Market Concentration & Deforestation: Evidence from the Brazilian Soy Industry

Ben Rommelaere*

April 20, 2023

Abstract

The Brazilian soy industry is a leading cause of deforestation in several major biomes, including the Amazon, yet little is known about how market structure affects land-use outcomes. This paper exploits the 2014 acquisition and merger of two major soy exporters by China's state-owned firm COFCO to estimate the causal effects of buyer concentration on farmgate soy prices, production, and deforestation. Using a municipality-level panel from 2006–2018 linking supply-chain data, administrative data, and remote sensing based deforestation records, I implement a staggered event-study design to estimate local impacts of the merger. Results show a short-run increase in farmgate prices where buyer concentration rose — an unexpected effect consistent with strategic mutual forbearance among oligopsonists. In contrast, when COFCO enters new markets, I find that increased competition leads to sustained price increases and lower deforestation. One likely mechanism is COFCO's deforestation-free sourcing commitments.

*Department of Economics, University of Toronto, Toronto, Canada, b.rommelaere@mail.utoronto.ca.
I would like to thank Eduardo Souza-Rodrigues for his invaluable guidance and supervision throughout this project. I am also grateful to Yanyou Chen, Victor Aguirregabiria, and the participants of the University of Toronto IO seminar for their valuable comments and suggestions.

1. Introduction

Between 2010 and 2020, the UNFAO estimates that an average of 10 million hectares was deforested annually. This deforestation is concentrated in a handful of countries, with Brazil accounting for 20% of total global deforestation. Deforestation has global consequences largely because it is a major driver of carbon emissions. Since the beginning of the 2000's about 10%-15% of global carbon emissions can be attributed to deforestation (Stern, 2008; IPCC, 2013).[1] Deforestation is driven by many factors that influence the profitability of land intensive economic activities; such as agricultural prices, transportation costs, and land characteristics.[2] However, optimal policy provision for slowing deforestation is complicated by that fact that many of the industries responsible for deforestation, namely agriculture, are highly concentrated. While agricultural goods like soy or cattle are produced by thousands of small farmers, these farmers sell into concentrated markets such as grain processing and exporting or meat packing facilities. There is currently a gap in the literature on deforestation policy assessing how much farmer profits and land-use decisions are affected by changes to market structure and the subsequent implications for future environmental policy. This paper addresses this gap by estimating the impact of a change market structure in the Brazilian soy industry on farmer prices and decisions to determine the potential role market structure ultimately has on land-use decisions.

As Brazil has the highest global rates of deforestation, it is critical to understand the efficacy of various policies designed to combat deforestation. Within Brazil, there are six separate and unique forest and savanna biomes. Within regions which produce soy, two biomes have experienced a greater extent of deforestation: the Cerrado and the Amazon. Both of these biomes are of global importance as they both contain large carbon stocks and are two of the most biodiverse regions on the planet (IPBES, 2019). In these biomes a key industry driving deforestation is the expanding agricultural frontier for soy cultivation (Araujo, et al, 2020). While it is clear that the growth in the Brazilian soy industry has the potential for large negative environmental impacts, regulation is complicated by the market structure of the industry. This is because the market is highly concentrated. In 2017, the latest year of Brazil's decennial agricultural census, roughly 250,000 farmers produced soy; however, these farmers sell their crops within concentrated local soy purchasing markets to

[1] Perhaps of equal importance, as argued by Dobson et al, 2020, addressing deforestation can help protect us from future zoonosis outbreaks – like Covid-19. In other words, deforestation also directly raises the risk of pandemics which in turn have global consequences and large economic costs. This is in addition to the loss ecosystem services forests provide in terms of biodiversity, water availability, and water quality.

[2] There is a long standing literature on using structural economic models of land use choices to assess the drivers of deforestation beginning with Pfaff, 1999, and most recently summarized in Souza-Rodrigues, 2019.

the few competing soy processing and exporting firms who operate nearby their farm.[3] In other words, farmers are selling their goods to oligopsonists who have market power over farmgate prices. The impact of changes to this market structure, such as a reduction in oligopsony power should have straight forward short-run effects and higher farmgate prices. However, how much prices respond due to changes in market structure and how changes in farmgate prices within the soy market alone affect deforestation is an open empirical question.

To address these questions, I exploit a change in the market structure of local soy purchasing markets due to the entry of COFCO, a Chinese state-owned company, which purchased and merged two mid-sized Brazilian soy exporting firms, Noble and Nidera, in the fourth quarter of 2014. At the time of purchase, these two firms each accounted for roughly 3% of soy exports at a national level. The purchase and merger of led to COFCO immediately becoming the fourth largest exporter in 2015 in the Brazilian soy industry - a market historically dominated by 4 major exporters. Within local soy purchasing markets the firms operated in, Nidera and Noble accounted for 12.5% and 5% respectively. In total this led to COFCO becoming the largest firm in just over 90% of the markets the two firms previously operated in. COFCO's entry as an exporter in the Brazilian soy industry creates three distinct sets of local soy purchasing markets, allowing us to study potential asymmetric differences between an increase and decrease in market concentration. The first group, is the set of markets where Nidera and Noble were competitors prior to COFCO's purchase and experience a reduction in the number of competing soy exporting firms. The second group, is the set of markets where only one of Nidera or Noble were present prior to COFCO's entry in 2014 and experience no change in the local competition levels. The final group is the set of markets that COFCO newly enters, first in 2015, and then in all subsequent years. These markets experience increased competition from a new market entrant. The impact of entry on market outcomes is studied using an event-study design where the control group is defined as nearby never treated markets and each group is considered to be treated upon COFCO's entry. The set of markets COFCO expands into lends itself to a staggered event-study design. As has been shown in the literature, estimation of the ATT for an staggered event-study is not straight forward in the presence of rolling treatment dates or dynamic effects. To account for these issues, this paper estimates the treatment effect for this group using the two-stage difference-in-difference (2sDiD) estimator developed by Gardner, 2021. The extent to which these discrete changes to market structure causally impact the prices farmers receive and subsequent land-use decisions is the key empirical question this paper seeks to understand.

To employ this empirical strategy, data from Trase, which links annual exporter purchases

[3] A typical municipality contains 300-600 soy farmers selling to 1-4 soy processing or exporting firms.

to specific municipalities, is used to determine which purchasing markets COFCO entered and delineate markets into one of these three groups.[4] This data is then linked to outcome data. First, annual data from Brazil's main statistical agency, IBGE, which tracks farmgate prices and production of soy at the municipal level is linked. Next annual data on land-use decisions, including changes to forest cover, from Mapbiomas, are linked at the municipal level. Using these linked datasets, I then estimate the causal effect of a change in market structure on soy prices, soy production, and deforestation.

The results suggest that for the group of local soy purchasing markets experiencing a reduction competition that there was a counterintuitive, but short-lived, increase farmgate prices. This result is found to be robust to both alternative definitions of the treatment and control group. Further, I find a positive price effect even in markets where the merger effectively created a monopsonist. For the group of local markets which experience no change in competition, farmgate prices are not estimated to change. This result rules out several potential explanations for the counterintuitive result for the reduction group, such as: the elimination of double marginalization, a demand shock, or suboptimal strategic decisions due to either subsidies or biased beliefs of competitor strategies. Finally, the group of local purchasing markets COFCO newly enters experience a persistent increase in farmgate prices; however, this is accompanied by a reduction in deforestation in these markets. One possible explanation for why deforestation is lower even though the profitability of soy production increased is due to strong environmental commitments by COFCO.

2. Relevant Literature

This paper builds on a growing literature that has begun to develop detailed models of specific industries to better analyze deforestation policy. To measure the cost effectiveness of various policies in the Brazilian Amazon, Souza-Rodrigues, 2019, models the trade-off between forested land and deforestation by heterogeneous farmers. This work is built on by Araujo, et al, 2020, who include adjustment costs between pasture, soy, and forested land to determine the optimal level of forestation in the Amazon. Araujo, et al, 2020, reveals that farmers in the Amazon partially internalize the carbon emissions from deforestation due to Brazil's command-and-control policies in the region. Yet neither of these papers address the fact that both cattle and soy markets are concentrated industries which in turn may impact the welfare costs of a given policy.

There is a long standing literature on the interaction between imperfectly competitive

[4]To gain traction is estimating this model and due to data availability limitations, I assume a purchasing market is the same as a municipality.

markets and environmental regulation. Early work by Oates and Strassmann, 1984, consider the welfare impact of a Pigouvian tax in a monopolized industry. They show that the welfare losses from production efficiencies are more than offset by the gains in environmental quality. Subsequent literature began to investigate second-best Pigouvian taxes to balance these welfare trade-offs (Simpson, 1995; Van Long and Soubeyran, 2005). Fowlie, 2009, extends on this work by modelling the interaction between imperfect competition and incomplete environmental regulation of industrial pollution. Ryan, 2012, and Ryan, Reguant, and Fowlie, 2016, study a related problem by investigating the impact of environmental policy on market structure. They study impact of the US Clean Air Act on the cement industry and show that this environmental policy increased the market distortions associated with the exercise of market power due to increasing fixed costs. The results compared to a perfectly competitive market show welfare losses from increased market power outweigh the environmental benefits of the policy. They show that polices need to be designed to mitigate both market power and emissions leakage to deliver positive welfare gains. This paper will build on this literature by beginning to investigate the role market power in the soy industry plays in deforestation and how this can impact the welfare costs of a given policy.

This paper also relates to a line of literature at the intersection of trade and climate change. Notably, Dominguez-Iino, 2022, who develops a framework to study the impact of environmental tariffs in imperfectly competitive supply chains for trade-exposed agricultural industries, including the Brazilian soy industry. The results demonstrates that that market power exacerbates the regressivity of global environmental tariffs, harming farmers in the poorest regions most. However, how market structure is affected by environmental policy, and how the impact of changes to market structure affect land use decisions is not addressed. Therefore, this paper will contribute to our understanding of the implications the market structure of this industry has for environmental policy.

Finally, there is a growing literature looking at the rapid expansion of Chinese firms in a number of capital intensive industries such as steel, solar panels, and shipbuilding. Most notably, Kalouptsidi, 2018, develops a model of the shipbuilding industry to determine Chinese government subsidies to national firms. The author finds that there were large subsidies to the domestic industry leading to substantial reallocation of shipbuilding across the world to China. Further, there is a growing interest in the impact of Chinese entry and investment in countries worldwide which is being driven by China's Belt and Road initiative (Pigato and Tang, 2015). This paper contributes to this literature by investigating the impact of a large Chinese state-owned enterprise into Brazil to study both prices obtained by local farmers while also measuring potential external costs on environmental quality.

3. Market Overview

Soybean demand is largely driven by demand for meat and as global demand for meat increases, so does demand for soybeans. This is because the vast majority of soybean production processed into feed for livestock, such as poultry and pigs, while only a small percentage, roughly 2%, of soybean protein is consumed directly by humans (USDA, 2020). When soybeans are processed there are two main outputs: soybean meal and soybean oil. Soybean meal is used as feed for livestock, while soybean oil is used as a major feedstock to produce biodiesel. For both livestock feed and biodiesel, the main product substitutes for soybeans are other oilseed crops, such as rapeseed oil or oil palm. Over half of global soybean production is consumed in China, while the U.S., E.U., Brazil and Argentina also represent large sources of soybean demand (USDA, 2022). As soybean are a commodity the price is set internationally on market exchanges such as the Chicago Board of Trade, which sells both spot and future options for soybeans. However, for reasons discussed below, this is not the price farmers receive for soybeans when the crop leaves their farm, which known as the farmgate price.

Soybean production largely occurs in Brazil, the U.S., and Argentina. These three countries account for roughly 80% of total soybean production, and each export a large share of their domestic production directly to China. In Brazil the soy market is large, at a commodity level soybeans are Brazil's top export and Brazil is both the largest soy producer and exporter in the world. Soybean production in Brazil has more than doubled in the past decade, largely driven by an increase in exports to China. In 2018, over 55% of soybean production was directly exported to China. In total 70% of Brazilian soy is exported unprocessed, while 30% is processed domestically; however, roughly half of processed soybean meal is also exported.[5] The rapid increase in soy production has been shown to have large environmental externalities since the industry is associated with increased levels of deforestation (Rajao et al., 2020).

To better understand this industry, it is important to detail the key decisions and decisions makers within the supply chain. The first important decision maker is the farmer who faces a planting decision at the beginning of each growing season. Here the farmer first choses to invest in a specific land-use decision, such as keeping land forested, developing it for pasture, or growing a specific category of crop, each of which require different capital investments. Temporary crops such as soy or corn can be planted and harvested in one season and often

[5]In 1996 Brazil passed the Kandir law which eliminated the difference in export tax between soybeans and crushed soybeans, which prior to the law encouraged more soybeans to be crushed locally. Within only a few years this led to large changes in how the industry operated and led to lowered levels of processed soybean exports vis-à-vis unprocessed soybeans. The Kandir law was responsible for idle capacity in the soybean crushing industry in Brazil for years (Goldsmith, 2008).

have some overlap in capital requirements. While on the other hand, permanent crops such as sugar require years to mature and require significantly different capital investments. Once a farmer decides which category of crop to invest in, they form expectations about prices at harvest time.[6] Given price expectations, farmers them decide which crop to plant (e.g. corn or soy), how much to grow, and then allocate inputs and factors of production to maximize profits. Across Brazil, soy is planted between September and November and is harvested from January to March depending on local climatic factors. After harvesting soybeans, farmers then sell their production to grain traders.

Grain traders operate extensive distribution networks facilitating the movement of soy from farms to domestic consumption or export markets. The first stage in this system is delivery to local storage facilities where farmers deliver their soybeans from their largely farms by trucks. In addition to storage, these facilities provide drying, product grading, sorting, and marketing services. Drying is an important service as much of the harvest arrives at moisture levels above what is required for stable storage of soybeans. Often these facilities are in a relatively short distance to the farm, in the U.S. the distance is 30-80km, while in Brazil data on catchment areas is limited, but is estimated to be a longer distance (Informa Economics, 2016). Across Brazil, there are approximately 3200 of these storage facilities, owned by a mix of exporters, farming cooperatives, along with many sole proprietorships. Further, roughly 17% of farmers have on-farm storage structures (Dominguez-Iino, 2022). This stage of the industry is not concentrated and only two firms own more than 2% of the national storage capacity.

