier. I use the characteristics of the municipality and proxies for the profitability of land before 1996 to control for these differentiated expansion trends. It is not possible to completely eliminate omitted variable bias or reverse causality. Furthermore, there could be measurement error in our permanent price shock variables. I then instrument the expected prices with the minimum straight-line distance to market (Souza-Rodrigues, 2015).

Variation in Permanent Price Changes with the Distance to Market

As the long first difference model relies on the cross-sectional variation in prices to estimate the price effects on the soy expansion, I use a simple price model to show how the interaction between the macroeconomic policies introduced in Brazil after 1996 and different transportation costs created spatial variation in soybean price increases. See Appendix B for a derivation of the price model. I assume that the farmer makes land-use decisions based on the expected prices at the farm gate. The farm gate price at time t is $P_t^{farm} = (1 - tax_t)(e_t P_t^{mkt} - c_t d)$. The tax, tax_t, exchange rate, e_t, market price,

P_t^{mkt}, and unit cost of transportation (cost to transport a ton of soy per kilometer), c_t, can vary over time. d is the distance to market. By way of a simplification, I assume that the transport cost, c_t, and global soy price, P_t^{mkt}, do not change over time. The two policy changes that generated a permanent price increase in Brazil are modeled through changes in tax_t and e_t. By using the differences in logs as an approximation for the percentage change in prices, I decompose the price shock into three components

$$(7) \quad \Delta P \cong Log\left(P_1^{farm}\right) - Log\left(P_0^{farm}\right)$$

$$= Log\left[\frac{1-tax_1}{1-tax_0}\right] - Log(1 - S_{TC}) + Log(1 + g - S_{TC})$$

where g is the growth rate in the market price, defined as $e_1 P_1^{mkt} = (1+g)e_0 P_0^{mkt}$, while S_{TC} is the transportation cost share of the market price, $S_{TC} = cd/e_0 P_0^{mkt}$. The first term on the right-hand side of equation (7) represents the effect of the tax change and does not vary with distance in this simplified model. In reality, the tax changes implemented in Brazil affect transportation services and vary across states, inducing further spatial price variation. I ignore this tax effect here for simplification. For a tax reduction from 20% to 10%, the price increase will be approximately 5%. The elimination of ICMS tax for exports would thus result in a price increase of almost 10%.

The second term, $-Log(1 - S_{TC})$, is a function of the transportation costs and varies with distance. Since the transport cost share is below one, this term will always be positive and will increase nonlinearly with the transport cost share S_{TC}. For example, for a transport cost share of 30%, the change in prices will be approximately 15%. If the transport cost share in frontier locations with poorly maintained roads reaches 50%, the resulting price change would be 30%.

The last term in equation (7), $Log(1 + g - S_{TC})$, varies with both the growth rate in the market price and the distance to market. This term would be negative only if the transport cost share is greater than the market price growth rate. I expect this term to be positive because in the 10-year period analyzed in this study, the growth rate is larger than the transport cost share in most locations due to the large change in exchanges rates in Brazil. The derivative of the price change ΔP with respect to the transport cost share, S_{TC}, is $\frac{\partial \Delta P}{\partial S_{TC}} = \frac{g}{1+g-S_{TC}} > 0$. As expected, the price shock increases nonlinearly with transportation costs and therefore the distance to market. In farms located close to the port, where transportation costs would be close to zero, a 60% increase in prices due to currency devaluation would result in an approximately 37% increase in farm gate prices, not considering the effect of changes in taxes. For a farm located in the agricultural frontier where the transport cost share could reach 50%, the same 60% increase in market prices would result in a price increase of approximately 55%. This spatial variation is likely to be higher because the quality of transportation infrastructure decreases nonlinearly with the distance to market. Transportation costs tend to increase rapidly with the distance to the agriculture frontier. The large spatial variation observed in the data is thus consistent with the variation implied by the price model.

Decomposing the Price Effect on Acreage Expansion

Finally, I extend the long first differences model to decompose the price effect into the direct effect and the indirect technological adoption effect (equation (3)). The expected change in acreage can be expressed in terms of the probability of expansion: $E[\Delta A_i|X_i, \Delta P_{MU}] = Pr[\Delta A_i > 0|X_i, \Delta P_{MU}] \times E[\Delta A_i|X_i, \Delta P_{MU}, \Delta A_i > 0]$, where the vector X_i includes all the observed farm characteristics. I estimate separately the choice to expand, which in the case of the savanna implies the adoption of NF soy, and the acreage expansion model conditional on expansion. What differentiates the decision to expand soy production in the savanna from the choice of total acreage converted into soy production is the significant upfront investment necessary to prepare savanna soils for farming. Hence, this upfront cost of adopting NF soy represents an economic hurdle for the adoption of NF soy technology. However, if the permanent price increase is sufficiently large to overcome this adoption hurdle, large amounts of land could be converted into soy production through the technological adoption channel. Differentiating the unconditional expected acreage function, $E[\Delta A_i|X_i, \Delta P_{MU}]$, with respect to the price changes, ΔP_{MU}:

$$(8)\quad \frac{\partial E[\Delta A_i]}{\partial \Delta P_{MU}} = \underbrace{\Pr[\Delta A_i > 0]\frac{\partial E[\Delta A_i|\Delta A_i > 0]}{\partial \Delta P_{MU}}}_{Direct\ Price\ Effect} + \underbrace{\frac{\partial \Pr[\Delta A_i > 0]}{\partial \Delta P_{MU}} E[\Delta A_i|\Delta A_i > 0]}_{\substack{Technology\ Adoption \\ Price\ Effect}}$$

I omit the conditioning variables X_i and ΔP_{MU} in equation (8) to simplify notation. The first term on the right hand side captures the price effect on the acreage expanded in census blocks that expanded production. The second term captures the technology adoption price effect through the price effect in the probability of soy expansion. In the savanna, we expect a large price effect through the technology adoption channel. However, the two-step hurdle model will not work well in the south because the adoption of GE soy does not require a significant upfront investment. I use the hurdle model to decompose the price effects on the soy expansion in the savanna.

I estimate the hurdle model in two steps. In the first step, I estimate a Probit model for the choice of expansion; in the second step, I estimate the conditional acreage response model. This selection model relies on a parametric assumption and exclusion restrictions for identification. The ideal exclusion restriction would capture the variation in the upfront investment in soil correction for technology adoption. Two plausible candidates are the distance to lime mines and share of forested land before 1996. Because lime is used in large quantities to correct soil acidity in the savanna, farmers located further from this input source will incur a large upfront cost. The conversion of forested land to soy production is particularly expensive because the process of cleaning the land is costly. The exclusion restriction assumption is conditional on a set of farm characteristics including measures of land suitability for soy production and state fixed effects.

Empirical Results: Heterogeneous Supply Response

The key empirical results of this article are threefold. First, the regional acreage response functions are different, becoming more elastic towards the less developed agricultural frontier. Second, the heterogeneous long-run supply response is a result of the different technological diffusion processes. Finally, the indirect price effect through the technological channel explains most of the heterogeneity in the long-run supply response. The empirical evidence suggests that forecasting the long-run agricultural supply response requires explicitly modeling the technological change. Although modeling the technological change beyond time trends is challenging, I find that using potential yield measures to proxy for the return to innovation can improve the supply analysis. I start the empirical analysis with a graphical illustration of the acreage response function in Brazil. I show how the different measures of potential yield explain the GE and NF soy diffusion processes. I then estimate separately the acreage response functions for the south and savanna regions.

Figure 6 shows the fitted values for a non-parametric acreage response function based on the long first differences model of equation (5). Each dot represents one census block in the two main soy-producing regions of Brazil. Appendix D describes the sieves first differences model. The concave sieves acreage response function flattens as we move from the developed south region to the agricultural frontier in the savanna. The more inelastic part of the function corresponds to the soy expansion in the southern and eastern parts of the country where GE soy diffused rapidly. As these regions are located closer to ports, the price changes were relatively smaller. By contrast, the function flattens as we move further inland towards the agricultural frontier of the savanna. The set of census blocks furthest to the right are from the state of Mato Grosso, the largest soy-producing state in Brazil, in the frontier of the savanna and Amazon biomes. Although the sieves acreage response function in figure 6 shows the non-linear nature of the soy supply response function in Brazil, it is highly elastic because I estimate it for the entire country without geographical fixed effects that capture differentiated trends at the state or mesoregion level.

Modeling the Return to GE and NF Soy Adoption

I incorporate the technological diffusion process into the estimation of the acreage response function by using two potential yield measures that capture the return to adopting GE and NF soy. The GAEZ potential yield measures are constructed independently of Brazilian farmers' actual choices of production system and therefore are plausible sources of exogenous variation in the return to agricultural intensification. The maximum potential yield measure, PY (high-medium), captures the natural soil suitability for soy production. PY (high-medium) is the marginal value of adopting advanced soy production systems and is a robust measure for the return to GE soy adoption in Brazil because GE soy was designed for highly productive soy plantations when introduced in 1996. By contrast, because the NF soy technological package was designed for marginal land in the savanna, PY (high-medium) does not capture the return to NF adoption. However, PY (medium-low) is negatively correlated with the nitrogen content of soil, and thus it captures the marginal value of adopting the nitrogen-fixing technology on savanna land[3].

Farmers choose land-use and production technology jointly. Table 2 shows the results for a joint choice model of soy expansion and soy technology. The soy expansion, ΔA, is the difference between the areas harvested with soy in 2006 and 1996. The GE soy technology is identified by using the type of seed variable from the IBGE census and the

NF technology is based on the technological package used for NF soy production in the savanna. Trading companies and input suppliers offer a technological package that includes technical assistance, debt, certified seeds, pH correction inputs, and soil management practices. The IBGE census contains variables for each component of the NF soy package. Columns 1 to 4 in Table 2 show the results for a Biprobit model for the choices of the soy expansion and adoption of GE soy in the south. Columns 5 to 8 show the results for the soy expansion and NF soy adoption in the savanna. I estimate regional models controlling with fixed effects and controls for baseline characteristics. Columns 1 and 2 for the south model and columns 5 and 6 for the savanna model use state biome fixed effects, whereas columns 3, 4, 7, and 8 use mesoregion fixed effects.

The choices of soy expansion and technology are strongly correlated as expected. The correlation between the error term in the two choice models, the rho coefficient in table 2, ranges from 0.77 to 0.87 and is statistically significant in all models. It is thus not possible to model separately the long-run acreage change from the technological adoption process. Moreover, standard supply response models implicitly assume that the set of available technologies and the return for different technologies are homogeneous across large geographical regions. This assumption is particularly problematic in the case of technologies designed for the conversion of marginal land to agriculture, such as the NF soy technology.

The most robust result in the Biprobit models as well as with the alternative specifications of technology choice in Brazil using the IBGE census data is the effect of the potential yield measures in the choice of GE and NF soy. PY (high-medium) is a robust predictor of the choice of adopting GE soy and expanding soy production in the south of Brazil and PY (medium-low) is a robust predictor of NF soy adoption in the savanna. PY (high-medium) is statistically insignificant in most specifications for the choice of NF soy. I tested different functional forms for the potential yield measures, including linear, quadratic, and dummy variables for the different quantiles of potential yield. The evidence from the choice and expansion models suggests a non-linear relationship between the potential yield measures and expansion and technological outcomes.

Finally, the Biprobit results presented in Table 2 show that the price effect in the soy expansion and GE soy adoption in the south is statistically insignificant. In the savanna, when we add the mesoregion fixed effects, the price effects also become statistically insignificant. The inclusion of a rich set of control variables and fixed effects absorbs much of the variability in price changes, measured at the municipality level. In the next sections, I estimate separate regional acreage response models by using the interaction effects between price changes and potential yields under different specifications and with the distance to the port as an instrument for the price changes. I find that after accounting for census block fixed effects, controlling for trends based on land productivity and the productivity of alternative crops, and controlling for geographical trends by using state and mesoregion fixed effects, the interaction terms between price and potential yield are robust across specifications.

Acreage Response Function with GE Soy Adoption in the South

I estimate the acreage response function for the south region by using the first differences model of equation (5). My preferred dependent variable is the change in the soy harvested area share of the census block area; however, I also use the model to estimate the change in the soy harvested area, change in the GE soy harvested area, and changes in the agricultural acreage share. The parameter of interest is the coefficient of the change in the log of the expected price variable, $\Delta Log(EP)$, and the interaction between the changes in price and dummies for the quantiles of the potential yield measures, $Dummy\ PY$. Table 3 shows the estimated acreage response function for the south, using the mesoregion fixed effects in all the specifications to capture trends at the mesoregion level. Brazil is divided into 137 mesoregions. The south region is the smallest in the country with only three states, 23 mesoregions, and 7,591 rural common census blocks.

Columns 1 and 2 in table 3 show the acreage response function without interactions between the price and potential yield. Column 1 uses a quadratic function for the potential yield variables, whereas column 2 uses the dummy variables. The functional form of the potential yield variables does not affect the estimated price effect. A 1% increase in the expected soy price increases the share of the soy harvested area in the census block by 3%. To compute the long-run elasticities, I first convert the shares into the changes in the soy harvested area by multiplying the coefficient of $\Delta Log(EP)$ by the average census block area in the south, 4,221 hectares. I divide the resulting expected change in soy acreage by the mean soy harvested area for 1996 and 2006, 582 hectares. This calculation is equivalent to multiplying the coefficient of $\Delta Log(EP)$ by 7.25.

