NHH

A Brazilian Soy Story - How International Soy Demand Affects Deforestation and Agricultural Land Use

Master Thesis at the Norwegian School of Economics, under supervision of
Associate Professor Harding, in the Economic Analysis major. *

Halvard Sandvik Jansen

June 2018

Abstract

In this paper, I use municipal soy trade data covering the years 2010 to 2015 to investigate the export market for Brazilian soy and what the expansion of soy exports leads to in terms of land use. The soy data were acquired from the Trase database and offer an unprecedented opportunity to map the international demand for soy to the municipal production and land use in Brazil. To the best of my knowledge, no equally detailed agricultural trade flow data has been available for research studies before this, lending originality to this study. For the econometric analysis, I use a fixed effects instrumental variable approach, with trade-weighted world income as an instrument for soy demand, to estimate the effect of soy demand on agricultural land use and deforestation. Unsurprisingly, I find a strong positive link between the Brazilian soy export market and the land use of exporting soy farms. This expansion of land has necessarily replaced other forms of land use. This paper is primarily an investigation of what alternative land uses have been restricted as a consequence of soy exports increasing. The main finding is that there is a significant negative elasticity between the land use of non-soy crops and the international soy demand. This implies that a significant share of the land use expansion of soy happens at the expense of other agricultural land use. However, I find no conclusive evidence that deforestation has been hastened by the increasing international demand for soy. This non-finding can be caused by either lacking power in statistical tests, by the effect of soy expansion on deforestation only being indirect due to displacement of other crops which again replace forests, or by Brazilian policies restricting the expansion of soy farming into forested territories being successful in curbing the negative externalities of soy farming. I also discuss the dominant role of China in the importing market, with a short analysis of what a trade war between the US and China would entail for Brazil.

Contents

Appendices **72**

1 Introduction

This thesis is an investigation of how the growth of Brazilian soy exports, with evidence from 2010 to 2015, leads to changes in land area devoted to different purposes in Brazilian municipalities. As far as I know, this study is the first to use municipal trade flow data to model international soy demand and the effects that changes in this demand has on land use and deforestation in Brazil. The remarkable detail level of the recent Trase (2018) data on municipal trade is what makes this possible.

The topic is of particular research interest due to several reasons. Firstly, tropical deforestation, or the removal of tropical forests, is a possible effect of the soy industry expanding recklessly and a definite cause of carbon emissions (Andersen *et al.*, 2002). Secondly, the Brazilian soy industry is vital to the Brazilian economy, and the negative externalities of its expansion, both in terms of how other industries are impacted and in terms of carbon emissions, have large economic significance. Thirdly, the international soy market is of particular geopolitical interest due to the dominant role of China as an importer and of Brazil, the United States and Argentina as the main suppliers of soy (United States Department of Agriculture, 2018b). At the time of writing this thesis, China looks poised to implement significant tariffs on US soy in retaliation to President Donald Trump's recently announced tariffs on Chinese goods (Donnan and Hancock, 2018). This trade tension, looking likely to develop into a full-scale trade war, will have substantial effects on the demand for both American and Brazilian soy, making the topic of Brazilian soy exports highly relevant to current events in global diplomacy and trade.

There are several research questions of interest in this thesis. For the econometric analysis of the Brazilian soy industry and its land use effects, there are two main research questions:

> 1: *How do changes in international demand for soy, with evidence from 2010-2015, affect the usage of land devoted to export-oriented soy farming?*
>
> 2: *To what extent does increasing international soy demand, through the expansion of land use, result in deforestation and the displacement of other forms of agricultural land use in Brazilian municipalities?*

The first research question directly concerns the land expansion of the soy industry. The second question is about how this expansion, if significantly tied to soy demand, can be decomposed into the displacement of different alternative uses of land.

Additionally, I investigate the Asian demand for Brazilian soy and the potential consequences of the aforementioned trade war between the United States and China. In light of this, there are two additional research questions to answer:

3: *To what extent has growing Asian demand been a driving force for the expansion of the Brazilian soy industry?*

4: *What are likely consequences of a trade war between the United States and China if sizable tariffs on American soy products are implemented?*

The empirical approach taken in this paper is a combination of fixed effects estimation to account for systematic differences between municipalities and an instrumental variable approach to solve the simultaneity problem caused by the intrinsic link between supply and demand. These methods are both presented and discussed in-depth in the methodology section.

Also discussed in the methodology section is the choice of instrumental variable for soy demand. In this thesis, I use the trade-weighted Gross Domestic Product (GDP) of importing countries to instrument the international demand for Brazilian soy products, isolating demand shocks in the system of simultaneous equations. The effectiveness and appropriateness of this instrument is evaluated both with a partial correlation analysis, a comparison with existing literature, and with an econometric fixed effects model.

The soy data used in the thesis is sourced from the Trase database (2018), and other sources include the IBGE database (SIDRA, 2018) for the land use of different crops and the World Bank's World Development Indicators (2017) for the GDP statistics of different countries. More details are provided in the data section.

The results of this paper's analysis are mostly in line with prior expectations. Firstly, to answer the first research question, I find clear evidence to suggest that international soy demand causes the land use for exported soy to expand. The estimated coefficient of elasticity between the land use for exported soy and the demanded quantity of soy was 0.766, indicating a highly elastic relationship between the variables. It also implies a certain productivity or yield increase due to soy demand since land use would have to respond one-to-one to exported volume in order for there to be no such effect.

I find no evidence of a causal link between international soy demand and deforestation or between international soy demand and the land use for domestically consumed soy. There may be many explanations for these non-findings, first of all the possibility that the links

between soy demand and these variables are weaker than one *a priori* would imagine. Especially with regard to deforestation, this particular negative effect of soy expansion may have been reduced drastically the last decade due to significant regulatory efforts raising the cost of expanding into forests for soy producers.

I find a significant negative relationship between international soy demand and the land use of non-soy agricultural crops. This was the *a priori* expected results, and the estimated elasticity coefficient of -0.322 indicates that a large share of the displacement effect caused by soy farms expanding affects non-soy agricultural crops. A major share of the displacement is however still unaccounted for at the end of the analysis, which indirectly implicates pasture land as a likely type of land use that is negatively affected by soy expansion. The reasoning behind this is partly that a causal relationship between soy expansion and the displacement of pasture land is hinted at in the literature, for example by a recent study by (Boerema *et al.*, 2016). The other reason for suspecting a negative relationship between soy demand and pasture land is that all land expansion necessarily must be accounted for, and pasture land is a likely "missing link" through the sheer process of elimination.

At the end of the analysis section, I find evidence to suggest that growing Asian demand has been the main determinant for growing Brazilian exports and the expansion of land use for soy farms. I also qualitatively analyze potential effects of a US-China trade war in terms of how it could affect the demand for Brazilian soy and the subsequent effects on land use in Brazil. The main conclusion is that demand for Brazilian soy would increase drastically if China were to commit to a tariff on imports of American soy. Through the results from the econometric analyses of this thesis, it is clear that this would result in a major increase in the land use for exported soy in Brazil as well as the displacement of non-soy crops. The effects on deforestation and the land use of domestically consumed soy are harder to pinpoint, but it does seem that multiple signs point in the direction of deforestation being indirectly linked with soy demand through the displacement of pastures. This is discussed towards the end of the paper and in appendix A.2, and I conclude that the existence and extent of such an effect will be an important question for further studies to investigate more thoroughly.

2   Background

There exists a sizable body of literature both on deforestation and agricultural land use in Brazil. However, studies on deforestation have largely been focused on Amazonia's primary

forests exclusively and there are as of yet no land use studies using municipal trade data to the best of my knowledge. Studies on deforestation in the Amazon forests include Andersen *et al.* (2002), Tyukavina et al (2017), Moran (1993), Binswanger (1991), Saatchi et al (1997), Margulis (2004) and Skole et al (1994) amongst others. There are other examples of deforestation studies, such as Kastens *et al.* (2017), where biomes (or major ecosystems) other than Amazonia are considered as well. Kastens *et al.* (2017) specifically evaluated the effect of the "Soy Moratorium" on soybean and deforestation dynamics in the Mato Grosso state, a state that contains both the Amazon ecosystem as well as Cerrado and Pantanal. In this paper, when I consider all Brazilian municipalities that both have soy production data and deforestation data available from Trase (2018), I will also cover several biomes other than Amazonia.

The role of international trade with regard to deforestation has also been discussed in the literature. Boerema *et al.* (2016) investigate the effect of soybean trade in terms of land use displacement using macro data from 1961 to 2008, mainly focusing on European imports. They find that soy expansion largely replaced other agricultural crops and pastures, but that especially pastures were replacing forests, meaning that the soybean industry indirectly contributed to deforestation in the relevant period. Brander and Taylor (1998) discuss the pros and cons of trade for exporters of renewable resources, and the role of trade on deforestation is discussed in a theoretical framework by Copeland and Taylor (2009). Faria and Almeida (2016) investigate the effect of openness to trade on deforestation, also estimating the effect of crop and pasture expansion on deforestation in Amazonia. Harding *et al.* (2018) find that high international prices on agricultural commodities adds pressure to tropical forests, soy being one of central agricultural commodities since the study also focuses on deforestation in Amazonia. The global soy market, specifically from the viewpoint of the United States as a major exporter, is investigated by Bolling *et al.* (2001) who state that the market can be seen as an example of oligopolistic competition among exporting nations. This indicates that trade policies in the soy market can be interpreted through the classical theories on imperfect competition in international trade as proposed by Dixit (1984).

Land use and acreage is a parameter that is often looked into alongside deforestation. For example, Verburg *et al.* (2014), Tyukavina *et al.* (2017) and Saatchi *et al.* (1997) explicitly look into the land use of pastures and agricultural crops in Amazonia in addition to addressing the issue of deforestation. There is also a paper by Barr *et al.* (2011) investigating land use elasticities in Brazil and the United States with focus on aggregate implied land

use elasticities with respect to prices. Berry and Schlenker (2011), on the other hand, investigate crop-yield elasticities in their empirical paper focused on bio-fuels and indirect land use. They investigate the relationship between yield, output prices and land use, which is such that the effect of prices on land use expansion is lessened by yield (or the productivity of the land) often increasing when prices increase (Berry and Schlenker, 2011). Lambin and Meyfroidt (2011) write about global land use and the scarcity of land. They specifically mention the South American Cerrado as one of the largest areas in the world without primary forests where agriculture can expand freely, but they also point out that this area is rich in biodiversity and that rapid expansion of agriculture will have substantial and negative environmental effects (Lambin and Meyfroidt, 2011). Displacement of land use, which is the migration of activities from one place to another, in turn affecting the land use there, is also discussed in the Lambin and Meyfroidt paper.

2.1 Soy Exports and Land Use

Brazil is currently the world's leading exporter of soybeans, followed by the United States and Argentina (United States Department of Agriculture, 2018c). The country's farming regions host a multitude of field crops. Dominant among them are corn, soybeans, wheat, rice and cotton, and they compete with each other, livestock and other agricultural crops for land area (Schnepf *et al.*, 2001). Schnepf *et al.* (2001) describe the rise of export-oriented soy production as an outcome of macroeconomic conditions stabilizing in Brazil, national agricultural policies becoming more export-friendly in conjunction with trade liberalization gradually removing barriers to trade in the 1990s.

Today, China has become the undisputed number one importer of soybeans, importing more than the rest of the world combined (United States Department of Agriculture, 2018b). In the 2016-2017 marketing year (September-August), the share of Brazilian soybean exports imported by China reached 75% according to recent data from the United States Department of Agriculture (2018a). In light of this, it is difficult to overstate the importance of Chinese demand and China's growing economy on the workings of the export-oriented Brazilian soy industry.

Figure 1 presents the distribution of soy exports in 2015, with colors representing the different biomes as seen in the leftmost column. The data used in this presentation, as well as the graphical presentation itself, are from the Trase database (2018) website, which is the
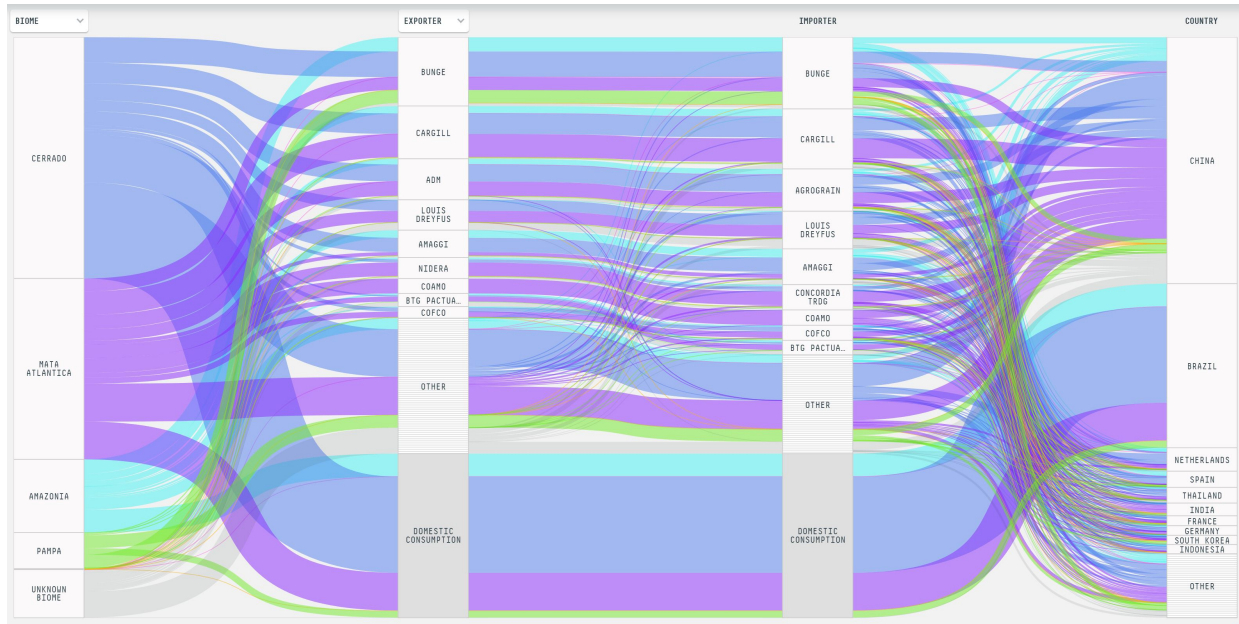
Figure 1: Graphical presentation by Trase (2018) of the distribution of Brazilian soy exports in 2015, colored by originating biome. Source: https://trase.earth/flows

main source of data in this paper. The middle columns show the companies that handle the exporting and importing of the soy, and the rightmost column presents the country that ends up importing or consuming the soy from each biome. The most obvious information one can retrieve from this graphical representation is that the Chinese market is the largest consumer of Brazilian soy, followed by the domestic market. Furthermore, as indicated by the left column, the bulk of soy production happens in the Cerrado biome in Brazil's center and in Mata Atlantica on the southeast coast. The Amazon biome, famous for its tropical rainforest, produces substantially less as the biome with the third highest production volume.

Figure 2 shows the growth of Brazilian export volume between 2010 and 2015, using the soy production data used in this paper's econometric analysis (Trase database, 2018). An increasing trend in soy export volume is evident, with an especially notable increase between 2014 and 2015. The figure also presents the development of Asian and European soy imports from Brazil, with Asian imports decomposed into Chinese and non-Chinese Asian importers. The imports of the rest of the world imports are also graphed, and it is evident that Brazilian soy exports are almost fully absorbed by the European and Asian market. However, the European and Asian soy imports from Brazil develop very differently between 2010-2015. The European import volumes are stable, even slightly declining, while

Brazilian Exports 2010–2015
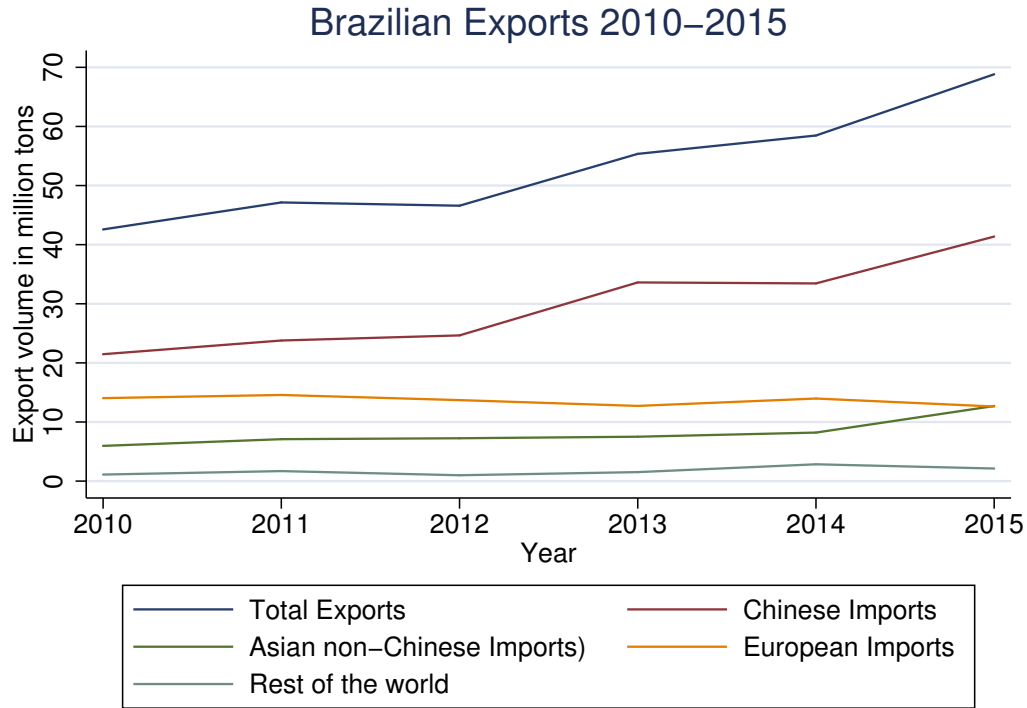
Figure 2: Brazilian export volume of soy between 2010-2015 in million tons, decomposed into importing regions. Source: https://trase.earth/data

Asian imports increased significantly over the period. Especially looking at the graphed development of Chinese imports in comparison to the development of total Brazilian exports, it seems that most of the variation in Brazilian export volumes are caused by China importing more. The non-Chinese Asian countries also see a marked increase over the 2010-2015 period, catching up to Europe in total soy imports as of 2015. The influence of the Asian market, with extra focus on China, will be discussed more thoroughly at the end of the analysis section.