While most shipments from farms go directly to these local storage facilities, not all do. River, rail, and port facilities may also receive grain directly from farmers if they are in a relatively short distance to the farm (Informa Economics, 2016). Further, there is a lack of adequate storage capacity in many soy growing regions, leading farmers to have to sell products at further distances (e.g. direct to rail, river, or port terminals), or risk large post-harvest losses of soybeans (Pera et al., 2017). After the soybean is received at these storage facilities it can be stored several months while it is marketed for sale to domestic processing facilities or grain traders with nearby logistics terminals. In an average year, about 70% is sold to exporters and 30% to domestic processing facilities (USDA, 2022). While in the U.S., many options exist for managing price risk through the purchase of future contracts, usage of these options are limited in Brazil, leading to short storage times and farmers selling their crops quickly after harvest (Informa Economics, 2016).

[6]This would be likely be based on future options prices, however, the use of to these options contracts themselves are limited in Brazil.

Once the farmer agrees on a price for their product, the soybean is then transported by truck to larger distribution facilities such as river, rail, or port facilities. From these terminals, the soybeans are either delivered to domestic processing facilities or exported. This segment of the market is highly concentrated. At a national level, the number of soy farmers in agricultural census was 236,345, while in the same year there were 171 exporters and 56 soy processing companies. These processing and exporting firms are a mixture of large multinational firms and farming cooperatives. Historically, the industry has been dominated by four firms ADM, Bunge, Cargil, and Louis Dreyfus who collectively accounted for over 40% of all soybean exports in a given year as well as 40% of domestic soybean processing capacity.[7] These trading firms gain oligopsony power over soy purchasing markets because grain transportation is both costly and time consuming (US DOJ, 1999). Farmers sell grain within a limited geographic area surrounding their farms and for a given farmer it is common for there to be a limited number of potential purchasers nearby. In a stylized analysis of the soy market, Dominguez-Iino, 2022, shows that the concentration ratio of the top exporting firms within a municipality is negatively correlated with farmgate prices and positively correlated with a proxy for exporter margins. This implies this market concentration is a key factor in the price farmers receive and the profitability of exporters in this market.

From this it is clear that soybean exporters and domestic processing facilities are the two other important decision makers in the soybean supply chain. Prior to a given years harvest, soybean exporters make a series of investment decisions. First, they invest in export capacity at one or multiple ports. Second, they choose both the capacity and the location of their logistics infrastructure within soy growing regions to gain access to local markets.[8] Then during harvest time, exporters compete for soybean production by offering farmers a price per tonne which is some discount from the international price accounting for logistics costs to export the soybean. As this part of the industry is highly concentrated, the ultimate price exporters pay farmers is based on strategic interactions with other exporting firms who invested in logistics infrastructure in the same local soy purchasing market.[9]

As mentioned above, transportation costs to the port are a key determinant in the price farmers receive and in key soy growing regions these costs are roughly 30% of the international price of soybeans (Pera et al, 2017). This implies that highway improvement projects along with the development of new rail and barge facilities can cause large decreases in transportation costs of soy. For example, according to industry estimates, the completion

[7]Another major player in the industry is Amaggi, who is both the largest soybean farmer in Brazil, the 5th largest exporter, and the 5th largest domestic processor by capacity.

[8]The problem for soybean processing facilities is similar, but investment costs and factor input ratios would be different.

[9]In general, exporters and soybean processing firms would compete on prices at this stage.

of a new highway in the Brazilian Amazon reduced transportation costs by 25% to 30% compared to old routes to southern ports. Further, as most of the soy is harvested at the same time, along with the noted storage capacity issues, pressure to move soy from the farm to port increases road congestion and the price of cargo transportation increases sharply around the time of harvest (Informa Economics, 2016). This highlights the value storage facilities have for farmers as they allow farmers to avoid paying for higher transportation costs.

A major recent development in the Brazilian soy industry was the entry of China's largest food processor, manufacturer, and trader, the state-owned agricultural company, China Oil and Foodstuffs Corporation (COFCO). As the industry has historically been dominated by 4 firms, the entry of this new and large competitor has reshaped the market structure of the industry. COFCO first purchased two mid-level soy traders, Noble and Nidera in late 2014, and merged them together.[10] Based on national-level market shares, the merged firm became the 4th largest exporter in the year following the consummation of the merger.[11] Thus, this entry of a large new competitor has the potential to impact price competition among exporters and ultimately impact the decisions of farmers.

4. Data

Using annual data at the municipal level for the 2006-2018 period, I compile a dataset of soy production, prices, exporter information, and land-cover changes. As noted above, soybeans are purchased by exporters within close proximity to the location soy is farmed. Therefore, this paper assumes that a soy purchasing market is reasonably approximated by behaviour within a municipality.[12] While accurate annual surveys exist to measure farmgate quantities and prices at the municipality level, data on soy purchases by grain traders from farmers is not directly available. As described below, I use data which links individual company purchases to municipalities of production. However, this data has non-negligible measurement issues, which is briefly described below and in detail in Appendix A. After removing municipalities with unreliable grain trader purchasing information, the dataset contains detailed information on soy purchases in 2,256 municipalities and a total of 38,352 observations, where the unit of

[10]In September, 2014, COFCO acquired majority shares of Noble Agriculture, then in October, 2014, purchases majority shares of Nidera. In March 2016, COFCO purchased the remaining shares of Noble Agriculture, and in February, 2017, purchased the outstanding shares of Nidera.

[11]Prior to the acquisition, by exporter quantity, Noble was the 10th largest exporter in 2014 and Nidera was the 7th largest exporter.

[12]The concepts of a soy purchasing market and a municipality are synonymous throughout this paper, while the terminology Brazilian soy market refers to the national aggregate market.

observation is a municipality-year.[13] As described below, only a subset of this dataset is used to analyze the impact of COFCO's entry into the Brazilian soy market.

4.1. Farmer production & prices

Annual data for soy production, farmer revenue, and yields for each year from 2006-2020, is obtained from the Brazilian Institute of Geography and Statistics (IBGE) Municipal Agricultural Production (PAM) survey. This data provides the total area of land soy is produced on, the tonnes of soy sold by farmers, the total value farmers received (in Brazilian Real), and yields (tonnes per hectare) for each soy producing municipality in each year. It is important to note that not all farm level production will reach export markets or domestic consumption. This is due to a variety of factors including shrinkage when soy is dried[14], damaged or contaminated product, and losses during transportation. This dataset is used to measure farmgate prices, which equals total revenue divided by production, and farmgate quantity measured as production quantity.

4.2. Purchaser volumes

To link soy exports to individual soy trading companies I use supply chain data compiled by Trase. This dataset is constructed from customs based export data from 2006-2018, and maps purchaser quantities and FOB value back to the source municipality.[15] The data construction follows a form of material flow analysis known as Spatially Explicit Information on Production to Consumption Systems (SEI-PCS), originally developed by Godar et al, 2015. Data from Trase is also used to obtain the location, entry date, ownership, and capacity of grain storage silo's and soy crushing facilities. While this dataset is used in many studies of Brazil's soy industry, it is not without issues.[16] This dataset is used to measure presence of exporters, the quantity of purchased soy by purchasing firm, or crushing facilities for domestic use, within purchasing markets in a given year. However, mapping these volumes based on customs records data and domestic demand back to the source municipality is subject to measurement error. In brief, this process relies on export data which contains a municipal

[13] As farmers decide annually whether to grow soy, it is possible that in some years, all soy farmers in a region stop growing soy. This leads to an unbalanced panel due to exit of all farmers within a market.

[14] Drying soy removes moisture content, and therefore reduces the weight of the final product that is exported

[15] A detailed explanation of the process of mapping export to source municipalities is provided in Appendix A.

[16] For examples of how this data has been employed elsewhere, see Rajao et al, 2020, or Dominguez-Iino, 2022.

location of taxation for individual company asset (i.e. local storage facility, warehouse, train yard, or port) linked to the exporters purchase of soy. This does not necessarily equate to the municipality where the soy was farmed in, to complete this mapping models based on Godar et al, 2015, are employed. In some cases, soy export volume measured in the customs data cannot be mapped to an municipal location where they soy was ultimately farmed. These issues are heavily concentrated in a several dozen municipalities where there is not sufficient, or any, exporter data, yet there is substantial soy production as measured by IBGE. Since this purchaser data is keystone to delineating treatment and control groups, as described below, I drop all municipalities where purchaser information is missing within the Trase data.[17] The process of determining which municipalities can be included is described in detail in Appendix A.[18]

4.3. Land-cover changes

Data on land-cover changes by municipality for each year from 2006-2020 is obtained from MapBiomas. MapBiomas provides detailed land-cover transition data for a large variety of land uses including soy, pasture, other agricultural goods and forested land. For the empirical analysis section, I aggregate to forest and non-forested land to measure the loss of forested land as a percent of total land area in the municipality.[19] MapBiomas tracks land-cover transitions for each pixel of Brazil at a 30m resolution; however, for the purpose of this analysis, I use their aggregated statistics at the municipality level. This data is used to construct the main measure of deforestation, which is deforestation as a percent of remaining forested area.

4.4. Additional Data

Additional data used to understand the soy market, such as the number of temporary farms and their size, is obtained from IBGE's 2006 and 2017 Census of Agriculture. National

[17] The main concern is that COFCO must be present in at least some of these municipalities as roughly 10% of COFCO's exported volumes don't map to any municipality. Therefore, COFCO is purchasing soy from some subset of these municipalities. In turn, COFCO could be active or never active in any of these municipalities.

[18] The algorithm employed by Trase has been continuously updated over time and with each iteration, the percent of export flows missing decreases. Discussions with the data provider suggest that the next iteration, to be released Q1 2022, should continue the trend of reducing missing export flows further. Additionally, the next release will extend Trase data up until 2020 which is my only constraint to update the analysis through 2020.

[19] To better measure the extent of deforestation in the Cerrado, where the main forest type is carbon-rich savanna, my definition of forested land includes both forests and savanna.

average transportation costs to port is obtained from the USDA's annual Brazil Soybean Transportation guides between 2006 and 2018 These which provide data on trucking costs per 100 miles driven. These costs are combined with straight-line distances from the center of each municipality to the main export port for that municipality (revealed in Trase's dataset) to provide a proxy for local transportation costs. USDA data is also used to measure freight costs from each port to typical import locations in the EU (Hamburg, Germany) and China (Shanghai). This data is mapped to municipalities based on each municipalities main port of export.[20]

4.5. Defining Treatment Groups

While COFCO formally entered the market in late 2014, this was after the majority of soy harvested in the year was exported. Thus, the first year COFCO is considered to be an active grain trader in the Brazilian soy market is 2015. To study the impact of COFCO's entry to Brazil, we create three distinct groups of soy purchasing markets by the type of shock to market structure caused by COFCO's entry. First, there are the set of markets which experienced a reduction in the number of competitors. I select markets for this group by observing the presence of both Noble and Nidera using Trase data from 2014. In these markets, it is expected that farmgate prices should go down prior to entry as competition is decreased. Second, there are markets which experienced no change in the level of competition, but do experience a change in ownership to a large state-run firm. I select markets in this group by observing the presence of either Noble and Nidera in 2014, but not both. In these markets it is a prior ambiguous what impact the change in ownership will have on farmgate prices. Third, there are the markets which COFCO newly enters after the completed purchase of the two firms. This group includes all markets that had neither Noble or Nidera presence prior to the purchase, but COFCO is active in 2015 or subsequent years. As these markets experience the presence of a new competitor, the average price receieved by farmers is expected to increase in these markets.

Table 1 shows the count of each treatment group, broken out by biome and whether COFCO subsequently exits the market prior to the end of the sample period. First, this shows that COFCO's entry in Brazil only affected soy purchasing markets within in three biomes: the Amazon, the Cerrado, and the Atlantic Forest. Second, it shows a disproportionately high level of early exit from municipalities within the Atlantic Forest. This is likely driven by the proximity of many of these markets to the port of Santos whereby COFCO can source

[20] The USDA reports do not cover all export ports, so the cost of freight needs to be imputed for these port. This is done based on regressions of shipping costs on time fixed-effects and shipping distances.

direct to the port from farmers in a range of purchasing markets.[21] To restrict analysis where there is more robust evidence of sustained purchasing market entry, I restrict each treatment group in the analysis to markets where there is no exit post-treatment.

Table 1: Count of Treated Markets

	Amazon		Cerrado		Atlantic Forest	
	Early Exit	No Exit	Early Exit	No Exit	Early Exit	No Exit
Reduction	0	2	1	5	10	37
No Change	0	3	4	8	101	40
2015	0	3	1	3	15	6
2016	1	2	1	1	8	3
2017	0	2	0	5	0	13
2018	0	5	0	19	0	28
Total	1	17	7	41	161	127

4.6. Selecting the Control Group

My baseline identification strategy is to exploit the discrete change market structure in the markets COFCO enters compared to a set of markets COFCO never enters. This requires the selection of a control group from the set of municipalities where soy was grown prior to treatment, where COFCO is never active in the sample period. As the sample of potential markets that are never entered into by COFCO is large, I select a subset of markets as my control group. This is to establish a group of markets where observable and unobservable characteristics are similar across groups. Thus, the control group is defined as the group of markets that are contiguous to markets in each individual treatment group, but are never entered into by COFCO. For example, the control group selected for the reduction in competition treatment group is all markets contiguous to any market in the reduction in competition treatment group. Doing so ensures that markets where COFCO enters are unrelated to outcomes prior to entry and that the two groups have similar trends prior to entry. The geographic proximity of these control markets to the treated markets means that they are similar both observable characteristics, such as distance to port and available nearby storage, but also unobservable characteristics and shocks, such as local changes to the transportation network, weather related shocks to yields, or potential unobserved closures of key local infrastructure. Finally, for the expansion treatment group, in addition

[21]Municipalities in the Atlantic Forest are also smaller in size. Future robustness checks will investigate altering the market definition in the Atlantic Forest to a more aggregated regional level, the meso-region, to keep the size of the purchasing market consistent across the country.

to the contiguous never treated group, I utilize the rolling treatment date to also include the not-yet-treated within the control group. Appendix B investigates the robustness of the results to an alternative definition of the control group.

