Università
Ca'Foscari
Venezia

Scuola Dottorale di Ateneo
Graduate School

Dottorato di ricerca
in Economia
Ciclo XXIV
Anno di discussione 2013

Essays on the Microeconomics of Bioenergy

SETTORE SCIENTIFICO-DISCIPLINARE DI AFFERENZA: SECS-P/01

Tesi di dottorato di Breno Pietracci, matricola 955551

Coordinatore del dottorato
Prof. Michele Bernasconi

Tutore del Dottorando
Prof. Agar Brugiavini

The undersigned Breno Pietracci, in his quality of doctoral candidate for a Ph.D. degree in Economics granted by the *Università Ca' Foscari Venezia* and the *Scuola Superiore di Economia* attests that the research exposed in this dissertation is original and that it has not been and it will not be used to pursue or attain any other academic degree of any level at any other academic institution, be it foreign or Italian.

Acknowledgments

I would like to thank Professor Agar Brugiavini for supervising this research, for her kindness, encouragement and incentives.

My gratitude to Professor Ignazio Musu for guidance, motivation, reading earlier versions of this research and helpful comments.

To Professors Mario Padula, Giacomo Pasini and Juliano Assunção for reading earlier versions of this work, helpful comments and discussions.

To all Professors at Università Ca'Foscari di Venezia from whom I had the pleasure to learn.

To Professor Ana Balcão Reis and Professor Maria Antonieta Cunha e Sá who were extremely helpful during my visiting period at Universidade Nova de Lisboa.

To Mrs. Lisa Negrello for all her support.

To my colleagues Caterina Cruciani, Arya Kumar Srurstidhar Chand and Matija Kovacic with whom I spent so many good hours inside and outside the university and who have also taught me so much.

To my colleagues Laura Concina and Yuri Pettinichi for sharing experiences.

A special thanks to Franco Pietracci and Francesca Dissera Pietracci, Lorenzo Pietracci and Merce Sanchez Prados for all support in Venice.

This thesis is dedicated to my wife, Lílian Sales de Oliveira Pietracci and to my parents, Luciano Pietracci and Solange Bellote Pietracci.

Contents

Introduction to the Dissertation

This thesis explores the microeconomics of bioenergy with focus on Brazil. The three chapters are interconnected by this thread.

Chapter 1 investigates empirically land use patterns for liquid biofuels production in Brazil using the neoclassical land use model. Sugarcane and soybean crops are analyzed as these are the main feedstocks employed in country's bioethanol and biodiesel production, respectively. The dataset consists of a panel, comprised of 306 micro-regions in the Brazilian Center-West, Southeast and South regions, with variables encompassing mainly the period from 1994 to 2010. The estimated agricultural production functions exhibit increasing returns to scale, mainly due to the land factor. The four theoretical results of the model are validated empirically, having the principal export port as the market center. Estimated bid-rent functions allow for calculating sugarcane land use own and cross elasticities for prices and transportations costs. The soybean bid-rent function is quite flat and greater than zero inside *Cerrado* and Amazon biomes. Reducing soybean transportation costs causes the least competition for land use between the two crops. The patterns of land use predicted in the neoclassical land use model, under hybrid land use for food and energy, is unchanged, although, hybrid land use affects significantly the variables driving model results and intensifies competition for land.

Chapter 2 investigates empirically, using regional variables, location and capacity decision drivers for ethanol and biodiesel mills in Brazil as of 2011. A cross-sectional dataset is employed comprised of 306 micro-regions in the Brazilian Center-West, Southeast and South regions. Probit and Tobit regressions are estimated to elicit location and capacity decisions, respectively. The Probit regression shows that ethanol mills are located in micro-regions with abundant sugarcane production, low feedstock price, high river density, high number of automobiles and near ethanol storage terminals. The Tobit regression indicates that the capacity of ethanol mills, measured by the number of workers in the industry, is determined by the same variables plus the number of cattle heads in the micro-region which captures land availability for sugarcane production expansion. Biodiesel mills are located in micro-regions with abundant soybean production, near soybean crushing mills and with high employment in raw vegetable oil production, a proxy for soybean crushing mills capacity. Biodiesel mills capacity in each micro-region depends on the same variables as location decisions. The biodiesel industry observes huge excess capacity with continued entry that can be explained by mislocation of initial entrants. First generation biorefineries can be located on the transition of land use from sugarcane to soybean.