A logical necessity of soy exports increasing is that it either needs to be accompanied by an expansion of export-dedicated soy farming area, increased productivity of existing soy agriculture or more likely a combination of these two outcomes. Additionally, an increase in the land use for exported soy must replace land use dedicated to alternative purposes (Schnepf *et al.*, 2001). This means that an increase in the export volume of soy must lead to either land expansion into forests, the displacement of land used for other agricultural purposes, increased productivity or the reallocation of soy farms from producing for the

11

domestic market to producing for the international market. Theoretically, it is also possible for the land expansion of soy farms to replace or displace the land use of urban territories, but this possibility is deemed unlikely and will not investigated in this paper. The primary purpose of the thesis, as far as the available data allow, is then to investigate and estimate these various possible effects of export volumes increasing in Brazilian municipalities.
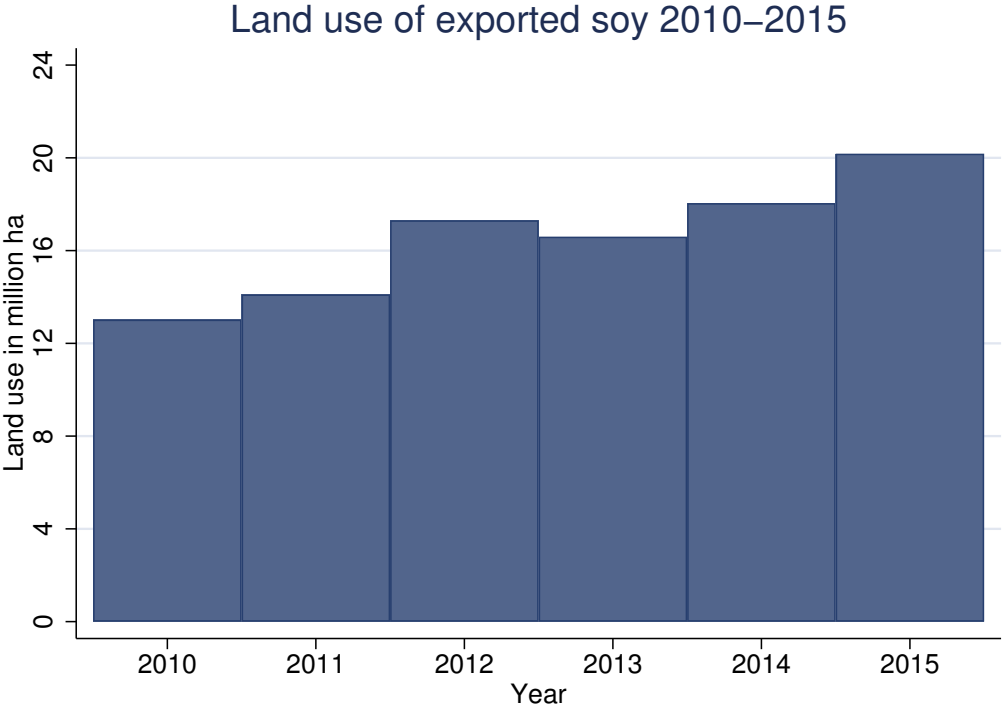
Figure 3: Agricultural land use of soy farmed for exports in 2010-2015 in million ha. Source: https://trase.earth/data

Figure 3 presents the land area used for producing exported soy from 2010-2015 (Trase database, 2018). Here, we see a similar development as soy exports, indicating that these variables are intrinsically connected as one would expect. Panel data analysis presented later in this paper supports this statement, finding an elasticity between soy export volume and related land use at about 0.766. This elasticity is quite high, indicating that most of the increase in export volume can be explained by an expansion of agricultural land dedicated to producing soy for the international market, while productivity or yield also has increased slightly. If productivity is equal to the production per area, then the growth of productivity must be equal to:

$$g_{productivity} = \frac{1 + g_{volume}}{1 + g_{acreage}} - 1 \tag{2.1.1}$$

The proof for this is simple, with indexes denoting period 1 and subsequent period 2 after growth, land use for exported soy being "acreage" and export volume being "volume":

$$Productivity_1 = \frac{volume_1}{acreage_1} \quad \& \quad Productivity_2 = \frac{volume_2}{acreage_2}$$

$$Productivity_2 = Productivity_1 \cdot (1 + g_{productivity}) = \frac{volume_1 \cdot (1 + g_{volume})}{acreage_1 \cdot (1 + g_{acreage})} \tag{2.1.2}$$

$$\downarrow$$

$$1 + g_{productivity} = \frac{1 + g_{volume}}{1 + g_{acreage}} \tag{2.1.3}$$

The implication of this is that the productivity gain is equal to zero if export-dedicated acreage grows one-to-one with export volume and one-to-one with export volume if export-dedicated acreage remains constant while export volume increases. It also implies that one quite easily can estimate the elasticity between production volume and productivity if one has an estimate for the elasticity between soy production and land use. An elasticity of 0.766 between the land use of exported soy and trade volume indicates an estimated productivity (or yield) increase of about 0.232% per 1% increase in export volume. This result can also be found using the logarithmic approximation that is used in the interpretation of econometric log-regressions, saying that $\log(1 + x) \approx x$ for small values of x. This transformation applied on equation 2.1.3 yields:

$$\log(1 + g_{productivity}) = \log \frac{1 + g_{volume}}{1 + g_{acreage}} = \log(1 + g_{volume}) - \log(1 + g_{acreage}) \tag{2.1.4}$$

$$\downarrow$$

$$g_{productivity} \approx g_{volume} - g_{acreage} \tag{2.1.5}$$

In other words, as seen in equation 2.1.5, the log-approximation of small growth indicates that the productivity growth can be approximated by taking the difference between the growth rate of trade volume and the land use growth of exported soy. In a log-log econometric model, this would be equivalent to estimating the difference between 1 and the elasticity coefficient between export volume and the land use of exported soy. This is because the

elasticity coefficients estimate the acreage response (in percent) to a 1% increase in export volume. Using an estimated elasticity coefficient of 0.766, the logarithmic approximation returns an estimated productivity growth of 0.234% per 1% growth in export volume.

2.2 Deforestation

Deforestation, or the clearing of forests, is an issue that consistently garners international attention (Andersen *et al.*, 2002). Few forests have been more controversial than the tropical forests of Brazil, and the deforestation of these forests has been an important topic for decades both because of the massive area the forests cover and because of their role as a source of biodiversity and as a major carbon sink (Andersen *et al.*, 2002). The expansion of soy farming into the Amazon biome has historically been regarded as one of the prime drivers of Brazilian deforestation, and this relationship is more ready to be investigated than ever as the trade data and data on deforestation becomes more and more detailed and complete. The link between agricultural output prices and deforestation has been analyzed multiple times in the literature. For example, Robalino and Herrera (2010) of the World Trade Organization (WTO) find that agricultural output prices affect deforestation positively, a finding supported by Verburg *et al.* (2014) who find that soy and beef prices are central determinants for deforestation in Brazil. Additionally, findings by Harding *et al.* (2018) suggest that increases in agricultural output prices in the 2002-2013 period may have accounted for about twenty-five thousand square kilometres of deforestation in Brazilian municipalities in Amazonia in the relevant time frame.

Deforestation is in itself a classical example of the tragedy of the commons (Harstad and Liski, 2013). One can describe it as a strategic interaction between multiple individual actors who choose their own extraction rate without fully taking negative externalities into account. The end result is that the sum of individual actors will extract considerably more than what is socially optimal (Harstad and Liski, 2013), leading to the tragedy of the commons where non-excludable resources are subject to over-exploitation and in extreme cases extinction. The tragedy of the commons can be seen as a problem caused by the resource-owning nation having limited control over property rights, a factor which Copeland and Taylor (2009) explores and analyzes through the lens of classical theories from resource economics on optimal resource extraction and steady-state resource levels for renewable resources.

Brazil, with the help of the international community, has implemented several measures

meant to slow down the rate of deforestation, some particularly targeted at the soy industry. In 2006, Brazil enacted the celebrated zero-deforestation "Soy Moratorium" pact where signees pledged to stop purchasing soybeans produced on deforested land in the Brazilian Amazon (Gibbs *et al.*, 2015). This has led to a considerable decrease in the deforestation rate of primary forest in the country over the past decade. However, the annual rate of soy expansion has been sizable in other regions and biomes that are not covered by the Moratorium. Gibbs *et al.* (2015) conclude that "further study is needed to assess potential leakage into the Cerrado and other countries and to quantify the avoided deforestation from [the Soy Moratorium]". Additionally, not all soy exporting companies are signatories in the moratorium, meaning that leakage via those companies may have hampered the effectiveness of the policy.

Deforestation can be also be reduced by campaigns aiming to negatively influence the demand for deforestation-causing projects such as the farming of soy or palm-oil. Campaigns such as this have become more and more prevalent in recent years, typically organized by non-governmental organizations and non-state actors and targeted at companies accused of actively harming the environment. For instance, the Rainforest Action Network (RAN) successfully organized a boycott against Mitsubishi to stop deforestation-causing practices in 1989-1998 (Rondinelli and Berry, 2000). Greenpeace is another notable actor, both organizing protests against Cargill, a soy producer accused of being responsible for massive soy-related deforestation in 2006 (Brannstrom *et al.*, 2012), and a PR-campaign against Nestlé in 2010 because of deforestation caused by palm oil production (Champoux *et al.*, 2012). Governments also have the power to directly influence markets through bans and trade restrictions. An example of a coordinated international effort is the EU parliament voting to ban the use of palm oil in biofuels in 2017, aiming to prevent EU's renewable transport targets from adversely contributing to deforestation.

Figure 4 is a graphical representation of deforestation from the Trase database (2018). The colors represent different biomes, where light blue represents Amazonia, dark blue represents the Cerrado and purple represents the Atlantic Forest (Mata Atlântica). The figure shows that most of the deforestation in Brazil is happening in the Cerrado biome, almost twice as much as in the Amazon biome. About 64% of total Brazilian deforestation in 2015 is registered in the Cerrado while about 36% is registered in the Amazon biome (Trase database, 2018). Although most deforestation is happening in the Cerrado, the amount of deforestation relative to soy activity (comparing with figure 1) is significantly higher in Ama-
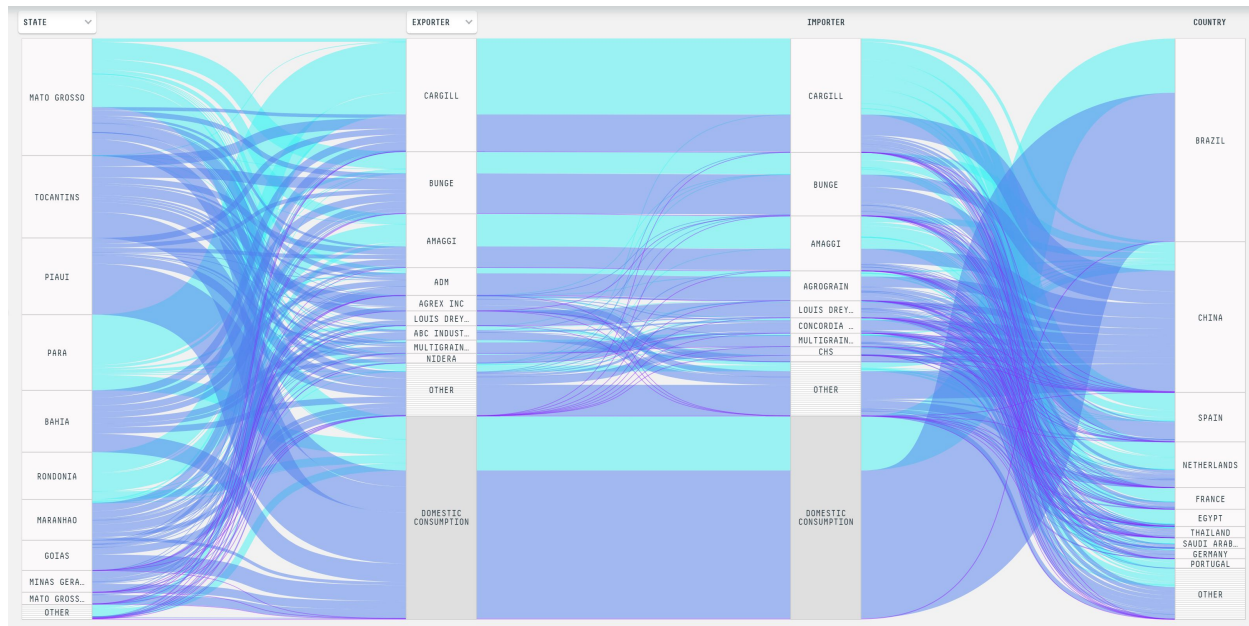
Figure 4: Deforestation in Brazilian states in 2015, colored by biome, and an estimate of which exporter-importer and consuming nation that is responsible for different parts of the deforestation. Source: https://trase.earth/flows

zonia. The leftmost part of the graphic ranks the Brazilian states according to hectares of deforestation in 2015, showing that Mato Grosso, the state focused on in the paper by Kastens *et al.* (2017), experienced the most deforestation in 2015. Mato Grosso is, as specified by Kastens *et al.* (2017), one of the few states that contains several biomes. This is presented in figure 4 as the Mato Grosso state experiencing both light blue (Amazonia biome) and dark blue (Cerrado) deforestation. The other soy-producing states in Amazonia that experience significant deforestation are Pará and Rondnia. Other than this, almost all deforestation happens in the Cerrado biome, notably in Tocantins, Piauí, Bahia, Maranhão, Goiás and the Cerrado part of Mato Grosso.

Additionally, the Trase supply chain graphic seems to suggest, without the method of calculation being entirely obvious, that domestic consumption and Chinese market imports are the main drivers for deforestation. Interestingly, the share of deforestation caused in the Cerrado biome seems to be higher for these countries relative to for example Spain, Portugal, France and Egypt who seem to be causing a higher share of deforestation in the Amazon biome according to the graphic.

Figure 5 shows how total deforestation has progressed in the Cerrado, Atlantic Forest and

16

Development in Total Deforestation 2010–2015
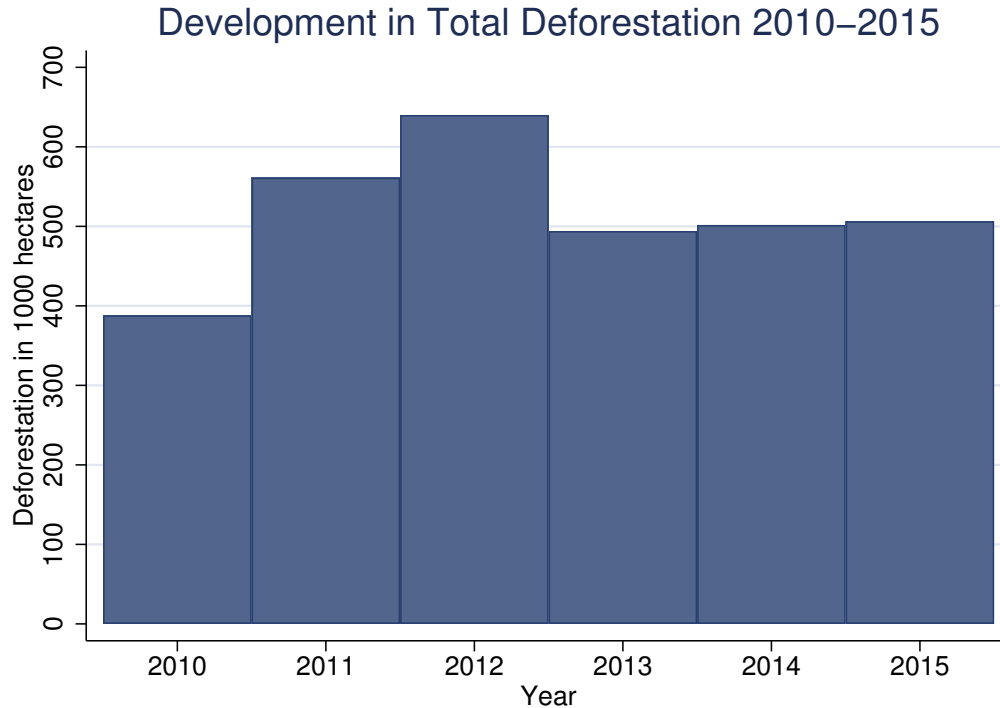
Figure 5: Total deforestation in Brazil in 2010-2015 in the Cerrado, Amazonia and Atlantic Forest biome based on data from the Trase database (2018). Source: https://trase.earth/data

Amazon biome in the relevant 2010-2015 period based on Trase's (2018) deforestation data. The aggregate deforestation seems to be increasing rapidly between 2010-2012 before flatting out between years 2013-2015 at a level slightly below the 2012 rate of deforestation. It is important to remember that deforestation is already a variable denoting change in land use, namely in forest cover. This means that a steady or even decreasing rate of deforestation means that forest cover is lost at a high rate if the starting rate of deforestation is high enough. However, one could assume that the rapidly increasing export volume of soy as seen from figure 2 could speed up the rate of deforestation. With the increase in export volume being especially high between years 2014-2015 and the rate of deforestation being stable, this connection is not immediately made upon looking at these figures. However, the econometric analysis of this paper is on the municipal level rather than the national, meaning that there are hundreds of municipal export-deforestation relationships to base an estimation on rather than just one.

17

2.3   Leakage

As stated in the introductory literature section, Lambin and Meyfroidt (2011) wrote a paper concerning the displacement of land use. One specific type of displacement is the leakage effect, which is displacement of land use caused by changes in policies. This can for example be trade policies in the form of tariffs or supply policies in the form of prohibiting certain activities in a certain area.

While there are multiple policies, such as the "Soy Moratorium", that may have contributed to the relationship between soy production and deforestation becoming less clear, the leakage effect may have caused the efficiency of the policy to be reduced (Gibbs *et al.*, 2015). The limited nature of almost all policies, either demand-side or supply-side, will often lead to leakage or adverse effects affecting the areas not targeted by a policy. One example is the earlier mentioned expansion of soy production in the Cerrado biome, which did not enjoy the same protection that Amazonia did after the "Soy Moratorium" limited the expansion of land use in that biome (Gibbs *et al.*, 2015). Another possible case is for a specific country to reduce its demand for soy in a municipality where deforestation is especially prominent. This could for example be motivated by campaigns against global warming being effective in influencing public opinion in the country. The problem is that the effect of this demand reduction is reduced significantly because the price of soy will fall in this municipality, and other countries not as concerned with deforestation will import more lower-priced soy from this municipality (Harstad and Mideksa, 2017). In such a way, it might be that the demand policies targeting individual countries with limited market power have limited effect on both aggregate soy production in Brazil as well as deforestation in vulnerable municipalities.

Another very recent example, directly relating to the global soy market, is the 2018 trade dispute between the United States and China. China announcing retaliatory tariffs on US soy (among other products) would reduce Chinese soy imports from the United States, at least in the short-run (Meredith, 2018). This would in turn likely lead to increased demand for Brazilian soy since Brazilian and American soybeans are virtually perfect substitutes (Sherman, 2018). Additionally, although China did not formally implement tariffs on soy in April 2018, Chinese traders cancelled almost all shipments from the US for that month (Cang *et al.*, 2018). According to a June 2018 article in the *Financial Times*, the current status of the situation is that China is poised to implement a 25% tariff on American soy on July 6th, 2018 (Hancock, 2018). This is the same date that US tariffs on a range of

Chinese goods are to be implemented, allegedly as punishment for long-standing intellectual property (IP) theft and as a way to reduce the American trade deficit with China (Donnan and Hancock, 2018).