Table 2 shows a summary of the different categorizations of municipalities within the three biomes affected by COFCO's entry. The first thing to note from this is that Trase exporter data is more likely to be missing in the Amazon. Second, we see that the majority of municipalities Nidera or Noble were active in and exited prior to COFCO's entry are in the Atlantic forest. The no exit row for each of the treatment groups represents the total number of markets included in the analysis, while the contiguous line represents the number of control markets for all treatment groups selected in the data.

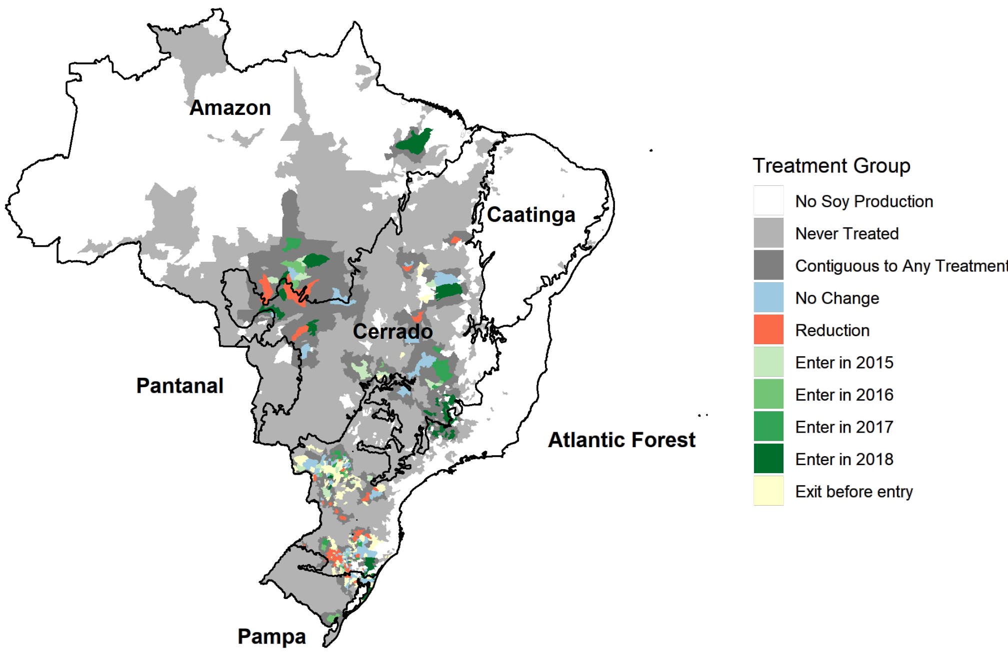
Table 2: Categorization of Municipalities

	Biome of Production		
	Amazon	Cerrado	Atlantic Forest
Soy Production	175	817	1365
Missing Trase Data	67	101	44
Never-Treated	90	668	1036
Contiguous	*11*	*135*	*208*
Non-Contiguous	*79*	*533*	*828*
Previously Active before COFCO Entry	0	4	99
Reduction	2	6	47
No Exit	*2*	*5*	*37*
No Change	3	9	66
No Exit	*3*	*8*	*40*
Entry (any year)	13	30	73
No Exit	*12*	*28*	*50*
No Soy Production	321	290	1359
Total Count of Municipalities	496	1107	2724

Figure 2 maps this control group along with each of the treatment groups analyzed in this paper. This map shows one additional group not yet mentioned, markets which Noble or Nidera were active in prior to 2014 which they never re-enter.[22] This group is dropped from the analysis and is not included in any treatment or control group. As shown in the next section, many of the markets entered in the Amazon and Cerrado have relatively large soy production.

[22]The group is labeled 'Exit before entry'.

Figure 1: Categorization of Municipalities



4.7. Summary Statistics

To better contextualize the results, I summarize temporal and spatial trends within the Brazilian soy industry. Figure 2(a) and Figure 2(b) highlight that soy production largely occurs in the three biomes affected by COFCO's entry: the Cerrado, the Atlantic Forest, and the Amazon. Each of these biomes are of global importance as they have large carbon stocks and are some of the most biodiverse regions on the planet (IBPES, 2022). Spatially, we see that soy production in 2018 was spread across the country; however, several municipalities are responsible for large volumes soy production - notably in the Amazon and Cerrado. Over the past decade we see that there has been increasing soy production in each of these biomes, but the rate of growth has been highest in the Amazon. Between 2004 and 2018, soy production in the Amazon went up over 300%, while in the Cerrado and Atlantic forest production doubled.

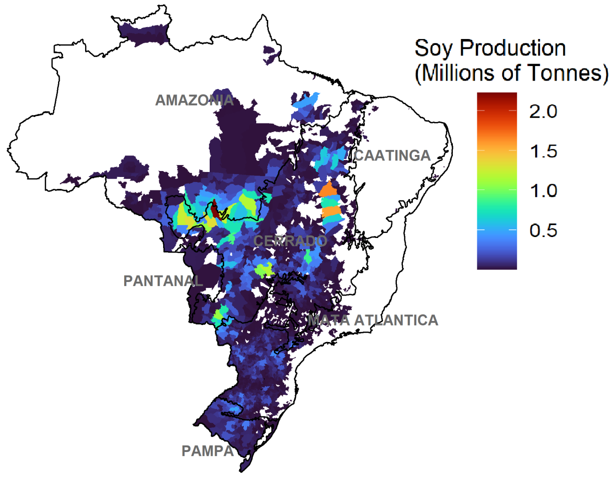
Figure 2(c) shows that the rate of deforestation in both the Amazon and Cerrado were falling in the early 2000's, but has since begun to increase. By 2018, within municipalities where soy was grown, annual deforestation reached 11,000 square kilometers of forest in the Amazon, roughly twice the size of the state of Delaware. While deforestation in the Amazon and Cerrado are much larger than in other biomes, the amount of deforestation in these biomes is not inconsequential. Specifically, the rate of deforestation in the Atlantic Forest has been relatively constant at 2,500 square kilometers a year over the past several decades. From 2010 to 2020, this amounts to a loss of forest in this biome equal to the size of state of Massachusetts. Table 3 shows the transition matrix for various field types across Brazil to better demonstrate the drivers of this deforestation. The first thing to note is that land-use is highly persistent, for example, the probability that a given pixel is forested given it was forested last year is above 98%. This also shows that the probability of conversion from forest to soy is the same as all other crops combined, excluding pasture. While pasture is the main form of agricultural land use in Brazil, it has been falling during the sample period, while soy land has expanded. Further, this expansion in soy land area is larger than all other agricultural crops combined. The type of land use most likely to be converted to soy is land used for other crops, followed by pasture and forest. The high probability of transition from other crops to soy may be capturing measurement error in the satellite data and is simply highlighting the role of crop rotation, or temporal differences in the growing season as soy and corn are often double cropped within a growing single season.[23]

As market concentration is posited to be an important determinants of farmgate prices, it is important to understand how concentration has changed over time. Figure 2(d), shows the weighted average Herfindahl–Hirschman Index (HHI) across local soy purchasing markets by forest biome, using market level soy production as weights. This shows that farmers sell into a highly concentrated markets. We also see that these markets have been becoming less concentrated throughout the study period in the Cerrado and the Amazon. Yet, even at the lowest point, the level of market concentration is much above 2500, the standard threshold used by the US DOJ, to delineate markets that are considered to be highly concentrated market. The trends in market concentration are reversed in 2018 in the Amazon and Cerrado which revert to the highest level in a decade. Storage capacity, another determinant of farmgate prices, has been on the rise in each of the biomes where soy is grown. In the decade preceding 2018, storage capacity doubled in the Amazon, increased by 30% in the Cerrado and 20% in the Atlantic Forest.

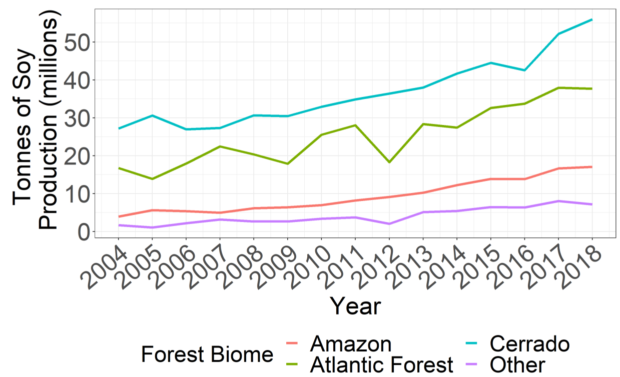
Table 4 presents key descriptive and control variables for the pre-treatment year of 2014 for each treatment and control group. Here we see that the treatment and control group are relatively similar on observables for the reduction in competition group. However, the control group has relatively more soy production, is more concentrated, and is less likely to have observations in the Cerrado; yet, differences are within one standard deviation of the

[23]DePaula and Fortes, 2019, show that in the Cerrado, roughly 50% of farmers double-crop their fields. That is they grow soy at the beginning of the season, and after the soy harvest they plant corn on the same field.

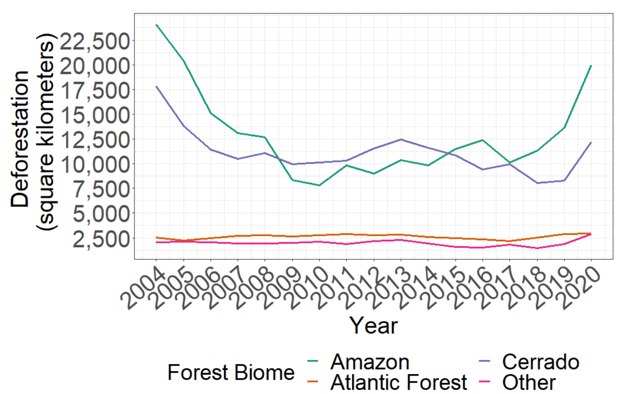
Figure 2: Temporal and Spatial Market Trends



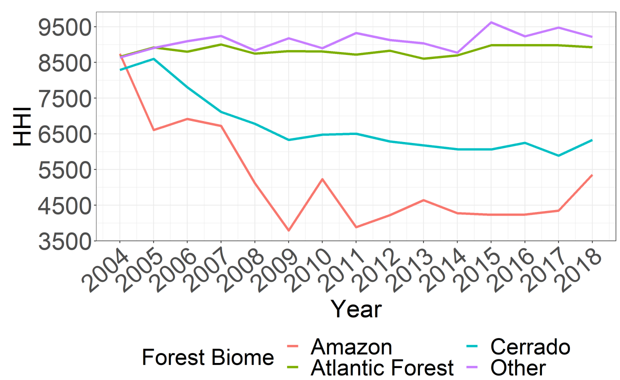
(a) 2018 Soy Production



(b) Annual Soy Production by Biome



(c) Deforestation in Soy Growing Municipalities



(d) Weighted Average Market Level HHI by Biome

Table 3: Transition Matrix Across Land Use Categories

	To						% of Land Area	
From	Soy	Pasture	Other Crops	Forest	Other Veg.	All Other	2006	2018
Soy	95.3	0.3	4.3	0.0	0.1	0.0	2.3	3.8
Pasture	0.2	95.9	2.1	1.5	0.2	0.1	19.1	18.3
Other Crops	2.6	3.4	90.4	2.6	0.8	0.2	7.2	8.2
Forest	0.2	0.6	0.2	98.6	0.4	0.0	61.5	60.1
Other Veg.	0.1	0.7	1.2	0.7	96.6	0.6	7.1	6.8
Other Land Uses	0.0	0.3	0.7	0.7	1.7	96.5	2.7	2.7

Cells represent the average annual probability (in %) of transitioning from the land use type listed in the rows to the land use type listed in the columns during the sample period. Other crops include cells identified as a mosaic of crops and pasture.

treated group. In the no change in competition group, the control group is also similar on observables but has slightly more soy production, less concentrated markets, and is more likely to be in the Cerrado. All differences are within one standard deviation of the treated group. Finally for the group of markets COFCO enters, we see notable differences among the treated markets based on the year of entry. Firstly, soy production is lower in the 2016, 2017 entry year. Second, the the percent of markets in the Amazon is much lower in the control group than all treated groups. On all other measures the never treated group is roughly the median of the treated groups. Although, Amazonian markets are underrepresented in the never-treated control group, as described below, when analyzing the causal effect of entry both the never-treated and not-yet-treated groups are utilized. This works to alleviate observable differences between these groups. Finally, this table further highlights that, on average, farmers in these markets face purchasers with market power as the average market has roughly 150-600 farmers but only 2-3 soy purchasers.