I find a long-run price elasticity of 0.25, about one-third that estimated by Hausman (2012), using the municipality data and annual variation in spot prices from the IBGE PAM dataset. Hausman also estimates regional elasticities for Brazil by using US-traded futures prices, finding a much more elastic acreage response; however, none of the regional elasticities is statistically significant[4]. In this article, we refer to the Hausman elasticities based on spot prices using the same price dataset from IBGE. Without the mesoregion fixed effects, my estimated price elasticity for the south increases to 0.49.

Columns 3 to 5 present the results for the acreage response function with interactions between the price change and potential yield. The parameter of interest is the interaction term between the change in price and dummies for the two top quantiles of the maximum potential yield, PY (high-low). This coefficient represents the differentiated price effect on land with higher returns for the adoption of GE soy and is robust across the specifications with the mesoregion fixed effects. The estimated elasticity for the top quantile of the maximum potential yield ranges from 0.51 to 0.62, twice as large as the regional elasticity[5]. Column 5 shows the results when I instrument the price changes with a quadratic function for the straight-line distance from the municipality to the port. The estimated elasticity, 0.46, is similar to the elasticity estimates when using OLS with the mesoregion fixed effects. The estimated IV model did not use the mesoregion fixed effects, as the instrument has no effect on price changes after accounting for the mesoregion trends.

In the first differences model, the coefficient for the time-invariant potential yield variables capture the effect of the introduction of a new technology. In the case of the south region of Brazil between 1996 and 2006, this represents the introduction of GE soy. This approach was used by Bustos, Caprettini, and Ponticelli (2016) to analyze the introduction of GE soy in Brazil. The coefficient of PY (high-low) Q2 in column 3 suggests a technological change effect of a 1.3% increase in the soy harvested area share for land in the second quantile of the maximum potential yield after accounting for the price effect. The other coefficients for the potential yield dummies are not statistically significant. When I add the interaction effect between the price and potential yield, the coefficients for the PY (high-low) dummies and price effects are no longer statistically significant. It is therefore not clear whether the introduction of GE soy had any direct effect on the soy expansion without taking into consideration price increases. To disentangle the direct and indirect price effects, I model the expansion of the GE soy harvested area.

In the south, from 1996 to 2006, the soy acreage and GE soy acreage increased by 1.3 and 2.6 million hectares, respectively and 50% of all soy planted in the south was GE soy by 2006. At least half of the diffusion of GE soy happened through the conversion of traditional soy plantations. Table 4 shows the estimated soy and GE soy expansion models. The dependent variable is the change in the total area harvested with soy and GE soy. The change in the area harvested with GE soy is equal to the area harvested with GE soy in 2006, as GE soy was only approved for commercialization in Brazil in 2001. I find a large and robust price effect on the diffusion of GE soy – even with the mesoregion fixed effects model. This effect corresponds to the $\frac{\partial T_{it}}{\partial P_{it}}$ term in equation (1), namely the price effect of the adoption of a new agricultural technology. The estimated price elasticity of the GE soy expansion ranges from 1.11 to 2.41 and is at least twice as large that the price elasticity of the total soy expansion. The maximum potential yield dummies have a significant and robust effect on the total soy and GE soy expansions, but the effect is three times larger for the latter.

In summary, these results suggest that the introduction of GE soy in combination with the price increase led to a large conversion of traditional soy plantations to GE soy in the south. However, this large and robust price effect on such technology diffusion did not translate into a large effect in the expansion of soy into new agricultural land. I find that the elasticity of the acreage response function not only varies with the potential yield, but is also only statistically significant in locations with the highest potential yield for soy production. The GE soy technology was not designed to and did not extend the agricultural frontier in Brazil up to 2006. In traditionally productive regions such as the south of Brazil, the indirect price effect in the acreage response function thus tends to be small.

Acreage Response Function with NF Soy Adoption in the Savanna

The acreage response function is more elastic in the savanna than in the south because of the indirect effect of price on the soy expansion through the technological change channel. Whereas in the south the indirect price effect is small, in the savanna it accounts for most of the supply response. The permanent price increase enabled the adoption of an expensive technology, namely NF soy, designed for the conversion of marginal pastureland into productive soy plantations.

Table 5 shows the results of the first differences model estimated for the savanna. The model is the same as that used for the south; however, the potential yield measure that captures the return to the adoption of NF soy is the medium potential yield, PY (medium-low). Columns 1 to 3 show the results for the models without price and potential yield interactions. Model 1 uses a quadratic functional form for the potential yield measures, whereas model 2 uses dummy variables. The estimated price elasticity is approximately 1.15. Both models include the log of farm size in 1996 and the log of the value of agricultural production in 1985 as control variables. Model 3 also includes the log of population density and the log of income per capita in 1985 to capture the differentiated trend based on the variation in local market demand. The elasticity decreases to 0.6 in model 3 when I add these baseline variables. In the savanna region, the mesoregion fixed effects are too restrictive, and thus they absorb the price variation. However, the interaction coefficients of the price and medium potential yield are statistically significant and robust across the specifications in those models with mesoregion fixed effects. The estimated elasticities are not significantly different from zero for the mesoregion fixed effects. These results show the nonlinear relationship between price changes and distance. As all the locations in the savanna region are far from ports, the variation in price changes within this region is small.

Columns 4 to 6 show the results for the models with interaction terms between the price change variable and dummies for the medium potential yield. For the savanna models, I only use dummies for the medium potential yield variables; however, all the specifications control for the maximum potential yield. The results are robust when using the linear, quadratic, and dummy specifications for the maximum potential yield. The price–potential yield interactions are statistically significant across all the specifications for the two top quantiles of the medium potential yield. The acreage response functions are more elastic in the top quantiles of the medium potential yield, particularly in the second quantile. The estimated price elasticity varies from 1.2 to 2.18 for the second quantile and from 0.74 to 1.21 for the top quantile.

Model 6 is estimated by using the straight-line distance to the port as an instrument for price changes. The key assumption in the IV model is that the distance to the port is unrelated to unobserved changes affecting the soy expansion, possibly due to the unobserved productivity of a demand shock. This assumption is more plausible when conditioning for the baseline characteristics that control for variation in land productivity and local demand. The larger elasticity estimated by using the IV model could reflect the attenuation bias due to the measurement error in the price change variable.

I also estimate the acreage response function for the Midwestern part of the Brazilian savanna, which accounts for most of the soy production in Brazil, reaches the Amazon frontier, and includes the states of Mato Grosso, Mato Grosso do Sul, and Goias. The state of Mato Grosso is the largest soy producer in Brazil and has land in both the savanna and the Amazon biomes. Because I do not separate the Amazon part of the Midwest savanna in the estimation, the elasticity reflects the price response in the agricultural frontier of the savanna and the savanna's Amazon border. The estimated elasticities in models 7 and 8 are three times larger than the price elasticities for the entire savanna region. Models 7 and 8 both include the baseline characteristics and model 8 includes the state fixed effects; nonetheless, the results are unaffected by the addition of these state fixed effects. The large price elasticities in the agricultural frontier of the savanna are consistent with the nonparametric sieves acreage response function shown in figure 6.

Hausman (2012, table 8) estimates a short-run spot price elasticity for the Midwest region of 0.99, statistically significant at the 1% level, and a long-run spot price elasticity of 2.31 (not statistically significant). For the Amazon border, Hausman (2012) estimates a short-run spot price elasticity of 0.703, statistically significant at the 5% level, and a long-run spot price elasticity of 2.032 (not statistically significant). The long-run price elasticities I estimate for the Midwest savanna are naturally larger than Hausman's short-run elasticities, but also larger than the long-run elasticities estimated for the Midwest and Amazon frontier. The time-series model captures the short-run effect of price changes on acreage well; however, the simulation of the long-run elasticities from the short-run price changes might miss the longer and more capital-intensive process of converting marginal land into agriculture. Furthermore, the aggregated estimates of the supply elasticities also average out the differentiated responses across land with different suitability to new technologies. The estimated elasticities from the time-series and long first differences models tend to converge in traditional agricultural regions such as the south, whereas the time-series model is likely to underestimate the long-run elasticity in the agricultural frontier driven by a specific technological diffusion process.

Finally, I separate the direct and indirect price effects in the soy expansion in the savanna region by using a selection model and the price effect decomposition in equation (8). The indirect price effect, the second term on the right-hand side of equation (8), is calculated by multiplying the marginal price effect on the probability of the soy expansion by the expected expansion conditional on soy expansion. I use a Probit model for the choice of the soy expansion in the savanna, assuming that all soy production on savanna land requires the use of NF soy owing to its acidic, nitrogen-deficient soil. The nonlinear Probit model generates heterogeneous price responses according to the land characteristics. My preferred Probit model is the equivalent of model 3 in table 5. I use a quadratic function for the potential yield measures and control for the baseline characteristics. Figure 7 shows the technological change component of the long-run price elasticity as a function of the medium potential yield.

The price elasticity doubles as the medium potential yield increases. A permanent increase in soy prices results in a larger soy expansion on land with a higher marginal value of NF soy adoption, as the medium potential yield measure captures the variation in the return to NF soy adoption. In the savanna, the indirect price effect shown in terms of the price elasticities in figure 5 accounts for most of the effect of prices on the soy expansion. I find that the change in the expected size of the soy expansion in the savanna is unresponsive to price increases. This result reflects the importance of economies of scale for soy production in the savanna. The farmer’s most important decision is whether to convert pastureland into soy plantation. When he or she decides to produce soy, the average acreage allocated to soy is more a function of the land characteristics than the market conditions. The savanna farmer explores the economies of scale of large soy plantations to compensate for the small profit margins due to the high transport costs and high capital expenditure incurred when converting pastureland and preparing the soil.

Land-use Changes

Agricultural supply elasticities are important to simulate the impact of biofuel and climate change policies on the conversion of natural vegetation into agriculture. Such policies can permanently increase commodity prices and thus affect farmers' land-use decisions across regions. I therefore use the long first differences model to estimate the effect of a permanent change in soy prices on the share of the land allocated to agriculture (i.e., cropland and pastureland). The dependent variable is the ratio of the share of the census block acreage allocated to cropland and pastureland. The average share of agricultural land in Brazil is 73% and this figure is consistent at the regional level[6]. Table 6 summarizes the price effect on total agricultural land for Brazil as well as for the south and savanna regions. All the models include the baseline characteristics. Further, models 1, 3, and 5 have the state fixed effects, whereas models 2, 4, and 6 have the mesoregion fixed effects. The elasticities are reported at the bottom of the table.

I find small but statistically significant price effects in the agricultural area share nationally and in the savanna region. The price effects are not statistically different from zero in the south region. In addition, when adding the mesoregion fixed effects, none of the price effects is statistically significant, although the estimates are comparable to the models with the state fixed effects. The price elasticity of agriculture acreage is 0.046 nationally and 0.128 in the savanna, both statistically significant at the 5% level (models 1 and 5).

I compare the estimated elasticities with those calculated by Barr and coauthors (2010) using time-series data at the country level. I report their estimated elasticities for two periods, 1997–1999 to 2001–2003 and 2004 to 2006, which combined match the 10-year period between the agricultural censuses (Barr et al., 2010, Table 18). My estimate for the price elasticity of agricultural land falls between the elasticities estimated by Barr and coauthors for these two periods. It is possible that the stricter monitoring of forestland or changes in credit constraints changed the total agricultural elasticity after 2004 (Barr et al., 2010). However, the most significant policy changes targeting deforestation were implemented after 2006 and the most land-intensive crops in Brazil such as soybean and corn are largely financed by private companies (e.g., trading and input suppliers). Furthermore, soy and corn acreage continued to grow rapidly throughout that decade. A methodological challenge, the measurement of pastureland, could also affect the estimation of agricultural elasticities.

The Brazilian agricultural census is suitable for estimating the elasticity of agricultural land because it has pastureland information. The available time-series data on agricultural land use in Brazil do not report pastureland. In the first differences model, I use the pastureland acreage at the census block level. Measuring pastureland is particularly important in Brazil because approximately 50% of all private land is allocated to managed and natural pasture. A common approach to address this data deficiency, as also used by Barr and coauthors, is to interpolate annual estimates of pastureland based on census data. I tested different interpolation methods using the IBGE PAM dataset and estimates for grazing intensity but found it difficult to reconcile the results with the observed aggregated crop acreage at the municipality level. The challenge is that the intensity of grazing is jointly determined with farmers' land-use choices and is heterogeneous across a large country such as Brazil. Hausman (2012) and Roberts and Schlenker (2013) estimate the price elasticities of crop acreage in Brazil by using time-series data without pastureland. The crop acreage elasticities should be larger than total agricultural elasticity as it is easier to convert pastureland than forestland into cropland. For example, Barr and coauthors (2010) estimate both elasticities and find that the crop elasticity is three times higher than the agricultural elasticity including pastureland.