Chapter 3 proposes a microeconomic framework for modeling several types of bioenergy production. Drawing upon the Brazilian experience, three bioenergy chains are considered in detail; (1) sugarcane-ethanol, (2) soybean-biodiesel and (3) forestry-firewood-charcoal-paper and pulp-black liquor. This single framework accommodates all these production possibilities with a nested structure for biomass output in the lower nest and biomass upgrading in the upper nest. It uses a multiple output production function to allow for the production of a vector of five Fs which is comprised of food, feed, fuels, fibers and forestry products. The idea that these microeconomic production structures forms bioenergy clusters from a statistical standpoint is put forward here. To test this concept, four k-means cluster analyses are performed using a dataset comprised of 306 micro-regions in the Brazilian Center-West, Southeast and South regions, with 15 variables referring exclusively to the nested multiple output production function averaged from 2006 to 2010. Results allow to assign each micro-region to one of the three types of bioenergy clusters considered plus one additional cluster unsuited for bioenergy production. Plotting results of cluster membership in a Geographic Information System, reveals that there is geographic proximity among micro-regions belonging to the same cluster. Furthermore, the forestry cluster shows that more than one agglomeration of bioenergy producing micro-regions, employing the same feedstock, can co-exist separated in space.

Chapter 1

The Neoclassical Land Use Model Applied to Liquid Biofuels Production in Brazil

1.1 Introduction

In the quest for more sustainable and less carbon intensive energy sources, first generation biofuels are consolidating as one of the most successful options in the renewable energy portfolio to displace fossil fuels in a non negligible scale.

Lee et al. (2008) [38], affirms that *"In the short term, producing liquid fuels from biomass is one of the only alternatives to petroleum-based products"*. Hausmann (2007) [28], estimate that biofuel production could match that of crude oil, acting as a cap for crude oil prices. At least for the ground transportation sector, according to the International Energy Agency - Bioenergy (2011) [6], there is more consensus that biofuels are the unique short-run climate change mitigation policy.

In fact, many governments are incentivizing fuel production from biomass with economic, environmental and social objectives.

Sorda et al. (2010) [50] investigates which policies are being implemented worldwide to foster biofuels production. Blending mandates, tax exemptions, subsidies, trade barriers and other interventions are boosting ethanol and biodiesel production growth. In 2010, global ethanol production amounted to 1.5 million barrels per day, biodiesel production amounted to 0.3 million barrels per day and crude oil production amounted to 82.1 million barrels per day.[1]

With increasing production, controversy has emerged regarding the impacts of biofuels on agricultural markets. Specifically, biofuels are suspected to affect land and food prices.

Another controversy has emerged on the actual climate benefits of biofuel production because crop expansions cause direct and indirect land use change that releases greenhouse gases stored in soils and biomass, especially if their production causes deforestation directly or indirectly.

Accordingly, one major concern regarding the promotion of biofuel production, stressed by Lee et al. (2008) [38], refers to the complexity of coordinating agricultural, energy and environmental policies.

In spite of controversies or amid them, biofuel production keeps growing at steady pace. From 2005 to 2010 world ethanol production increased at an average rate of 20 % per year, while biodiesel production increased at an average rate of 42.4 % per year.

This paper investigates empirically land use patterns for liquid biofuels production in Brazil using the neoclassical land use model. Sugarcane and soybean crops are analyzed as these are the main feedstocks employed in the country's bioethanol and biodiesel production, respectively.

The patterns of land use predicted in the neoclassical land use model, under hybrid land use for food and energy, is unchanged, although, hybrid land use affects significantly the variables driving model results and intensifies competition for land.

The constructed dataset consists of a panel, comprised of 306 micro-regions in the Brazilian Center-West, Southeast and South regions, with variables encompassing mainly the period from 1994 to 2010.

The agricultural production function is estimated and exhibits increasing returns to scale, mainly due to the land factor. This is the first result of the paper.

Next, the four results of the neoclassical land use model are tested empirically. Two major export ports, Santos and Paranaguá, are chosen as market centers. Santos port

[1] Data on biofuel production from US Energy Information Administration and for crude oil production from British Petroleum Statistical Review of World Energy.

is the most important for sugar and ethanol exports while Paranaguá port is the most important for soybean, soybean meal and soybean oil exports.