Looking at the international market for soy, Brazil's exports responding quickly to lower demand of American soy would not be surprising. The international export market of soy is dominated by three countries, namely the United States, Brazil and Argentina (United States Department of Agriculture, 2018c). The United States Department of Agriculture (2018c) informs that Brazil is the leading exporter of soybeans while Argentina is the leading exporter of soybean meal, a processed soy product. The world imports are in turn dominated by one country, China, that in 2016/2017 absorbed 52.6% of world exports (United States Department of Agriculture, 2018b). This means that a Chinese trade policy or any shift in China's demand would have a particularly large effect on the integrated soy market. China needs to meet its domestic demand for soy, and buying more from its leading supplier Brazil is a natural option when tariffs and political mistrust makes American soy less attractive to Chinese companies. When data for soy production and exports from Brazil for 2018 are available in a few years, one could expect to find an effect of this dispute on the exports of Brazil, which again could affect deforestation and agricultural land use in Brazil.

3 Data

This paper uses longitudinal data, meaning panel data, with a time span of 6 years from 2010 to 2015. The cross-sectional dimension is quite large, with data on all municipalities in Brazil that export soy. The main data source is the newly available Trase data set on supply chains of agricultural commodities compiled by Stockholm Environment Institute (SEI) and Global Canopy from customs declarations, tax records, bills of lading, company self-declarations and national registries (Trase database, 2017). For example, the soy production and land use data stems from the Brazilian Bureau of Statistics (IGBE) and its annual agricultural survey called *Produção Agrícola Municipal* (PAM, or the Municipal survey on the Production of Agricultural goods). The Trase data set links municipality-level production (quantity, price and land use) of soy products in Brazil and Paraguay to information on the importing company, the exporting company and the destination country. Data collection is still in progress by SEI, and their plan is to expand the data set both in terms of years, countries and commodities on the municipality level. It is important to note that the Trase data

aggregates all soy products into one export figure between municipality and import country, with the denomination of "Soy equivalent tons". Processed products, like soy oil and soybean meal, are in the export figures lumped together with soybeans, although Trase calculates how many tons of soybean are input into the products. There is however a potential loss of information affecting both prices and financial flows since processed products might have higher prices per ton of soybean input compared to the shipments of soybeans themselves, and this data set will not take that into account.

I chose to restrict the data to municipalities that had non-zero total export volumes for all six years. This was both due to the possibility that zero values are missing because of incomplete reporting (that they are not true zeros) and because of my preference both to have balanced panel data regressions and to use logarithms to get estimates for elasticities. This proved not to be a very strict restriction, and the full dataset used in land use regressions ended up having soy data from 1570 out of the 5570 Brazilian municipalities.

The Trase data on soy trade volume (in tons), financial flows (in current US dollars) and land use (in hectares) are available on the individual trade flow level, with information about both the exporting and importing company, the port at which the trade is registered, which municipality the soy is produced in and which country is importing the shipment. For my analysis, the data were aggregated twice. This is because the trade-weighted GDP had to be calculated on the country-level and the econometric analysis is on the municipal level. The aggregation was done through the use of Stata version 15 (StataCorp, 2018), and means that some of the potential detail level that the Trase database (2018) offers is lost.

Because of how the Trase database (2018) has coupled all soy trade flows to the related land use in every municipality it stems from, I can separate the land use for soy dedicated to the domestic consumption and the land use for soy dedicated to soy exports. The split between land use dedicated to exports and land use dedicated to domestic consumption was indicated by the Trase database (2018) based on total agricultural soy land use, yields in different municipalities in conjunction with the exported volume traced back to each municipality. The main motivation for doing this is that I can separately assess the effect of changes in international soy demand on the land use of domestically consumed soy and on the land use of exported soy, as these are expected to respond differently to international demand.

The chosen GDP measure was real 2005 US dollars, and the data were gathered from the World Development Indicator database of the World Bank (World Bank, 2017). Most soy-

importing countries were represented in this database for 2010-2015 except Taiwan because it is excluded for political reasons and Venezuela because 2015 data were missing. Countries with very low amounts of trade flows, less than 50 registered trades over the six year period, were also dropped from the analysis because of their limited relevance and the added difficulty of getting proper GDP data on the smaller nations and island groups of the world. This should not have impacted the results of this paper significantly.

The weights for trade-weighted GDP were constructed based on the Trase data (2018) and represent, for every municipality, the share of exports that each country imports in the initial year.

The data on deforestation in Brazilian municipalities were also supplied by Trase on request, and stems from three separate sources: PRODES for the Amazon biome, LAPIG for the Cerrado biome and SOS for Mata Atlantica (Trase database, 2017). These abbreviations represent different deforestation projects focused on the documentation of deforestation in the relevant biomes. The fact that the deforestation data ultimately stem from different sources might lead to some systematic measurement errors if the method for estimating deforestation is slightly different for the different projects. The alternative, however, is restricting the data set significantly as there does not seem to be any single source that covers the whole of Brazil from 2010 to 2015 as thoroughly as the combination of sources that Trase provides. By using the Trase data, I have deforestation data from 203 municipalities over 6 years, totalling 1218 observations. The three states in this final dataset with the most municipalities recorded having deforestation are Mato Grosso, Tocantins and Goías. These states are according to Figure 4 the states with respectively the highest, second highest and eighth highest aggregate deforestation in 2015.

Data on the land use of other agricultural crops were gathered from SIDRA. SIDRA is the Brazilian Bureau of Statistics' (IBGE) database for statistical tables, meaning that the data on soy land use and production and the data on other crops ultimately come from the same source, the annual agricultural survey PAM (SIDRA, 2018). The non-soy agricultural data cover almost all of the 1570 soy-producing municipalities as only a few soy-producing municipalities do not have supplementary crops registered in the PAM data for all 6 years.

The data used in the supplementary analysis of cattle farming in the appendix were also gathered from SIDRA (2018). The specific database is called the "Pesquisa Pecuaria Municipal" (PPM), which is the IBGE section for research on the livestock industry on the municipal level.

4 Methodology

In this paper, I have chosen to use an econometric approach to estimate the effects that variations in Brazilian soy exports have on different forms of land use. This entails that I set up several models with different dependent variables, representing the form of land use, trying to find the effect of the same independent variable, international soy demand, in all of them.

To estimate the causal effect of soy demand on land use, I am using a combination of two main strategies. The first strategy is fixed effects estimation, accounting for systematic time-invariant differences between the municipalities. The second strategy is the implementation of an instrumental variable with the two-stage least squares (2SLS) approach, with the intention of solving the system of simultaneous equations between the national supply and the international demand of Brazilian soy. Additionally, I use clustered standard errors on the municipal level in order to minimize the risk of biased standard errors (Hansen, 2007).

This methodology section consists of five main parts. First, the main econometric model is introduced, showing how the effect of international soy demand on land use is modelled in this paper. Then, since the soy demand is a pivotal part of this paper's analysis, I demonstrate the different ways it can be measured. The third and fourth subsection respectively describe the purpose and theory behind fixed effects estimation and the instrumental variable approach, the two main econometric approaches of this paper. Lastly, after motivating the importance of implementing instrumental variables in systems of simultaneous equation, I describe the simultaneity problem regarding soy demand and supply and how it is solved through the implementation of trade-weighted GDP as an instrument for soy demand.

4.1 The Main Econometric Model

The main model for land use is:

$$\log LU_{it} = \alpha_i + \beta \cdot \log S_{it}^d + u_{it} \tag{4.1.1}$$

In this model, the dependent variable is the logarithm of municipality i's land use at period t, written as LU_{it}. This land use can be the land use dedicated to producing exported soy, to producing soy for domestic consumption or for producing other agricultural goods. α_i is called the unobserved time-invariant individual effect and captures any unobserved

time-invariant variables on the municipality level. The last term, u_{it}, contains both fixed errors v_i caused by unobserved time-invariant variables and a random error component ϵ_{it} (Wooldridge, 2010). The regression model estimates the average elasticity of soy demand (S_{it}^d) to the land use of the relevant agricultural good in Brazilian municipalities between 2010 and 2015. This elasticity is represented by the β-coefficient. If the β-coefficient is statistically significant, the null hypothesis that there is no link between soy demand and the relevant type of land use can be rejected. Then, the elasticity coefficient β is such that an increase in the soy exports of 1% from a municipality increases the use of land dedicated to soy exports by β% in that municipality.

In order to convert the unobserved effects model into a fixed effects model, one time-demeans the model in such a way that all unit-fixed effects disappear (Wooldridge, 2010). This means that both α_i and the unobserved fixed error v_i disappear from the analysis. This directly tackles the issue of omitted time-invariant variables and is equivalent to adding dummies for every unit i. The downside is that all independent variables need to vary over time in order to not be differenced away (Wooldridge, 2010).

Time-fixed effects, the average effects of being in a certain year, are also controlled for in many of the panel data regressions and implemented through the simple step of adding time dummies to the regressions. One can say that they control for factors that affect all municipalities equally in the different years - meaning that certain year-specific phenomena can be controlled for to a degree. Time-fixed effects are often controlled for in modern panel data analysis, for example in the paper by Acemoglu *et al.* (2008) which will serve as a comparison for the instrumental approach taken in this paper.

As stated earlier, I will run this model with several kinds of dependent variables. For example, when estimating the effect of soy demand on deforestation, I will be using the logarithm of deforestation as the dependent variable. In that case, we have the model:

$$\log DF_{it} = \alpha_i + \beta \cdot \log S_{it}^d + u_{it} \tag{4.1.2}$$

Equation 4.1.2 shows the regression model for deforestation, $\log DF_{it}$ being the logarithm of deforestation in municipality i in period t. Analogous to the land use example, the object of interest is the β-coefficient that represents the elasticity between the soy demand (S_{it}^d) and deforestation. Since the soy demand is the explanatory variable of interest, the estimation or measurement of the soy demand is central to this paper.

4.2 Measuring Soy Demand

Soy demand can be modelled and measured in several different ways. *A priori*, I expect that the relationship between land use and international soy demand is largely quantity-driven. This is due to the obvious connection between land use and trading volume, since the motivation for increasing the acreage must be to increase the volume produced and exported. The expected relationship between importing countries' demand, price and the resulting land use is less straightforward than what is the case with export volume. For price, one would expect that an increase in demand leads to increased prices, but lower prices might also lead to increased international demand. It might also be the case that the average price per shipment decreases when the average size of bulk orders increases or China gains more and more of the importing market share. If this were a supply-side model, price could be a good motivator for land use expansion, but demand-wise I suspect that price rather is a negative influence on quantity demand than a good measurement of demand itself. Additionally, a very large share of the variation in prices can be explained by time-fixed effects since price is expected to develop fairly equally in different municipalities (since soybeans are an example of a fairly homogeneous good).

Thus, I propose two main measures of soy demand, one as the financial flows from soy trade in US dollars and one as the export volume of soy in tons. The relationship between these variables is then that financial flows are equal to the trade volume multiplied by price for all trades registered at customs in Brazilian ports.

$$S_{it}^{d(FF)} = \sum_j p_{ijt}Q_{ijt} = \bar{p}_{it} * Q_{it} \tag{4.2.1}$$

This equation describes the demand for soy measured as financial flows to municipality i in year t from all countries j. The separate parts p_{ijt} and Q_{ijt} are shipment prices and trade volume for every combination of municipality, country and year. The indexes ijt denote that flows are between municipality i and country j in period t, and that they are aggregated over the countries j in order to represent the demand each municipality faces. I can also use only the trade volume as a representation of demand. The equation for trade volume aggregation is then:

$$S_{it}^{d(vol)} = \sum_j Q_{ijt} \tag{4.2.2}$$

This means that the representation of municipal soy demand as trade volume is simply the sum of all trade flows between municipality i and country j in period t, summing over countries j in order to get the municipal demand for period t.

In my analysis, I will be using the logarithms of soy demand to estimate elasticities between soy demand and different forms of land use. When doing this, the effect of prices on financial flows becomes additive:

$$Log(S_{it}^{d(FF)}) = log(\bar{p}_{it}) + log(Q_{it}) \qquad (4.2.3)$$

This implies that any difference between the models where soy demand is measured as financial flows and models where it is measured as export volume must be caused by prices having a differentiating effect. To illustrate the effect of prices, I will also add a third model of soy demand that picks up the correlations between prices, the instrument and the dependent variables. The measurement of price can be interpreted as the average shipment price in municipality i at time t, and is mathematically formulated as:

$$S_{it}^{d(p)} = \frac{S_{it}^{d1}}{S_{it}^{d2}} = \frac{\sum_j p_{ijt} Q_{ijt}}{\sum_j Q_{ijt}} = \sum_j p_{ijt} \frac{Q_{ijt}}{Q_{it}} = \bar{p}_{it} \qquad (4.2.4)$$

4.3 Fixed Effects Estimation and Clustered Standard Errors

In this paper, I use fixed effects estimation to control for systematic and time-invariant differences between municipalities. The first to offer a detailed explanation of this topic was Eisenhart (1947) in his article on *The Assumptions Underlying the Analysis of Variance*. Eisenhart classified variance into two groups, namely "Class 1" and "Class 2" variance. The former is what Eisenhart called the variance between population means, fixed variance, meaning that different populations have different average values of certain parameters (Eisenhart, 1947). The analysis of "Class 2" variance is "the detection and estimation of components of (random) variance associated with a composite population" (Eisenhart, 1947), and one tests whether a component of variance in one variable (the dependent variable in econometrics) can be ascribed to the variance in another variable (the independent variable) rather than testing whether there are fixed differences between the means. Eisenhart (1947) suggests that the analysis of "Class 2" variance in a sense is the true analysis of variance. In a fixed effects model, one is able to specifically target the "Class 2" variance by controlling for the

individual-specific unobserved constant α_i for unit i. By doing this, one controls for fixed differences between the means, which is the source of "Class 1" variance, and it is then possible to estimate and analyze "Class 2" variance between the dependent and independent variable.

$$y_{it} = \alpha_i + (\sum_j (\beta_j \cdot x_{jt})) + u_{it} \tag{4.3.1}$$

Equation 4.3.1 presents a general unit-effects model. According to Wooldridge (2010), the fixed-effects model is attained by time-demeaning the unobserved effects model like this:

$$\ddot{y}_{it} = y_{it} - \bar{y}_i = (\alpha_i - \alpha_i) + (\sum_j \beta_j^* \cdot (x_{jt} - \bar{x}_j)) + (u_{it} - \bar{u}_i) = (\sum_j \beta_j^* \cdot \ddot{x}_{jt}) + \ddot{u}_{it} \tag{4.3.2}$$

By time-demeaning the variables, one gets rid of all individual unobserved time-invariant effects α_i and any fixed components present in the error term. Wooldridge (2010) explains that the estimates from the fixed effects method will be the same as if one introduced dummies for every individual unit, which in this paper would be municipalities. This is because the effect of introducing an individual-specific constant like a municipality-dummy also will pick up all "Class 1" variance caused by differences between means.

The main reason for why a fixed effects model is the appropriate tool to use in this paper is that the groups of units in the paper, namely the municipalities, face different and non-random treatments and time-invariant factors that affect the soy production. Municipalities are systematically different, and by using a fixed effects model one can control for or time-demean away all fixed differences between municipalities and more directly be able to pinpoint the effect of soy demand varying over time. By doing this, one focuses entirely on the variance within municipalities rather than between them (Wooldridge, 2010). In practice, a Hausman test is the most common tool to use in order to identify whether a fixed effects model is needed or not (Clark and Linzer, 2015).

The Hausman test originates from a famous 1978 article on the *Specification Tests in Econometrics* (Hausman, 1978) and tests whether a random effects estimator is inconsistent and a fixed effects estimator is needed. Wooldridge (2010) explains that the test is based on evaluating the difference between the random and fixed effects estimators, and that a statistically significant difference is taken as evidence against the random effects estimation being a consistent model. Furthermore, the fixed effects model is consistent even in the event

26

that the Hausman test does not reject the null hypothesis of there not being any systematic differences between the panel data groups. The argument for using random effects models in that case is that they are known to be more efficient when there are no systematic differences between units, meaning that fewer observations are needed for unbiased estimation and that the random effects estimator in that sense is a stronger tool due to less variance being differenced away (Wooldridge, 2010).

A Hausman test on a panel data regression of soy exports on land use rejects the null hypothesis that there is no difference between the random effects estimator and the fixed effects estimator [1]. This indicates that the individual-fixed effects, not controlled for fully in the random-effects estimation, affect the system and lead to inconsistent estimators (Wooldridge, 2010). If an estimator is inconsistent, its expected value will not converge on the true population parameter even for large numbers of observations (Wooldridge, 2010). Since there are both empirical indications of systematic differences between municipalities as well as an *a priori* argument for these differences existing, I have chosen to use fixed effects estimation in this paper.

According to Angrist and Pischke (2009), fixed effects estimation is also quite useful for removing omitted variable bias, conditional on the premise that individual unobserved effects are time-invariant. In panel data with fixed effects estimation, only within-group variation over time in observed variables is used to estimate coefficients. This leads to one of the weaknesses of fixed effects estimation, namely that it can be hard to identify the effect of variables that experience little change over time (Angrist and Pischke, 2009). This is due the fixed effects approach removing a lot of "good" variation along with the "bad", variation that could be useful in order to identify potentially causal relationships. Additionally, Angrist and Pischke (2009) state that measurement errors varying over time can lead to significant bias in fixed effects model estimators. They elaborate, however, that this problem can be tackled through taking an instrumental variable approach like I have chosen to do in this paper.

Clustered standard errors are also added to the panel data models. Abadie *et al.* (2017) explain that this often is a good idea in the experimental design situation where clusters are non-randomly sampled and treatment effects are heterogeneous. In this paper, it is clear that municipalities and differences between them are non-random and systematic, indicating that clustered standard errors should be included according to the standards of Abadie *et al.*

[1]See appendix section B

(2017). The main point of clustering standard errors is to make the statistical inference about estimators robust to heteroscedasticity and serial correlation (Stock and Watson, 2008). The failure to include clustered standard errors can then lead to biased standard errors, leading to potentially incorrect inference about the statistical significance of estimators (Hansen, 2007).

4.4 The Instrumental Variable Approach

Choosing a proper instrumental variable is seldom an easy task, but the usefulness of such variables have made them an ubiquitous part of modern social science literature. First and foremost, instrumental variables are powerful tools when facing systems of simultaneous equations like one often does when quantities and prices are involved. This finding was first reported in Appendix B of Phillip Wright's book on *Animal Tariffs and Vegetable Oils*, published in 1928 (Stock and Trebbi, 2003). Goldberger (1972) informs that this appendix was co-written by Phillip's son Sewall, also an accomplished statistician. According to Goldberger (1972), the Wrights quite succinctly describe the solution to the simultaneity problem for demand and supply, which is to introduce an additional factor that can describe demand conditions without affecting cost conditions or describe the cost conditions without affecting demand conditions (Wright, 1928). In other words, in the relevant simultaneous equation, the instrument needs to be uncorrelated with the error term while it exogenously estimates the targeted instrumented variable.

Angrist and Krueger (2001) explain that the second purpose of instrumental variables is to reduce measurement errors. Furthermore, instrumental variables lead to consistent, but not necessarily unbiased estimates of coefficients. This indicates that instrumental variable approaches like the two-stage least squares benefit from having large sample sizes, as consistent predictors, unlike unbiased predictors, need large samples in order for the estimator to converge on the population parameter (Angrist and Krueger, 2001).