Table 4: Pre-Treatment Summary Statistics

		Reduction		No Change				Entry		
Variable		Treated	Control	Treated	Control	2015	2016	2017	2018	Control
Count of Purchasing Markets		44	90	49	109	12	6	20	52	155
% in Amazon		4%	5%	4%	3%	25%	33%	11%	16%	3%
% in Cerrado		19%	11%	11%	39%	25%	17%	28%	39%	43%
Farmgate Price	*2006 BRL/Tonne*	596.2	601.4	605.9	592.6	583.4	550.7	597.1	590.7	596.48
		(39.9)	(41.9)	(33.6)	(50.1)	(71.5)	(61.2)	(45.1)	(36.4)	(39.2)
Farmgate Quantity	*1000 Tonnes*	99.8	136.1	65.4	78.2	130.8	84.6	33.9	140.0	75.32
		(221)	(342)	(123)	(139)	(218)	(88)	(67)	(277)	(171)
Deforestation	*% of Forest Area*	2.23%	2.05%	2.15%	1.90%	1.92%	1.65%	2.11%	2.43%	2.30%
		(0.009)	(0.008)	(0.014)	(0.008)	(0.01)	(0.005)	(0.009)	(0.011)	(0.013)
Soy Area	*% of Total*	23.0%	27.3%	18.9%	18.0%	23.1%	20.2%	14.2%	12.1%	17.0%
		(0.188)	(0.261)	(0.192)	(0.199)	(0.234)	(0.192)	(0.217)	(0.205)	(0.222)
# of Soy Farms*		436	502	300	260	625	303	217	149	260
		(651)	(439)	(256)	(298)	(1130)	(482)	(237)	(153)	(364)
# Of Exporting Firms		3.2	1.9	2.1	1.6	2.3	1.8	1.2	2.7	1.6
		(2.86)	(1.78)	(1.65)	(1.19)	(1.87)	(0.75)	(0.43)	(2.55)	(1.32)
Local Market HHI		4,319	5,947	6,807	4,547	6,976	6,022	7,222	7,098	6,419
		(4684)	(1560)	(4243)	(4765)	(3913)	(4899)	(4609)	(3869)	(4235)
Distance to Port	*Miles*	402	322	309	404	450	481	382	352	406
		(256)	(234)	(171)	(197)	(239)	(291)	(165)	(263)	(204)

5. Empirical Approach

To estimate the effect of the discrete change in market structure caused by COFCO's entry, I implement a event-study design to identify and estimate the causal effect. Figure 3(a) shows the discontinuous shock to market shares national level from COFCO's entry. COFCO immediately became the 4th largest exporter in the Brazilian soy industry in 2015, but afterward COFCO's national market share fell post-entry.[24] Upon entry, the local soy purchasing markets where Nidera and Noble were previously active are impacted differently based on the prior history of competition between these firms, as described above. Figure 3(b) shows the weighted average market share of all markets Nidera, Noble, or COFCO were ever active in. From this it is clear that at a local level, the combining of Nidera and Noble led to COFCO becoming, on average, the largest grain trader in the local soy purchasing markets it operated in. The market share here also declines after entry, but this is in part also driven by entry into new markets where their market share is lower. This shows that the change in market structure for these soy purchasing markets was large, especially in markets where Noble and Nidera previously competed. Figure 3(c) shows that the two purchased firm, Noble and Nidera, were expanding prior to COFCO's purchase. Specifically, 2014, and the year following COFCO's entry, 2015, saw large increases in soy purchases.[25] Total purchases by the merged firm however peaked in 2015. As seen in Figure 3(d), the new merged firm changed purchasing patterns and began expanding into new markets across Brazil. COFCO became active a number of new logistics hubs and soy purchasing markets following entry; tripling its presence in the Amazon and steadily increasing its presence in the Cerrado. It is this change in soy sourcing behaviour that leads to local soy purchasing markets experiencing increased competition from a new entrant.

5.1. Estimation Strategy

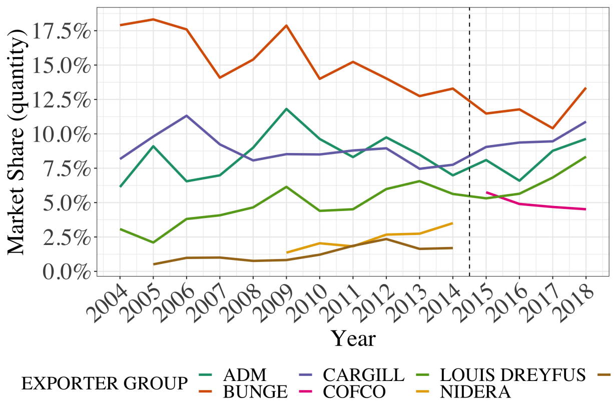
To identify the causal effect of a change in market structure, I individually examine each of the three treatment groups described in the previous section separately. The estimation strategy for the reduction and no change groups are the same. However, due to the rolling treatment timing of the expansion group I will employ a related, yet alternative estimation strategy to circumvent known issues in the canonical model. This strategy will be used to estimate the effect of COFCO's entry on farmgate prices, farmgate production, and deforestation.[26]

[24]The fall in market share post entry could be due to learning effects, among other reasons. COFCO experienced documented integration issues between the two firms, as well as legal issues with Nidera due to accounting irregularities discovered after the merger was complete. Due to these issues, COFCO put Nidera up for sale in 2017; however, they found no suitable bidder and did not end up divesting Nidera's assets.

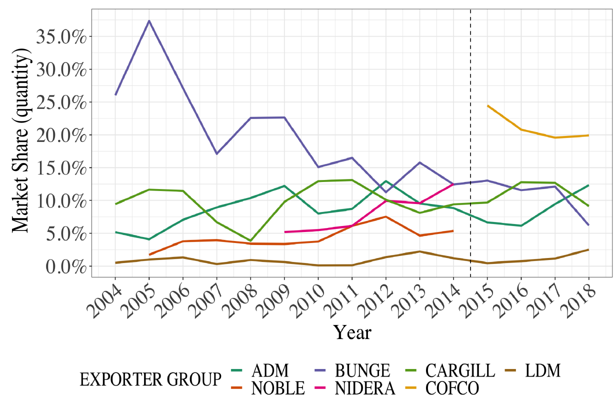
[25]It is possible that the large increase in 2014 was driven by an anticipation effect; whereby COFCO began importing larger flows from these firms the year they were in negotiation to purchase each firm.

[26]Note that the model is estimated separately for each treatment group, as opposed to pooling these groups into one estimating equation, as each treatment group is compared to a different control group.

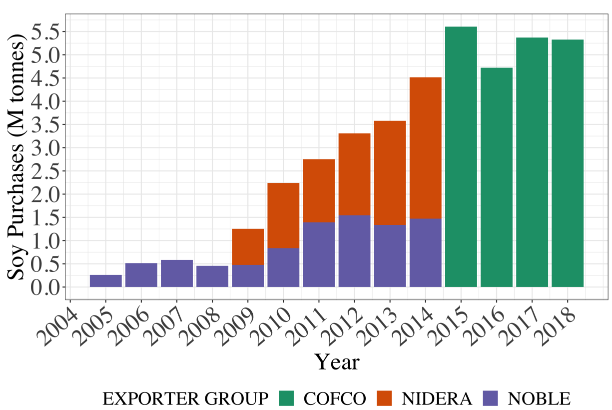
Figure 3: Impact of COFCO Entry on Soy Market Structure



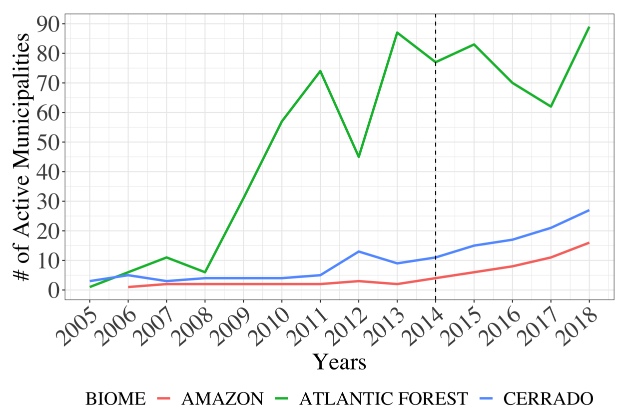
(a) National Market Share by Exporter



(b) Local Market Share by Exporter



(c) Purchases Pre- and Post-Acquisition



(d) Active Municipalities

5.1.1 Canonical TWFE Event-Study

Both the reduction and no change treatment groups are impacted by a change in market structure in 2015, the first year COFCO can make purchasing decisions over the soy harvest. Therefore, the canonical TWFE regression model will provide an unbiased estimate of the treatment effect. For each of these treatment groups, and their respective control group, I then estimate the following equation:

$$y_{it} = \sum_{t \neq 2014} \delta_t D_i 1[Year = t] + \beta' X_{it} + \alpha_i + \alpha_t + \epsilon_{it}$$

where y_{it} is the outcome of interest (i.e. prices, production, or deforestation), α_i is a market fixed-effect, α_t is a time fixed-effect, D_i is a treatment indicator variable equal to 1 if the market is entered into by COFCO, X is a vector of additional controls, and ϵ_{it} is our error term. In this model, δ_t estimates the ATT for each year post treatment. The estimates of δ_t allow for tests of pre-treatment trends that may violate the necessary parallel trends assumption required to identify the ATT as well as providing an understanding of the potential dynamic effects of entry. The vector of additional controls X include: shipping prices to the municipalities main export port, ocean freight costs from that port to Shanghai, China and Hamburg, Germany, and the municipal level of storage capacity. Further, X includes state, biome, and regional fixed effects and state, biome, and regional linear time trends.[27] The inclusion of these fixed effects and time trends flexibly capture unobserved changes in policies or the transportation network that occur over time and may otherwise impact farmer decisions. Finally, ϵ_{it} is clustered at the region-year level, as errors are likely to be correlated at this level to due common region-level shocks.

5.1.2 Identifying Assumptions for the Canonical TWFE Model

For the canonical TWFE event-study model to estimate the causal effect of a change in market structure, a number of identifying conditions need to be met. The key assumption is that of parallel trends. In this case, I assume that, in the absence of COFCO's entry, treatment and control markets would have followed common trends in the relevant outcome variables. A violation of this assumption will bias the results. A feature of the event-study

[27]The Brazilian statistical agency groups together municipalities into larger meso-regions. I use this meso-region definition to include sub-state regional effects and time-trends. The idea here is that these regional effects in turn capture the impact of changing infrastructure conditions that impact prices, quantities, and deforestation within the region. Further, policies governing the soy industry can differ by biome (e.g. the Forest Code) and state (e.g. financing support to farmers).

specification is that we will be able to observe if this assumption holds in the pre-period by testing if $\delta_t = 0$ in each period prior to entry. In general, the parallel trends assumption removes any potential selection bias when studying the effect of treatment on the outcome variables. The usage of a geographically similar control group is to ensure that the two groups are similar in terms of unobservables so that the selection bias is eliminated.

The second identifying assumption is that of no confounding treatments. There must be no other differential changes between the treated markets relative to their control group at the onset of entry, other than the entry of COFCO itself. The biggest concern is that COFCO selects which markets to enter based on expected changes in transportation costs due to the completion of infrastructure projects.[28] By including a rich set of fixed effects and regional, state, and biome time trends my regression specification removes any differential effect that may be present along these dimensions.

Finally, specific to this paper, a third implicit assumption is being made that markets and the strategies of firms across markets are not connected. That is, changes to strategies or entry into a given market do not affect the strategies or behaviour of firms in other markets. This assumption would be violated if entry into a given market leads to economies of scale or if there is multimarket competition among firms. For tractability, this paper assumes that markets and firm behaviour across markets is disconnected.

5.1.3 Event-Study with Rolling Treatment

The expansion group is the set of markets that COFCO first enters in 2015, and then in each subsequent year. Thus, this treatment group lends itself to staggered difference-in-difference design, whereby the treatment time is varying for the group of markets that are treated at different times. Adjusting the canonical model to this set up yields the following regression:

$$y_{it} = \sum_{t \neq 2014} \delta_t D_{it} 1[Year = t] + \beta' X + \alpha_i + \alpha_t + \epsilon_{it}$$

where the main difference is that now the treatment indicator variable, D_{it}, indicates if the market is treated by time t. There is a growing literature shows that this standard TWFE difference-in-difference estimator and related event-study design may not identify the

[28]While COFCO did not directly select which markets to enter, they did select which exporters to purchase. For this type of selection issue to be meaningful in this context, COFCO would have needed to know the markets each firm was active in prior to entry. The public availability of Trase data suggests this would have been common knowledge.

causal parameter of interest in this setting (Borusyak and Jaravel, 2017; de Chaisemartin and D'Haultfoeuille, 2020; Goodman-Bacon, 2018; Abraham and Sun, 2018; and Athey and Imbens, 2021). These papers show that the parameter identified and estimated by the standard models actually represents a weighted average of the dynamic treatment effects. Further, these papers illustrate that these weights can even be negative in certain cases, such as selective treatment timing. In the empirical context of this paper, this is very important as COFCO's market entry strategy may be changing over time due to selection or updating beliefs about the market. In other words, we should expect heterogeneous treatment effects in this context.

A decomposition of the weighting of the underlying treatment effects by Gardner, 2021, highlights the essence of the problem. When including period and unit fixed effects the above TWFE model projects the heterogeneous treatment effects onto the fixed effects. To circumvent this issue, I employ the two-stage difference-in-difference estimator (2sDiD) developed by Gardner, 2021. The logic of the estimator is simple: first estimate the fixed effects separately on untreated units, so that they aren't projected onto the treatment effects. That is, the first stage estimates the only the fixed effects, $\hat{\alpha}_i$ and $\hat{\alpha}_t$, and covariate parameters $\hat{\beta}$, when $D_{it} = 0$. This gives consistent estimates for the fixed effects and estimate of the impact of covariates under the parallel trends assumption. Next, the outcome variable is residualized by removing the estimated effect from the measured outcomes: $\hat{y}_{it} = y_{it} - \hat{\beta}'X + \hat{\alpha}_i + \hat{\alpha}_t$. Next the second stage a consistent and unbiased estimate of the average treatment effect is obtained by regressing this transformed outcome variable on our treatment status: $\hat{y}_{it} = \sum_{t \neq 2014} \delta_t D_{it} 1[Year = t] + u_{it}$. For inference, the estimated standard errors in the second stage will be incorrect as the dependent variable is a generated regressor. Therefore, the model is estimated using GMM to obtain asymptotically correct standard errors. Since this solution retains the regression framework of the canonical TWFE model, the same regressors and level of clustering can be included for this group.

5.1.4 Identifying Assumptions for the Rolling Treatment Model

The identifying assumptions for this estimation strategy are very similar, but with some notable differences. First, the treatment and control groups must have common support with respect to the fixed effects and covariates for these values to be identified in the first stage. In this context, we would fail to identify the fixed effect for markets where COFCO is the first firm to ever purchase soy from the entered market. Put another way, the treatment effect cannot be identified for markets that COFCO enters which were previously not growing soy. In total this leads to two markets being dropped for the 2018 treatment group when estimating this model. Second, we must have parallel trends to consistently estimate the parameters in the first step. Put another way, this procedure only works so long as the fixed effects are consistently estimated in the first stage. However, these fixed effects could have bias if the parallel trends assumption fails or if there are anticipation effects. Therefore a violation of either of these assumptions would bias our results. Thus, we still need to investigate the estimated treatment effect in the pre-treatment period to look for evidence against parallel trends. Finally, as before, there must be no confounding treatments. As argued above, by including a rich set of covariates, we can alleviate these concerns.