Santos port is validated as the unique market center as all results stemming from the neoclassical land use model holds and the pattern of land use can be understood as circular rings from it. Closer to the port, land is employed in sugarcane production due to its much higher output per hectare and higher transportation costs followed by the soybean land use zone which has lower output per hectare and lower transportation costs. This is the second result of the paper, which shows that one port in Brazil explains land use patterns in almost half of the country.

Including data on transportation costs for sugarcane, refined sugar, soybean and wages, bid-rent functions are calculated for the years 2008, 2009 and 2010. From bid-rent functions, sugarcane land use own and cross elasticities for prices and transportation costs are computed. Land rent elasticity of a hectare at 2,500 km from Santos port is computed with respect to soybean price and transportation cost.

From elasticities, it is more likely that sugarcane land use zone expands into the soybean land use zone, or pushes soybean area inland.

As soybean transportation cost has little impact on land use competition and on the land rent of the outermost hectare, reducing logistic costs for this industry should be pursued. Each 1% decrease in soybean transportation costs, captures 0.05 % of land use for sugarcane and increases land rent of the outermost hectare by 0.72 %, *ceteris paribus*. Logistics in general, and transportation and port infra-structure in particular, can be significantly improved in Brazil reducing export bottlenecks.

While the concern has been on food versus fuel competition for land, Brazil can be considered as the first case where fuel versus fuel competition for land can occur as sugarcane and soybean crops compete with each other for space.

The extent to which this will actually happen will depend on several factors such as global and local economic growth, the evolution of agricultural and energy commodities prices, the rate of growth in yields per hectare resulting from technological change and international trade framework.

With respect to the later, Brazil is embracing policies to expand its agricultural and biofuels production. One of the most sought objectives of its trade policy is to reduce global agricultural trade barriers and subsidies, which will potentially cause a boom in demand for its agricultural products and thus land.

Brazil has a particular greenhouse has emissions profile according to its second national communication to the UNFCCC, coordinated by the Ministry of Science and Technology (2010) [40]. Land use change and Forestry is the main source of carbon dioxide while agriculture is the main source of methane and nitrous oxide. Converting these measures to CO_2 equivalent using their global-warming potential (GWP), these two sectors represent the bulk of greenhouse gas emissions in the country.

Results will help understanding land use dynamics under hybrid use for food and energy. It can help formulating policies to promote biofuel industry expansion, with further substitution of fossil fuels, while protecting sensitive biomes, avoiding deforestation and mitigating emissions from land use change.

Finally, they can be used to promote biofuels in other countries or regions, especially those that have similar climate conditions like some central African countries.

1.2 The Sugarcane and Soybean Sectors

As of 2010, Brazil ranked as the second major world producer of ethanol and biodiesel and thus, an unique case of producing large quantities of both biofuels.

Sugarcane and soybean crops are chosen because they are currently the building blocks of the Brazilian biofuels industry. Part of the output of these two agribusiness serve as the principal feedstocks used in the production of ethanol and biodiesel, respectively.

In Brazil, all ethanol is produced using sugarcane as feedstock. Sugarcane can be used to produce either sugar of several qualities, but mainly raw or refined sugar, or ethanol of two qualities, hydrated ethanol used in neat ethanol or in flexible fuel engines, or anhydrous ethanol used as a blend in gasoline. Thus, sugar and ethanol produced from sugarcane compete with each other as output of both can vary, to some extent, according to producers decisions.

The biodiesel program started in 2005 with an optional blending of 2% that would become mandatory in January, 2008. Since the onset of the biodiesel program up to the end of 2010 data on feedstock use shows that soybean oil is by far the most used feedstock. Summing these 6 years soybean oil represented 81.6 % of the feedstocks used, followed by tallow with a share of 13.2 %, cottonseed oil with 2.7 % and other fatty acids with 2.5 %.[2]

For this reason the focus is restricted on these two crops or feedstocks, sugarcane used for ethanol production and soybean used for biodiesel production. The problem analyzed here consists of a two crop land use pattern in the plane.

Additionally, these sectors produce the country's major exported agricultural commodities; sugar and ethanol and soybean and its products.

In 2009, Brazil was the world top exporter of raw and refined sugar and non food alcohol and the second world exporter of soybean, behind US, and soybean oil, behind Argentina.[3]

Regarding the country trade balance in 2010, soybean and its products ranked as the third main export product, while sugar and ethanol ranked fourth amounting to 17 and 13 billion USD approximately, near 20 % of the country's total exports.[4]

The theoretical neoclassical land use model rely on the fact that an unique market center exists where prices are determined exogenously. Thus, the initial step is to select one single point in space which will represent the empirical counterpart of the market center. In the case of agricultural commodities, markets can be cities, food processing plants or ports for international trade.