The third use of instrumental variables is to tackle the issue of omitted variables. The most famous example, also used by Angrist and Krueger (2001), is the effect of ability and education on pay. Ability is unobservable and ends up in the error term of a regression of education on pay. Additionally, ability and education are known to be somewhat correlated, meaning that the independent regressor is systematically correlated with the error term as well. This is solved by introducing an instrumental variable that predicts education, but is

not correlated with the ability of the individual and thereby the error term of the regression.

4.5   Instrumenting Brazilian Soy Exports

In the case of soy exports in Brazil, the main purpose of introducing an instrument is the problem of there being a system of simultaneous equations through the dynamics of supply and demand. The demand of soy is the main focus of this paper, and it will need to be isolated from this system of simultaneous equations.

An instrument that can properly estimate soy demand without being strongly correlated with soy supply is necessary for causal inference, like the Wrights proposed according to Goldberger (1972). In this paper, the simultaneous equations are caused by a complicated web of correlations between land use, demand and supply. The simplest connection between demand and supply is that they are both usually measured by the exact same variable, that is the quantity sold or the trading volume. In addition, both are affected by prices. Notably, demanded quantity is known to generally increase if prices fall while supply will increase when a surge in demand increases prices and projected profit margins increase. In other words:

$$
\begin{aligned}
S^d &= f(p) \\
S^s &= g(p) \\
p &= z(S^d - S^s)
\end{aligned}
\tag{4.5.1}
$$

As equation 4.5.1 presents, both supply and demand can be seen as a function of prices, and prices are affected by excess demand or excess supply. This means that changes in demand and changes in supply must be intrinsically linked.

The endogeneity of supply and demand is even more obvious when land use is introduced into the system of simultaneous equations. If one is to estimate the effect of increasing soy demand on land use, one essentially tries to estimate part of the relationship between demand and supply. This is because supply is affected by the amount of available land, among other time-varying factors such as rainfall and active workforce. Demand, on the other hand, will mostly be decided by a combination of the purchasing power and preferences of the importers as well as the quality and price of the product. The quality and price of the product are in turn affected by the cost conditions and production capabilities of the supply-side. In other words, the demand is highly endogenous with production volume and land use, and

an instrument is therefore necessary in order to properly test the effect of demand changing. In an econometric model, where i denotes municipality, t year and j the importing country:

$$LU_{it} = \alpha_i + \beta \cdot S_{it}^d + u_{it} \tag{4.5.2}$$

$$S_{it}^d = \alpha_i + \sum_j (\beta_j PurchasingPower_{jt}) + \sum_j (\gamma_j p_{ijt}) + u_{it} \tag{4.5.3}$$

$$S_{it}^s = \beta_0 + \beta_1 \cdot LU_{it} + \beta_2 \cdot Rainfall_{it} + \beta_3 \cdot Workforce_{it} + u_{it} \tag{4.5.4}$$

Here, I simply write the statistical equivalent of my earlier argument, that demand and supply are interconnected and simultaneously decided through different factors while both are represented by the measures of demand constructed in subsection 4.2. The demand equation 4.5.3 argues that the conditions in importing countries j as well as the price they face in municipality i at time t are important determinants for demand. Since supply and demand are measured similarly and land use affects supply, the main regression model, shown in equation 4.5.2, has multiple omitted variables and a biased and inconsistent demand variable coefficient because the supply-affecting variables have an effect on the production volume and therefore export volume, financial flows and also prices over time. One could say that the soy demand and the residual u_{it} of 4.5.2 are correlated because the factors deciding the soy supply are omitted from the model even though they affect soy supply and therefore indirectly affect the soy demand.

The solution, as written by the Wrights (1928), is to implement an instrument that can estimate the demand (S^d) without affecting the cost conditions (S^s). By doing this, implementing an instrumental variable z_{it} that is exogenous and therefore uncorrelated with S^s and u_{it}, one ends up with the first-stage instrumental variable equation with unit-fixed effects:

$$\hat{S}_{it}^d = \alpha_i + \beta \cdot z_{it} + u_{it} \tag{4.5.5}$$

Then, it is possible to estimate the potentially causal effect of soy demand on land use by implementing the estimation of soy demand into the original land use model. In the end, the main regression model ends up being a two-stage least squares model with fixed effects and trade-weighted GDP instrumenting soy demand, what Wooldrigde (2010) calls a FEIV (Fixed Effects Instrumental Variables) model:

$$\log LU_{it} = \alpha_i + \beta \cdot \log \hat{S}_{it}^d + u_{it} \tag{4.5.6}$$

The difference here, compared to equation 4.5.2 for land use, is that soy demand is now instrumented through a first-stage regression. $\hat{S}_{it}^d$ is simply the soy demand estimated through the first-stage regression, represented by equation (4.5.3).

In this paper, the chosen instrument for soy demand is trade-weighted GDP with weights set in the initial period. The main motivation for this is that changes in trade-weighted GDP should be a relatively good estimate for the purchasing power of the importers, weighted by relevance to the Brazilian soy export market. This should *a priori* be an important determinant for demand, as stated through equation 4.5.3 and the related discussion. The weights of the trade-weighted GDP, calculated separately for each Brazilian municipality, are the share of exports from a municipality i that country j imports. The GDP measure is in real 2005 US dollars so that relative inflation is less of an issue, and the trade-weighted GDP that each municipality faces is the sum of all importing countries' weighted GDP. Essentially, for municipality i and country j:

$$\log(\hat{S}_{it}^d) = \alpha_i + \beta \cdot \log(\sum_j (w_{ij}^{2010} \cdot GDP_{jt})) + u_{it} \tag{4.5.7}$$

In equation 4.5.7, showing this paper's main first-stage regression, $\hat{S}_{it}^d$ represents the soy demand that municipality i faces in period t. w_{ij}^{2010} are the weights determined by the share of each municipality i's export volume that each country j imports in 2010, while GDP_{jt} is country j's GDP in period t. The trade-weighted GDP is then the sum of all countries j's Gross Domestic Products weighted by w_{ij}^{2010}. α_i is the municipality-fixed effects, capturing all time-invariant heterogeneity between the municipalities when controlled for by the fixed-effects estimation method. β is a coefficient that measures the elasticity between the trade-weighted GDP and the dependent soy demand variable. The general interpretation when using a model like this is that a 1% increase in the independent log-variable, which in this case is trade-weighted GDP, leads to a β% increase in the dependent log-variable. The model can also be expanded to include time-fixed effects, which I have done in the later panel data analysis. When controlling for time-fixed effects, I add dummies for every year, catching the average effect that being in a certain year has on the municipal soy exports. In this way, I more accurately isolate the effect that changes in the trade-weighted GDP have on changes in the municipal soy exports. The time-fixed effects, if highly significant, can for

example be caused by macroeconomic conditions or other year-specific effects that impact every municipality simultaneously.

The method I used to create the country-specific weights between was to, for every municipality, divide country-specific soy imports on total municipal exports. Mathematically: 1

$$w_{ij}^{2010} = \frac{volume_{ij}^{2010}}{\sum_j volume_{ij}^{2010}} \tag{4.5.8}$$

In this way, the weights w_{ij} all individually add up to 1 when summing over all countries j, meaning that w_{ij} represents the share of total soy export volume from municipality i that country j imported in 2010.

5   Analysis

5.1   Evaluation of Trade-weighted GDP as an Instrument

5.1.1   Qualitative Comparison with Recent Literature

The use of trade-weighted GDP as an instrument has some support in the literature. For example, Acemoglu et al (2008), in a paper investigating the link between national income and democracy, use trade-weighted world GDP as an instrument for each country's GDP (and later each country's GDP per capita since GDP per capita is a direct function of GDP). Mathematically, with $\hat{Y}_{it-1}$ being the estimated national income of country i in period $t-1$, $\hat{y}_{it-1}$ being the analogous estimate for GDP per capita:

$$\hat{Y}_{it-1} = \left(\sum_{j=1, j \neq i}^{N} w_{ij} Y_{jt-1} \right) + \epsilon_{it-1} \tag{5.1.1}$$

$$\hat{y}_{it-1} = f(\hat{Y}_{it-1}) = \pi^F \hat{Y}_{it-1} + \sum_k \beta_k x_{k,it-1} + \epsilon_{it-1} \tag{5.1.2}$$

In these equations, Acemoglu *et al.* (2008) describe the GDP of a nation as the trade-weighted sum of all other nations' national income. π^F represents the relationship between estimated trade-weighted GDP and GDP per capita and the $\beta_k \cdot x_k$ for all indices k represents all other independent variables x_k in the system with their own estimated coefficients β_k. Several independent variables were specified by Acemoglu *et al.* (2008), but no other variables

than the trade-weighted GDP are of particular interest when making a comparison to this paper.

One key difference between the approach taken in this paper and the one taken by Acemoglu et al is that the instrumented variable in this paper is international demand for Brazilian soy on the municipal level. In contrast, Acemoglu et al (2008) use different nations' GDP per capita as the variable instrumented by trade-weighted GDP. Secondly, I chose to fix the weights as the share of soy imports from municipalities in the initial year, while Acemoglu et al (2008) chose weights w_{ij} that represent the share of total imports from every other country $j \neq i$ to the exporting country i in the period 1980-1989. The rationale behind fixing the weights is to let shocks to GDP be the only factor causing changes in the instrumental variable. This would mean that a significant relationship between trade-weighted GDP and soy production would indicate that growth in the importing nations' economies explains increases in soy production and exports to some degree. If weights were not fixed, changes in the instrument would also be affected by the composition of importing countries and would not necessarily signify a demand shock happening in the importing countries. The downside of this approach is that variation in the composition of importing countries could provide useful information for analysis, and fixing weights will mean that one operates under the assumption that this variation either is not large over the relevant period or not significant to the relationship between demand and acreage (Arora and Vamvakidis, 2004). In this paper, fixing the weights in the initial year should be a reasonable approximation because the relatively short time frame of the data limits the amount of variation in trade composition that municipalities experience over the period.

The trade-weighted world income seems to have been quite successful at estimating the GDP and GDP per capita of the exporting country (Acemoglu *et al.*, 2008). This is a mixed blessing for this paper. First of all, this could mean that the trade-weighted GDP is directly correlated with Brazilian GDP, which could be problematic when using it to estimate the international demand of soy. This is because the Brazilian GDP is intrinsically affected by the production and exports of soy, meaning that the problem of endogeneity is not fully accounted for. However, because the weights are the share of soy imports rather than the share of total imports, the soy-trade-weighted GDP should relatively speaking be a much worse estimator for total Brazilian GDP compared to its main channel of influence, namely soy exports. The main defense of trade-weighted GDP as an instrument, with weights decided by the share of soy imports in the initial year for every municipality, must be that

the instrument exclusively (or at least primarily) affects the national GDP through the changes it causes in soy exports.

5.1.2 Partial Correlation Analysis
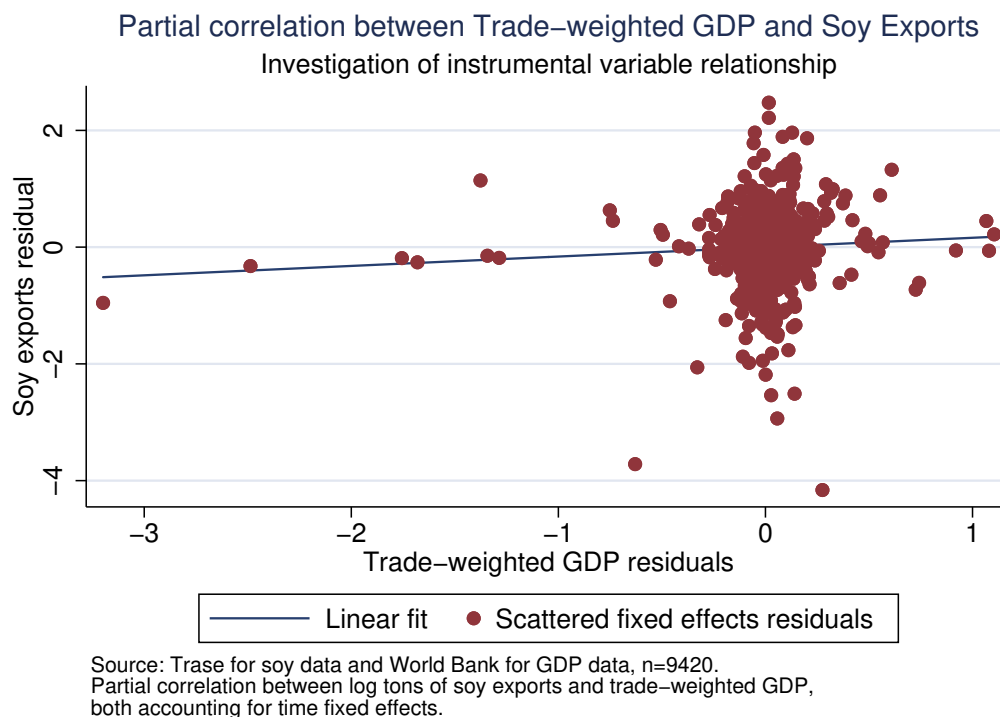
Figure 6: Partial correlation between log soy exports (volume) and log weighted GDP, using residuals from fixed effects regressions. Sources: https://trase.earth/data & http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators

Figure 6 shows a partial correlation analysis between trade-weighted world income and soy export volume. This was done by regressing log soy export volume and log trade-weighted world income on municipality- and time-fixed effects before plotting the non-fixed component (ϵ_{it}) of the residuals of these panel data regressions against each other in a scatter plot. The residuals contain the variance of log export volume and log trade-weighted GDP that is not explained by the municipality- and time-fixed effects. I find that there is a slight positive correlation between the residuals, indicating that there is a positive correlation between the underlying variables when I account for the fixed effects. As the correlation is

limited, mostly due to trade-weighted GDP having few observations with significant error terms, it seems that the relationship between trade-weighted GDP and soy export volume is not very strong. Trade-weighted GDP seems to be explained fairly well by the fixed effects themselves, meaning that average municipal and time effects account for much of the variation in the variable. This is potentially a worrying sign for its usage in later analysis, as variance is the key component in measuring the effect that one variable changing has on another. However, the partial correlation analysis does indicate non-zero correlation, and the cross-sectional sample size is large, meaning that trade-weighted GDP will still be able to explain soy export to a certain degree. This is confirmed in the following subsection's fixed effects estimation of soy demand instrumented by trade-weighted GDP.

5.1.3 An Econometric Estimation of Soy Demand

Table 1: Estimates from first-stage FEIV regressions of the log of export volumes, financial flows, and prices of soy on log trade-weighted GDP in Brazilian municipalities

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Volume	Volume 2	FF	FF 2	Prices	Prices 2
logwGDP	0.0941***	0.0481***	0.0923***	0.0345**	-0.0018	-0.0136***
	(0.0177)	(0.0138)	(0.0181)	(0.0144)	(0.00319)	(0.00265)
Municipality-FE	Yes	Yes	Yes	Yes	Yes	Yes
Time-FE	No	Yes	No	Yes	No	Yes
F	28.37	12.15	25.91	5.74	0.328	26.34
Observations	9420	9420	9420	9420	9420	9420

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Three panel data models with fixed effects for 1570 Brazilian municipalities between 2010-2015. Coefficients for time-fixed effects and the intercept are suppressed as they are not of particular interest to the analysis. Dependent variables for model 1 & 2 are log of trade volume, while model 3 & 4 use log of financial flows as a dependent variable and models 5 & 6 use log of prices. Even-numbered models control for time-fixed effects (using year dummies) as well as municipality-fixed effects, while odd-numbered models only control for municipality-fixed effects. Standard errors are clustered on the municipal level in all models.

Table 1 presents six versions of the first-stage instrumental variable regression of trade-weighted GDP on municipal soy exports. All models control for municipality-fixed effects and the even-numbered models also control for time-fixed effects. The standard errors are clustered on the municipality level in all six models in order to deal with serial correlation within the municipalities. Clustering makes the standard errors more robust to heteroscedasticity, and in fixed effects estimation it is advised to cluster when the treatment effects are heterogeneous between clusters (municipalities) (Abadie *et al.*, 2017). Hansen (2007) specify that clustering is necessary because of group-level random shocks causing correlation between observations within the groups for all time periods, meaning that the observations become serially correlated.

The first two models use municipal soy trade volume as dependent variables, with the difference of model 2 controlling for time-fixed effects while model 1 only uses municipality-fixed effects. To test the strength of the instrument, I use the F-statistic testing the null hypothesis that the coefficient for trade-weighted GDP is equal to zero. The rule-of-thumb proposed by Staiger and Stock (1994) is that an instrument should have an F-statistic above 10 in order to not be considered a "weak" instrument. In models 1 & 2, the F-statistic exceeds this lower boundary. This indicates that, while time-fixed effects weaken the apparent strength of the instrument slightly, trade-weighted GDP seemingly explains municipal soy trade volume well enough to be used in second-stage analysis.

An interesting result from the first two models in table 1 is that the coefficient of elasticity decreases when controlling for time-fixed effects, while the standard errors are stable in magnitude across the two models. In model 1, without time-fixed effects, a 1% increase in trade-weighted GDP leads to a 0.094% increase in soy exports from the relevant municipality. In model 2, this elasticity is about halved, meaning that a 1% increase in trade-weighted GDP is associated with a 0.048% increase in soy export volume.

A positive relationship between the trade-weighted GDP and the export volume was expected since a demand shock in the form of an income increase for an importing country should lead to a higher volume of soy being imported, and the response should be stronger the closer the trade relationship is between municipality i and the country j experiencing the shock to national income. The closeness of the trade relationship is of course represented by the 2010 import share w_{ij}^{2010} that country j has of municipality i's exports. The model 2 results also make sense given the slight positive correlation found in figure 6's partial correlation analysis. The linear fit in figure 6 between the residuals of trade-weighted GDP

and export volume controlling for municipality- and time-fixed effects is in fact equivalent to running a fixed effects regression of export volume on trade-weighted GDP with municipality- and time-fixed effects.

Similar results are found in models 3 & 4, where the logarithm of financial flows in US dollars are used as dependent variables. When controlling for time-fixed effects, the effect of increasing trade-weighted GDP by 1% is a 0.0345% increase in financial flows. However, the F-statistic is below the rule-of-thumb lower bound of 10, indicating that the trade-weighted GDP is a weak instrument for financial flows when time-fixed effects are controlled for (Staiger and Stock, 1994). Staiger and Stock (1994) find that weak instruments can lead to biased second-stage coefficient estimates even in large samples, so it is important to be a careful when interpreting the two-stage least squares coefficients of a model with a weak instrument. Trade-weighted GDP seemingly estimates financial flows better when time-fixed effects are not controlled for, as the F-statistic for the elasticity coefficient is 25.91, well above the rule-of-thumb limit for weak instruments. However, the fact that introducing time-fixed effects has a significant effect on the elasticity estimates implies that time-fixed effects are present and should be controlled for. This means that the even-numbered models with time-fixed effects in a sense are the more complete models and are generally preferred over the models without time-fixed effects.