The price elasticities of agricultural land are useful to simulate land-use changes in response to permanent price increases. However, the simulation must be analyzed cautiously as it assumes no additional technological change, land-use policy change, and it does not account for general equilibrium effects. I simulate the change in land use in Brazil and in the savanna, using the range of elasticities estimated from models 1 and 2 for the former and models 5 and 6 for the latter. I then multiply the simulated price increase, 10%, by the elasticities and by total agricultural land in Brazil and in the Midwest savanna in 2006 based on the census (204.5 million and 104 million hectares, respectively). In 2006, about 70% (76%) of all agricultural land in Brazil (the savanna) was pastureland (IBGE, 2006)[7]. At the national level, a 10% permanent price increase would raise the total agricultural area by between 520 and 936 thousand hectares. In the savanna, a 10% permanent price increase would increase the total agricultural area by between 761 and 953 thousand hectares. The difference in the simulated land-use change using the country and regional elasticities show the importance of accounting for spatial heterogeneity. The application of country-level estimates for the supply elasticity to land-use simulation understates the potential for land-use change in the most ecologically sensitive regions.

Although I find statistically significant land-use effects from the price increases, it is useful to provide a reference for these estimates. According to estimates from the Food and Administration Organization (FAO) and Brazilian Institute for Space Research (INPE), the cumulative deforestation in the Amazon from 1996 to 2006 was 19.2 million hectares (FAO, 2014). Hence, the simulated deforestation based on a 10% permanent price increase for the entire country represents only 5% of Amazon deforestation in those 10 years.

Robustness Checks

In this section, I report the results of the robustness tests for the spatial correlation and alternative measurements of permanent price chances. Appendix E presents the complete results for the alternative specifications of the supply response functions estimated in this analysis.

Spatial Correlation

A potential problem when estimating supply functions using micro datasets is spatial correlation in the unobserved determinants of the soy expansion. Neighboring farmers are likely to be subject to the same demand and supply shocks. Moreover, the technological and land-use choices of a farmer could influence the choice of his/her neighbors. The unit of observation in my two-period panel dataset is the common census block, which contains on average 40 farms. I test the statistical significance of the estimated price and technological change effects firstly by changing the clustering variable to allow for correlation in areas larger than the municipality and secondly by applying the standard error correction for spatial correlation proposed by Conley (1999) (see also Fetzer, 2014; Hsiang, 2010). Under Conley's approach, spatial dependence is modeled as a function of the distance between agents. Hence, in my case, this represents the straight-line distance between rural census blocks.

Table 7 compares the soy response function in Brazil by using municipality, microregion, and mesoregion as the clustering variables. The estimated coefficients of the price effects are statistically significant at the 1% level in all specifications without price potential yield interactions. When I add the interaction effects, the direct effect of price changes becomes statistically insignificant in the models with the state fixed effects. However, the coefficients of the price–potential yield interaction (second quantile) are statistically significant across all specifications.

Table 8 shows the estimated soy response function in Brazil, using Conley's standard errors. For models with Δ soy area share as the dependent variable, the price effects are statistically significant at the 1% level for the four cutoff distances tested. Conley's standard errors for the price effects increase with the cutoff distance when the dependent variable is Δ soy area but are still statistically significant at the 500 km threshold.

Measurement of the Permanent Price Change

I exploit two macroeconomic policy changes to measure the permanent increase in soy prices in Brazil to estimate the long-run soy acreage function. My preferred measure of the permanent price change uses long averages (i.e., six years) before and after the introduction of these two policies to separate the effect of permanent changes from the annual variation in prices. I test the robustness of the estimated long-run price effects to the alternative measures of price changes, varying the averaging period.

Table 9 presents the estimated price effects, using four measures of permanent price changes, namely the baseline measure and the alternatives ΔP1, ΔP2, and ΔP3. Appendix A describes the variables. Each line in the table corresponds to one acreage response function for Brazil. I report the estimated price effects for two dependent variables and for models with the baseline control variables and state fixed effects. The estimated price effects are economically and statistically significant across the alternative measures of price changes, dependent variables, and specifications. The only exception is the price effect estimated by using ΔP2 with the baseline characteristics and without the state fixed effects. However, the price effects for ΔP2 are significant and consistent with the other estimates in the specification with the state fixed effects.

The estimated price effects when using these alternative measures of price changes are consistent within each specification. A 1% permanent increase in soy prices would increase the share of the soy area by approximately 2% in each census block (column (3)). If we use the Δ soy area dependent variable, the soy acreage would expand by about

500 ha (column (6)). The price effects decrease as we add controls and fixed effects, with the exception of the models with the ΔP3 measure of price changes. ΔP3 uses averages for only three years before and after the policy changes and is therefore more sensitive to short-term variation in local prices across Brazil. The price effects estimated by using the ΔP3 measure are higher than those under the alternative definitions. The price effects estimated by using the baseline and ΔP1 and ΔP2 measures are not statistically different.

Conclusion

I use a novel agricultural dataset combining farm-level data for 1.5 million commercial farms from the 1996 and 2006 Brazilian agricultural census surveys (IBGE, 1996, 2006) to model the effect of a permanent price increase and two technological diffusion processes on the agricultural supply response function of soy in Brazil. I find that the heterogeneity in the long-run supply response is a result of the different technological diffusion processes and is explained by the indirect price effect on technology adoption. The permanent increase in soy prices enabled the adoption of an expensive technology, namely NF soy, designed for the conversion of marginal pastureland into productive soy plantations in the Brazilian savanna. The presented empirical evidence highlights the importance of modeling technological diffusion processes to predict the long-run agricultural supply response. The long-run response of agricultural supply is more elastic on the agricultural frontier where innovation is designed for the conversion of marginal land into productive farming. Technology adoption, however, can also change agricultural supply through intensification. For example, the recent expansion of soy/corn double cropping systems in Brazil will change the agricultural supply functions for these crops. Price increases could then lead to further agricultural intensification, which is the subject of ongoing research.

Endnotes

[1] The IBGE PAM panel dataset is publicly available and was used by Hausman (2012) to estimate the price elasticities of the soy acreage by using soy spot prices.

[2] Hausman's long-run price elasticities for the savanna are not statistically significant.

[3] The correlation between the PY (high-medium) and PY (high-low) potential yield measures is 0.94, meaning that the empirical results are robust to alternating these two variables. The correlation between PY (high-low) and PY (medium-low) is 0.32. Please see DePaula (2017) for a production function model of the potential yield measures.

[4] Table 8 in Hausman summarizes the regional price elasticities for soybean acreage in Brazil.

[5] The estimated elasticity for the top quantile is measured in relation to the first quantile of the maximum potential yield PY (high-low) and the first quantile of the medium potential yield PY (medium-low). In the south, there are 2,626 census blocks in the top quantile of PY (high-low) and the bottom quantile of PY (medium-low) and the mean maximum potential yield in these census blocks is 3.75.

[6] The census dataset surveys farmland only; therefore, only private land is included in the calculation of the share of the land allocated to different land uses. Public forestland is not accounted for in this analysis.

[7] The data for total agricultural land and pastureland are based on the 1.5 million commercial farms used in this analysis. There are over five million farms in Brazil and these totals might change if we included smaller farms.

References

Ainsworth, E.A., Yendrek, C.R., Skoneczka, J.A. and Long, S.P., 2012. Accelerating yield potential in soybean: potential targets for biotechnological improvement. *Plant, cell & environment*, *35*(1), pp.38-52.

Alves, B.J., Boddey, R.M. and Urquiaga, S., 2003. The success of BNF in soybean in Brazil. *Plant and soil*, *252*(1), pp.1-9.

Alves, E., Souza, G.D.S. and Rocha, D.D.P., 2012. Lucratividade da agricultura. *Revista de Política Agrícola*, *21*(2), pp.45-63.

Barr, K.J., Babcock, B.A., Carriquiry, M.A., Nassar, A.M. and Harfuch, L., 2011. Agricultural land elasticities in the United States and Brazil. *Applied Economic Perspectives and Policy*, *33*(3), pp.449-462.

Bustos, P., Caprettini, B. and Ponticelli, J., 2016. Agricultural productivity and structural transformation: Evidence from Brazil. *American Economic Review*, *106*(6), pp.1320-65.

Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models. *Handbook of econometrics*, *6*, pp.5549-5632.

Cochrane, W.W., 1955. Conceptualizing the supply relation in agriculture. *Journal of Farm Economics*, *37*(5), pp.1161-1176.

Conley, T.G. and Molinari, F., 2007. Spatial correlation robust inference with errors in location or distance. *Journal of Econometrics*, *140*(1), pp.76-96.

Companhia Nacional de Abastecimento (CONAB), 2018. Retrieved from https://portaldeinformacoes.conab.gov.br/

Costinot, A., Donaldson, D. and Smith, C., 2016. Evolving comparative advantage and the impact of climate change in agricultural markets: Evidence from 1.7 million fields around the world. *Journal of Political Economy*, *124*(1), pp.205-248.

De Melo, F.H., 1999. Plano Real e a Agricultura Brasileira: Perspectivas. *Revista de Economia Política*, *19*(4), p.76.

DePaula, Guilherme, 2017. Three Essays in Environmental and Agricultural Economics. PhD dissertation, Yale University.

Fetzer, T., 2014. Can workfare programs moderate violence? Evidence from India. Working paper.

Fischer, G., Nachtergaele, F.O., Prieler, S., Teixeira, E., Tóth, G., Van Velthuizen, H., Verelst, L. and Wiberg, D., 2012. Global Agro-ecological Zones (GAEZ v3. 0)- Model Documentation.

Food and Agriculture Organization of the United Nations (FAO) and International Institute for Applied Systems Analysis (IIASA). Global Agro-ecological Zones dataset (GAEZ v3.0), 2017. Retrieved from http://www.gaez.iiasa.ac.at/

Food and Agriculture Organization of the United Nations (FAO). 2014. Global Forest Resources Assessment 2015 (FRA 2015). Country Report. Brazil. Rome, 2014.

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*, pp.931-953.

Hausman, C., 2012. Biofuels and land use change: sugarcane and soybean acreage response in Brazil. *Environmental and Resource Economics, 51*(2), pp.163-187.

Hendricks, N.P., Janzen, J.P. and Smith, A., 2014. Futures prices in supply analysis: Are instrumental variables necessary?. *American Journal of Agricultural Economics, 97*(1), pp.22-39.

Hsiang, S.M., 2010. Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proceedings of the National Academy of sciences*, *107*(35), pp.15367-15372.

Hungria, M., Campo, R.J. and MENDES, I.D.C., 2001. *Fixação biológica do nitrogênio na cultura da soja*. Embrapa Soja; Brasília, DF: Embrapa Cerrados.

Instituto Brasileiro de Geografia e Estatística (IBGE), 1996. Censo Agropecuário - Ano 1995/1996. Confidential data accessed at Centro de Documentação e Disseminação de Informações (CDDI), Rio de Janeiro, Brazil.

———. 1998. Documentation - Censo Agropecuário - Ano 1995/1996. Número 1. Brasil.

———. 2006. Censo Agropecuário - Ano 2006, IBGE, Rio de Janeiro, Confidential data accessed at Centro de Documentação e Disseminação de Informações (CDDI), Rio de Janeiro, Brazil.

———. 2012. Documentation - Censo Agropecuário - Ano 2006. Segunda Apuração. Brasil, Grandes Regiões e Unidades da Fereração.

———. 2015. Pesquisa Agricola Municipal (PAM). Retrieved from at www.sidra.ibge.gov.br/bda/pesquisas/pam/

———. 2016. Documentation - Pesquisa Agricola Municipal. Culturas Temporárias e Permanents. Volume 43. Brasil

———. 2017. Censo Agropecuario by Microregion – Anos 1975, 1985, 1995/1996, 2006. Dataset compiled by Brazilian Agricultural Research Agency (EMBRAPA).

Instituto de Pesquisa Econômica Aplicada (IPEA), 2018. Retrieved from www.ipeadata.gov.br

Junior, V.J.W., 2011. *Dinâmicas e estratégias das agroindústrias de soja no Brasil* (Vol. 4). Editora E-papers.

Just, R.E., 1993. Articles and notes discovering production and supply relationships: Present status and future opportunities. *Review of Marketing and Agricultural Economics*, *61*(1).

———. 2000. Some guiding principles for empirical production research in agriculture. *Agricultural and Resource Economics Review*, *29*(2), pp.138-158.

Kume, H. and Piani, G., 1997. O ICMS sobre as exportações brasileiras: uma estimativa da perda fiscal e do impacto sobre as vendas externas.

Nagavarapu, S., 2010. Implications of unleashing Brazilian ethanol: trading off renewable fuel for how much forest and savanna land. In *Working Paper*.

Nerlove, M., 1956. Estimates of the elasticities of supply of selected agricultural commodities. *Journal of Farm Economics*, *38*(2), pp.496-509.

Nunn, N. and Qian, N., 2011. The potato's contribution to population and urbanization: evidence from a historical experiment. *The Quarterly Journal of Economics*, *126*(2), pp.593-650.

Qaim, M. and Zilberman, D., 2003. Yield effects of genetically modified crops in developing countries. *Science*, *299*(5608), pp.900-902.

Rezende, C.L., 2008. Pacta sunt servanda? Quebra dos contratos de soja ver*de* . Doctoral dissertation, Universidade de São Paulo.

Rezende, F., 2012. ICMS, gênese, mutações, atualidade e caminhos para a recuperação.