6. Results

This section summarizes and discusses the results for each treatment group using the control group defined in Section 4.6. Appendix C investigates the robustness of these results to an alternative definition of the control group and finds no qualitative change in the estimated effects.

6.1. Reduction in Competition Group

Table 5 summarizes the impact of COFCO's entry on the group of markets experiencing a reduction in competition. For each variable of interest, the odd numbered column represents the canonical model with no additional control variables, while even numbered columns include all controls described above. Figure 4 plots the coefficients from estimation with controls. This shows that for farmgate prices, the inclusion of control increases the plausibility of the parallel trends assumption. With controls, there is still evidence against parallel trends for prices as the impact in 2012 is weakly significant. This issue aside, the model estimates an significant increase in farmgate prices in the first year by roughly 6% and a weakly significant increase of 4% four years after the merger. However, there is no clear trend in prices post-entry. On the other hand, there appears to be an upward trend in farmgate quantity post-entry, but these results are not significant. Finally no effect on deforestation is detected for this group.

6.2. No Change Group

Table 6 summarizes the impact of COFCO's entry on the group of markets experiencing a no change in competition, while Figure 5 plots the coefficients from estimating the model with controls. This shows that across all three variables, the inclusion of these controls increases the plausibility of the parallel trends assumption. With controls, there is still evidence against parallel trends for deforestation as the impact in 2007 is weakly significant, and negative in most periods. The model estimates a null effect on prices post-entry. An upward trend in farmgate quantity is estimated post-entry, but these results are not significant. Finally, deforestation is estimated to decrease three years post-entry, but this needs to be interpreted with caution due to issues with parallel trends.

6.3. Expansion Group

Table 7 summarizes the impact of COFCO's expansion into new markets, while Figure 6 plots the coefficients from estimating the model with all controls. From this we can see that there is a clear trend towards higher prices after COFCO enters. After 2 years prices increase by 3% and after 3 years prices increase by 5.7%. While there appears to be an estimated increase in quantity produced, the estimates are not precise enough make the inference that COFCO's entry increased demand for soy. Finally, there is evidence that 4 years after entry, deforestation levels decrease in the markets COFCO expands into.

6.4. Discussion

The most striking result is that the estimated effect of COFCO's entry on prices are not in the expected direction. For the group experiencing a reduction in competition, the increase in oligopsony power should lead to a reduction in prices. However, we find a positive price effect in the year of entry. One potential reason for this could be due to the definition of this treatment group which only requires Nidera and Noble were competing in 2014, but not in any other previous years. In Appendix B an alternative specification is used to see if the results are different in markets with a longer history of competition between Noble and Nidera. However, this shows that the counterintuitive result is not driven by the treatment group definition. Moreover, the estimated price increase from the increase in oligopsony power is even found to be higher for markets which have a longer competitive history between Noble and Nidera. Another potential factor is that there could only be negative price effects in markets where the merger of Noble and Nidera created a monopsonist or duopsonist. In Appendix B, heterogeneity of the result is investigated by the pre-treatment competitor count to investigate if there are negative price effects in these markets. These results show that there is still a positive price for local soy purchasing markets where COFCO became a monopsonist and no effect in markets where COFCO became a duopsonist. In other words, this does not add clarity to what is driving this result.

Another potential cause for this result could be that COFCO's entry led to a demand shock which increased prices. However, the estimate on total farmgate production while trending upwards is not significant. Further, if the results were driven by a demand shock, then the results for the no change in competition group should be the same, yet this is not what is found. One could also posit that prices increase even though competition is reduced as COFCO is a heavily subsidized firm making suboptimal decisions as in Kaloupstidi, 2018. Or, that more generally, the results represent learning by a new market entrant as in Aguirregabiria et al, 2020. However, neither of these hypotheses can be reconciled with the null result observed for the no change in competition group. Entry into Brazil as an exporter also led to COFCO becoming an importer of Brazilian soy for their processing plants in China. This implies that this result could be driven by an elimination of double marginalization. Once again, this would imply that the no change group should experience a price increase as well, which is not the case.

Another cause for these results could be due misspecification of the geographic boundaries of a market leading to bias in the results. If the geographic boundaries are larger than a municipality and the control group consists of contiguous markets to treated markets, then the control group could include treated units. Appendix C investigates this by specifying an alternative control group of markets that are within the same meso-region, but are not-contiguous to any treated market. This approach ensures that the control group is similar to treated markets on unobservables while removing the potential that this form of misspecification is driving this result. The results from this robustness exercise are qualitatively the same as those reported above for all treatment groups. Therefore, this does not appear to be a major driver of the counterintuitive results.

Perhaps the most compelling explanation for this result is that the assumption that markets are not connected is violated and that there is multimarket competition among firms. In this environment, when firms meet their competitors in multiple markets this can lead to them coordinating their strategies across markets (Sengul et al., 2015). The outcome of this form of competition is known as mutual forbearance, where competition is reduced in each individual market. In this context, COFCO would have an incentive to avoid undercutting their rivals in the short run in the local purchasing markets where they gained market power. This is because rival firms can respond in future periods within that market, or any other market the firms compete in. In this framework, the contemporaneous price increase after COFCO merged Nidera and Noble could viewed as a signal to rival firms that COFCO intends to maintain the strategy of mutual forbearance across local soy purchasing markets.

Finally, it is important to point out that none of the estimates suggest COFCO's entry had negative environmental effects. In fact, we see the opposite. Specifically for the expansion group the estimates show a reduction in the rate of deforestation after COFCO enters these soy purchasing markets. This does not suggest that the soy industry does not drive deforestation - but rather that the shock to market structure caused by COFCO did not increase deforestation rates. One reason for this, and the estimated decrease in deforestation in the expansion group, is COFCO's stated strong commitment to deforestation free soy.

7. Conclusion

This paper provides the first estimates of the impact of a change in market power of soy grain traders in the Brazilian soy industry and the impacts on farmers. I study the entry of COFCO as a grain trader, who purchased and merged two mid-sized grain trading firms, Noble and Nidera. I find that markets experiencing a reduction in competition due to the merger, had a counter-intuitive temporary increase in farmgate prices. However, quantities and deforestation are not found to respond to COFCO's entry. The results for other treatment groups do not help us better understand the mechanism behind this counterintuitive result. Encouragingly, the results suggest that there were no obvious negative environmental effects upon China's entry into this market.

Future work for this project should focus on several areas, including a reformulation the definition of a market. While it would be more in line with how the soy market is organized to use the logistics hubs, as defined in Trase data, as the definition of a soy purchasing market, this is not without is challenges. Notably, it is not possible to directly link farmgate prices to a logistics hub and would require a series of assumptions. The plausibility of changing this definition to logistics hubs will be investigated in future work. Additionally, using updated Trase data when released, may change the results due to refinements in their ability to map exports to the municipality of soybean harvest. Further, utilizing better data for the control variables may help improve the precision of the estimates while simultaneously reducing the risk of measuring coinciding events. Most important would be compiling better data for transportation costs to account for how a exogenously changing transportation infrastructure is impacting farmer decisions. Further, utilizing property registration data which contain geographic boundaries of all farms in Brazil in conjunction with satellite imagery would allow for much more accurate measurements of the farmer side of this problem, including the number of active soy farms in each year, the size of those soy farms, and changes in on-farm storage.

Finally, to better understand developments beyond COFCO's entry, future work should develop a structural model of the industry. This model shall take into account the farmers profit maximization problem which decides if to grow soy or other crops, how much acreage of soy to grow, and investment into on farm storage. This model shall also take into account the exporter profit maximization problem which decides where to purchase soy, where to invest in export assets, and what prices to pay farmers. The results from this paper suggest that the appropriate model of competition among firms in this industry may be one of multimarket competition. A more detailed modelling of profits for both farmers and exporters will allow for a wider array of interactions between market power and environmental externalities in this industry to be studied.

Table 5: Event-Study Results for Reduction in Competition Group

	Farmgate Prices		Farmgate Quantities		Deforestation	
	(1)	(2)	(3)	(4)	(5)	(6)
2006	-.0101	-.0393	-5713	-16702	.00109	-.000601
	(.0131)	(.0464)	(21104)	(16087)	(.00179)	(.00363)
2007	.1**	.0809	-10441	-15151	.000785	-.000855
	(.0419)	(.101)	(21576)	(18254)	(.00324)	(.00473)
2008	.0132	-.00153	-4971	-9985	.000966	-.000623
	(.0091)	(.0336)	(17431)	(14223)	(.00255)	(.00359)
2009	.0287	.0173	-1865	-12405	-.000966	-.00243
	(.0176)	(.0308)	(16175)	(13094)	(.00233)	(.0039)
2010	-.0217*	-.0284	-2238	-7256	-.00098	-.00221
	(.0108)	(.026)	(12145)	(9713)	(.00227)	(.00381)
2011	-.0205	-.0288	5318	1049	.00161	.000901
	(.0184)	(.0264)	(3604)	(4704)	(.00198)	(.00294)
2012	.0426***	.0356*	-1860	-4802	.00153	.00115
	(.012)	(.0192)	(9324)	(14684)	(.00235)	(.00347)
2013	.00322	.00599	-1099	-5496	.000797	.000345
	(.0103)	(.0173)	(4437)	(5460)	(.000997)	(.00137)
2015	.0518***	.0606***	2654	2142	.00118	.000865
	(.0095)	(.0121)	(5157)	-4640	(.003)	(.00599)
2016	-.00679	.00791	10545	11841	-.000887	-.00099
	(.0129)	(.021)	(6569)	(19210)	(.00105)	(.00235)
2017	-.00308	.00862	11951	20556	-.00361	-.00341
	(.0228)	(.0255)	(12538)	(12611)	(.00211)	(.0033)
2018	.0311	.0443*	6234	17401	-.00168	-.00142
	(.0286)	(.0213)	(11970)	(11901)	(.00224)	(.00389)
Municipality FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Meso Region FE		✓		✓		✓
Biome FE		✓		✓		✓
Municipality x Year		✓		✓		✓
Meso Region x Year		✓		✓		✓
Biome x Year		✓		✓		✓
Shipping Costs		✓		✓		✓
Storage Capacity		✓		✓		✓
Mean of Dep. Var.	1.6	1.6	101663	101663	.0223	.0223
Observations	1723	1723	1723	1723	1723	1723

Farmgate prices are measured as the natural logarithm of real 2006 farmer revenue divided by farmgate quantity. Farmgate quantities are measured in tonnes. Deforestation is measured as gross deforestation of forest or savanna divided by total remaining forest in the municipality. Standard errors are reported in brackets, clustered by meso-region and year. * p≤0.10, ** p≤0.05, *** p≤0.01.