Models 5 & 6 use the logarithm of prices as dependent variables. The results from this analysis are a bit more puzzling, but it is important to remember that these models were meant to act as an explanation for why the models for financial flows and for trade volume differ. Looking at table 1, it is apparent that the coefficients in the price model is approximately equal to the difference between the coefficients in the financial flows model and in the trade volume model. To exemplify, the difference between the first-stage coefficients in models 1 & 3 is (0.0923 - 0.0941 = -0.0018), which is the first-stage coefficient for the corresponding price model 5 without time-fixed effects. Similarly for model 6, the difference between model 2 and model 4's estimated elasticity coefficients is (0.0345 - 0.0481 = -0.0136). This means that the price model explains the difference between the financial flows models and the corresponding trade volume models, and statistically significant results indicate that prices create enough noise in the system that the trade volume estimates and financial flow estimates of soy demand differ significantly. An economic interpretation of the actual elasticity estimates might not be very meaningful, as it is not intuitive that an increase in trade-weighted GDP, indicating an income and therefore demand increase in importing

37

countries, should lower prices of soy.

Much in the same way that the price models 5 & 6 can be seen as the differentiating factor between the financial flows and trade volume demand models, financial flows models 3 & 4 combine information about prices and export volume since the log of financial flows is equal to the sum of log prices and log trade volume. Seeing prices and trade volume often have inverse relationships in terms of demand, being negatively correlated because higher trade volumes lead to lower prices and higher prices lead to less volume demanded, it is fairly intuitive that the estimates for elasticity between financial flows and trade-weighted GDP falls somewhere in between the elasticity estimates using export volume and prices. It is also evident that the relationship between trade-weighted GDP and export volume dominates the relationship between trade-weighted GDP and prices. Since this is the case, and because the F-statistic for the elasticity between financial flows and trade-weighted GDP falls short of the Staiger & Stock rule-of-thumb (1994), I conclude that model 2 with trade-weighted GDP estimating export volume and year-dummies for time-fixed effects is the strongest out of these six first-stage regression models.

5.2 The Land Use of Exported Soy

In this subsection, I investigate the effect of international soy demand on the land use of exported soy. My intent is to provide the answer for research question 1 by analyzing whether international soy demand causes expansion of the Brazilian soy industry. *A priori*, this effect is expected to be positive and large. However, when formally investigating the relationship in an econometric model, the null hypothesis is that there is no such effect:

$H_0 = $ There is no link between international soy demand and the land use of exported soy - the elasticity between the variables is zero.

$H_A = $ There is a significant relationship between soy demand and the land use of exported soy, meaning that the elasticity coefficient representing this relationship is significantly different from zero.

Table 2 presents the results from the second-stage regression of the land use of exported soy on soy demand. The six models use three different estimators for the soy demand analogous to the first-stage estimates in table 1. The first two models use the first-stage estimates of log export volume using trade-weighted GDP as an instrument. Trade-weighted

Table 2: Second-stage estimates from FEIV regressions of the log of land use of exported soy on log export volumes, financial flows, and prices of soy (instrumented by trade-weighted GDP) in 1570 Brazilian municipalities between 2010-2015.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Volume	Volume 2	FF	FF 2	Prices	Prices 2
logexpvol	0.927***	0.766***				
	(0.0496)	(0.105)				
logFF			0.946***	1.068***		
			(0.0514)	(0.149)		
logprice					-47.80	-2.712**
					(85.21)	(1.242)
Municipality-FE	Yes	Yes	Yes	Yes	Yes	Yes
Time-FE	No	Yes	No	Yes	No	Yes
F (IV)	28.37	12.15	25.91	5.74	0.328	26.34
Observations	9420	9420	9420	9420	9420	9420

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Second-stage of the two-stage least squares panel data models for land use of exported soy. The models include municipality fixed effects for 1570 Brazilian municipalities between 2010-2015. Like in table 1, coefficients for time-fixed effects and the intercept are suppressed. The dependent variable for all models is the log of land use dedicated to soy exports. Models 1 & 2 use log of trade volume as the instrumented independent variable, while model 3 & 4 use log of financial flows and models 5 & 6 use log of prices. Even-numbered models control for time-fixed effects (using year dummies) as well as municipality-fixed effects, while odd-numbered models only control for municipality-fixed effects. The reported statistics are the F-statistic from the first-stage regression as well as the number of observations. The number of observations is 9420 in all models, corresponding to 6 years of observations from 1570 municipalities. Standard errors are clustered on the municipal level in all models.

GDP also instruments the log of financial flows in models 3 & 4 and the log of prices in models 5 & 6.

Models 1 & 2 show that changes in the export volume of soy seem to be a good indicator for the changes in land use dedicated to soy exports. This is not at all surprising. In model 1, with only the municipality-fixed effects and no time-fixed effects, the elasticity between export volume and land use dedicated to soy exports is 0.927 and statistically significant on the 1% significance level. In other words, model 1 indicates that land use is highly responsive to changes in soy demand, represented by export volume instrumented by trade-weighted GDP. A 1% increase in export volume is expected to lead to a 0.927% increase in land use dedicated to soy exports, according to this model. In model 2, when time-fixed effects are controlled for, the elasticity is still high, but slightly more modest in magnitude. In this model, the elasticity between soy export volume and land use dedicated to soy exports is 0.766, and also here the coefficient is significantly different from zero on the 1% significance level.

Models 3 & 4 yield coefficients that can be interpreted as the elasticities between financial flows from soy exports and the land use dedicated so soy exports. Here, the elasticities are close to one-to-one, which is called a unit elasticity (Antras, 2004). The estimated elasticity coefficients are 0.946 in the model without time-fixed effects and 1.068 in the model with time-fixed effects. A null hypothesis of the coefficients being equal to zero can be rejected at the 1% level in both model 3 & 4, while a null hypothesis of there being a unit elasticity (meaning that the true elasticity is 1) would not even be rejected at the 5% level as 1 is well within the 95% confidence interval of the elasticity coefficients in both models. If the elasticity between financial flows and land use dedicated to soy exports indeed is a unit elasticity, export-dedicated land use of soy would increase 1% in response to a 1% increase in financial flows from soy exports.

However, model 4 has a glaring weakness in the sense that the first-stage regression indicates that the log of trade-weighted GDP is a weak instrument for the log of financial flows. As I prefer to control for time-fixed effects due to their obvious relevance both in the first and second-stage regression, the comparison between using financial flows and export volume must lead to a comparison between models 2 & 4. Because of the failure of log trade-weighted GDP instrumenting log financial flows, model 2 using the instrumented export volume seems like the better model.

The last two models, using log price as the independent variable instrumented by the

log of trade-weighted GDP, seem substantially weaker. Referring to the discussion of the first-stage model presented in table 1, the first-stage results for the price models seem to lack an unambiguous interpretation in an economic sense. They were initially meant to be used strictly as an explanation for why the financial flows models and the trade volume models differ. If taken at face value, model 6 presents statistically significant results showing that increases in price has a remarkably large negative effect on the land use of exported soy. This does seem a bit counter-intuitive, and is probably caused by the puzzling first-stage results indicating a negative relationship between trade-weighted world income and prices. It is, however, plausible that higher prices actually slow down the land expansion of soy from a demand perspective. As stated in the beginning of the background section, a study by Berry and Schlenker (2011) finds that higher prices have shown signs to increase yields or the productivity of the land, which again causes the effect of prices on land use to be affected negatively. In the end, I must conclude that the price models provide slightly ambiguous results, and that a direct interpretation of the coefficients does not seem to be expedient.

One can also compare the effects of trade-weighted GDP increasing by 1% using models 2, 4 & 6 in tables 1 & 2, representing the compound effect of income shocks in importing countries. The method is quite simple, as table 1 shows the relationship between trade-weighted GDP and the relevant instrumented variable and table 2 shows the relationship between the instrumented variables and land use dedicated to soy exports. For model 2, a 1% increase in trade-weighted GDP leads to a 0.0481% increase in trade volume. Plugging this into table 2 means that one analyzes the effect of trade-weighted GDP increasing by 1% if one analyzes the effect of trade volume increasing by 0.0481%. In table 2, trade volume increasing by 0.0481% leads to land use dedicated to soy exports increasing by $0.0481 \cdot 0.766 \approx 0.037\%$. Analogously for model 4, a 1% increase in trade-weighted GDP ultimately results in a $0.0345 \cdot 1.068 \approx 0.037\%$ increase in the land use dedicated to soy exports. For model 6, a 1% increase in trade-weighted GDP, instrumenting prices, leads to a $(-0.0136) \cdot (-2.712) \approx 0.037\%$ increase in land use dedicated to soy exports. In other words, the relationship between increases in trade-weighted GDP and increases in land use dedicated to soy exports is the same in all three models and is independent of what variable I instrument. Again, this underlines the inter-relatedness of the models.

In conclusion, the null hypothesis of there being no link between the land use of exported soy and international soy demand can resoundingly be rejected. The relationship between the variables is statistically significant, and the chosen model, model 2, estimates an elasticity

coefficient of 0.766. This implies that the relationship between soy demand and the land use of exported soy is highly elastic and has large economic significance. Throughout the next few subsections, the topic of relevance is the identification of potential negative externalities the land expansion of export-oriented soy farms have on alternative forms of land use. The first outcome variable of interest is deforestation and how it responds to soy demand.

5.3 Deforestation

In this subsection, I investigate the direct effects of international soy demand on deforestation in Brazilian municipalities. In other words, this concerns the land expansion of soy farms into forests and is the first attempt at answering research question 2. The research question, as stated in the introduction, is about the composition and extent of the different displacement effects caused by the land expansion of the export-oriented soy industry. Land expansion of soy into forests, causing deforestation, is one of these possible displacement effects, and in the following econometric analysis I investigate whether this effect can be found in the data from soy-producing Brazilian municipalities between 2010 and 2015. As stated in the methodology section, the goal is to estimate an elasticity coefficient between deforestation and soy demand and see whether it is possible to reject the null hypothesis that this coefficient is zero.

H_0 = There is no link between international soy demand and deforestation - the elasticity coefficient between the variables is zero.

H_A = There is a significant relationship between soy demand and deforestation and the elasticity coefficient between the variables is significantly different from zero.

Table 3 presents the results when estimating the effect of soy demand on deforestation. The number of observations is significantly lower in this deforestation study, 1218 to 9420 in the full sample of soy-producing municipalities, making it harder to find statistically significant results. The results using the same 6 models as earlier, split up into three pairs of models using trade volume, financial flows and prices to represent soy demand, are lacking in terms of statistical significance. None of the models have elasticity coefficients from the second-stage regression that are significantly different from zero. This means that one cannot conclude that there is a relationship between the different estimations of soy demand and deforestation. However, the sign of the coefficients or the direction of the estimated effects

Table 3: Estimates from FEIV regressions of the log of deforestation on log export volumes, financial flows, and prices of soy (instrumented by trade-weighted GDP) in Brazilian municipalities

	Model 1 Volume	Model 2 Volume 2	Model 3 FF	Model 4 FF 2	Model 5 Prices	Model 6 Prices 2
logexpvol	0.167	1.035				
	(0.195)	(0.933)				
logFF			0.170	1.155		
			(0.200)	(1.151)		
logprice					-9.826	-9.972
					(15.38)	(12.66)
Municipality-FE	Yes	Yes	Yes	Yes	Yes	Yes
Time-FE	No	Yes	No	Yes	No	Yes
First-stage	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
logwGDP (IV)	0.611***	0.177*	0.601***	0.159	-0.0104	-0.0184
	(0.189)	(0.0980)	(0.193)	(0.106)	(0.0150)	(0.0219)
F (IV)	10.45	3.26	9.70	2.25	0.48	0.71
Observations	1218	1218	1218	1218	1218	1218

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Two-stage least squares model of deforestation on instrumented export volume, financial flows and prices. The models include municipality-fixed effects for 203 Brazilian municipalities between 2010-2015. The first-stage results are reported under the second-stage results, with estimated coefficients for the elasticities between log export volume, financial volume or prices and the trade-weighted GDP of the trading partners for each municipality. The instrumental approach and the application of fixed effects is analogous to the combination of tables 1 & 2. The reported statistics are the F-statistic from the first-stage regression and the number of observations. The number of observations is 1218 in all models, corresponding to 6 years of observations from 203 municipalities. Standard errors are clustered on the municipal level in all models.

point in the *a priori* expected direction. The estimated (non-significant) relationship between log export volume and deforestation is positive, so is the relationship between financial flows and deforestation. The opposite is true for the non-significant relationship between prices and deforestation. In other words, the directions of the estimated effects, although insignificant for all models, point in the same direction as the equivalent models for land use of exported soy. This is probably not a coincidence, since soy agriculture expanding can lead to the displacement of forests, which necessarily would cause deforestation. However, with first-stage F-values as low as they are in table 3, especially in the models with time-fixed effects, one cannot infer too much about the relationship between deforestation and soy demand.

An in-depth interpretation of the elasticities is not very interesting given the lack of statistical significance in the models, but the sizable magnitudes suggests that the economic significance of a relationship between export volume and deforestation in model 2 could be high if it were found. The problem is, of course, that the estimation in table 3 cannot reject the null hypothesis that the effect of increasing soy demand on deforestation is zero due to the high uncertainty of the estimates. That said, this does not disprove the alternative hypothesis that there is a link between the variables. This is an important distinction. Finding no evidence of a causal relationship does not imply that this causal relationship does not exist. A very informal analogy could be the fact that a man cannot disprove the existence of his keys by not finding them in his pockets.

Although the failure to reject the null hypothesis that the elasticity between soy demand and deforestation is zero does not prove that there is no link between soy demand and deforestation, it opens for this possibility. In other words, the true relationship between soy demand and deforestation could be that they are not causally related. As stated in the background section, Brazil has implemented several environmental policies meant to curb the effect of soy agriculture on deforestation. Additionally, there is the international pressure to reduce carbon emissions that keeps soy exporters on their toes. The combined effects of social pressure and domestic laws may have made the true relationship between international soy demand and deforestation in Brazilian municipalities less clear. While table 3 does not prove that this is the case, it is a possible explanation for why there is no conclusive evidence of a link between the variables. Another possible culprit, pointing to a potential design failure of the models in this sub-sample, is the lackluster results from the first-stage regressions when controlling for time-fixed effects.

Model 2, which seemed to be the most promising model according to the previous land

use analysis, falls short in the first-stage instrumental variable regression in table 3. One plausible reason for why this is the case can be that the relationship between trade-weighted GDP and export volume became harder to estimate when the amount of observations was lowered from 9420 to 1218. As seen in the partial correlation analysis in figure 6, the correlation between trade-weighted GDP and trade volume is modest when one accounts for municipality-fixed and time-fixed effects. A large number of observations could therefore be of major assistance in identifying this variation. Model 1 is the only model with an F-statistic that exceeds Staiger and Stock's (1994) rule-of-thumb lower bound of 10. In other words, all other first-stage regressions point to the log of trade-weighted GDP being a weak instrument for the log of export volume, financial flows and prices of soy. As stated earlier, time-fixed effects are preferred when they are jointly significant since they explain aggregate trends that can be caused by other underlying variables. This means that model 2 is preferred over model 1, and the apparent statistical significance of the relationship between trade-weighted GDP and export volume in model 1 does not hold when controlling for time-fixed effects in model 2.

Another possible reason for why the first-stage regressions fail is that there are significant time-varying differences between the subset of municipalities that experience deforestation and the total group of 1570 municipalities used in the land use estimations. It could be that the relationship between trade-weighted GDP and export volume that holds in the full sample does not hold in the subset of municipalities that experience deforestation. For example, it could be that restrictive supply policies like the "Soy Moratorium" mentioned in section 2.2 has weakened the relationship between trade volume and the national income of importing countries in municipalities that experience deforestation. It might also be that importing countries are affected by an anti-deforestation sentiment. This could cause the potential for deforestation in the relevant municipalities to affect the importing behavior of soy-importing countries in such a way that the GDP of an importing country becomes a poor indicator for how much soy it wants to import from a specific municipality.

Finally, as seen in figure 5, deforestation was quite stable between 2013-2015. This is perhaps also reflected in the results of table 3, as only time-varying variables can be estimated well in a fixed effects model. The stagnation of deforestation also means that finding a relationship between the changes in soy exports and the changes in deforestation becomes challenging. The clear evidence of the soy industry growing coupled with the lack of changes in deforestation also lends strength to the hypothesis that soy expansion does not

cause deforestation to increase. In the end, I must conclude that the null hypothesis of there being no causal link between deforestation and growing soy demand cannot be rejected.

5.4 The Land Use of Domestically Consumed Soy

This subsection is an investigation of whether international soy demand affects the land use of domestically consumed soy. Analogously to forest cover, the land use for domestically consumed soy is expected to decrease when the land use for exported soy increases. This is because it represents an alternative use of land, making this another form of land use displacement relevant to research question 2. However, it might be the case that the expected effect is limited due to Brazil being its own main supplier of soybeans and that domestic demand has limited response to changes in international demand. The formulation of the null and alternative hypotheses is analogous to the previous subsections:

H_0 = There is no link between international soy demand and the land use of domestically consumed soy, and the elasticity coefficient between the variables is zero.

H_A = There is a significant relationship between soy demand and the land use of domestically consumed soy, and the elasticity coefficient between the variables is significantly different from zero.

Table 4 presents an estimation of the effect that international soy demand has on land use of soy destined for domestic consumption. The split between land use dedicated to exports and land use dedicated to domestic consumption was indicated by the Trase database (2018) based on total agricultural soy land use, yields in different municipalities in conjunction with the exported volume traced back to each municipality. The estimation in this table exclusively uses the acreage in hectares for soy farms producing for the domestic market as the land use variable, and can also be seen as the difference between the total land use of soy farms and the land use of exported soy. The models of table 4 test the correlation between exported trade volume, financial flows from soy exports or prices (in respectively model pairs 1 & 2, 3 & 4 and 5 & 6) and the land use of domestically consumed soy.

The 494 municipalities present in this sub-sample are necessarily the municipalities that both have soy production destined for domestic consumption and for exports in the years 2010-2015. Seeing as the number of municipalities with soy exports for every year 2010-2015

Table 4: Estimates from FEIV regressions of the log of land use of domestically consumed soy on log export volumes, financial flows, and prices of soy (instrumented by trade-weighted GDP) in Brazilian municipalities

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Volume	Volume 2	FF	FF 2	Prices	Prices 2
logexpvol	0.547***	-0.334				
	(0.133)	(1.156)				
logFF			0.537***	-0.352		
			(0.132)	(1.247)		
logprice					31.98	6.328
					(53.44)	(36.75)
Municipality-FE	Yes	Yes	Yes	Yes	Yes	Yes
Time-FE	No	Yes	No	Yes	No	Yes
First-stage	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
logwGDP	0.252***	0.0403	0.256***	0.0382	0.00430	-0.00213
	(0.0771)	(0.0348)	(0.0804)	(0.0351)	(0.00747)	(0.0117)
F (IV)	10.68	1.34	10.15	1.18	0.33	0.03
Observations	2964	2964	2964	2964	2964	2964

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This is a two-stage least squares model of land use dedicated to domestic soy consumption on instrumented export volume, financial flows and prices. The models include municipality-fixed effects for 494 Brazilian municipalities between 2010-2015. The first-stage results are reported under the second-stage results, with estimated coefficients for the first-stage elasticities. The fixed-effects instrumental variable setup is completely analogous to the deforestation and other land use models. The reported statistics are the F-statistics from the first-stage regressions and the number of observations. The number of observations is 2964 in all models, corresponding to 6 years of observations from 494 municipalities. Standard errors are clustered on the municipal level in all models.

numbered 1570, this reduction of viable observations underlines how export-oriented the Brazilian soy industry has become. Most soy-producing municipalities seem to only produce for the international market.