Roberts, M.J. and Schlenker, W., 2013. Identifying supply and demand elasticities of agricultural commodities: Implications for the US ethanol mandate. *American Economic Review*, *103*(6), pp.2265-95.

Silva, F.P., 2012. Financiamento da cadeia de gãos no Brasil - o papel das tradings e fornecedores de insumos.

Soares, M.R.D.C., 2007. Lei Kandir: breve histórico. *Consultoria Legislativa. Brasília: House of Representatives*, *15*.

Souza-Rodrigues, E.A., 2015. *Deforestation in the Amazon: A unified framework for estimation and policy analysis*. working paper.

Tomek, W.G. and Robinson, K.L., 1981. *Agricultural product prices*.

Figures and Tables

A. Soy Yield B. Soy Harvested Area

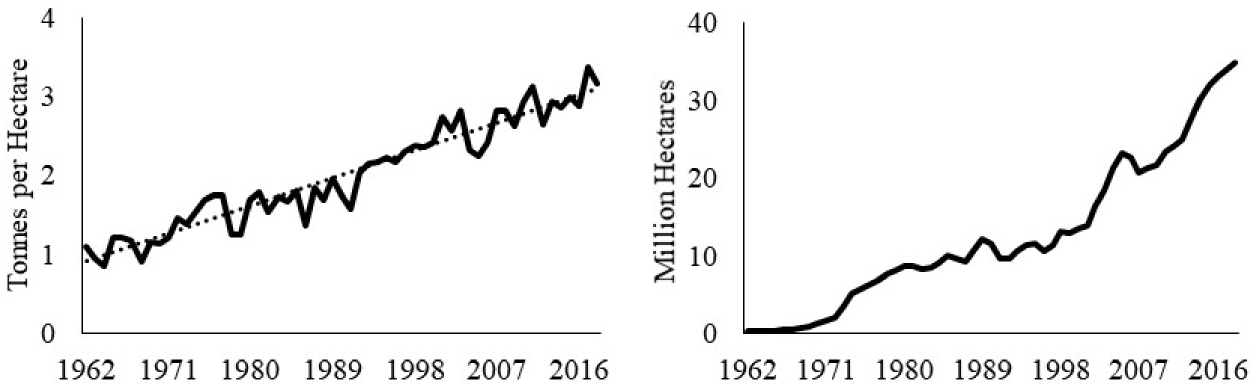
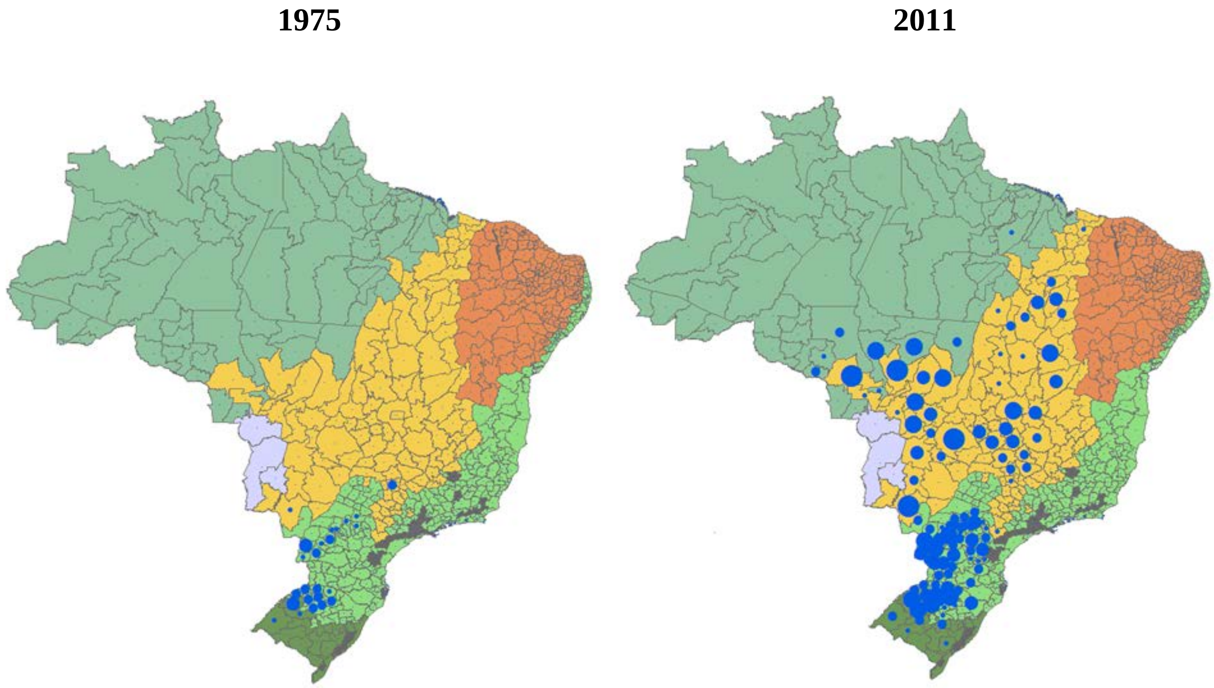


Figure 1. Expansion of soy production to the Brazilian savanna.

Note: Maps created by the author using the data from the Brazilian Institute for Applied Economic Research (IPEA, 2018) and from the Brazilian National Agricultural Supply Company (CONAB, 2018).



Legend. Biome: Amazon; Savanna; Caatinga; Atlantic Forest; Pantanal; Pampa

Soy Production (million tons): • < 0.2; • < 0.5; • < 1.0; • < 3.0; • < 5.0;

Figure 2. Expansion of soy production to the Brazilian savanna.

Note: Maps created by the author by using the IBGE Agricultural Census (IBGE Census, 2017) and IBGE Annual Production Survey (IBGE PAM, 2015).

A. Average state tax on the movement of soy export production (ICMS)

B. Index of the effective real exchange rate of the Real currency

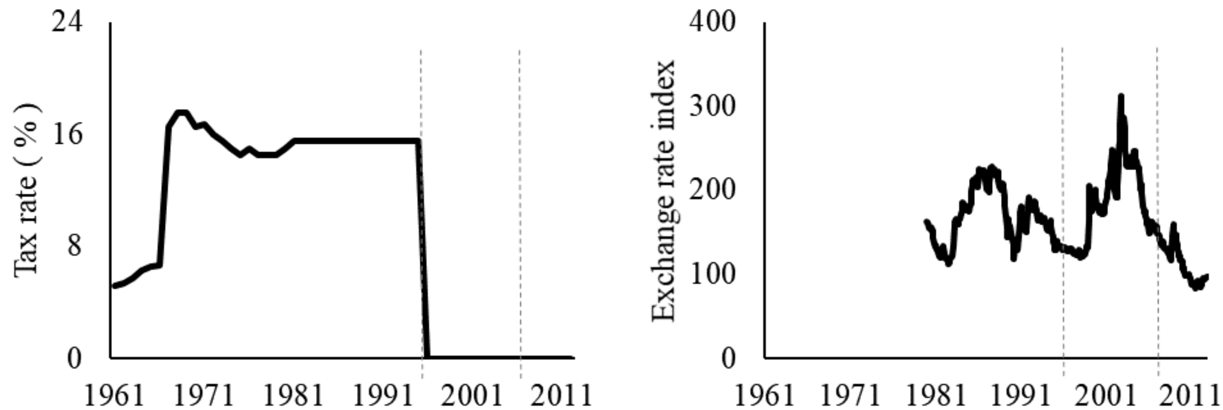


Figure 3. Changes in tax and exchange rates in Brazil between 1996 and 2006.

Note: Graphs 3A and 3B were prepared by the author by using data from Rezende (2012) and the Brazilian Institute of Applied Research (IPEA, 2018), respectively.

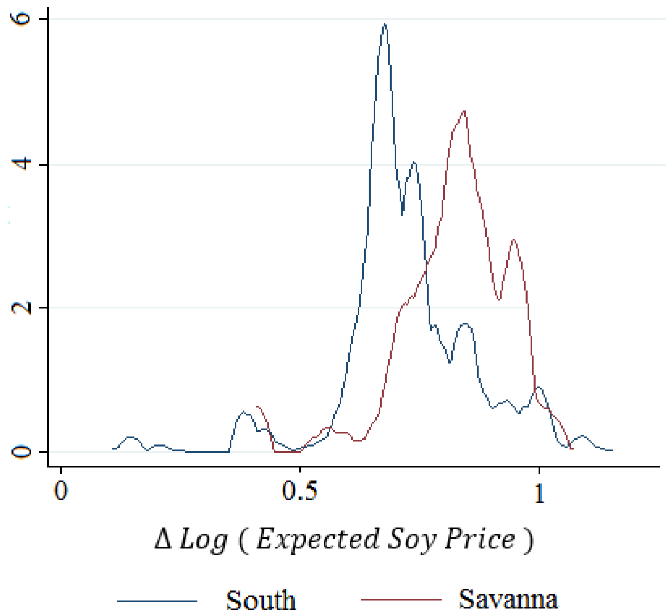


Figure 4. Density functions for price changes in Brazil.

Note: Changes are measured as the differences in the log of expected prices between 2006 and 1996. The large price changes result from the introduction of the Kandir law in 1996 and change in exchange rate policy in 1999. Density function for the savanna includes states of Mato Grosso, Mato Grosso do Sul, e Goias.

Table 1. Summary Statistics – Brazil

Variables:	1996	2006	Δ
A. Totals:			
Soy area harvested (million hectares)	8.3	13.7	5.4
Soy production (million tons)	19.0	35.9	16.9
Total farm area (million hectares)	294.2	281.4	(12.9)
Soy area share (soy harvested area / total area)	0.03	0.05	0.02
GE Soy area harvested (million hectares)	0.0	3.2	3.2
B. Mean and standard deviation:			
Soy area harvested (hectares)	238	401	187
	(1,776)	(3,039)	(1,877)
Agricultural area share (agr. area / total area)	0.74	0.73	0.01
	0.21	0.23	0.21
Expected soy price (2012 $Real per Kg)	0.42	0.83	0.42
	(0.07)	(0.07)	(0.08)
Maximum potential yield soy (high inputs) (tons/hectare)	3.1		
	(1.0)		
Delta Potential yield soy (high - low) (tons/hectare)	2.8		
	(0.9)		
Delta Potential yield soy (medium - low) (tons/hectare)	0.5		
	(0.3)		
Distance to port (kilometres)	442		
	(325)		
Distance do lime mine (kilometres)	191		
	(157)		
Baseline characteristics - Year 1985:			
Production value (2006 1,000 $Reals per hectare)	1.09		
	(1.28)		
Populational density (habitants/square km)	53.01		
	(155.70)		
Income per capita (2006 1,000 $Reals per capita)	6.82		
	(4.99)		
Share of pasture land	0.42		
	(0.21)		
Share of crop land	0.27		
	(0.17)		
Share of forest land	0.20		
	(0.16)		
Number of common census blocks	34,115	34,115	
Number of commercial farms (million)	1.59	1.50	-0.09

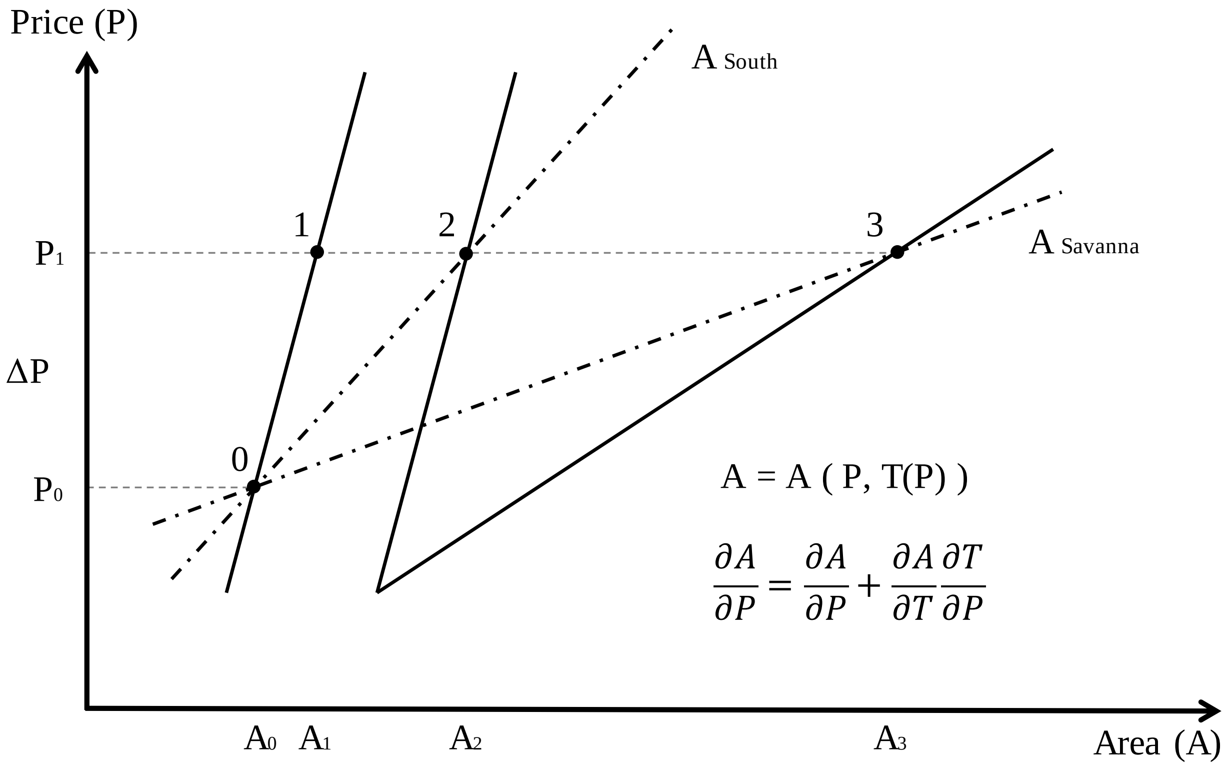


Figure 5. Heterogeneity in the acreage response function.