Figure 4: Event-Study Results for Reduction in Competition Group

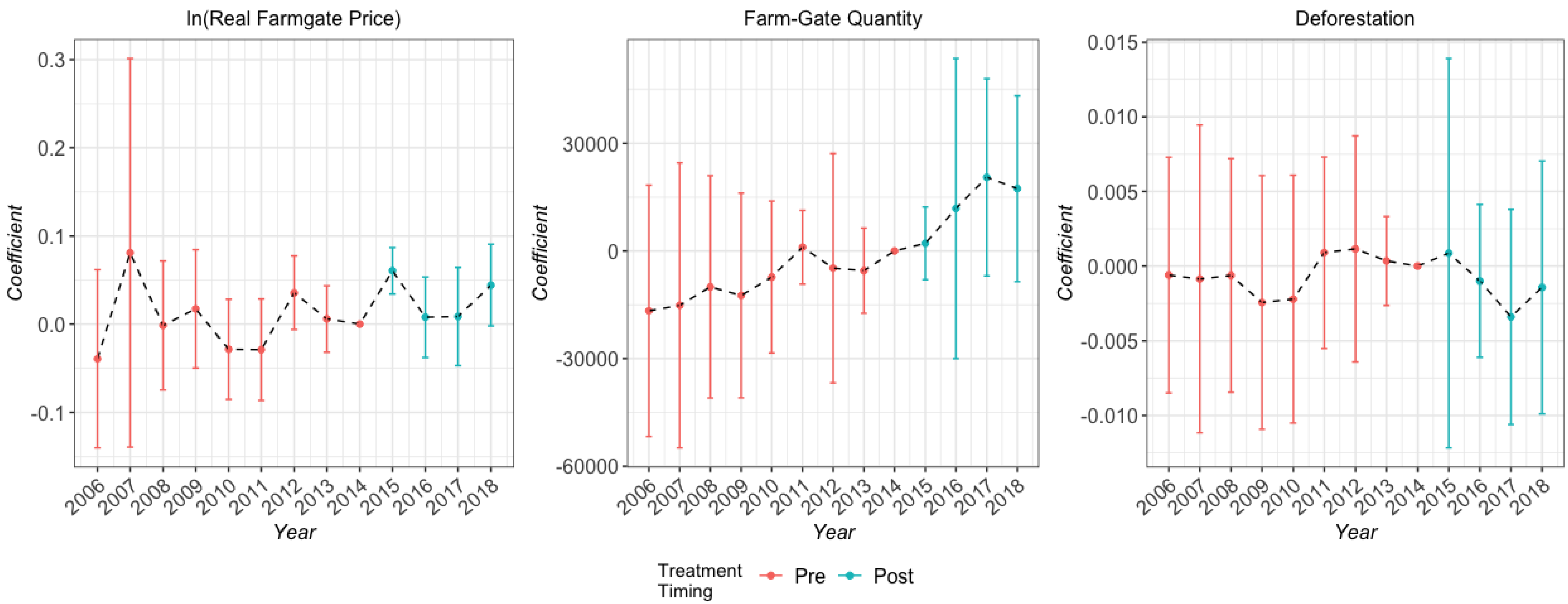


Table 6: Event-Study Results for No Change in Competition Group

	Farmgate Prices		Farmgate Quantities		Deforestation	
	(1)	(2)	(3)	(4)	(5)	(6)
2006	-.00609	-.00551	9928*	6475	-.00313***	-.00326
	(.00982)	(.0167)	(5188)	(9043)	(.000969)	(.0028)
2007	.00965	.011	8205	4193	-.00431***	-.00425*
	(.027)	(.0549)	(5701)	(8309)	(.000992)	(.00197)
2008	-.0155	-.0127	5932	4502	-.00124	-.00105
	(.0121)	(.0215)	(4820)	(9768)	(.000963)	(.00165)
2009	.0349***	.0372	4911	4152	-.00225*	-.00233
	(.00852)	(.0248)	(4919)	(5961)	(.00112)	(.00196)
2010	-.0467**	-.0432	2463	-1593	-.00256	-.00258
	(.0199)	(.0363)	(3709)	(5466)	(.00154)	(.00279)
2011	-.023	-.0199	2269	-997	-.00139**	-.00158
	(.0131)	(.0218)	(2081)	(5631)	(.000625)	(.00111)
2012	.0208	.0252	3583	7879	-.00167	-.00177
	(.0175)	(.0308)	(2707)	(5460)	(.00104)	(.00191)
2013	.00232	.00439	-2106	-4875	-.000193	-.000238
	(.0121)	(.0224)	(2401)	(3626)	(.00085)	(.00141)
2015	-.0115	-.014	2696	1711	-.000623	-.000422
	(.01)	(.0238)	(2109)	(2527)	(.0015)	(.00196)
2016	-.00603	-.000144	-141	369	-.000949	-.00127
	(.00543)	(.0188)	(2303)	(6220)	(.000713)	(.00102)
2017	-.012	-.0118	1860	7158	-.00406***	-.00379**
	(.00805)	(.0203)	(4772)	(7618)	(.000961)	(.00124)
2018	.0013	-.00131	7514	12688	-.00189	-.00129
	(.00914)	(.0196)	(5757)	(8747)	(.00106)	(.00129)
Municipality FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Meso Region FE		✓		✓		✓
Biome FE		✓		✓		✓
Municipality x Year		✓		✓		✓
Meso Region x Year		✓		✓		✓
Biome x Year		✓		✓		✓
Shipping Costs		✓		✓		✓
Storage Capacity		✓		✓		✓
Mean of Dep. Var.	1.6	1.6	64716	64716	.0198	.0198
Observations	2431	2431	2431	2431	2431	2431

Farmgate prices are measured as the natural logarithm of real 2006 farmer revenue divided by farmgate quantity. Farmgate quantities are measured in tonnes. Deforestation is measured as gross deforestation of forest or savanna divided by total remaining forest in the municipality. Standard errors are reported in brackets, clustered by meso-region and year. * p≤0.10, ** p≤0.05, *** p≤0.01.

Figure 5: Event-Study Results for No Change in Competition Group

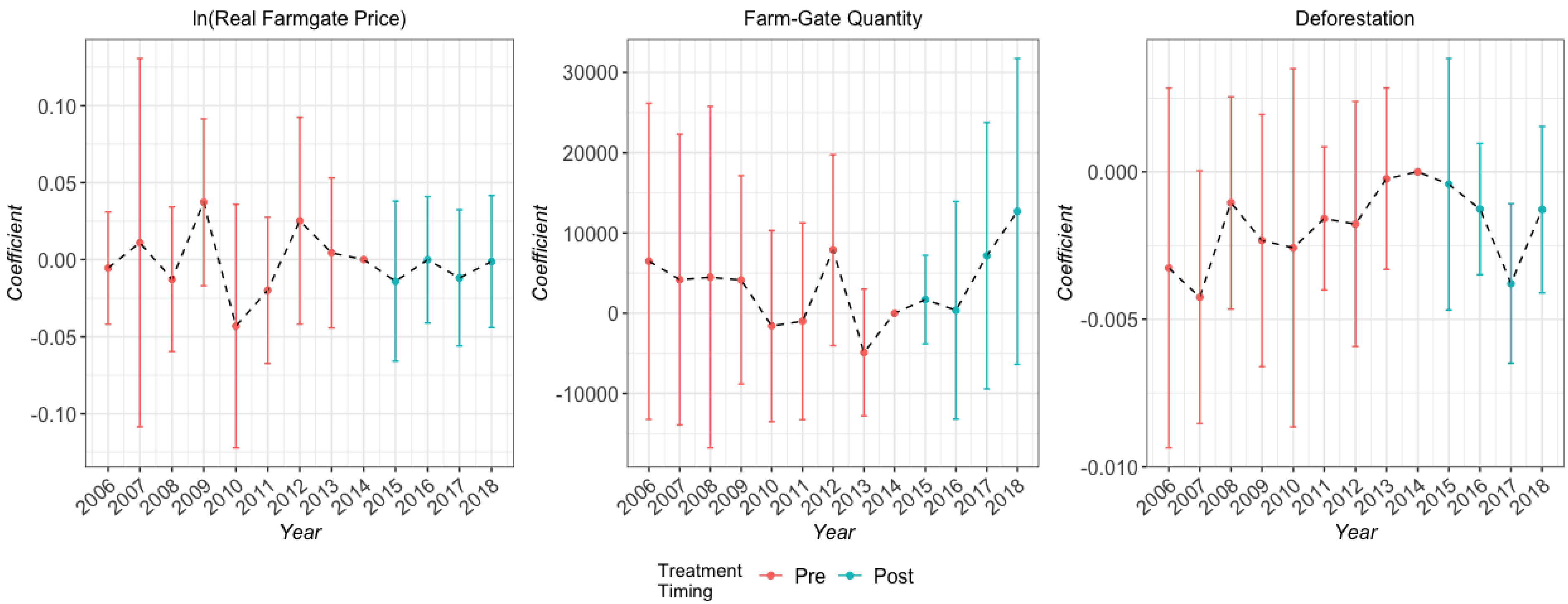
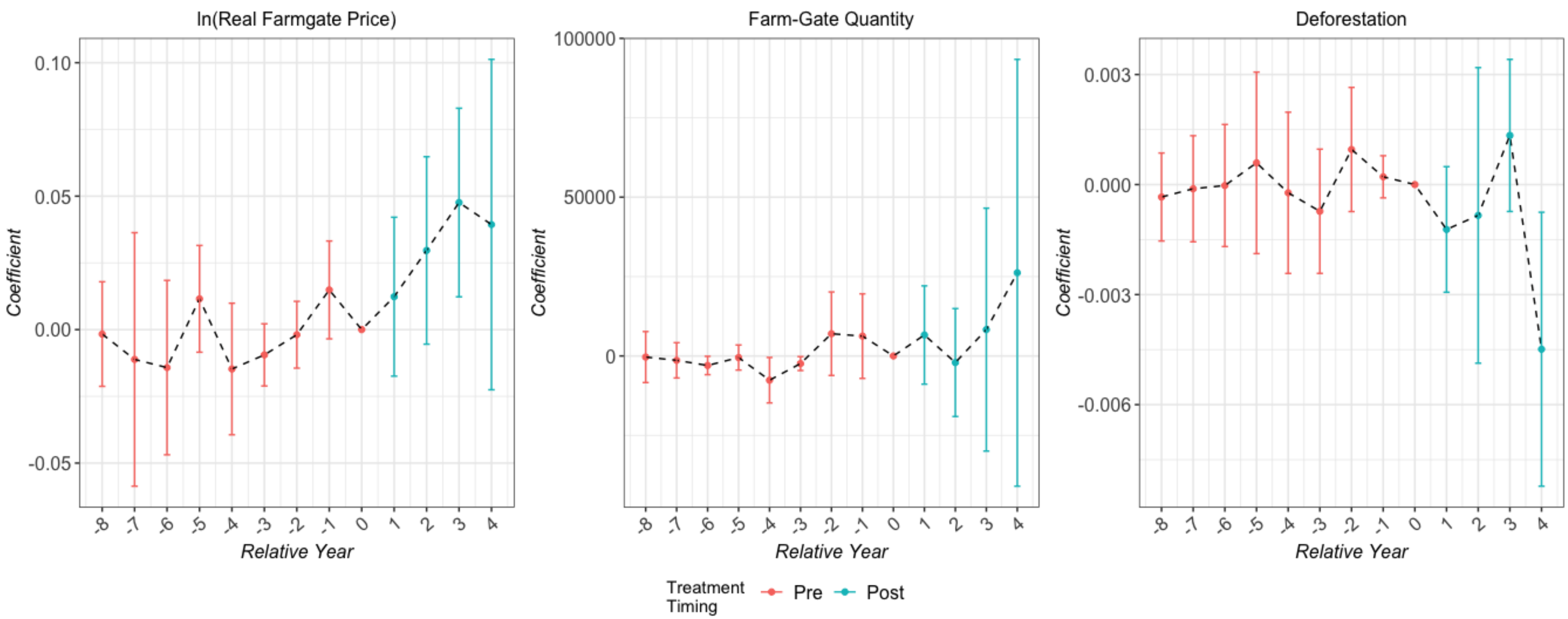


Table 7: Event-Study Results for Market Entry Group

	Farmgate Prices		Farmgate Quantities		Deforestation	
	(1)	(2)	(3)	(4)	(5)	(6)
-7	-0.013	-0.011	-2273	-1375	-0.0001	-0.00011
	(0.024)	(0.024)	(2756)	(2834)	(0.0007)	(0.00074)
-6	-0.015	-0.014	-4452***	-2998**	-0.0001	-0.00003
	(0.017)	(0.017)	(1600)	(1477)	(0.0008)	(0.00085)
-5	0.011	0.012	-796	-499	0.0005	0.00059
	(0.011)	(0.01)	(2299)	(2014)	(0.0013)	(0.00126)
-4	-0.014	-0.015	-8627**	-7628**	-0.0003	-0.00023
	(0.012)	(0.013)	(3821)	(3650)	(0.0012)	(0.00112)
-3	-0.009	-0.009	-1646*	-2396**	-0.0008	-0.00073
	(0.006)	(0.006)	(996)	(1116)	(0.0009)	(0.00086)
-2	-0.014	-0.015	-8627	-7628	-0.0003	-0.00023
	(0.012)	(0.013)	(3821)	(3650)	(0.0012)	(0.00112)
-1	0.015	0.015	8338	6246	0.0002	0.00021
	(0.011)	(0.009)	(9123)	(6791.575)	(0.0003)	(0.00029)
0	0	0	0	0	0	0
	0	0	0	0	(0)	(0)
1	0.006	0.012	9535	6599	-0.0024***	-0.00122
	(0.017)	(0.015)	(9906)	(7895)	(0.0004)	(0.00087)
2	0.031*	0.03*	3443	-2070	-0.0007	-0.00084
	(0.016)	(0.018)	(9928)	(8669)	(0.0018)	(0.00206)
3	0.057***	0.048***	25067	8292	0.0015	0.00134
	(0.016)	(0.018)	(23490)	(19512)	(0.0013)	(0.00106)
4	0.052**	0.039	44040	26187	-0.0043**	-0.00449**
	(0.023)	(0.032)	(38987)	(34284)	(0.0017)	(0.00191)
Municipality FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Meso Region FE		✓		✓		✓
Biome FE		✓		✓		✓
Municipality x Year		✓		✓		✓
Meso Region x Year		✓		✓		✓
Biome x Year		✓		✓		✓
Shipping Costs		✓		✓		✓
Storage Capacity		✓		✓		✓
Mean of Dep. Var.	1.78	1.78	86370	86370	0.0231	0.0231
Observations	2598	2598	2598	2598	2598	2598

Farmgate prices are measured as the natural logarithm of real 2006 farmer revenue divided by farmgate quantity. Farmgate quantities are measured in tonnes. Deforestation is measured as gross deforestation of forest or savanna divided by total remaining forest in the municipality. Standard errors are reported in brackets, clustered by meso-region and year. * p≤0.10, ** p≤0.05, *** p≤0.01.

Figure 6: Event-Study Results for Expansion Group



References

Aguirregabiria, V. and Magesan, A. (2020). Identification and estimation of dynamic games when players' beliefs are not in equilibrium. *The Review of Economic Studies*, 87(2):582–625.

Araujo, R., Costa, F., and Sant'Anna, M. (2020). Efficient forestation in the brazilian amazon: Evidence from a dynamic model. *Working Paper*.

Athey, S. and Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, 226(1):62–79.

Borusyak, K. and Jaravel, X. (2017). Revisiting event study designs. *Available at SSRN 2826228*.

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

DePaula, G. and Fortes, A. (2019). The adaptation of soy-corn double-cropping to the brazilian savanna. Technical report.

Dobson, A. P., Pimm, S. L., Hannah, L., Kaufman, L., Ahumada, J. A., Ando, A. W., Bernstein, A., Busch, J., Daszak, P., and Engelmann, J. (2020). Ecology and economics for pandemic prevention. *Science*, 369(6502):379–381.

Domínguez-Iino, T. (2022). Efficiency and redistribution in environmental policy:an equilibrium analysis of agricultural supply chains.

Fowlie, M., Reguant, M., and Ryan, S. P. (2016). Market-based emissions regulation and industry dynamics. *Journal of Political Economy*, 124(1):249–302.

Fowlie, M. L. (2009). Incomplete environmental regulation, imperfect competition, and emissions leakage. *American Economic Journal: Economic Policy*, 1(2):72–112.

Freyaldenhoven, S., Hansen, C., and Shapiro, J. M. (2019). Pre-event trends in the panel event-study design. *American Economic Review*, 109(9):3307–38.

Gardner, J. (2021). Did2s: Two-stage difference-in-differences. *arXiv preprint:2109.05913*.

Godar, J., Persson, U. M., Tizado, E. J., and Meyfroidt, P. (2015). Towards more accurate and policy relevant footprint analyses: tracing fine-scale socio-environmental impacts of production to consumption. *Ecological Economics*, 112:25–35.

Goldsmith, P. D. (2008). Economics of soybean production, marketing, and utilization. pages 117–150. Elsevier.

Gollnow, F., Hissa, L. d. B. V., Rufin, P., and Lakes, T. (2018). Property-level direct and indirect deforestation for soybean production in the amazon region of mato grosso, brazil. *Land use policy*, 78:377–385.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Harding, T., Herzberg, J., and Kuralbayeva, K. (2021). Commodity prices and robust environmental regulation: Evidence from deforestation in brazil. *Journal of Environmental Economics and Management*, 108:102452.