Preference is given to ports as they represent the point in space where free on board commodity prices are exogenously determined, i.e. world prices. World prices can be further interpreted as the opportunity cost of selling a commodity in the domestic market.

Nevertheless, this simplification represents an approximation of reality as the bulk of sugarcane is actually sold to sugar and ethanol mills, while soybean can be sold to soy crushing mills or directly exported. Among all Brazilian ports, the most relevant for the two crops are selected.

Sugar and ethanol are mainly exported through Santos port and soybean and its

[2]Data from Agência Nacional do Petróleo; National Petroleum Agency; www.anp.gov.br

[3]Data from FAOstat. http://faostat3.fao.org/home/index.html

[4]Data from Brazilian Ministry of Development, Industry and Trade; www.desenvolvimento.gov.br

products are mainly exported through Paranaguá port. Not by coincidence these two ports are close to each other and distance 515 km by road (310 Km in a straight line) approximately. All model results are tested having these two ports as the market center.

Santos port acts as the gravity center for sugarcane production, while Paranaguá port acts as the gravity center for soybean production. These two ports can be considered as the *foci* of an ellipse that generates the spatial pattern of land use zones.

Land use in Brazil for sugarcane and soybean are steadily increasing and amounts, as of 2010, to approximately 9 million hectares (9 thousand Km^2) and 23 million hectares (23 thousand Km^2) respectively. These figures represent 1.05 % and 2.7 % of the 851.5 million hectares of the national territory.

As of 2010, Brazil ranked as the first sugarcane producer in the world, followed by India, China, Thailand and Mexico. In the same year, Brazil ranked as the second soybean producer behind the United States and followed by Argentina, China and India.[5] Among the top 5 producers, in the same year, Brazil exhibited the highest yield per hectare for both crops.

Brazil has two consolidated regions of sugarcane production. One in the coastline of the Northeastern region characterized by the Atlantic Forest biome. The other in the inland of the Center-West, Southeast and South regions characterized by Atlantic Forest and Savannah biomes. The later exhibit more suitable conditions for sugarcane production, mainly in terms of soil quality and amount of rainfall which reduces or eliminates the need for irrigation, and thus higher productivity.

Concerned with the future sustainability of sugarcane, sugar and ethanol industry, the Federal Government launched the sugarcane agro-ecological zoning program, coordinated by the Brazilian Agricultural Research Corporation (2008) [13], aimed at indicating most adequate areas for sugarcane production expansion, where producers will be entitled to apply for federal credit. Areas were selected according to several sustainability criteria and inclinations lower than 12%, which allows for mechanization and the phase-out of pre-harvest burning, a major environmental concern regarding the industry and source of greenhouse gases.

[5] Data from FAOstat. http://faostat3.fao.org/home/index.html

This program forecasts that sugarcane production will double by 2017 and land use is expected to reach 14.5 million hectares or 1.7 % of the national territory. According to the agro-ecological zoning guidelines, land use for sugarcane could reach in a sustainable way 7.5 % of the territory or 64.7 million hectares. The sugarcane AEZ prioritizes production expansion in areas that are either abandoned or under inefficient grazing, with low cattle heads per hectare.

The sugarcane sector is facing challenges to meet increasing demand since the 2008 financial crisis. Investments in sugarcane supply have not kept pace with both higher international sugar prices and increased domestic ethanol demand, especially outside the harvesting season.

Brazil has also two areas of soybean production. The first is the South region, were soybean was first introduced, as the colder climate was more adequate to its production. But with research and development in new soybean varieties suitable for lower latitudes production expanded into the Center-West region which is characterized by the Savannah and Amazon biome.

It is interesting to note that Argentina, Bolivia, Paraguay and Uruguay rank amongst the top 10 world soybean producers, making one wonders if South America can become a major biodiesel producer and exporter.

Also concerned with sustainability, the soybean industry through its trade union, the Brazilian Association of Vegetable Oil Industries (2007) [11], has a self-imposed restriction or sustainability criteria, coordinated with other stakeholders, forbidding marketing of soybean originated from Amazon deforested areas after July, 24th, 2006, the so-called soybean moratorium.