Note: Different types of technologies, GE and NE, lead to different indirect price effects and price elasticities of acreage.

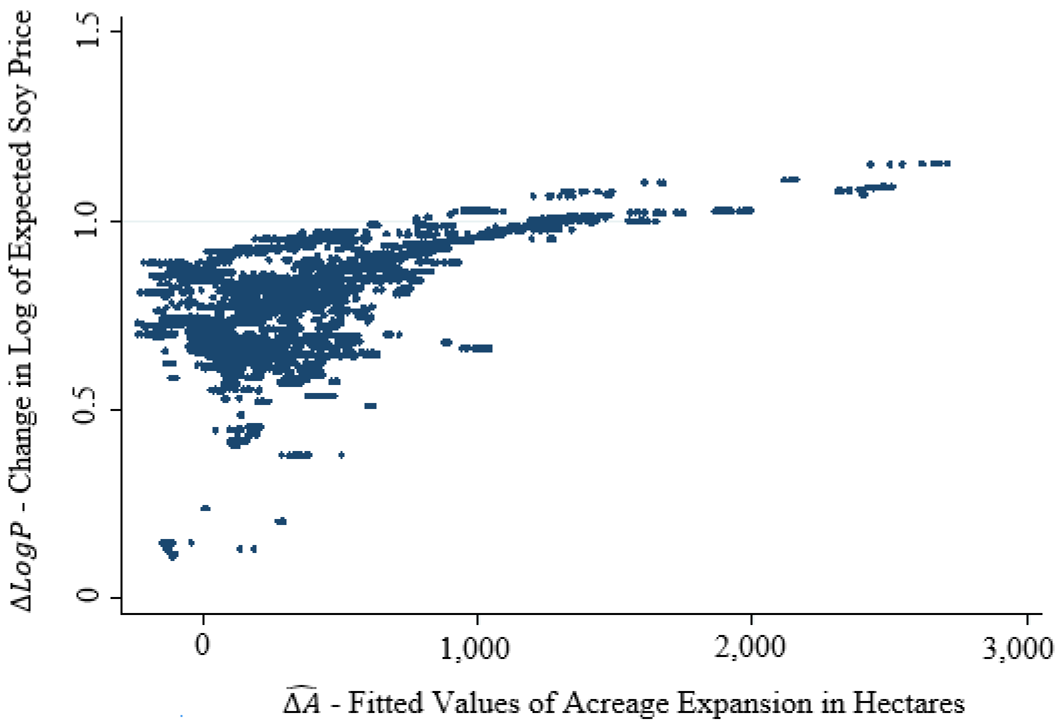


Figure 6. Sieves acreage response function.

Note: Fitted values for the non-parametric long first differences model. Each dot represents a census block in the south and savanna regions of Brazil.

Table 2. Biprobit Model for the Joint Choice of Soy Expansion and Production Technology

	South				Savanna			
	Expansion	GE Soy	Expansion	GE Soy	Expansion	NF Soy	Expansion	NF Soy
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δlog(EP)	0.206	-0.277	0.513	0.207	1.988***	2.098***	0.0168	0.598
	(0.306)	(0.367)	(0.345)	(0.396)	(0.733)	(0.787)	(0.749)	(0.847)
PY(high - medium)	0.694***	0.617**	0.620***	0.618**	0.150	0.667*	-0.0834	0.581
	(0.251)	(0.295)	(0.214)	(0.275)	(0.237)	(0.383)	(0.204)	(0.392)
PY(high - medium) squared	-0.129***	-0.115**	-0.111**	-0.0907*	0.0250	-0.121	0.0535	-0.152
	(0.0485)	(0.0573)	(0.0440)	(0.0537)	(0.0642)	(0.0972)	(0.0577)	(0.0986)
PY(medium - Low)	0.545	-0.00717	2.120***	2.332***	3.000***	6.422***	3.395***	7.416***
	(0.376)	(0.372)	(0.415)	(0.417)	(0.697)	(1.131)	(0.718)	(1.281)
PY(medium - Low) squared	-0.311	0.0640	-1.326***	-1.324***	-1.649***	-4.225***	-2.110***	-5.122***
	(0.296)	(0.300)	(0.315)	(0.323)	(0.470)	(0.736)	(0.471)	(0.789)
rho (correlation between choices)	0.792***		0.772***		0.828***		0.875***	
	(0.0372)		(0.0377)		(0.0723)		(0.0737)	
Baseline Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State and Biome fixed effects	Yes	Yes	No	No	Yes	Yes	No	No
Meso region fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Observations (census blocks)	7,949	7,949	7,949	7,949	2,444	2,444	2,445	2,445

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. All standard errors clustered at the municipality level. Savanna models include only states from the Midwest region: Mato Grosso, Mato Grosso do Sul, and Goias. Models (1), (2), (5), and (6) include the following control variables: price change for substitute crops (sugarcane, coffee, and rice), log of the agricultural production value in 1985, log of the average farm size in 1996, log of population density in 1985, share of crop land in the municipality in 1985, share of pasture land in the municipality in 1985, and dummy for soy production in the census block in 1990. Models (3), (4), (7), and (8) include the following control variables: log of the agricultural production value in 1985 and log of population density in 1985.

Table 3. First Differences Model for the Acreage Response Function in the South

Dependent variable is Δ soy area share	South				
	OLS				IV
Variables	(1)	(2)	(3)	(4)	(5)
Δlog(EP)	0.0338**	0.0383**	0.00528	-0.0141	-0.111***
	(0.0163)	(0.0165)	(0.0397)	(0.0442)	(0.0275)
Dummy PY(high - low) Q2		0.0130***	-0.0313	-0.0381	-0.00406
		(0.00473)	(0.0378)	(0.0423)	(0.0122)
Dummy PY(high - low) Q3		0.00303	-0.0460	-0.0470	-0.0912***
		(0.00514)	(0.0294)	(0.0321)	(0.0193)
Dummy PY(medium - low) Q2		0.000816	0.0437*	0.0767***	-0.0953***
		(0.00497)	(0.0223)	(0.0240)	(0.0148)
Dummy PY(medium - low) Q3		-0.00373	0.0533**	0.0823***	-0.0909***
		(0.00689)	(0.0267)	(0.0283)	(0.0156)
Δlog(EP) x Dummy PY(high - low) Q2			0.0624	0.0788	0.0116
			(0.0535)	(0.0601)	(0.0208)
Δlog(EP) x Dummy PY(high - low) Q3			0.0705*	0.0851*	0.174***
			(0.0423)	(0.0471)	(0.0320)
Δlog(EP) x Dummy PY(medium - low) Q2			-0.0581*	-0.0805**	0.158***
			(0.0301)	(0.0328)	(0.0238)
Δlog(EP) x Dummy PY(medium - low) Q3			-0.0784**	-0.0844**	0.143***
			(0.0352)	(0.0372)	(0.0243)
Quadratic functions of potential yield	Yes	No	No	No	No
Baseline Characteristics	No	No	No	Yes	Yes
Mesoregion fixed effects	Yes	Yes	Yes	Yes	No
Observations (census blocks)	7,591	7,591	7,591	7,397	7,397
R2	0.149	0.143	0.145	0.135	0.082
1st stage F-stat					14.49
Long-run price elasticity of soy acreage:					
Region average	0.25	0.28			
Top quantile of pontential yield			0.51	0.62	0.46
Long-run price elasticity (Hausman, 2012)	0.72				

Note: *** p<0.01, ** p<0.05, * p<0.1. All standard errors clustered at the municipality level.

Table 4. First Differences Model for the Soy Expansion and GE Soy Expansion in the South

	South				
	Δ Soy Area		Δ GE Soy Area		
Variables	(1)	(2)	(3)	(4)	(5)
Δlog(EP)	358.3**	89.15	1,400***	804.2**	644.3*
	(172.2)	(139.2)	(397.4)	(368.1)	(361.4)
Dummy PY(high - low) Q2	86.99**	81.70*	275.8***	284.7***	320.3***
	(38.60)	(46.74)	(65.36)	(59.78)	(66.74)
Dummy PY(high - low) Q3	128.3***	68.01	244.1***	70.20	129.4**
	(38.18)	(49.90)	(51.22)	(46.16)	(50.97)
Dummy PY(medium - low) Q2	-60.93*	-22.79	1.512	99.40	106.9
	(32.63)	(36.91)	(57.13)	(63.81)	(69.21)
Dummy PY(medium - low) Q3	-121.8**	-58.05	-362.9***	-208.7**	-193.5**
	(55.72)	(58.64)	(85.56)	(90.83)	(96.95)
Baseline Characteristics	Yes	Yes	No	No	Yes
Meso region fixed effects	No	Yes	No	Yes	Yes
Observations (census blocks)	7,591	7,397	7,591	7,591	7,397
R2	0.057	0.103	0.107	0.169	0.174
Long-run price elasticity of soy acreage:	0.62		2.41	1.38	1.11
Long-run price elasticity (Hausman, 2012)	0.72				

Note: *** p<0.01, ** p<0.05, * p<0.1. All standard errors clustered at the municipality level.

Table 5. First Differences Model for the Acreage Response Function in the Savanna

Dependent variable is Δ soy area share	Savanna						Midwest Savanna	
	OLS					IV	OLS	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δlog(EP)	0.0382***	0.0371***	0.0195***	-0.0113**	-0.0349***	-0.0295***	0.111***	0.101***
	(0.00647)	(0.00634)	(0.00703)	(0.00515)	(0.00695)	(0.0114)	(0.0326)	(0.0328)
Dummy PY(high - low) Q2		-0.000214	-0.00223**	-0.000457	-0.00250**	-0.00211**		
		(0.000950)	(0.00103)	(0.000942)	(0.00103)	(0.00102)		
Dummy PY(high - low) Q3		0.00605***	0.00461***	0.00520***	0.00373***	0.00355**		
		(0.00150)	(0.00144)	(0.00143)	(0.00138)	(0.00142)		
Dummy PY(medium - low) Q2		0.00564***	0.00677***	-0.0339***	-0.0352***	-0.0510***		
		(0.000878)	(0.000950)	(0.00608)	(0.00614)	(0.00866)		
Dummy PY(medium - low) Q3		0.00328***	0.00393***	-0.0249***	-0.0289***	-0.0275***		
		(0.00101)	(0.00104)	(0.00452)	(0.00467)	(0.00863)		
Δlog(EP) x Dummy PY(medium - low) Q2				0.0693***	0.0738***	0.1000***		
				(0.0111)	(0.0113)	(0.0158)		
Δlog(EP) x Dummy PY(medium - low) Q3				0.0505***	0.0589***	0.0568***		
				(0.00820)	(0.00865)	(0.0154)		
Quadratic functions of potential yield	Yes	No	No	No	No	No	Yes	Yes
Baseline Characteristics	No	No	Yes	No	Yes	Yes	Yes	Yes
State fixed effects	No	No	No	No	No	No	No	Yes
Observations (census blocks)	13,661	13,661	13,385	13,661	13,385	13,385	2,172	2,172
R2	0.071	0.076	0.091	0.085	0.102	0.097	0.110	0.130
1st stage F-stat						42.52		
Long-run price elasticity of soy acreage:								
Region average	1.18	1.15	0.60				3.44	3.13
Second quantile of pontential yield				1.80	1.20	2.18		
Top quantile of pontential yield				1.21	0.74	0.85		

Note: *** p<0.01, ** p<0.05, * p<0.1. All standard errors clustered at the municipality level.