Hayenga, M. and Wisner, R. (2000). Cargill's acquisition of continental grain's grain merchandising business. *Applied Economic Perspectives and Policy*, 22(1):252–266.

Informa Economics, I. (2016). Farm to market a soybean's journey from field to consumer.

IPBES (2019). Global assessment report on biodiversity and ecosystem services of the intergovernmental science-policy platform on biodiversity and ecosystem services.

IPCC (2013). Climate change 2013: The physical science basis.

Kalouptsidi, M. (2018). Detection and impact of industrial subsidies: The case of chinese shipbuilding. *The Review of Economic Studies*, 85(2):1111–1158.

Long, N. V. and Soubeyran, A. (2005). Selective penalization of polluters: an inf-convolution approach. *Economic Theory*, 25(2):421–454.

Oates, W. E. and Strassmann, D. L. (1984). Effluent fees and market structure. *Journal of Public Economics*, 24(1):29–46.

Pera, T. and Caixeta-Filho, J. (2017). Information system for grain storage in brazil. *Food and Nutrition Journal*.

Pfaff, A. S. P. (1999). What drives deforestation in the brazilian amazon?: Evidence from satellite and socioeconomic data. *Journal of environmental economics and management*, 37(1):26–43.

Pigato, M. (2015). China and africa: expanding economic ties in an evolving global context./pigato m., tang w.

Rajão, R., Soares-Filho, B., Nunes, F., Börner, J., Machado, L., Assis, D., Oliveira, A., Pinto, L., Ribeiro, V., and Rausch, L. (2020). The rotten apples of brazil's agribusiness. *Science*, 369(6501):246–248.

Ryan, S. P. (2012). The costs of environmental regulation in a concentrated industry. *Econometrica*, 80(3):1019–1061.

Salin, D. L. and Somwaru, A. (2018). The impact of infrastructure and transportation costs on us soybean market share: An updated analysis from 1992-2017.

Sengul, M. and Dimitriadis, S. (2015). Multimarket competition. *Journal of Organization Design*, 4(3):18–30.

Sexton, R. J. and Xia, T. (2018). Increasing concentration in the agricultural supply chain: implications for market power and sector performance. *Annual Review of Resource Economics*, 10:229–251.

Simpson, R. D. (1995). Optimal pollution taxation in a cournot duopoly. *Environmental and Resource Economics*, 6(4):359–369.

Souza-Rodrigues, E. (2019). Deforestation in the amazon: A unified framework for estimation and policy analysis. *The Review of Economic Studies*, 86(6):2713–2744.

Stern, N. (2008). The economics of climate change. *American Economic Review*, 98(2):1–37.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Trase (2018). Supply chain mapping in trase: Summary of data and methods. Technical report.

U.S. Department of Justice, A. D. (2000). Final judgement: U.s. v. cargill, inc. and continental grain co. Technical report.

USDA (2020). The impact of brazil's infrastructure and transportation costs on u.s. soybean market share: An updated analysis from 1992-2019.

USDA (2022). Oilseeds: World markets and trade.

A. Trase Data Discussion

The Trase data on purchases by grain trading firms within municipalities is critical for the analysis of this paper, and any detailed economic or environmental analysis of the soy industry. Yet as the data is not directly observed in administrative data, there are several notable issues.

A.1. How Trase Data is Produced

First, to understand the issues, it is important to discuss what Trase does to produce this dataset. Trase maps supply chains for the Brazilian soy industry using an approach called Spatially Explicit Information on Production to Consumption Systems (SEI-PCS) based on Godar et al., 2015. This approach models the structure of the supply chain using exports data where a municipal tax location is recorded for all individual customs records. This allows Trase to link soy flows to export markets by trading companies to purchases at subnational logistic hubs. Logistic hubs, located in specific municipalities, are identified as supply chain nodes where a given commodity is produced, stored, or processed, before being transported to an export facility.

The main goal of Trase's model is to determine the likely municipality in which soy was produced. The supply chain is modelled for raw and processed products that can be directly linked to the original farmed commodity. Highly transformed subproducts are excluded due to the difficulties tracing complex transformations. Processed products are converted to soy equivalent values using conversion factors that account for crushing ratios and processing waste volumes. Once converted, all exports are mapped back to the municipal locations of taxation which can correspond to farms, silos, crushing facilities or wholesale retailing. The location of municipal taxation is then defined as a logistics hub. These logistic hubs are then linked with the municipality where the soy was most likely produced, and ultimately purchased from using a linear programming approach.

The linear programming approach used by Trase minimizes the distance covered from production to logistic hub, optimizing the allocation of the traded volumes per logistic hub as well as the spatially explicit domestic demand calculated from official sources. This optimization is constrained by the spatial distribution of assets linked to the trading company. The algorithm ensures that total production matches total exports and domestic consumption. Further, the relationship between production areas and logistic hubs is modeled to reflect the available information on the operations of each trader in the region if such data exists. Trase then validates the results of the SEI-PCS products with company information, where possible. Initial validations by some leading soy traders confirm that this approach to the allocation of production municipalities is robust and accurate (Trase, 2018).

The difficulty in this mapping is that individual farms and logistic hubs are connected by a complex web of relationships, including production, harvesting and storage that occur before the trader purchases a given amount of soy. The optimization model distributes available supply, total known production per municipality, to known demand. Demand is the sum of both export and domestic demand, as defined by the known exported volume from each logistic hub and processing facility. The distribution of domestic demand is estimated by mapping the location and capacity of individual processing facilities in the country. While some traders have contracts with specific farmers, the large size of the spot market relative to total traded volumes, the fact that soy volumes sourced from different farms are typically bulked in intermediary storage facilities, and the fact that fluctuations in prices drives a high level of inter-annual variability in farmer-trader relationships, all make the mapping of specific volumes from individual production regions to individual logistic hubs a difficult and costly exercise, even for traders directly involved in these operations (Trase, 2018).

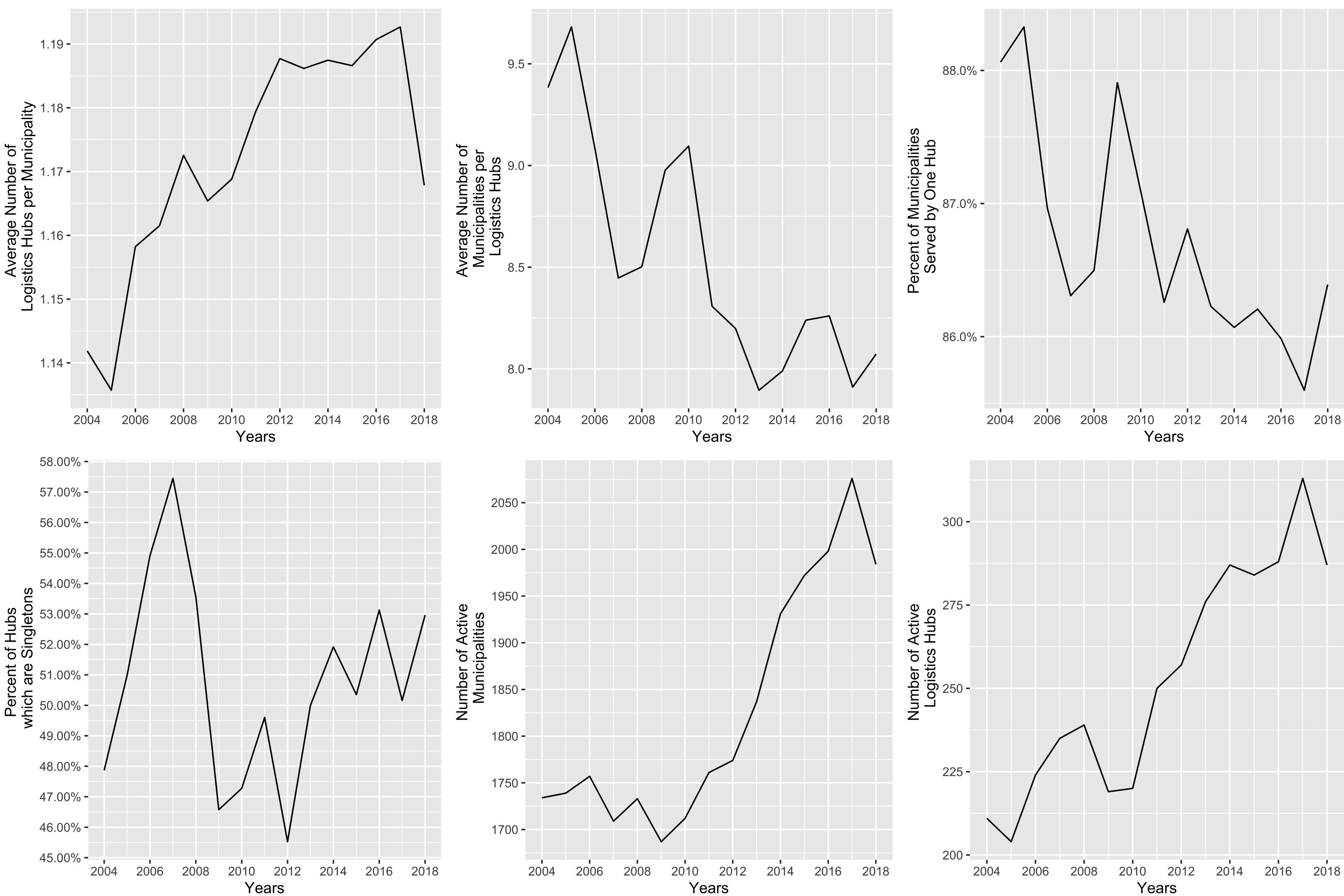
A.2. Logistics Hubs & Municipality Network

The connections from logistics hub to municipality in the data is a central feature of the data, so understanding and visualizing this network is helpful for better understanding the structure of the market. Figure A.1 shows that, both the number of active municipalities and logistics hubs in a given year has been increasing since 2010. In 2018, 2000 municipalities provided soy to roughly 300 logistics hubs. Further, over 85% of municipalities are served by only one logistics hub. On average, this leads to about 1.15 logistics hubs serving a typical municipality in each year.[29] On the other hand, the average logistics hub is found to be serving 8-9 municipalities within a year.[30] Finally, about 50% of the observed logistics hubs would be considered singletons in the network in a given year – that is, the municipality the logistics hub is located in is the only municipality that hub purchases soy from.

[29]The max logistics hubs serving a municipality in the data in any year is 6; in most years the max is 4.

[30]Several logistics hubs, such as Sao Paolo where several company headquarters are, or Santos, the largest port for soy exports, serve well over 200 municipalities in a given year.

Figure A.1: Logistics Hubs & Municipality Connections



A.3. Issues with Trase Supply Chain Mapping

Due to the difficulties in this mapping, not all the shipments registered in the original trade data used by Trase can be linked conclusively to a logistic hub in or near a soy-producing area. In a given year these unmapped shipments account for between 5% and 18% of total exported volumes. The percentage of exports not mapped to municipalities is highest in 2006-2008, and 2018, and the average not mapped during the 2006-2018 sample period used in the empirical analysis section is 7.5%. These unmapped shipments mostly correspond to shipments associated with trading companies that may not physically handle the commodity or have operations in far from soy production areas (Trase, 2020).

For the major exporters of Brazilian soy, ADM, Bunge, Cargil, Louis Dreyfus, and COFCO, the amount of exports not traced to a given municipality also varies over time. This missing competitor data complicates any analysis of the soy industry as one cannot understand the full competitive structure without it. However, as shown below, this missing data is concentrated in a small number of municipalities. Thus, the approach this paper takes is to only use municipalities where the mapping process successfully accounts for expected purchases from that location. In this paper, this ensures that I correctly identify the presence of COFCO within a given municipality. If municipalities with a large percent of missing purchaser data were utilized, I would be unable to delineate these municipalities into the control or treatment group accurately.

A.4. Defining a Suitable Analysis Municipality

In this paper, I must isolate the municipalities where COFCO and its predecessors, Nidera and Noble, were active. This requires I exclude all municipalities where it is unclear if these companies were present. Comparing mapped trader purchases in Trase to the IBGE's production value for each municipality shows that missing flows are not uniform across regions but are concentrated in specific locations. To classify if a municipality is missing an unexpectedly large volume of measured soy production I allow for expected discrepancies between export values and production due to shipping losses and storage shrinkage.

I considered production of a given municipality to be fully accounted for by Trase if the following conditions are met. First, I define expected Trase flows from a municipality-year to be at least 75% of measured production.[31] Next a municipality is defined to be missing sufficient purchase data if Trase volumes are below expected purchases from that

[31] The overall results are not very sensitive to increasing this value to $\geq$90%. The main issues stems from municipalities missing virtually all their production in Trase's purchaser data.