Similarly to table 3 presenting the deforestation results, table 4 provides lacking conclusive results about the role of soy demand on the land use of domestically consumed soy. Models 1 & 3, that do not account for time-fixed effects, are the only models where the trade-weighted GDP seems to estimate the soy demand variable, respectively export volume and financial flows, well enough for the F-statistic to exceed Staiger and Stock's (1994) rule of thumb value of 10. In other words, trade-weighted GDP is a weak instrument in this sub-sample and fails in models 2,4,5 and 6. Models 1 & 3, seem to predict a quite strong positive relationship between trade-weighted GDP and respectively soy export volume and financial flows and a strong positive relationship soy demand and land use for domestically consumed soy. However, when accounting for time-fixed effects in models 2 & 4, both the statistical significance in the first- and the second-stage estimations falls apart.

As in the earlier tables, I am mainly interested in the even-numbered models that account for time-fixed effects. Unfortunately, the statistical support for these models is very weak in table 4. The sign of the elasticity coefficients between export volume and land use and between financial flows and land use are negative as one would expect, but the standard errors are almost four times the size of the coefficients. This means that no real inference can be made. The instrument being weak in the first-stage regressions also undermines the validity of table 4's results.

Models 5 & 6 in table 4, using the log prices of soy shipments as the representation of soy demand, provide no results of interest.

In conclusion, table 4 presents no evidence of a causal relationship between international soy demand and the land use of domestically consumed soy, meaning that the null hypothesis of there being no such relationship cannot be rejected. A possible explanation for this non-finding is that the sample was too small for the trade-weighted GDP to fully function as an instrument for soy demand. Alternatively, there might be a more complicated link between land use for domestically consumed soy and export volumes because soy farms producing for domestic consumption are still soy farms identical to soy farms producing for export. There may also be measurement errors in the coupling of soy production to domestic consumption from the Trase database (2018). Another plausible explanation is that domestic demand is largely unaffected by the global demand of soy, and the size of the Brazilian soy industry may

make importing large quantities of soy slightly redundant. If this is the case, international demand may have limited effect on the production for domestic consumption other than positive externalities in the form of productivity boosts. An implication of this is that international demand is more likely to affect other crops meant for international consumption negatively as the increase in demand for one good is likely to accompany a shift in preference away from a substitute good.

5.5   The Land Use of Non-soy Agricultural Crops

In this subsection, I investigate whether there is a significant relationship between international soy demand and the land use of non-soy agricultural crops within Brazilian municipalities. The non-soy agricultural land use was calculated by adding together the acreage of several different crops in years 2010-2015, sourced from IBGE's PAM survey (SIDRA, 2018). The types of crops that were included in the non-soy land use were corn, rice, beans, pineapples, oranges, coconuts, passion fruit, coffee, cocoa, bananas, sugarcanes, tomatoes, papayas, limes and watermelons.

Moreover, this is the last subsection to cover research question 2 about the displacement effects of increasing soy demand through the expansion of soy farms, and the elasticity coefficient of interest is between the land use of non-soy crops and soy demand:

$H_0 =$ There is no link between international soy demand and the land use of non-soy agricultural crops, and the elasticity coefficient between the variables is zero.

$H_A =$ There is a significant relationship between soy demand and the land use of non-soy agricultural crops, meaning that the elasticity coefficient between the variables is significantly different from zero.

Table 5 presents the results from the first- and second-stage FEIV-regression of non-soy agricultural land use on soy demand instrumented by trade-weighted GDP. The reason why there are 9390 observations in this table is because 1565 out of the 1570 soy-producing municipalities in the full sample also produce at least one of the non-soy crops in all the years 2010-2015.

Comparing the first-stage regression results in table 5 with the first-stage regression in table 1, it is evident that they are close to identical. This is because it is almost exactly the

Table 5: Estimates from FEIV regressions of the log of land use of non-soy crops on log export volumes, financial flows, and prices of soy (instrumented by trade-weighted GDP) in Brazilian municipalities between 2010 and 2015.

	Model 1 Volume	Model 2 Volume 2	Model 3 FF	Model 4 FF 2	Model 5 Prices	Model 6 Prices 2
logexpvol	-0.169***	-0.322**				
	(0.0521)	(0.130)				
logFF			-0.172***	-0.448**		
			(0.0536)	(0.227)		
logprice					8.838	1.141***
					(15.98)	(0.382)
Municipality-FE	Yes	Yes	Yes	Yes	Yes	Yes
Time-FE	No	Yes	No	Yes	No	Yes
First-stage	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
logwGDP	0.0941***	0.0483***	0.0923***	0.0346**	-0.00180	-0.0136***
	(0.0177)	(0.0138)	(0.0181)	(0.0144)	(0.00319)	(0.00265)
F (IV)	28.37	12.25	25.92	5.77	0.319	26.33
Observations	9390	9390	9390	9390	9390	9390

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This is a two-stage least squares model of the log of land use of non-soy agricultural crops regressed on the instrumented log export volume, financial flows and prices. The models include municipality-fixed effects for 1565 Brazilian municipalities between 2010-2015. The first-stage results are reported under the second-stage results, with estimated coefficients for the first-stage elasticities. The fixed-effects instrumental variable setup is completely analogous to the deforestation and other land use models, with time-fixed effects added in even-numbered models. The reported statistics are the F-statistics from the first-stage regressions and the number of observations. The number of observations is 9390 in all models, corresponding to 6 years of observations from 1565 municipalities. Standard errors are clustered on the municipal level in all models.

same regression, only with 9390 observations instead of 9420. This is the reason why the first-stage coefficient estimates, standard errors and therefore also the F-statistics (being the squared ration between the coefficient estimate and the accompanying standard error) are almost exactly the same in the two tables.

The first-stage results for model 2 and 6 present the approximately the same relationship between trade-weighted GDP and respectively trade volume and prices as they did in table 1. Trade-weighted GDP is still positively correlated with trade volume and negatively correlated with prices, according to the first-stage regression. The results are highly significant, on the 1% level, indicating that the instrument works fairly well at estimating soy demand both represented by trade volume and prices. Like in table 1, the financial flows model fails the weak instrument test when controlling for time-fixed effects. The F-statistic of model 3, not controlling for time-fixed effects, looks promising. However, when controlling for time-fixed effects in model 4, the F-statistic falls to 5.77, which is well below Staiger and Stock's (1994) rule-of-thumb that an instrument should have a first-stage F-statistic above 10 in order to not be considered weak. As weak instruments can lead to bias in second-stage coefficient estimation even in large samples, this means that one needs to be careful when interpreting the second-stage results (Staiger and Stock, 1994).

The second-stage results in models 1 & 2 estimate a negative elasticity between export volume and the land use of non-soy crops. This was the expected result as soy exports increasing is correlated with land use for exported soy increasing, which should crowd-out or displace the land use dedicated to other crops. It is even possible that the evidence of a negative relationship between the land use of soy farming and other crops is part of the reason why the table 3 failed to establish a statistically significant relationship between soy demand and deforestation. The "leakage" or displacement caused by export-oriented soy farming expanding can replace either forests or other agricultural use, which means that the clear evidence of non-soy crops being replaced can be an explanation for why forests seemingly are not. It is possible that there exists an indirect effect of soy expansion on forest clearing through the displacement of other crops, but that the models used in this paper are not able to estimate this effect.

Model 2, controlling for time-fixed effects and still having a decent first-stage F-statistic, is the preferred trade volume model and estimates a negative elasticity between export volume and the land use of non-soy crops at −0.322. The result is statistically significant on the 5% level, indicating that the null hypothesis of there being no link between soy demand and

the land use of non-soy crops can be rejected. The interpretation of the negative elasticity coefficient is that a 1% increase in the export volume of soy leads to a 0.322% decrease in the land use of other crops. Comparing this result with table 2, it seems that a 1% increase in the export volume of soy leads to a larger increase in the land use of exported soy compared to the decrease in the land use of non-soy crops. This is also what one would expect, but it is important to remember that these are elasticities (measured in percentage-wise responses) and not absolute measures of area (measured in hectares). In order to know the relative effect with regards to areas, one needs to know the relative land use of exported soy to the land use of non-soy crops.
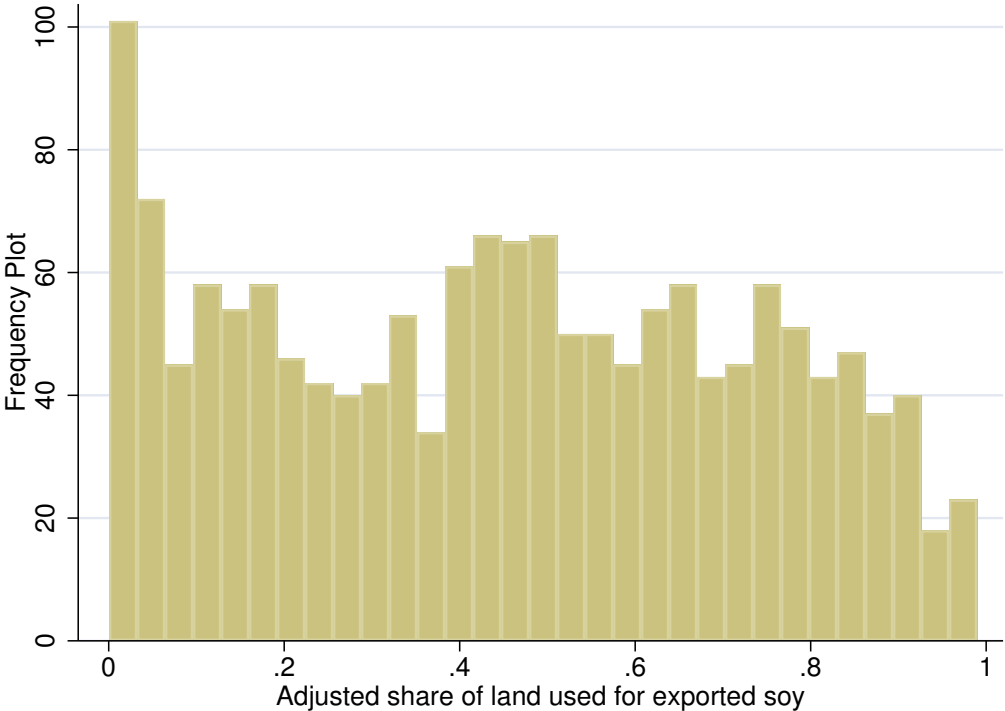
Figure 7: A frequency plot or histogram (with number/frequency of municipality-year observations in Y-axis) of the 2015 acreage of exported soy in Brazilian municipalities relative to the sum of the 2015 acreage of exported soy and the 2015 acreage of non-soy crops. Source: https://trase.earth/data

Figure 7 shows that the 2015 land use for exported soy is approximately the same as the land use for the sum of all these non-soy crops. This means that comparing the magnitudes of the elasticities in table 5 with the elasticities in table 2 directly is not a bad approximation

in terms of the actual area in hectares that both elasticities refer to. In other words, when model 2 in table 2 indicates that a 1% increase in soy exports leads to a 0.766% increase in land use of exported soy and model 2 in table 5 indicates that the same increase in soy exports leads to a 0.322% decrease in the land use of non-soy crops, the absolute increase in acreage for exported soy is expected to be a bit more than twice as large as the reduced acreage of non-soy crops. The exact ratio between the elasticity coefficients is $\frac{-0.322}{0.766} \approx 0.42$, indicating that about 42% (with a rather large uncertainty) of the positive land use response (in hectares) that the exported soy has with increasing soy demand results in displacement of agricultural land for non-soy crops. This might even be a somewhat conservative estimate, as the histogram in figure 7 is slightly left-heavy, indicating that the land use of non-soy crops in total and on average is slightly higher than the land use of exported soy. In subsection 5.7, the estimated displacement of non-soy crops is updated to account for about half of the land use expansion of soy when speaking in terms of absolute acreage[2].

The negative estimated elasticity between trade volume and the land use of non-soy crops, in conjunction with the positive relationship between trade-weighted GDP and trade-volume from the first-stage regression, indicates that there is a negative compound relationship between trade-weighted GDP (when instrumenting trade volume) and the land use of non-soy crops. Analogously to what was the case in the land use model for exported soy, the compound relationship between trade-weighted GDP and the land use of non-soy crops is identical in models 2, 4 and 6. In other words, multiplying the coefficients from the first-stage regressions with the coefficients from the second-stage regressions result in exactly the same compound effect of increases in trade-weighted GDP on the land use. This indicates that the models yield similar results on this point, no matter the choice of instrumented variable. The approximate compound effect in table 5 is $0.0483 \cdot (-0.322) \approx -0.0156$, meaning that a 1% increase in trade-weighted GDP leads to a 0.0483% increase in soy export volume which again leads to a 0.0156% decrease in the land use of non-soy crops.

As stated earlier, one needs to be careful when interpreting the second-stage coefficients of a two-stage least squares model with a weak instrument in the first-stage regression. However, the F-statistic of 5.77 that model 4 has, although falling below the rule-of-thumb boundary, is not so low that the second-stage results should be totally unreliable. This means that the estimated coefficient of -0.448, although potentially slightly biased, should not be

---

[2] $\frac{-0.322}{0.766} * \frac{24.5}{20.2} \approx -0.509$, with 24.5 and 20.2 being the total land use in million hectares of respectively non-soy crops and exported soy (in the dataset) as reported in the US-China analysis of subsection 5.7

a poor approximation of the true elasticity between financial flows and non-soy agricultural land use. It makes sense intuitively that the elasticity is negative, for the same reason that the elasticity of trade volume should be negative, because higher financial flows from soy exports is correlated with an expansion of land for exported soy, some of which will replace the acreage of other agricultural crops. The magnitude of the estimated elasticity coefficient also makes sense intuitively relative to the estimate from table 2, as the relationship between the table 2 elasticity and the table 5 elasticity is approximately the same as what is the case for the trade volume elasticities. For trade volume, the estimated elasticities were 0.766 and -0.322, the positive table 2 coefficient 2.379 times larger in magnitude. For financial flows, the estimated elasticities in table 2 and table 5 are respectively 1.068 and -0.448, meaning that the ratio between them is approximately 2.384. In other words, the relative increase in land use for exported soy in comparison to the decrease in land use for non-soy crops is fairly similar in response to changes in soy demand no matter if the instrumented variable is log trade volume or log financial flows.

The models using prices, models 5 & 6, are still mostly of interest because of their interpretation as the differentiating factor between the trade volume model and the financial flow model for soy demand. As in table 1, the first-stage regressions in table 5 show a negative correlation between trade-weighted GDP and prices, which is nonsense if one were to argue for trade-weighted GDP to causally affect prices. This means that the second-stage results, although seemingly highly significant, must face some doubt if one were to attempt an economic interpretation of the results. In the case of soy prices and alternative land use, it is indeed possible that higher soy prices lead to non-soy crops being in relative demand, leading to expansion of non-soy agriculture. However, I do think that the first-stage regressions mostly reflect the prices and quantities being inversely related in matters of demand, leading to the second-stage model for prices giving little grounds for causal inference about the isolated effect of changes in price. Intuitively, it is not the soy price decreasing that leads to an expansion of soy agriculture and displacement of non-soy crops. The natural order of events is that the first mover of demand is the international soy market demanding a higher quantity from the Brazilian soy industry and that the land use of exported soy must expand to meet this demand. This expansion again happens at a cost, namely that alternative uses of land, such as acreage for non-soy crops, are displaced.

To summarize, the preferred model, model 2, indicates that there is a significant and negative relationship between the land use of non-soy agricultural crops and international

soy demand. This means that the null hypothesis of there being no link between international soy demand and the land use of non-soy crops can be rejected. This was the *a priori* expected result, and it confirms one of the central results from the study by Boerema *et al.* (2016) - the expansion of the soy industry has directly led to the displacement of other agricultural goods.

5.6 The Role of Asian Demand

As seen and discussed in the background section, Asia, and China in particular, play a major role in the international trade of Brazilian soy. Looking at figures 1 and 2 in subsection 2.1, it is natural to believe that soy demand is largely driven by the Asian countries. In this subsection, I will motivate the importance of Asian demand for the Brazilian soy industry before discussing the potential outcomes of suddenly increasing Asian demand due to trade tensions between China and the United States in subsection 5.7.

Table 6 presents the land use of exported soy regressed on the instrumented trade volume with time-fixed effects. The difference between the models, and what is new compared to the analysis in subsection 5.2, is that I have tried to decompose the trade-weighted GDP into Asian and non-Asian trade-weighted GDP. By doing this, I aim to find econometric evidence for the Asian countries affecting the soy demand more in the 2010-2015 period compared to the non-Asian countries. Table 6 presents compelling evidence that this is the case.

First of all, the instruments are now altered. I constructed a trade-weighted GDP for Asian countries with weights based on the share of exports to Asia that every Asian country imports from every Brazilian municipality. Analogously for the non-Asian trade-weighted GDP, the weights are now calculated by dividing the country-specific imports (for non-Asian countries) on the total imports by non-Asian countries for every Brazilian municipality. In essence, I constructed a dummy denoting whether a country is Asian or not and estimated the Asian imports by multiplying this dummy with all municipality-country trade-flows. The non-Asian imports are necessarily the difference between total imports by the international market and the Asian imports.

The first-stage regressions are quite telling. Only the changes in the Asian trade-weighted GDP seem to lead to changes in the Brazilian export volume, and the relationship is positive as one would expect. The near-zero estimated effect of non-Asian trade-weighted GDP indicates that growing Asian demand is the dominant reason for Brazilian export volumes

Table 6: Estimates from FEIV regressions of the log of land use of exported soy on the international soy demand in log trade volume instrumented by the log of Asian and Non-Asian trade-weighted GDP.

	Double Instrument	Only Asia	Only non-Asia
Second-Stage	Land Use 1	Land Use 2	Land Use 3
logexpvol	0.565***	0.582***	0.243
	(0.128)	(0.129)	(0.643)
First-stage	Trade Volume 1	Trade Volume 2	Trade Volume 3
log(Asian wGDP)	0.0101***	0.0104***	
	(0.0030)	(0.0029)	
log(non-Asian wGDP)	-0.0006		-0.0010
	(0.0009)		(0.0009)
Municipality-FE	Yes	Yes	Yes
Time-FE	Yes	Yes	Yes
F (IV)	6.80	12.69	1.14
Observations	9420	9420	9420

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This is a two-stage least squares model of the log of land use of exported soy regressed on the instrumented log export volume in Brazilian municipalities. The instruments are the varying factor between the models. The first model uses both Asian-trade-weighted GDP and non-Asian-trade-weighted GDP to estimate export volume, while the second model uses only the Asian trade-weighted GDP and the third model uses only the non-Asian trade-weighted GDP. Weights are the initial share of the Asian or non-Asian imports from a municipality. The models include municipality-fixed effects for 1570 Brazilian municipalities between 2010-2015 and all models control for time-fixed effects. The first-stage results are reported under the second-stage results, with estimated coefficients for the first-stage elasticities. The reported statistics are the F-statistics from the first-stage regressions and the number of observations. The number of observations is 9420 in all models, corresponding to 6 years of observations from 1570 municipalities. Standard errors are clustered on the municipal level in all models.

increasing. The first-two models, with and without the non-Asian trade-weighted GDP, are quite similar in their estimates. However, the second model's higher F-statistic seems to indicate that the Asian trade-weighted GDP works at least as well when it is the only instrument present.