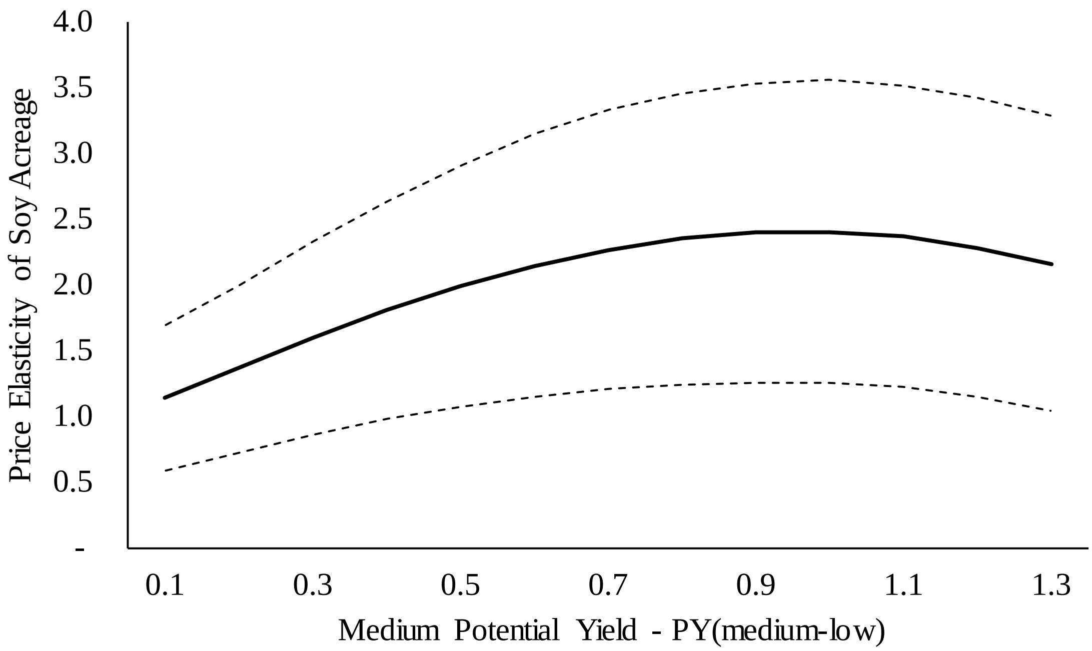


Figure 7. Technological change component of the price elasticities of soy acreage in the savanna.

Note: The price elasticities are estimated by using the probability model of the soy expansion in the savanna, corresponding to the second term in equation (8).

Table 6. First Differences Model for the Land-use Response Function in Brazil

Dependent variable is Δ agricultural area share	Brazil		South		Midwest Savanna	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Δlog(EP)	0.0337**	0.0187	0.0227	-0.00164	0.0942**	0.0752
	(0.0166)	(0.0180)	(0.0244)	(0.0236)	(0.0405)	(0.0492)
Baseline Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	No	Yes	No	Yes	No
Mesoregion fixed effects	No	Yes	No	Yes	No	Yes
Observations (census blocks)	32,537	33,576	7,717	7,912	2,172	2,172
R2	0.063	0.078	0.039	0.043	0.155	0.151
Price elasticity of agricultural acreage:	0.046	0.025	0.030	(0.002)	0.128	0.102
Price elasticity of agricultural acreage (Barr et. al., 2010):						
1997-1999 to 2001-2003	0.201					
2004 to 2006	0.007					
Price elasticity of crop acreage (Hausman, 2012):	0.110					
Price elasticity of crop acreage (Roberts and Schlenker, 2013):	0.174 - 0.261					

Note: *** p<0.01, ** p<0.05, * p<0.1. All standard errors clustered at the municipality level.

Table 7. Soy Response Function in Brazil with the Alternative Clustering Variables

Dependent variable is Δ soy area share	Brazil					
Clustering:	Municipio		Microregion		Mesoregion	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Models without price PY interaction:						
Δlog(EP)	0.0672***	0.0340***	0.0672***	0.0340**	0.0672***	0.0340***
	(0.00482)	(0.00838)	(0.00899)	(0.0141)	(0.0111)	(0.0115)
Models with price PY interaction:						
Δlog(EP)	0.0549***	0.0214*	0.0549***	0.0214	0.0549***	0.0214
	(0.00753)	(0.0126)	(0.0129)	(0.0200)	(0.0160)	(0.0151)
Δlog(EP) x Dummy PY(medium - low) Q2	0.0492***	0.0351***	0.0492***	0.0351*	0.0492**	0.0351**
	(0.0111)	(0.0116)	(0.0190)	(0.0187)	(0.0195)	(0.0170)
Δlog(EP) x Dummy PY(medium - low) Q3	0.00933	0.00611	0.00933	0.00611	0.00933	0.00611
	(0.00991)	(0.0114)	(0.0167)	(0.0170)	(0.0194)	(0.0136)
State fixed effects	No	Yes	No	Yes	No	Yes

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors for models (1) and (2) are clustered at the municipality level. There are 5,562 municipalities in Brazil. Standard errors for models (3) and (4) are clustered at the microregion level. There are 558 microregions in Brazil. Standard errors for models (5) and (6) are clustered at the mesoregion level. There are 1375 mesoregions in Brazil. Al models include the log of the agricultural production value in 1985 and the log of the average farm size in 1996. The models with the interaction terms also control for the maximum potential yield of soy.

Table 8. Soy Response Function in Brazil with Standard Errors Corrected for Spatial Correlation

	Brazil							
Dependent variable	Δ soy area share				Δ soy area			
Conley cutoff distance (kilometers)	50	100	200	500	50	100	200	500
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δlog(EP)	0.0671***	0.0671***	0.0671***	0.0671***	460.8***	460.8**	460.8**	460.8*
	(0.00695)	(0.00972)	(0.0116)	(0.0130)	(169.0)	(194.3)	(221.3)	(272.1)
PY(medium - low)	-0.00497***	-0.00497**	-0.00497*	-0.00497**	482.5***	482.5**	482.5**	482.5*
	(0.00172)	(0.00224)	(0.00267)	(0.00208)	(133.3)	(197.8)	(197.8)	(248.7)
PY(medium - low) squared	0.00377***	0.00377***	0.00377***	0.00377***	-487.1***	-487.1***	-487.1***	-487.1***
	(0.000695)	(0.000917)	(0.00111)	(0.000936)	(91.02)	(138.6)	(138.6)	(183.3)
Log of farm size in 1996	0.00388***	0.00388***	0.00388***	0.00388***	213.1***	213.1***	213.1***	213.1***
	(0.000676)	(0.000907)	(0.00110)	(0.00122)	(33.04)	(36.10)	(46.16)	(60.42)
Log agricultural production value / ha in 1985	0.00385***	0.00385***	0.00385*	0.00385	-142.1***	-142.1***	-142.1***	-142.1***
	(0.000827)	(0.00129)	(0.00196)	(0.00288)	(33.78)	(37.71)	(45.82)	(52.29)
Observations (census blocks)	34,025	34,025	34,025	34,025	32,869	32,869	32,869	32,869
R2	0.144	0.144	0.144	0.144	0.057	0.057	0.057	0.057

Note: *** p<0.01, ** p<0.05, * p<0.1. Models (5) to (8) also include the full set of baseline characteristics, potential yield for alternative crops, and dummy variable for soy production in 1990.

Table 9. Soy Response Function in Brazil with Different Measures of Permanent Price Changes

	Brazil					
Dependent variable	Δ soy area share			Δ soy area		
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Δlog(EP) : baseline	0.0672***	0.0340***	0.0234***	1,166***	663.9***	489.5**
	(0.00482)	(0.00838)	(0.00843)	(162.8)	(194.8)	(204.2)
Δlog(EP) : ΔP1	0.0599***	0.0326***	0.0266***	1,072***	599.6***	562.4***
	(0.00464)	(0.00687)	(0.00825)	(145.9)	(147.4)	(176.5)
Δlog(EP) : ΔP2	0.0712***	0.0130	0.0209**	856.7***	443.7*	436.8*
	(0.00725)	(0.00938)	(0.00876)	(148.7)	(229.8)	(240.3)
Δlog(EP) : ΔP3	0.0521***	0.0556***	0.0525***	562.2***	734.6**	776.8**
	(0.00859)	(0.0113)	(0.0115)	(204.0)	(292.9)	(309.0)
Baseline Characteristics	No	No	Yes	No	No	Yes
State fixed effects	No	Yes	Yes	No	Yes	Yes

Note: *** p<0.01, ** p<0.05, * p<0.1. All standard errors clustered at the municipality level. Each line in the table represents the price effect of a different regression model. I test the alternative measurements for permanent price changes: ΔP1: average for 1990 to 1994 and average for 1996 to 2000. ΔP2: average for 1990 to 1994 and average for 2000 to 2004. ΔP3: average for 1994 to 1996 and average for 2000 to 2002.

Appendix A – Description of Variables

Variable	Description	Dataset / Unit of Observation
Soy acreage expansion	Difference between the soy area harvested in 2006 and 1996 in Brazilian commercial farms, measured in hectares.	IBGE Census / Census block
GE soy acreage expansion	Difference between the soy area harvested using genetically modified soy in 2006 and 1996 in Brazilian commercial farms, measured in hectares. I use the type of seed variable of the IBGE census to determine the area harvested with GE seeds.	IBGE Census / Census block
Soy area share	Soy harvested area divided by the total area in a rural common census block.	IBGE Census / Census block
Agriculture area share	Total area allocated to agriculture production, which includes cropland and pastureland, divided by the total area in a rural common census block.	IBGE Census / Census block
Soy potential yield	Potential yield of rain fed soy production under low, intermediate, and high input use, measured in tons per hectare. I also use potential yield measures for corn, cotton, sugarcane, and rice. The potential yield variables were averaged at the census block level.	GAEZ IIASA/FAO / Census block

Expected soy spot price	My preferred measure of the expected soy price is the average of the spot price in the municipality in the previous six years. For the census year 1996 I use the average price for 1990 to 1996. For the census year 2006 I use the average price for 1999 to 2004. All prices were deflated using the IGP-DI inflation index computed by Fundacao Getulio Vargas.	IBGE PAM / Municipality
Alternative measures of expected soy spot price	For robustness test, I also used the following measures for the expected soy price. $\Delta P1$: average for 1990 to 1994 and average for 1996 to 2000. $\Delta P2$: average for 1990 to 1994 and average for 2000 to 2004. $\Delta P3$: average for 1994 to 1996 and average for 2000 to 2002. $\Delta P4$: average for 1994 to 1996 and average for 2000 to 2001. $\Delta P5$: average for years 1990 to 1995 and average for 1999 to 2002.	IBGE PAM/ Municipality
Expected price of alternative crops	I computed expected prices for sugarcane, corn, cotton, and rice using the same formulas described above for calculation of the expected price of soy.	IBGE PAM/ Municipality

Log of population density in 1985	Log of population density of a microregion in Brazil in year 985. Population density is measured in 1985. Population density is measured in number of habitants per squared quilometer. There are 558 microregions in Brazil/	IBGE Census – Embrapa / Microregion
Log of the value of agricultural production in 1985	Log of the average agricultural production value of a microregion in Brazil. Agricultural production value is measured in 2006 1,000 Reals per hectare.	IBGE Census – Embrapa / Microregion
Log of income per capital	Log of average income per capita of a microregion in Brazil. The income per capital is measured in 2006 1,000 Reals per habitant.	IBGE Census – Embrapa / Microregion
Share of cropland in 1985	Total area harvested with crops in 1985 in a microregion divided by the total area of the microregion. All harvested areas are measured in hectares.	IBGE Census – Embrapa / Microregion

Share of pastureland in 1985	Total area allocated to pasture in 1985 in a microregion divided by the total area of the microregion. All areas are measured in hectares.	IBGE Census – Embrapa / Microregion
Share of forestland in 1985	Total area of natural vegetation in 1985 in a microregion divided by the total area of the microregion. All areas are measured in hectares.	IBGE Census – Embrapa / Microregion
Log Farm size	Log of the average farm size of the rural common census block in the 1996 census. Farm size is measured in hectares.	IBGE Census / Census block
Dummy for soy production in 1990	1 if the soy area harvested in a municipality in year 1990 is greater than zero. Zero otherwise.	IBGE PAM/ Municipality
Distance to port	Distance from each microregion to the closets port. Includes ports in the coast as well as ports in major rivers. Computed using geographical information system. The distance to	IBGE GIS - Embrapa/ Microregion

	port is measured in kilometers. Geographical shape for ports in Brazil is available by IBGE.	
Distance to lime mine	Distance from each census block to the closets lime mine. Include lime mines of medium and large size. Computed using geographical information system. The distance to the lime mine is measured in kilometers. Geographical shape for lime mines in Brazil is available by IBGE.	IBGE GIS - Embrapa/ Census Block
Soil characteristics	Average index for soil characteristics measured at the Municipality level. The following soil characteristics were integrated into the census dataset: nitrogen content, pH index, organic matter index; Potassium index, Aluminum index, share of clay soil, share of silt soil, and share of sand soil.	Embrapa/ Municipality
Price risk	Measure of price volatility. I follow Hausman (2012) in computing a weighted average of the squared deviation of average price from current price. The wages are 0.5 for the most recent year, and 0.3 and 0.2 for the previous years.	IBGE PAM / Municipality