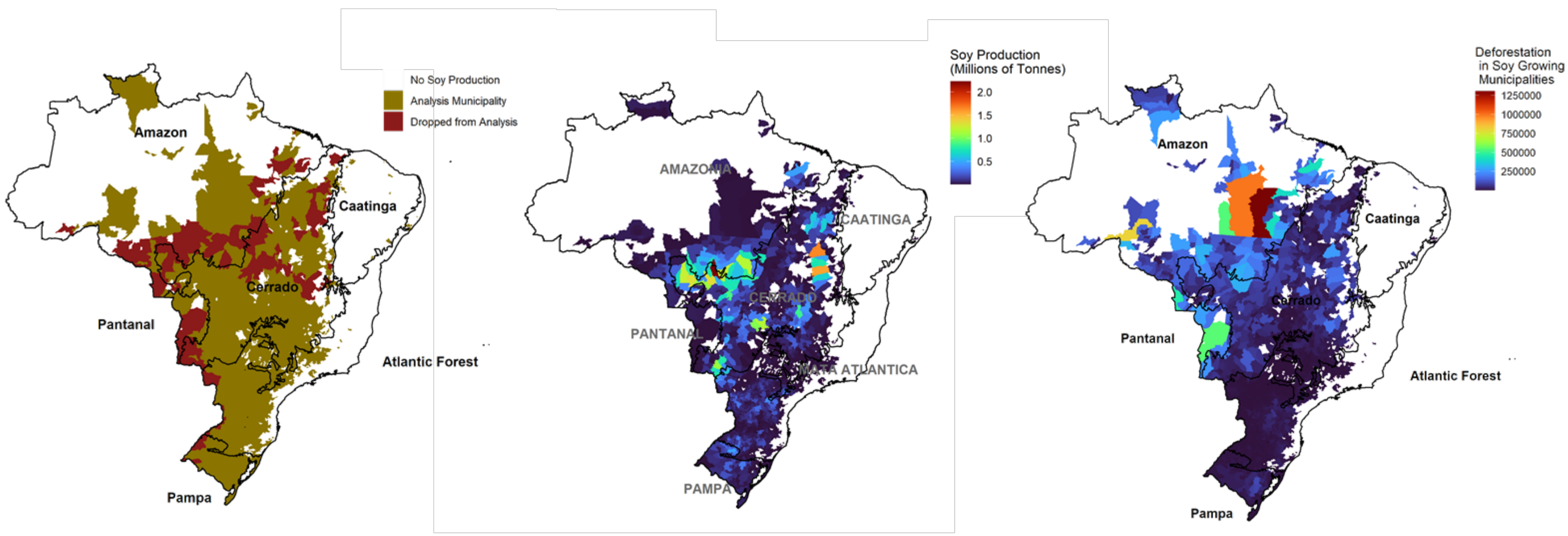
municipality. Out of the 2,496 municipalities that ever produce soy in the sample period, based on this measure, 2,052 are found to have no missing production volumes.[32] This implies 444 municipalities are missing significant production volume in at least one year in the sample period, yet not all must be excluded.

I consider two additional groups of municipalities for inclusion in the analysis. First is the set of municipalities where Trase data is only missing in a small number of years in the pre-treatment period and Nidera/Noble are never found to be present in any other years. Using a maximum of 3 missing pre-treatment years allows for the inclusion of an additional 64 municipalities to the analysis. Finally, due to the increase in missing Trase flows in 2018, I also allow for the inclusion of municipalities that are only missing Trase data in 2018. While these municipalities could be subject to COFCO entry in this year, it is unlikely as the majority have been historically linked to a singular firm. This leads to the inclusion of another 123 municipalities. The remaining 252 municipalities are excluded from the analysis.

Figure A.2 displays the geographic location of these municipalities along production volume in 2018, and total deforestation over the sample period. This shows that municipalities are more likely to have missing data in the arc of deforestation and other high deforestation regions.[33] Further, several municipalities with high volumes of soy production are excluded based on these definitions. The causes of this seemingly spatially correlated missing data should be studied further.

[32] If the threshold is increased to 90%, then 2,044 are found to have no missing production volumes.
[33] Although the municipalities with the highest rate of deforestation don't seem to have missing data.

Figure A.2: Analysis Municipalities, Soy Production, and Deforestation



B. Heterogeneous Effects for Reduction Group

B.1. Empirical Strategy to Study Heterogeneity

For the reduction in competition treatment group, underlying heterogeneity may be driving the counterintuitive result. To investigate this, the treatment group is stratified into smaller sub-groups, $g \in G$ to test if the main results can be explained by underlying heterogeneity. This leads to an extension of the main regression model to the following:

$$y_{itg} = \sum_{t \neq 2014} \sum_{g \in G} \delta_{tg} D_i 1[Year = t] 1[i \in g] + \beta' X_{itg} + \alpha_i + \alpha_t + \epsilon_{it}$$

In this specification, the estimated treatment effects, δ_{tg}, now vary for each time period and sub-group, g. Using this model, I investigate heterogeneity across two dimensions. First, I investigate potential misspecification of the treatment group driven by the selection rule. Since I select markets in this treatment group based on competition observed only in 2014 this may not be identifying markets with a robust history of competition. To investigate this issue, I stratify the reduction group into two subgroups: one with all markets where Nidera and Noble only competed in 2014, and one with markets where they have a longer history of competition. This allows me to estimate the effect of a change in market structure differentially depending on the length of the competitive history between the firms. Second, I investigate if the expected negative price effect occurs when the merger creates a monopsonist or duopsonist. To test this hypothesis, I break the treatment group into four sub-groups: markets where the number of competitors goes from 2 to 1, 3 to 2, 4 to 3, and many to many. This allows me to test if the effect differs in markets where a monopsonist or duopsonist is created.

B.2. Results

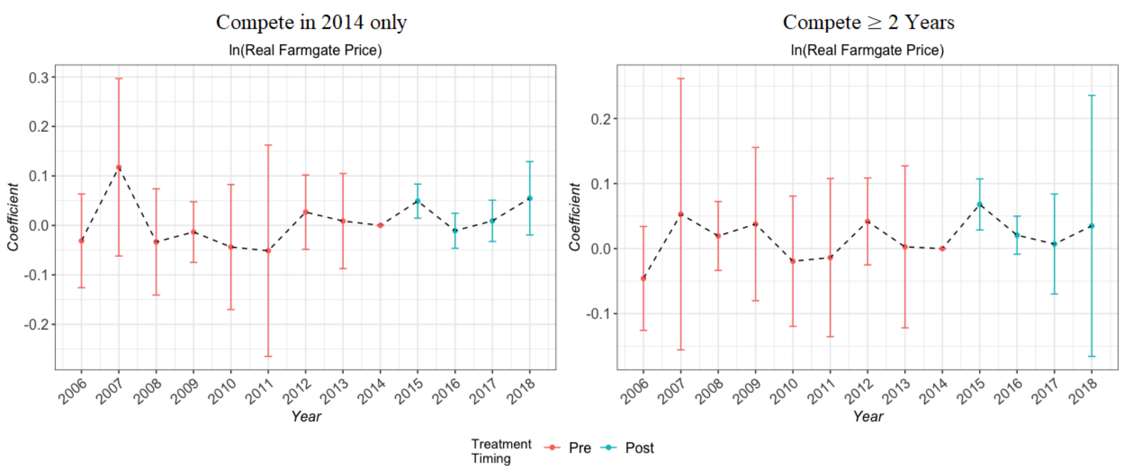
When stratifying by the history of competition between the firms, I find that markets with a longer history of competition experience a higher price increase post-merger. This suggests the main result is not driven by markets where the two firms were only competing in the pre-treatment year.

When looking at heterogeneity by pre-treatment competitor counts, I find a positive price increase in markets where a monopsonist is formed. Further, I find positive price effects where there were many competitors prior to entry.[34] This result does not add clarity to the

[34]It is important to note that there are violations of parallel trends for some groups, so these results are

counterintuitive price increase experienced in these markets.

Figure B.1: Reduction in Competition Group by Pre-Entry Competitive History



C. Alternative Control Group

To check the robustness of the results to the specification of the control group I develop a new definition for each control group. I define this new control group as all markets within the same meso-region, but not-contiguous to a treated market. This ensures that the two groups are similar on observables and unobservables, while minimizing the likelihood control units are treated due to market boundary misspecification. The model is then estimated for each treatment group and outcome variable for the model with all controls. The results presented below show a similar qualitative result, but now the price increase in the reduction group is estimated to be only 4% in the first year. All other results are similar to the main results section.

to be interpreted with caution.

Figure B.2: Reduction in Competition Group by Competitor Counts

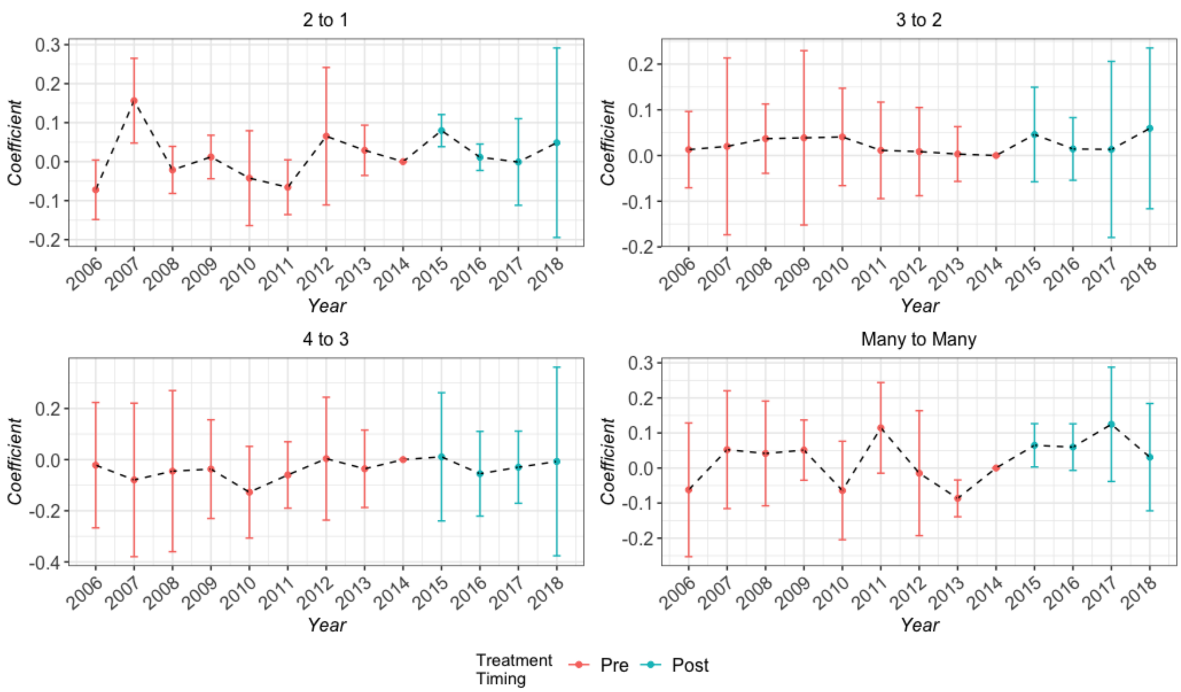
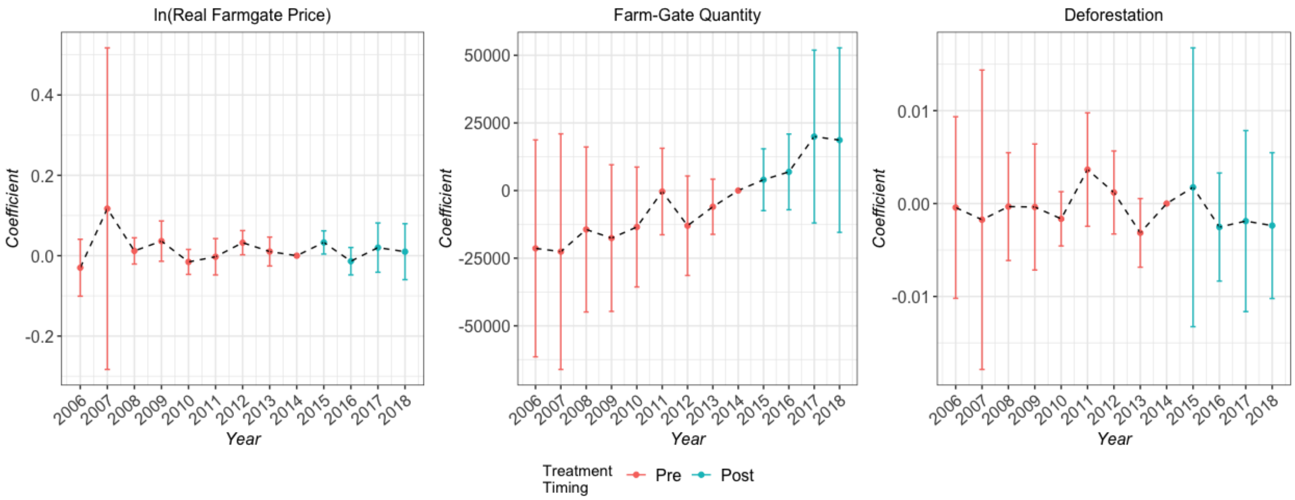
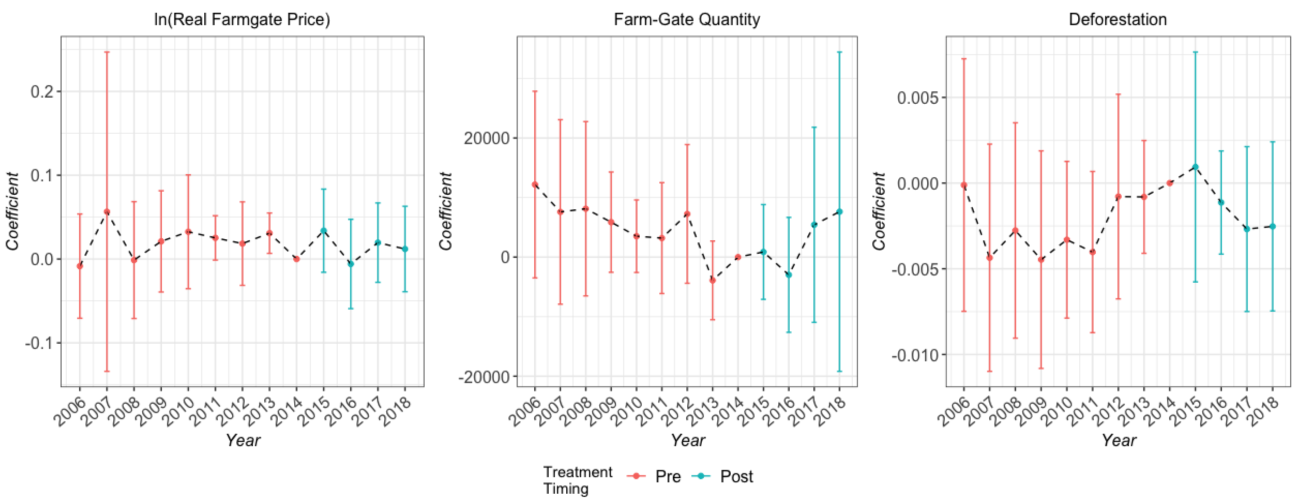


Figure C.1: Alternative Control Group Results

(a) Reduction in Competition Group



(b) No Change in Competition Group



(c) Market Entry Group