The second-stage results make intuitive sense as they, like table 2, indicate a quite strong positive relationship between the export volume of soy and the land use of exported soy. The estimated elasticity coefficients are slightly lower, which is slightly puzzling, but this model was mostly made to point out the heterogeneity between Asian and Non-Asian importers in terms of their demand for soy. From the lack of any significant relationship between non-Asian trade-weighted GDP and the Brazilian exports, as well the failure of Brazilian exports to explain land use expansion in the third model, it is evident that Asian demand is the main determinant causing Brazilian export numbers to increase and the related land use to expand.

To get the full sample of 1570 soy-producing municipalities, I log-transformed region-specific trade-weighted GDP plus one in order to make missing log-values into zeros. The problem was that not every municipality has both Asian and Non-Asian importers. However, these zeros are true zeros in the sense that they are based on the actual trade flows and reflect the fact that there either are no Asian or no non-Asian importers that import from the municipality. As the municipality obviously exports for all years 2010-2015 since the export volumes are non-zero for all years, there must be either only Asian importers, only non-Asian importers or a combination of Asian and Non-Asian importers that import from every municipality. Letting one of the instruments be zero should reflect the reality that there are no importers from the specific region in the relevant time frame even though every importing country has the opportunity to import from every municipality. Additionally, the $\log(x+1)$ transformation is a good approximation of the logarithm of a variable when the values of the variable are large. This is definitely the case for national incomes in real 2005 US dollars, meaning that the transformation should be relatively unproblematic.

5.7 The Effects of a US-China Trade War on Exports and Land Use in Brazil

In the 2016-2017 marketing year, the United States and Brazil were the two main exporters of soybeans to China. The United States Department of Agriculture (2018b) estimates that

39.4% of Chinese soybean imports (36.84 million tons) were imported from the US and 48.5% (45.34 million tons) were imported from Brazil. If China were to reduce their demand for US soy by implementing a sizable tariff or encourage Chinese traders to stop all imports of American soy, the effects on the Brazilian soy industry are likely to be of substantial magnitude.

It may be possible to measure the projected "lost trade" between the United States and China as a positive shock to Chinese demand for non-American soy. The problem is that the relationship between trade-weighted GDP and a more direct demand shock for soy is not that obvious, making the first-stage model in this paper's econometric analysis hard to apply to this scenario. The "freed" capital from importing less or stopping all imports from the US would be a very small share of total Chinese GDP, but a significant share of yearly Chinese soy expenditure. It is probably more reasonable to view the increased import demand from China as an exogenous shift in demand for Brazilian soy. A very rough approximation of the decrease in soybean imports from America is that they will be replaced by the other exporters in proportion to their share of the Chinese imports that are not from the United States. In the short-run, this is probably not feasible if the increased demand is too high, but the Brazilian capacity to increase exports is quite high as seen from macro-level evidence in figure 2. The United States Department of Agriculture (2018a) states that Brazil was able to increase their exports by about 10 million tons between marketing year 2015-2016 and marketing year 2016-2017. This implies that Brazil might be able to absorb a significant share of growing Chinese demand if a share or even all of the American 36.84 million tons of soy exports to China cease to be traded. Over a few years, not to say a decade, Brazil may be able increase their production capacity enough to meet the entirety of Chinese demand that the US does not supply.

Quick calculations based on data from the United States Department of Agriculture (2018b) show that out of all Chinese soybean imports not from the United states, 80% are imported from Brazil[3]. The report from the United States Department of Agriculture (2018b) on the Chinese agricultural industry and import needs also states that China has limited ability to expand its domestic production of soybeans and that demand for imports is growing at a faster rate than China's production capabilities. This indicates that almost all of the soy imports "lost" if China-US relations sour must be replaced by the international export market for soybeans. This is of course where Brazil and other South American countries can

---

[3]Calculated by import shares: $\frac{48.5}{100-39.4} \approx 80\%$

rise up to meet Chinese soy demand. If going by the rough approximation of non-American export shares to China, Brazil stands to meet an increased yearly import demand of up to 31.5 million tons if American soy exports to China cease[4]. As stated above, it is unlikely that Brazil alone can expand its soy export to meet an increased demand of this magnitude in the short term. On the other hand, other South American actors, or for example Canada, may be able to increase their share of the Chinese imports relative to Brazil. In other words, Brazil may quickly end up supplying less than 80% of the non-American Chinese soy imports.

It is likely that an asymmetric tariff made by China will have multiple strategy-adjusting effects for the actors in the international soy market. All actors would necessarily have to adjust their both their importing and exporting strategy if such a tariff were to be implemented. A Chinese tariff of 25% on American soy would, first of all, very likely lead to a major reduction in prices for American soy. This would simultaneously make American soy relatively more attractive for non-Chinese soy importers and South American soy relatively more attractive for Chinese importers (because the American soybean price will not decrease enough to fully counteract the 25% tariff). The implication of this is that a share of the non-Chinese importers of Brazilian soy will choose to import the now cheaper American soy. According to the United States Department of Agriculture (2018a) report on the Brazilian soy industry, the total non-Chinese import volume of soybeans was estimated to be about 17 million tons in the marketing year 2016-2017 out of a total export volume of 68.3 million tons. One of the results of a major Chinese tariff on American soy would then be an even more dominant position of China as an importer of Brazilian soy, since China shifts more of its demand towards Brazil and other importers shift some of their demand towards the US.

Any estimate I make of the actual demand effect a Chinese-US tariff would have on Brazilian exports would be an educated guess because of the numerous interaction and feedbacks in the system. It is even quite plausible that Brazil starts to import the cheaper US soybeans for domestic consumption in order to better meet the now higher Chinese demand. A reasonable outcome, or an illustrative example, of what might happen in the event of a sizable Chinese tariff on US soybeans, is that yearly Chinese imports of US soybeans are about halved. This means that China, in addition to their already growing import demand, would have to import about 18.4 additional million tons of soybeans from non-US sources in order to meet the consumption demand of the Chinese population and soy crushing industry. If the effect on other exporting countries is proportionate to the

---

[4]Calculation:$39.4 \cdot 0.8 \approx 31.5$

current importing composition of China, Brazil stands to meet an increased soy demand of about 15.75 million tons in this scenario. However, due to American soybean prices decreasing because of their excess supply, non-Chinese importers may increasingly choose to import American soy. Out of the 17 million tons Brazilian soybeans imported by non-Chinese importers, perhaps about 4-5 million tons of soy imports will be shifted to the now cheaper American soy. This means that Brazil now has an excess yearly import demand of 10-11 million tons of soybeans. This increased demand is, as evidenced by the extent of the increased production and export volumes of recent years, achievable for Brazil to meet over the course of a year.

Taking the illustrative example further - what would happen to land use in Brazil if met with an incrementally increased soybean demand of 10 million tons? Here, I can apply the econometric results from section 5 in this paper. First, I need to estimate what percentage increase in demand 10 million tons represents before utilizing the estimated elasticity coefficients to discuss the land use effects. Since estimated soybean exports for Brazil were 68.3 million tons in 2016-2017 according to the United States Department of Agriculture (2018a), an increase of 10 million tons constitutes about 14.6%. Using the significant elasticities between export volume and land use found in table 2 and table 5, my econometric analysis indicates that this will increase the land use for exported soy in Brazil by 11.2% and decrease the land use for non-soy crops by 4.7%. The Trase data (2018) indicates that total land use for exported soy was 20.2 million hectares in 2015, which must serve as the most recent estimate. An increase of 11.2% indicates an expansion of 2.26 million hectares of land used for producing soy for exports. A decrease of 4.7% in the land use for non-soy crops, which in the soy-producing municipalities have a total of 24.5 million hectares, indicates a contraction of 1.15 million hectares in the relevant municipalities covered in this paper.

If land use for exported increases by 2.26 million hectares and the land use for non-soy crops decrease by 1.15 million hectares, there are now 1.11 million hectares of estimated land expansion unaccounted for. This needs to replace some alternative use of land, but no causal link has been made between soy exports and land use for domestically consumed soy or between soy exports and deforestation. The answer is probably land use for pastures and livestock. As Boerema *et al.* (2016) state in their paper, macro-level data indicate that the expansion of the soy industry has happened at the expense of land used for pastures. In addition, the displacement of the livestock industry may be a direct cause for deforestation (Tyukavina *et al.*, 2017). In other words, the increase of Asian demand for soy may well

lead to the displacement of pastures of about 1 million hectares, which may indirectly lead to forests being replaced for pastures. Since I do not have data for the municipal land use of pastures, this is mostly based on conclusions in earlier studies. However, it is not unlikely that soybeans indirectly affect deforestation through the displacement of other agricultural goods, and livestock look likely to be the "missing link" of this study. Table 8 in appendix A.2 motivates a link between cattle pastures and deforestation in Brazilian municipalities, meaning that a negative elasticity between soy demand and the land use of cattle pastures, if those data were available on the municipal level and a significant relationship were found, would indicate that soy demand indirectly causes deforestation. If it indeed is the case that increased soy demand affects deforestation positively through the displacement of pastures, a Chinese tariff on US soy is highly likely to indirectly cause increased deforestation in Brazilian forests.

6   Discussion of External Validity

External validity, and the extent that the results of a study have generalizability, is a central aspect of a study. In this paper, the sample of soy data for the Brazilian soy industry and municipal trade flows are quite complete for the period 2010-2015. This means that the sample is a good approximation of the target population of soy-producing Brazilian municipalities in the years 2010-2015. The results concerning the effects of expanding the Brazilian soy export industry should then be valid and precise within this time frame if the statistical models are correctly constructed and estimators are unbiased. What I hope, and what can lend meaning to confidence intervals and other measures of uncertainty, is that the results are generalizable to a larger target population that this smaller sample can represent. In the case of this paper, the obvious and smallest degree of generalizability hoped for is that the results can be viable for the land use effects, in Brazil, of international soy demand outside of the 2010-2015 time frame. Other than this, it is possible that the land use response to increasing international soy demand is fairly representative for what land use responses would be for other crops in Brazil facing increasing demand from the international market. It is also possible that other nations with large agricultural industries would behave somewhat similarly when faced with increased international demand of an agricultural good. The decomposition of land use displacement is of course dependent on what kinds of land use are already present in a country. A country with no forests will of course not experience any

deforestation no matter how much the acreage of an agricultural crop expands. However, in the dynamics of displacement there are some elements that should be universal. For instance, the expansion of acreage due to increased international demand should lead to the decrease of acreage for alternative purposes, and this negative externality should be spread out over many types of land use.

6.1   Limitations of the Study

One of the obvious drawbacks of this study is that there are no municipal data for the land use of pastures. According to recent papers in the literature, such as Boerema *et al.* (2016), the expansion of the soy industry seems to have largely replaced pastoral land. This, of course, indicates that there should be a negative elasticity between instrumented soy exports and pastoral land use that my models cannot estimate due to lack of data. Additionally, Tyukavina *et al.* (2017) find that agroindustrial clearing of forests for pastures was the main contributor to deforestation in the Amazon forest. The findings in appendix section A.2 also corroborate the literature in this sense, indicating that deforestation and cattle farming in Brazilian municipalities are positively linked. This indicates that proper data on the municipal land use of pastures may have been able to both provide more information about the decomposition of land use displacement that increasing soy demand causes and about the indirect effects of soy expansion on deforestation. In addition to this, the Trase database (2018) is expanding to map the international trade flows of the Brazilian livestock industry as well, which could be valuable for the creation of a more complete demand-side model for deforestation in future studies.

Another potential pitfall of the study is also the limited variation seen in trade-weighted GDP. Although the instrument succeeds at estimating soy demand in the larger samples, the relationship does not hold for all parts of the analysis - notably when estimating soy demand in municipalities experiencing deforestation. Bound *et al.* (1995) state that when the correlation between an instrument and an endogenous variable is weak, the following second-stage regression will provide inconsistent and biased estimates even for large samples. It might be that another instrumental variable would have performed better, like the trade-weighted world population or trade-weighted tariff levels of the importing countries. However, these variables might share the same weakness as trade-weighted GDP by having limited variation, at least in a time-frame as short as six years. Further studies would need

to look into possible alternative instruments for soy demand.

7   Conclusion

The main objective of this paper was to analyze the effects of international soy demand on the land use in Brazilian municipalities. From the macro-level analysis in section 2 as well as the econometric analysis in section 5, it is evident that both international soy demand, the export volumes of soy and the land used for exported soy has grown significantly during the 2010-2015 time frame. It is also quite clear that growing demand in Asia, chiefly in China, is the main determinant leading the Brazilian soy export industry to expand.

The econometric models in this paper found a high and positive correlation between soy demand and the land use of exported soy. The estimated elasticity coefficient between the land use of exported and trade volume was estimated to be 0.766, indicating that land use is highly elastic to quantity demand. It also indirectly implies that productivity increases slightly with soy demand, as discussed in subsection 2.1, and the implied elasticity of productivity to changes in export volume is about 0.23.

The two main models for soy demand, using export volume and financial flows respectively, consistently returned similar, but somewhat different results. For example, the model for financial flows also indicated a positive relationship between soy demand and the land use of exported soy. However, this analysis was a bit weaker as the first-stage regression failed the F-test for weak instruments. As argued in the analysis, the discrepancy between the export volume model and the financial flows model must be caused by soy prices affecting the financial flows, and the failure of first-stage regressions indicates that the corresponding second-stage results are unreliable.

The main finding concerning the displacement of land use as a result of the soy industry expanding was that there is clear evidence for a substantial negative relationship between international soy demand and the land use of non-soy agricultural crops. This was the *a priori* expected finding since soy crops and non-soy crops compete for the use of arable land. The first-stage regressions indicate a positive and significant relationship between trade-weighted GDP and the different soy demand measures, while the second-stage results indicate a negative relationship between export volume of soy and the land use of non-soy crops, with an estimated elasticity of -0.322. This indicates that the displacement of non-soy agricultural crops may account for about half of the land displaced by increasing soy

demand. This is due to a combination of the relative elasticities, 0.322 being about 42% of 0.766, and the fact that the land use of non-soy crops in total is slightly higher than the land use of exported soy in the soy-producing municipalities relevant to this paper's analysis.

From the analysis in this paper, there is no conclusive evidence that increases in international soy demand leads to increased deforestation. The estimated coefficients in the second-stage regression were in the direction one would expect, deforestation increasing with soy demand, but the first-stage regressions indicated that trade-weighted GDP fails as an instrument for soy demand in the relevant sub-sample and no statistically significant results were found in the second-stage regressions.

The effect of soy expansion on deforestation, if it exists, could potentially be indirect through the displacement of the livestock industry. There are no land use data available for pastures on the municipal level in Brazil for now, but there seems to be evidence both in the literature and in the basic econometric approach taken in the appendix of this paper of a significant positive relationship between the size of the livestock industry and deforestation in Brazilian municipalities. The lack of municipal land use data may only be a temporary problem as the Brazilian Bureau of Statistics continuously improves data collection on different agricultural industries. This means that future studies might be able to investigate the effect of soy demand on pasture land and through this the indirect effect of soy demand on deforestation.

Another possibility is that deforestation and soy demand truly are not linked in Brazilian municipalities. A possible explanation is that the link between soy farming and deforestation has weakened in recent years due to stricter regulations regarding where soy farms can expand to as well as societal pressure due to environmental concern. The "Soy Moratorium" pact of 2006, restricting expansion into the Amazon biome, is an example of a policy that might have had a lasting impact on the relationship between soy farming and deforestation (Gibbs *et al.*, 2015). The costs of expanding soy farming land into forests may have increased for soy producers due to regulatory policies and increased social costs, meaning that other alternative uses of land are less costly and thus preferable to displace for soy farmers.

I find no statistically significant link between land use for domestically consumed soy and international soy demand either. One would perhaps expect a small negative correlation due to displacement effects, but while the estimated coefficients were in the *a priori* expected direction when controlling for time-fixed effects, the standard errors of the estimates were too large to reject the null hypothesis that there is no link between domestic soy land use

and international soy demand.

In total, this means that the only statistically significant effects of soy demand on land use is the positive effect of increasing demand on the land expansion of exported soy and the negative effect on the land use of non-soy agricultural crops. Since the estimated displacement of non-soy crops is significantly smaller in magnitude compared to the land expansion of soy, there are still components of the total land displacement effect that must be accounted for. In addition to a productivity increase, which there is implied evidence for, the soy demand must also have caused either deforestation, the reallocation of soy production for domestic consumption or the displacement of unobserved types of land use like pasture land and secondary forests. As there is no statistical evidence for soy expansion directly causing deforestation or the reallocation of soy production, the most likely "missing link" of displacement seems to be the livestock industry. In addition to the non-findings for the observed land use types, the existing literature consistently suggests that the expanding soy industry largely has displaced pasture land.

Because of China importing more than half of Brazilian soy exports, I specifically analyze the role of growing Asian demand and China as a dominant actor in the global soybean trade. Econometric analysis confirms that growth in Asian demand is the primary factor leading to recent increases in Brazilian soy exports. In fact, no other regions than Asia seem to have a growing demand for Brazilian soy in the 2010-2015 time frame.

In light of the rising trade tensions between the United States and China, I qualitatively analyze the effects that such a trade-war might have for the Brazilian soy industry and the usage of land. The main conclusion is that Brazil will face a considerably higher international demand due to China's pivotal role in the global soy trade. This higher demand will result in an expansion of land dedicated to exported soy, likely in the millions of hectares. As this land needs to replace other alternative uses of land, land use for non-soy crops will likely decrease significantly. As there is no causal link found between soy exports and deforestation or land use for domestic soy, no strong claims can be made for these effects. It is, however, likely that land use for the livestock industry will be affected, but lacking data makes this conclusion a combination of an appeal to the literature and to the process of elimination.

When, or if, a trade war between the United States and China has taken place and a few years have gone by, it will be of major research interest to examine the effects of this trade war on the Brazilian soy industry and on deforestation. If municipal land use data is available for the livestock industry, it may be possible to measure an indirect effect of the

tariff on deforestation through increasing soy demand potentially causing pasture land to be displaced into forested land.

On an ending note, the Trase (2018) data on the level of importing and exporting companies and ports could be the basis of interesting and detailed analysis for future studies. For example, it is feasible to test the effect of PR-campaigns, laws and zero deforestation commitments on the behavior of firms in terms of exporting strategy and practices leading to deforestation. There is unquestionably a multitude of possible research themes using municipality-, country- and company-level data on the Brazilian soy industry, and it will be quite interesting to observe how the field develops in the near future - both in terms of investigating the determinants of deforestation and in terms of gaining a better understanding of important agricultural industries like the Brazilian soy industry and how developments in international trade may affect it.