Table A1. Summary Statistics – South

Variables:	1996	2006	Δ
A. Totals:			
Soy area harvested (million hectares)	3.9	5.2	1.3
Soy production (million tons)	8.7	12.8	4.1
Total farm area (million hectares)	35.8	32.0	(3.8)
Soy area share (soy harvested area / total area)	0.11	0.16	0.05
GE Soy area harvested (million hectares)	0.0	2.6	2.6
B. Mean and standard deviation:			
Soy area harvested (hectares)	479	685	217
	(1,423)	(1,935)	(833)
Agricultural area share (agr. area / total area)	0.77	0.74	0.05
	0.18	0.20	0.14
Expected soy price (2012 $Real per Kg)	0.41	0.85	0.44
	0.04	0.05	0.06
Maximum potential yield soy (high inputs) (tons/hectare)	3.7		
	(0.5)		
Delta Potential yield soy (high - low) (tons/hectare)	3.4		
	(0.5)		
Delta Potential yield soy (medium - low) (tons/hectare)	0.5		
	(0.3)		
Distance to port (kilometres)	429		
	(230)		
Distance do lime mine (kilometres)	209		
	(104)		
Baseline characteristics - Year 1985:			
Production value (2006 1,000 $Reals per hectare)	1.64		
	(0.89)		
Populational density (habitants/square km)	44.06		
	(60.34)		
Income per capita (2006 1,000 $Reals per capita)	9.97		
	(3.84)		
Share of pasture land	0.35		
	(0.20)		
Share of crop land	0.42		
	(0.20)		
Share of forest land	0.15		
	(0.11)		
Number of common census blocks	7,591	7,591	
Number of commercial farms (million)	0.34	0.29	-0.05

Table A2. Summary Statistics – Savanna

Variables:	1996	2006	Δ
A. Totals:			
Soy area harvested (million hectares)	4.0	8.1	4.1
Soy production (million tons)	9.4	22.1	12.7
Total farm area (million hectares)	188.4	180.3	(8.1)
Soy area share (soy harvested area / total area)	0.02	0.05	0.02
GE Soy area harvested (million hectares)	0.0	0.6	0.6
B. Mean and standard deviation:			
Soy area harvested (hectares)	291	595	338
	(2,577)	(4,550)	(2,884)
Agricultural area share (agr. area / total area)	0.76	0.72	0.01
	0.19	0.22	0.21
Expected soy price (2012 $Real per Kg)	0.41	0.77	0.37
	0.09	0.07	0.10
Maximum potential yield soy (high inputs) (tons/hectare)	3.0		
	(0.9)		
Delta Potential yield soy (high - low) (tons/hectare)	2.7		
	(0.8)		
Delta Potential yield soy (medium - low) (tons/hectare)	0.6		
	(0.3)		
Distance to port (kilometres)	549		
	(372)		
Distance do lime mine (kilometres)	189		
	(127)		
Baseline characteristics - Year 1985:			
Production value (2006 1,000 $Reals per hectare)	0.62		
	0.76		
Populational density (habitants/square km)	28.21		
	65.44		
Income per capita (2006 1,000 $Reals per capita)	5.20		
	4.18		
Share of pasture land	0.50		
	0.18		
Share of crop land	0.17		
	0.09		
Share of forest land	0.19		
	0.11		
Number of common census blocks	13,667	13,667	
Number of commercial farms (million)	0.74	0.74	0.00

Description of IBGE Agricultural Census Dataset

Every ten years IBGE surveys over 5 million farmers in Brazil to create the agricultural census dataset. The agricultural census survey was designed based on the international standard for agricultural censuses developed by the Food and Agriculture Organization of the United Nations and based on information needs from institutional users and scholars. The farm-level version of the census is confidential but can be accessed for academic research at the IBGE Center for Documentation and Dissemination of Information (CDDI) in Rio de Janeiro, Brazil, following IBGE's protocols for the protection of the confidentiality of the information.

In this study, I use the 1995/1996 census, referred to in this study as the 1996 IBGE census, and the 2006 census. One difference between the 1996 and 2006 surveys is the reference period. The 1996 survey followed the agricultural calendar, from August 8 1995 to July 31 1996, whereas the 2006 survey used the calendar year, from January 1 2006 to December 31 2006. The difference in the reference period has no impact on this analysis as I use harvested acreage to compute agricultural expansion. Soy is harvested in Brazil between February and April. I thus compare the soy area harvested for the 2005/2006 and 1995/1996 agricultural seasons.

The novelty of my dataset is the integration of the two census surveys at the census block level. The census block is the survey unit for the agricultural census. Most census blocks are the same across surveys. However, there are cases where a census blocks are divided or combined between censuses. I used IBGE documentation of changes in census blocks to create the common census block unit. IBGE has extensive technical documentation for the agricultural census. For detailed information about the 2006 agricultural census, see IBGE (2012). For the 1996 agricultural census, see IBGE (1998).

Description of IBGE Municipal Agricultural Production Survey (PAM)

IBGE also surveys agricultural production for 41 crops at the municipality level monthly and annually. The Municipal Agricultural Production survey (PAM) contains annual information on planted area, harvested area, production, average yield, production value, and the average price paid to the producer for 41 crops at each municipality. The PAM data is publicly available through the IBGE website (SIDRA) and is used for calculation of producer price indexes in Brazil. For example, Fundação Getulio Vargas uses the agricultural production value reported in the PAM survey to compute the index of wholesale prices (IPA) and the general price index (IGP-M). The producer price, also referred in this study as the spot price, is the annual average price paid to the producer for a crop in a municipality weighted by the monthly quantities commercialized during the year. State and local agencies complete the PAM survey in the first three months of the year and submit the results to the IBGE Agriculture Department for validation using the monthly agricultural surveys and historical data. For a detailed explanation of the survey and the data validation process, see IBGE (2016).

Appendix B – Spatial Variation in Price Shocks

The farm gate price at time t is $P_t^{farm} = (1 - tax_t)(e_t P_t^{mkt} - c_t d)$. I use the log approximation to show how the price shocks, measured in terms of percentage changes in prices, vary spatially when taxes and market prices change:

$$(B1) \quad \Delta P = \frac{P_1^{farm} - P_0^{farm}}{P_0^{farm}} \cong Log\left(P_1^{farm}\right) - Log\left(P_0^{farm}\right) =$$

$$= Log\left[(1 - tax_1)\left(e_1 P_1^{mkt} - c_1 d\right)\right] - \ Log\left[(1 - tax_0)\left(e_0 P_0^{mkt} - c_0 d\right)\right]$$

It is know that the error of the log approximation increases with the percentage change. I ignore this error as my objective is to provide an intuition for the spatial variation in price shocks.

I use the properties of logarithms to rewrite equation B1

$$(B2) \quad \Delta P = Log\left[\frac{1 - tax_1}{1 - tax_0}\right] + Log[\frac{e_1 P_1^{mkt} - c_1 d}{e_0 P_0^{mkt} - c_0 d}]$$

Defining the market price growth rate g as $e_1 P_1^{mkt} = (1 + g) e_0 P_0^{mkt}$, the transportation cost share of the market price as $S_{TC} = cd/e_0 P_0^{mkt}$, and assuming that the transport cost does not change over time, $c_0 = c_1 = c$, I can simplify the second term in the right hand side of B2:

$$(B3) \quad Log\left[\frac{e_1 P_1^{mkt} - c_1 d}{e_0 P_0^{mkt} - c_0 d}\right] = Log\left[\frac{(1 + g) e_0 P_0^{mkt} - cd}{e_0 P_0^{mkt} - cd}\right] = Log\left[\frac{(1 + g) - S_{TC}}{1 - S_{TC}}\right]$$

$$= Log(1 + g - S_{TC}) - Log(1 - S_{TC})$$

The price shocks can be decomposed into three terms substituting equation B3 into equation B2:

$$(B4) \quad \Delta P = Log\left[\frac{1 - tax_1}{1 - tax_0}\right] - Log(1 - S_{TC}) \; + Log(1 + g - S_{TC})$$

The first term represents the effect of the tax change and does not vary with distance in this simplified model. The second term is a function of transportation costs and varies with distance. This term will always be positive, as the transport cost share is lower than one, and increases with the transport cost share. The last term varies with both the growth rate in the market price and distance to market. This term is also positive as the transport cost share is lower than the market price growth rate.

Taking the derivative of equation B4 with respect to the transport cost share:

$$(B5) \quad \frac{\partial \Delta P}{\partial S_{TC}}(g, S_{TC}) = \frac{1}{1 - S_{TC}} - \frac{1}{1 + g - S_{TC}} = \frac{g}{1 + g - S_{TC}} > 0$$

As expected the price shock increases nonlinearly with transportation costs and therefore distance to market. The derivative of price change with respect to distance to market is:

$$(B6) \quad \frac{\partial \Delta P}{\partial d}(g, S_{TC}) = \frac{g}{1+g-S_{TC}}\left(\frac{c}{e_0 P_0^{mkt}}\right)$$

Appendix C – Map of Rural Common Census Blocks in Brazil

Appendix D – Sieves Acreage Response Function

I estimate a Sieves semi-parametric version of equation 5 to explore the heterogeneity in the acreage response function modeled using the micro census data, to examine the robustness of the acreage supply function, and to provide a simple graphical interpretation of the acreage response function. For details on Sieves models and estimation see Chen (2007).

The sieves first-differences model is:

$$(D1) \quad \Delta A_i = \delta_1 + \delta_2 d_{UF} + G(\Delta P_i) + \theta_3 PY_i + \theta_4 X_{it} + \Delta u_i$$

The price response function in equation (D1), $G(\Delta P_i)$, is unknown, so I use the sieves approximation:

$$(D2) \quad G_k\left(\Delta P_i, \tilde{\theta}_3\right) = \sum_{j=0}^{a} \Delta P^j \tilde{\theta}_{jk} + \sum_{j=1}^{m} \theta_{j+m} (\Delta P - v_j)^a \, 1\left[\Delta P \geq v_j\right]$$

The sieves approximation is a polynomial spline where a is the order of the polynomial, m is the number of knots, and v_j represents the j^k knot. The sieves model in equation (D2) splits the support of the price changes into m splices, and uses a polynomial approximation for each segment. The semi-parametric sieves model with splines is estimated using ordinary least squares. Figure 6 shows the fitted values from the sieves acreage response function. The vertical axis shows prices changes in terms of differences in log of expected prices and the horizontal axis measures soy expansion in hectares. Every dot in Figure 6 is one rural common census block in Brazil.

Appendix E – Regression Results – Table E1 - Biprobit Models

	South				Savanna			
	Expansion	GE Soy	Expansion	GE Soy	Expansion	NF Soy	Expansion	NF Soy
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δlog(EP)	0.206	-0.277	0.513	0.207	1.988***	2.098***	0.0168	0.598
	(0.306)	(0.367)	(0.345)	(0.396)	(0.733)	(0.787)	(0.749)	(0.847)
PY(high - medium)	0.694***	0.617**	0.620***	0.618**	0.150	0.667*	-0.0834	0.581
	(0.251)	(0.295)	(0.214)	(0.275)	(0.237)	(0.383)	(0.204)	(0.392)
PY(high - medium) squared	-0.129***	-0.115**	-0.111**	-0.0907*	0.0250	-0.121	0.0535	-0.152
	(0.0485)	(0.0573)	(0.0440)	(0.0537)	(0.0642)	(0.0972)	(0.0577)	(0.0986)
PY(medium - Low)	0.545	-0.00717	2.120***	2.332***	3.000***	6.422***	3.395***	7.416***
	(0.376)	(0.372)	(0.415)	(0.417)	(0.697)	(1.131)	(0.718)	(1.281)
PY(medium - Low) squared	-0.311	0.0640	-1.326***	-1.324***	-1.649***	-4.225***	-2.110***	-5.122***
	(0.296)	(0.300)	(0.315)	(0.323)	(0.470)	(0.736)	(0.471)	(0.789)
Δ Price Risk	-0.0586*	-0.0715**	0.0322	0.0276	-0.0105	0.0493	-0.0726	-0.0147
	(0.0305)	(0.0361)	(0.0298)	(0.0358)	(0.0697)	(0.0781)	(0.0700)	(0.0832)
Δlog(EP) - sugarcane	-0.0695	-0.146	0.0269	0.0518	-0.0667	-0.444**	0.201	-0.0516
	(0.103)	(0.120)	(0.113)	(0.131)	(0.121)	(0.207)	(0.135)	(0.220)
Δlog(EP) - corn	-0.0135	0.282	0.419	0.836**	-1.327***	-1.001*	-1.102***	-1.051**
	(0.261)	(0.305)	(0.316)	(0.344)	(0.352)	(0.511)	(0.324)	(0.536)
Δlog(EP) - cotton	-1.279	-2.502*	-0.822	-3.067*	-0.490	-1.670*	-0.499	-1.641
	(1.193)	(1.381)	(1.498)	(1.600)	(0.742)	(0.854)	(0.838)	(1.084)
Δlog(EP) - rice	-0.228*	0.0154	-0.522***	-0.550***	1.082**	1.973***	1.584***	2.607***
	(0.138)	(0.169)	(0.180)	(0.204)	(0.464)	(0.650)	(0.506)	(0.713)
Log value agr. prod. / ha 1985	-0.392***	-0.760***	0.362***	0.132	-0.239	-0.283	0.339*	0.521**
	(0.0972)	(0.107)	(0.108)	(0.118)	(0.183)	(0.263)	(0.192)	(0.232)
Log population density 1985	-0.408***	-0.281***	-0.477***	-0.206**	-0.196**	-0.188	-0.00352	-0.0902
	(0.0684)	(0.0763)	(0.0862)	(0.0878)	(0.0915)	(0.125)	(0.119)	(0.173)
Soy area share 1996	-1.020***	1.328***	-0.842***	1.523***				
	(0.139)	(0.206)	(0.137)	(0.223)				
Log of farm size 1996	0.118***	0.151***			0.218***	0.175***		
	(0.0292)	(0.0332)			(0.0339)	(0.0375)		
Cropland share 1985	5.352***	6.609***			8.130***	8.815***		
	(0.427)	(0.500)			(1.540)	(1.944)		
Pastureland share 1985	2.471***	2.220***			-1.242**	-2.462***		
	(0.324)	(0.369)			(0.542)	(0.654)		
Dummy for soy production in 1990	0.527***	0.403***			0.847***	0.816***		
	(0.0612)	(0.0696)			(0.102)	(0.142)		
rho (correlation between choices)	0.792***		0.772***		0.828***		0.875***	
	(0.0372)		(0.0377)		(0.0723)		(0.0737)	
State and Biome fixed effects	Yes	Yes	No	No	Yes	Yes	No	No
Meso region fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Observations (census blocks)	7,949	7,949	7,949	7,949	2,444	2,444	2,445	2,445

Note: *** p<0.01, ** p<0.05, * p<0.1. All standard errors clustered at the municipality level.

Table E2 - First Differences Model for the Acreage Response Function in the South

Dependent variable is Δ soy area share	South					
	OLS					IV
Variables	(1)	(2)	(2b)	(3)	(4)	(5)
Δlog(EP)	0.0338**	0.0383**	0.0255	0.00528	-0.0141	-0.111***
	(0.0163)	(0.0165)	(0.0172)	(0.0397)	(0.0442)	(0.0275)
Dummy PY(high - low) Q2		0.0130***	0.0152***	-0.0313	-0.0381	-0.00406
		(0.00473)	(0.00519)	(0.0378)	(0.0423)	(0.0122)
Dummy PY(high - low) Q3		0.00303	0.00469	-0.0460	-0.0470	-0.0912***
		(0.00514)	(0.00542)	(0.0294)	(0.0321)	(0.0193)
Dummy PY(medium - low) Q2		0.000816	-0.00260	0.0437*	0.0767***	-0.0953***
		(0.00497)	(0.00514)	(0.0223)	(0.0240)	(0.0148)
Dummy PY(medium - low) Q3		-0.00373	-0.00914	0.0533**	0.0823***	-0.0909***
		(0.00689)	(0.00701)	(0.0267)	(0.0283)	(0.0156)
Δlog(EP) x Dummy PY(high - low) Q2				0.0624	0.0788	0.0116
				(0.0535)	(0.0601)	(0.0208)
Δlog(EP) x Dummy PY(high - low) Q3				0.0705*	0.0851*	0.174***
				(0.0423)	(0.0471)	(0.0320)
Δlog(EP) x Dummy PY(medium - low) Q2				-0.0581*	-0.0805**	0.158***
				(0.0301)	(0.0328)	(0.0238)
Δlog(EP) x Dummy PY(medium - low) Q3				-0.0784**	-0.0844**	0.143***
				(0.0352)	(0.0372)	(0.0243)
PY(high - low)	0.0360***					
	(0.00642)					
PY(high - low) squared	-0.00998***					
	(0.00179)					
PY(medium - low)	-0.0137					
	(0.0283)					
PY(medium - low) squared	-0.0352*					
	(0.0181)					
Log of farm size in 1996	0.00310*	0.00353**	0.00393**	0.00370**	0.00733***	0.00263***
	(0.00178)	(0.00177)	(0.00179)	(0.00173)	(0.00179)	(0.000550)
Log agricultural production value / ha in 1985	-0.0155***	-0.0144***	-0.00568	-0.0152***	-0.00215	0.00822***
	(0.00496)	(0.00481)	(0.00561)	(0.00481)	(0.00566)	(0.00119)
PY(high) - sugarcane	0.00488***	0.00621***	0.00682***	0.00595***		0.00315***
	(0.00155)	(0.00152)	(0.00163)	(0.00151)		(0.000690)
PY(high) - coffee	0.00922	0.00586	0.00633	0.00614		-0.00755***
	(0.00777)	(0.00764)	(0.00828)	(0.00751)		(0.00262)
PY(high) - cotton	0.0561*	-0.00456	-0.000136	-0.00161		-0.000306
	(0.0317)	(0.0283)	(0.0306)	(0.0285)		(0.00503)
Log populational density in 1985			-0.00756**		-0.0112***	-0.00683***
			(0.00317)		(0.00347)	(0.00126)
Log income per capita in 1985			-0.0156**		-0.00695	-0.00427***
			(0.00665)		(0.00679)	(0.00153)
Mesoregion fixed effects	Yes	Yes	Yes	Yes	Yes	No
Observations (census blocks)	7,591	7,591	7,397	7,591	7,397	7,397
R2	0.149	0.143	0.145	0.145	0.135	0.082
1st stage F-stat						14.49

Note: *** p<0.01, ** p<0.05, * p<0.1. All standard errors clustered at the municipality level.

Table E3 – First Differences Model for the Soy and GE Soy Expansion in the South

	South				
	Δ Soy Area		Δ GE Soy Area		
Variables	(1)	(2)	(3)	(4)	(5)
Δlog(EP)	358.3**	89.15	1,400***	804.2**	644.3*
	(172.2)	(139.2)	(397.4)	(368.1)	(361.4)
Dummy PY(high - low) Q2	86.99**	81.70*	275.8***	284.7***	320.3***
	(38.60)	(46.74)	(65.36)	(59.78)	(66.74)
Dummy PY(high - low) Q3	128.3***	68.01	244.1***	70.20	129.4**
	(38.18)	(49.90)	(51.22)	(46.16)	(50.97)
Dummy PY(medium - low) Q2	-60.93*	-22.79	1.512	99.40	106.9
	(32.63)	(36.91)	(57.13)	(63.81)	(69.21)
Dummy PY(medium - low) Q3	-121.8**	-58.05	-362.9***	-208.7**	-193.5**
	(55.72)	(58.64)	(85.56)	(90.83)	(96.95)
Log of farm size in 1996	163.0***	149.2***	291.9***	341.1***	340.9***
	(22.45)	(21.50)	(47.11)	(49.97)	(50.54)
Log agricultural production value / ha in 1985	-75.43**	-2.090	42.85	-84.62	-0.384
	(29.25)	(42.21)	(50.96)	(53.89)	(67.01)
Log populational density in 1985		-87.77***			-187.0***
		(26.77)			(47.50)
Log income per capita in 1985		15.50			15.50
		(41.92)			(41.92)
PY(high) - sugarcane	81.55***	60.91***	210.6***	96.68***	105.2***
	(16.34)	(16.62)	(27.52)	(20.44)	(22.70)
PY(high) - coffee	1.444	70.05	-367.3***	-138.6***	-205.6***
	(29.56)	(58.42)	(46.61)	(49.40)	(53.06)
PY(high) - cotton	-822.3***	-561.0*	-1,870***	-201.2	-423.0
	(271.4)	(287.5)	(415.5)	(334.8)	(373.9)
Meso region fixed effects	No	Yes	No	Yes	Yes
Observations (census blocks)	7,591	7,397	7,591	7,591	7,397
R2	0.057	0.103	0.107	0.169	0.174

Note: *** p<0.01, ** p<0.05, * p<0.1. All standard errors clustered at the municipality level.

Table E4 - First Differences Model for the Acreage Response Function in the Savanna

Dependent variable is Δ soy area share	Savanna						Midwest Savanna	
	OLS					IV	OLS	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δlog(EP)	0.0382***	0.0371***	0.0195***	-0.0113**	-0.0349***	-0.0295***	0.111***	0.101***
	(0.00647)	(0.00634)	(0.00703)	(0.00515)	(0.00695)	(0.0114)	(0.0326)	(0.0328)
Dummy PY (high - low) Q2		-0.000214	-0.00223**	-0.000457	-0.00250**	-0.00211**		
		(0.000950)	(0.00103)	(0.000942)	(0.00103)	(0.00102)		
Dummy PY (high - low) Q3		0.00605***	0.00461***	0.00520***	0.00373***	0.00355**		
		(0.00150)	(0.00144)	(0.00143)	(0.00138)	(0.00142)		
Dummy PY (medium - low) Q2		0.00564***	0.00677***	-0.0339***	-0.0352***	-0.0510***		
		(0.000878)	(0.000950)	(0.00608)	(0.00614)	(0.00866)		
Dummy PY (medium - low) Q3		0.00328***	0.00393***	-0.0249***	-0.0289***	-0.0275***		
		(0.00101)	(0.00104)	(0.00452)	(0.00467)	(0.00863)		
Δlog(EP) x Dummy PY(medium - low) Q2				0.0693***	0.0738***	0.1000***		
				(0.0111)	(0.0113)	(0.0158)		
Δlog(EP) x Dummy PY(medium - low) Q3				0.0505***	0.0589***	0.0568***		
				(0.00820)	(0.00865)	(0.0154)		
PY (high - medium)							-0.0106*	-0.00917
							(0.00622)	(0.00620)
PY (high - medium) squared							0.00466**	0.00374*
							(0.00206)	(0.00200)
PY (medium - low)							0.0704***	0.0928***
							(0.0214)	(0.0203)
PY (medium - low) squared							-0.0430***	-0.0653***
							(0.0165)	(0.0165)
Log agricultural production value / ha in 1985	-0.000243	-0.000209	0.00166*	2.80e-05	0.00234**	0.00271***	0.0279***	0.0205***
	(0.000569)	(0.000557)	(0.000987)	(0.000547)	(0.00100)	(0.00103)	(0.00562)	(0.00608)
Log populational density in 1985			-0.00619***		-0.00668***	-0.00595***	-0.0121***	-0.0129***
			(0.000975)		(0.000990)	(0.00111)	(0.00328)	(0.00334)
State fixed effects	No	No	No	No	No	No	No	Yes

Table E5 - First Differences Model for the Land-use Response Function in Brazil

Dependent variable is Δ agricultural area share	Brazil		South		Midwest Savanna	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Δlog(EP)	0.0337**	0.0187	0.0227	-0.00164	0.0942**	0.0752
	(0.0166)	(0.0180)	(0.0244)	(0.0236)	(0.0405)	(0.0492)
PY(high - medium)	-0.0227***	-0.00972	0.00622	0.00305	-0.0487*	-0.0492*
	(0.00847)	(0.00880)	(0.0220)	(0.0185)	(0.0253)	(0.0275)
PY(high - medium) squared	0.00320	-0.000163	-0.00270	-0.00256	0.0163**	0.0141*
	(0.00223)	(0.00231)	(0.00448)	(0.00386)	(0.00684)	(0.00804)
PY(medium - Low)	0.0183	0.00551	0.177***	0.113***	0.0779	0.00897
	(0.0212)	(0.0246)	(0.0296)	(0.0389)	(0.0708)	(0.0841)
PY(medium - Low) squared	-0.0135	-0.0175	-0.111***	-0.104***	-0.0444	-0.00585
	(0.0145)	(0.0144)	(0.0236)	(0.0265)	(0.0452)	(0.0475)
Δ Price Risk	-0.000119	0.000674	0.000583	0.00125	0.00147	-0.00738
	(0.00194)	(0.00188)	(0.00242)	(0.00240)	(0.00654)	(0.00638)
Δlog(EP) - sugarcane	-0.00202	-0.00252	-0.0126*	0.0105	-0.00308	0.0105
	(0.00637)	(0.00669)	(0.00705)	(0.00795)	(0.0138)	(0.0132)
Δlog(EP) - corn	-0.0266**	-0.0260**	0.0144	-0.00324	-0.00778	0.0783**
	(0.0120)	(0.0108)	(0.0228)	(0.0235)	(0.0279)	(0.0336)
Δlog(EP) - cotton	0.0862***	0.0397	0.254***	0.159*	-0.0602	0.0361
	(0.0245)	(0.0283)	(0.0674)	(0.0877)	(0.0661)	(0.0704)
Δlog(EP) - rice	-0.0116	-0.00143	0.00861	0.000912	-0.0123	-0.120***
	(0.00990)	(0.0117)	(0.00930)	(0.0123)	(0.0486)	(0.0400)
Log value agr. prod. / ha 1985	0.0343***	0.0102**	0.00634	0.0112*	0.0391**	-0.0256**
	(0.00504)	(0.00400)	(0.00854)	(0.00590)	(0.0159)	(0.0110)
Log population density 1985	-0.0106***		0.00162		-0.0108	
	(0.00338)		(0.00556)		(0.00729)	
Log of farm size 1996	0.0198***		0.0158***		0.0102*	
	(0.00223)		(0.00365)		(0.00582)	
Cropland share 1985	-0.0973***		0.0161		-0.386***	
	(0.0280)		(0.0424)		(0.145)	
Pastureland share 1985	-0.0861***		-0.0863**		-0.223***	
	(0.0212)		(0.0360)		(0.0509)	
Dummy for soy production in 1990	-0.00489		-0.00565		-0.00983	
	(0.00369)		(0.00524)		(0.00815)	
State fixed effects	Yes	No	Yes	No	Yes	No
Mesoregion fixed effects	No	Yes	No	Yes	No	Yes
Observations (census blocks)	32,537	33,576	7,717	7,912	2,172	2,172
R2	0.063	0.078	0.039	0.043	0.155	0.151

Note: *** p<0.01, ** p<0.05, * p<0.1. All standard errors clustered at the municipality level.