References

ABADIE, A., ATHEY, S., IMBENS, G. W. and WOOLDRIDGE, J. (2017). When Should You Adjust Standard Errors for Clustering? NBER Working Paper No. 24003.

ACEMOGLU, D., JOHNSON, S., ROBINSON, J. A. and YARED, P. (2008). Income and Democracy. *American Economic Review*, **98** (3), 808–42.

ANDERSEN, L. E., GRANGER, C. W., REIS, E. J., WEINHOLD, D. and WUNDER, S. (2002). *The Dynamics of Deforestation and Economic Growth in the Brazilian Amazon*. Cambridge, UK: Cambridge University Press.

ANGRIST, J. and KRUEGER, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic perspectives*, **15** (4), 69–85.

— and PISCHKE, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. New Jersey: Princeton University Press.

ANTRAS, P. (2004). Is the US Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution. *Contributions in Macroeconomics*, **4** (1), `https://doi.org/10.2202/1534-6005.1161`.

ARORA, M. V. B. and VAMVAKIDIS, M. A. (2004). *How Much Do Trading Partners Matter for Economic Growth?* IMF Working Paper No. 04/26, International Monetary Fund.

BARR, K. J., BABCOCK, B. A., CARRIQUIRY, M. A., NASSAR, A. M. and HARFUCH, L. (2011). Agricultural Land Elasticities in the United States and Brazil. *Applied Economic Perspectives and Policy*, **33** (3), 449–462.

BERRY, S. and SCHLENKER, W. (2011). Technical Report for the ICCT: Empirical Evidence on Crop Yield Elasticities. *International Council on Clean Transportation*.

BINSWANGER, H. P. (1991). Brazilian Policies that Encourage Deforestation in the Amazon. *World Development*, **19** (7), 821 – 829, doi:10.1016/0305-750X(91)90135-5.

BOEREMA, A., PEETERS, A., SWOLFS, S., VANDEVENNE, F., JACOBS, S., STAES, J. and MEIRE, P. (2016). Soybean Trade: Balancing Environmental and Socio-Economic Impacts of an Intercontinental Market. *PloS one*, **11** (5), doi:10.1371/journal.pone.0155222.

BOLLING, C., SOMWARU, A., KRUSE, J. B. *et al.* (2001). The United States in the Global Soybean Market: Where Do We Go from here? Presented at the *American Agricultural Economics Association Meeting*, Chicago, IL, August 5.-8., 2001.

BOUND, J., JAEGER, D. A. and BAKER, R. M. (1995). Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak. *Journal of the American Statistical Association*, **90** (430), 443–450.

BRANDER, J. A. and TAYLOR, M. S. (1998). Open Access Renewable Resources: Trade and Trade Policy in a Two-country Model. *Journal of International Economics*, **44** (2), 181–209.

BRANNSTROM, C., RAUSCH, L., BROWN, J. C., DE ANDRADE, R. M. T. and MICCOLIS, A. (2012). Compliance and Market Exclusion in Brazilian Agriculture: Analysis and Implications for Soft Governance. *Land use policy*, **29** (2), 357–366.

CANG, A., DURISIN, M. and PARKER, M. (2018). China Is Already Canceling U.S. Soybean Orders. `https://www.bloomberg.com/news/articles/2018-05-04/china-is-already-canceling-u-s-soybean-orders`. *Bloomberg*, published on: 05-04-2018 and accessed on 06-01-2018.

CHAMPOUX, V., DURGEE, J. and MCGLYNN, L. (2012). Corporate Facebook pages: When Fans Attack. *Journal of Business Strategy*, **33** (2), 22–30.

CLARK, T. S. and LINZER, D. A. (2015). Should I Use Fixed or Random Effects? *Political Science Research and Methods*, **3** (2), 399–408.

COPELAND, B. R. and TAYLOR, M. S. (2009). Trade, Tragedy, and the Commons. *American Economic Review*, **99** (3), 725–49.

DIXIT, A. (1984). International Trade Policy for Oligopolistic Industries. *The Economic Journal*, **94**, 1–16.

DONNAN, S. and HANCOCK, T. (2018). China retaliates against US tariffs on $50bn of imports. `https://www.ft.com/content/a8c670b8-7094-11e8-92d3-6c13e5c92914`. *Financial Times*, published on: 06-15-2018 and accessed on 06-16-2018.

EISENHART, C. (1947). The Assumptions Underlying the Analysis of Variance. *Biometrics*, **3** (1), 1–21.

FARIA, W. R. and ALMEIDA, A. N. (2016). Relationship between Openness to Trade and Deforestation: Empirical Evidence from the Brazilian Amazon. *Ecological Economics*, **121**, 85–97.

GIBBS, H. K., RAUSCH, L., MUNGER, J., SCHELLY, I., MORTON, D. C., NOOJIPADY, P., SOARES-FILHO, B., BARRETO, P., MICOL, L. and WALKER, N. F. (2015). Brazilquoterights Soy Moratorium. *Science*, **347** (6220), 377–378.

GOLDBERGER, A. S. (1972). Structural Equation Methods in the Social Sciences. *Econometrica*, **40** (6), 979–1001.

HANCOCK, T. (2018). China to impose duties on \$34bn of US imports from July 6. `https://www.ft.com/content/e4ac35be-70ca-11e8-852d-d8b934ff5ffa`. *Financial Times*, published on: 06-15-2018 and accessed on 06-16-2018.

HANSEN, C. B. (2007). Generalized Least Squares Inference in Panel and Multilevel Models with Serial Correlation and Fixed Effects. *Journal of Econometrics*, **140** (2), 670–694.

HARDING, T., HERZBERG, J. and KURALBAYEVA, K. (2018). Environmental Regulation and Commodity Prices: Evidence from Deforestation in Brazil.

HARRIS, R. and TZAVALIS, E. (1999). Inference for Unit Roots in Dynamic Panels Where the Time Dimension is Fixed. *Journal of Econometrics*, **91** (2), 201–226.

HARSTAD, B. and LISKI, M. (2013). Games and Resources. *Encyclopedia of Energy, Natural Resource, and Environmental Economics*, **2**, 299–308.

— and MIDEKSA, T. K. (2017). Conservation Contracts and Political Regimes. *The Review of Economic Studies*.

HAUSMAN, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, **46** (6), 1251–1271.

KASTENS, J. H., BROWN, J. C., COUTINHO, A. C., BISHOP, C. R. and ESQUERDO, J. C. D. (2017). Soy Moratorium Impacts on Soybean and Deforestation Dynamics in Mato Grosso, Brazil. *PloS one*, **12** (4), `https://doi.org/10.1371/journal.pone.0176168`.

LAMBIN, E. F. and MEYFROIDT, P. (2011). Global Land Use Change, Economic Globalization, and the Looming Land Scarcity. *Proceedings of the National Academy of Sciences*, **108** (9), 3465–3472.

Margulis, S. (2004). Causes of Deforestation of the Brazilian Amazon. World Bank Working Paper; No. 22. Washington, DC: World Bank. `https://openknowledge.worldbank.org/handle/10986/15060`.

Meredith, S. (2018). China announces new tariffs on 106 US products, including soy, cars and chemicals. `https://www.cnbc.com/2018/04/04/china-new-us-tariffs-including-soy-cars-and-chemicals.html`. *CNBC*, published on: 04-04-2018 and accessed on 06-01-2018.

Moran, E. F. (1993). Deforestation and Land Use in the Brazilian Amazon. *Human Ecology*, **21** (1), 1–21.

Robalino, J. and Herrera, L. D. (2010). *Trade and deforestation: A Literature Review*. Tech. rep., WTO Staff Working Paper, ERSD-2010-04.

Rondinelli, D. A. and Berry, M. A. (2000). Environmental Citizenship in Multinational Corporations: Social Responsibility and Sustainable Development. *European Management Journal*, **18** (1), 70–84.

Saatchi, S. S., Soares, J. V. and Alves, D. S. (1997). Mapping Deforestation and Land Use in Amazon Rainforest by Using SIR-C Imagery. *Remote Sensing of Environment*, **59** (2), 191–202.

Schnepf, R. D., Dohlman, E. N. and Bolling, H. C. (2001). Agriculture in Brazil and Argentina: Developments and Prospects for Major Field Crops. *International Agriculture and Trade Outlook*, No. (WRS-013) 85 pp, United States Department of Agriculture, Economic Research Service, `https://www.ers.usda.gov/webdocs/publications/40339/15081_wrs013_1_.pdf?v=41057`.

Sherman, N. (2018). Who is losing out from Trump's tariffs? `http://www.bbc.com/news/business-43731798`. Published on: 04-13-2018 and accessed on 06-01-2018.

SIDRA (2018). `https://sidra.ibge.gov.br/tabela/5457`. Accessed: 2018-03-25.

Skole, D. L., Chomentowski, W., Salas, W. A. and Nobre, A. D. (1994). Physical and Human Dimensions of Deforestation in Amazonia. *BioScience*, **44** (5), 314–322.

Staiger, D. O. and Stock, J. H. (1994). Instrumental Variables Regression with Weak Instruments. NBER Technical Working Paper No. 151. Later published in *Econometrica*, 65(3), 557-586, in 1997.

STATACORP (2018). Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.

STOCK, J. H. and TREBBI, F. (2003). Retrospectives: Who Invented Instrumental Variable Regression? *Journal of Economic Perspectives*, **17** (3), 177–194.

— and WATSON, M. W. (2008). Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica*, **76** (1), 155–174.

TRASE DATABASE (2017). Trase data sourceshttps://trase.earth. Accessed: 2018-01-20.

TRASE DATABASE (2018). https://trase.earth. Accessed: 2018-01-20.

TYUKAVINA, A., HANSEN, M. C., POTAPOV, P. V., STEHMAN, S. V., SMITH-RODRIGUEZ, K., OKPA, C. and AGUILAR, R. (2017). Types and Rates of Forest Disturbance in Brazilian Legal Amazon, 2000-2013. *Science Advances*, **3** (4), e1601047.

UNITED STATES DEPARTMENT OF AGRICULTURE (2018a). Brazil: Oilseeds and Products Annual. https://www.fas.usda.gov/data/{Brazil}-oilseeds-and-products-update-16. Accessed: 2018-06-12.

UNITED STATES DEPARTMENT OF AGRICULTURE (2018b). China: Oilseeds and products annual. https://www.fas.usda.gov/data/china-oilseeds-and-products-annual-3. Accessed: 2018-05-30.

UNITED STATES DEPARTMENT OF AGRICULTURE (2018c). USDA Agricultural Projections to 2027. https://www.ers.usda.gov/webdocs/publications/87459/oce-2018-1.pdf?v=43146. Accessed on: 2018-05-31.

VERBURG, R., RODRIGUES FILHO, S., LINDOSO, D., S. DEBORTOLI, N., LITRE, G. and BURSZTYN, M. (2014). The Impact of Commodity Price and Conservation Policy Scenarios on Deforestation and Agricultural Land Use in a Frontier Area within the Amazon. *Land Use Policy*, **37**, 1426.

WOOLDRIDGE, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT press.

— (2015). *Introductory Econometrics: A Modern Approach*. Nelson Education.

WORLD BANK (2017). World Development Indicators. http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators#. Accessed: 2018-01-14.

WRIGHT, P. G. (1928). *Tariff on Animal and Vegetable Oils*. New York: MacMillan.

Appendices

A   Supplementary Analysis

A.1   Testing the Effect of Asian Imports on Deforestation

In this appendix subsection, I try to split up the Brazilian exports of soy products into Asian and Non-Asian imports, each instrumented by respectively the trade-weighted GDP of Asian and Non-Asian countries (the instruments in table 6), and see if this split has an effect on the estimation of deforestation. To summarize the findings, the results are still not significant, but the first-stage regression looks much better and there is a stronger indication of a link between Asian Imports and Deforestation compared to the findings in table 3.

As table 7 shows, there is still no direct evidence of soy expansion causing deforestation, but the indication and significance level of a positive second-stage relationship is clearer in this table compared to table 3. In other words, the standard errors for the elasticity coefficients for Asian imports are small enough for the coefficient to barely not be significant on the 10% level, which was not the case in table 3. The combination of Asian imports and Asian trade-weighted GDP seem to estimate deforestation better than the more general model used in table 3, but the evidence of a causal link between soy demand and deforestation is still lacking.

Table 7: Estimates from FEIV regressions of the log of deforestation on log Asian and non-Asian imports (instrumented by region-specific trade-weighted GDP) in Brazilian municipalities

	Double Instrumented Asian Imports	Only Asia	Only Non-Asia
	Deforestation 1	Deforestation 2	Deforestation 3
log(Asian Imports)	0.0557	0.0572	
	(0.0357)	(0.0357)	
log(Non-Asian Imports)			-0.0387
			(0.0728)
First-stage	Asian imports	Asian imports	Non-Asian imports
log(Asian wGDP)	0.233***	0.233***	
	(0.014)	(0.014)	
log(non-Asian wGDP)	0.002		0.110***
	(0.0005)		(0.016)
Municipality-FE	Yes	Yes	Yes
Time-FE	Yes	Yes	Yes
F (IV)	145.92	290.03	49.95
Observations	1218	1218	1218

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This is a two-stage least squares model of log deforestation regressed on the instrumented log Asian Imports and log of Non-Asian imports in Brazilian municipalities. The instruments are the same as in table 6. The first model uses both Asian-trade-weighted GDP and non-Asian-trade-weighted GDP to estimate Asian Imports, while the second model uses only the Asian trade-weighted GDP and the third model uses only the non-Asian trade-weighted GDP. Weights are the initial share of the Asian or non-Asian imports from a municipality. The models include municipality-fixed effects for 1570 Brazilian municipalities between 2010-2015 and all models control for time-fixed effects. The first-stage results are reported under the second-stage results, with estimated coefficients for the first-stage elasticities. The reported statistics are the F-statistics from the first-stage regressions and the number of observations. The number of observations is 1218 in all models, corresponding to 6 years of observations from 203 municipalities. Standard errors are clustered on the municipal level in all models.

A.2   Motivating the link between cattle pastures and deforestation

In this part of the appendix, I motivate the link between the livestock industry and defor-estation. I found no viable instrument for the number of cattle in different municipalities - trade-weighted GDP not being possible because there are no municipality-country trade flow data available. The inclusion of an instrumental variable would make the economet-ric argument stronger, since the number of cattle in a municipality may well be part of a system of supply-and-demand equations. However, these results do provide some indica-tion of the relationship between the cattle industry and deforestation, and the results are clear. All models point to a strong positive relationship between the number of cattle in Brazilian municipalities and deforestation. The three different models presented in table 8 are pooled OLS with time-fixed effects, random effect panel data regression with time-fixed effects and fixed effects regression with municipality-fixed and time-fixed effects. The data were gathered from IBGE, or the Brazilian Bureau of statistics(SIDRA, 2018).

Table 8: Estimates from regressions of log deforestation on log cattle heads in Brazilian municipalities

	Pooled OLS	Random Effects	Fixed Effects
	Deforestation 1	Deforestation 2	Deforestation 3
logCattle	0.269***	0.231***	0.126***
	(0.0268)	(0.0357)	(0.0469)
r2 (within for FE)	0.0929	0.0927	0.0346
Observations	3738	3738	3738

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This is a table presenting different regression models of log deforestation on log heads of cattle in Brazilian municipalities. The first model is a pooled OLS model, the second a random effects panel data model and the third a municipality-fixed effects model. The models include 623 Brazilian municipalities between 2010-2015 and all models control for time-fixed effects. The number of observations is 3738 in all models, corresponding to 6 years of observations from 623 municipalities. Standard errors are clustered on the municipal level in all models.

The results of table 8 indicate that there is a positive relationship between the number

of cattle in the livestock industry of a municipality and the amount of deforestation in Brazilian municipalities. This further strengthens the argument that a negative relationship between soy demand and the land use of pastures would indicate an indirect negative effect on deforestation since pastures seem to replace forests in accordance to both earlier studies as well as the results in table 8. The municipality-fixed effects model is the stronger model as it takes fixed differences between municipalities into account, fixed differences that obviously are relevant given the difference in coefficient estimates. The inclusion of the Pooled OLS model also illustrates what happens with the random effects model when the individual-fixed effects are not controlled for (in the case that they should have been) - there will quickly be bias towards the OLS-estimator.

B   Tests

B.1   Hausman Test

The starting equation is, with time-fixed effects denoted as i.YEAR:

$$LU_{it} = \alpha_i + \beta \cdot S_{it}^d + \text{i.YEAR} + u_{it} \tag{B.1.1}$$

Using the Hausman test on RE vs FE in this equation, I test for systematic difference in the relationship between land use of exported soy and export volume based on whether municipality-fixed effects are used or not. The Hausman test returned a chi-squared value of 1275.39, indicating a need for fixed effects to be used (although an instrumental variable can also lessen the difference between the estimators). In the case of the instrumental variable equation, using the instrumented soy demand, the Hausman test did not reject the null hypothesis of no difference between the municipalities in terms of how soy demand affects land use for exported soy (estimated at 0.65 and 0.766). Under those circumstances, the choice between FEIV and REIV, that is fixed effects and random effects instrumental variable regression, is mostly a design choice as both models will be consistent if the Hausman test does not reject systematic differences. Wooldridge (2010) states that the actual magnitude of the estimated difference is important as well, and the difference between 0.65 and 0.76 is not inconsequential even though the Hausman test did not reject its null hypothesis. In this paper, since some models do indicate systematic differences between municipalities and I wish to be consistent in my choice of models, the design choice I made was to use fixed

effects instrumental variable regressions in almost all models. I would also argue that model design, if possible, should be done before observing the data so that one does not alter one's approach in such a way that false positives are more likely to occur (that one actively seeks statistically significant results rather than have good *a priori* reasons for testing the data the way one does).

B.2   Non-stationarity tests

As the time frame of this study is quite short, at 6 years, testing for the stationarity of variables might be a bit excessive. In any case, I did do some quick stationarity tests to make sure that there is no issue using the different variables non-differenced.

Table 9: Harris-Tzavalis (HT) test for non-stationarity

	Test statistic	Z-value	Stationary?
logLAND	0.529	-3.930	Yes
logwGDP	-0.227	-75.006	Yes
logDF	-0.094	-22.473	Yes
logexpvol	0.424	-13.814	Yes
log(non-soy)	0.354	-20.379	Yes
log(domestic)	-0.106	-35.684	Yes
logCattle	0.119	-42.450	Yes

Table 9 presents the results of the non-stationarity tests using the Harris-Tzavalis or HT test. This test, like most non-stationarity tests, operates with the null hypothesis that the variable is non-stationary, meaning that the panel data contain unit roots (Harris and Tzavalis, 1999). This would make the use of not-differenced versions of the variables invalid because of the well-known tendency to find spurious relationships when using non-stationary variables (Wooldridge, 2015). For all relevant variables in this paper, this null hypothesis is rejected, implying that all variables are stationary within the relevant municipality-time panels and that they safely can be used in econometric analysis without further alterations.