Finally, the agribusiness is extremely influential politically in Brazil due to its economic importance.

1.3 Literature Review

In order to address the research question, the focus is on agriculture location literature and land use, both theoretical and empirical.

The initial step in understanding locational drivers of first generation liquid biofuels production is to analyze location theories related to its biomass feedstocks, i.e. the formation of monoculture agricultural zones, known as von Thünen rings.

Beckmann (1999) [4], page 122, defines von Thünen rings as *"... a sequence of zones ..."* such that each zone is *"... a circular ring, in which land is used for one particular product or method of cultivation exclusively"*. Thus, von Thünen rings consist of a sequence of finite diameter annuli from a city or market center where, in each annulus, full specialization in one and only one agriculture crop occurs.

The theoretical definition of full specialization is seldom observed in reality. One of the most important techniques to maintain or regenerate soil quality is crop rotation. Indeed, Samuelson (1983) [47] mentions that crop rotation *"... enhance land's steady state productivity"*. This imply that, at any given distance from the market, the same hectare can be used for different crops or left fallow over time.

Also, land use for livestock is expected to be connected to other agricultural activities that provides its feed. The possibility that complementary crop and livestock activities overlap in space exists due to production or technological linkages as stressed by Fujita and Thisse (2002) [19] and/or, according to Herrero et al. (2010) [31], because they give more resilience to price takers small holder farmers.

Thus, regions can specialize, but not exclusively, in one agricultural product forming what is commonly called agricultural belts, e.g. the corn and cotton belt in United States, which has a less stringent definition.

For example, the US Environmental Protection Agency defines the corn belt as *"The area of the United States where corn is a principal cash crop, including Iowa, Indiana, most of Illinois, and parts of Kansas, Missouri, Nebraska, South Dakota, Minnesota, Ohio, and Wisconsin"*.[6]

Beckmann (1972) [3] proposes a theoretical neoclassical model of land use that explains the formation of these agricultural rings. It consists of a von Thünen model of land use with a neoclassical production function. While Beckmann (1972) [3] used a constant returns to scale production function, Renaud (1972) [46] demonstrated that the model main conclusions hold under different degrees of homogeneity of the production function.

[6] http://www.epa.gov/agriculture/ag101/cropglossary.html

The model exhibits sharp predictions regarding the pattern of land use. First, in any land use zone there should be declining employment per hectare, yield per hectare and land rents as a function of the distance from a single point in space, which in the model is considered to be a city but can refer to any market, e.g. a major port of export or an agroindustrial plant for feedstock processing. The crucial assumption is that the market price is unique and refers to a single point in space where physical delivery of the agricultural commodity or consumption will take place. Second, there will be specialization in land use. Third, considering all land use zones labor per hectare is a decreasing and continuous function of distance to the market. And fourth, land use zones will be organized by decreasing transportation costs of the yield per hectare.

One of the caveats of the neoclassical land use model is to assume that there are no differences in soil quality. This is very unlikely to hold as soil productivity varies according to nutrient characteristics, irrigation and declivity which enables higher or lower degrees of mechanization. Additionally, the model is based on a pure agricultural environment, on the existence of a mono-centric market and without any sort of spatial interference such as other markets, cities or raw material processing plants.

Jones and Krummel (1987) [34], add more realism to the pure agricultural location problem with atomistic farmers by introducing the location decision problem of a plantation, defined as an *"agricultural activity in which on farm assembly costs of the harvested crop are significant"* coupled with a *"punctiform ... large operational unit with increasing returns to scale ... in which some processing of raw crop must be conducted prior to shipment to market"*. This intermediate step between fields and market, caused by the introduction of an assembling or processing facility, enables the possibility of observing reverse hauls of raw material which in turn affect land rents creating a saw-toothed pattern around each facility. The authors also introduce the size of the processing plant as a choice variable which, for a given transformation technology and yield per hectare, determines the area of the plantation.

As the production of liquid biofuels requires biomass feedstocks produced from agriculture, increased attention has been devoted to understand how it affects land use patterns. In particular, the concern has been with the competition for land for food and energy production.

In this stream, Reilly and Paltsev (2008) [45] describe the methodology used to incorporate bioenergy production technologies into a economic-climate computable general equilibrium model, the MIT Emissions Prediction and Policy Analysis, and investigates scenarios of bioenergy output and competition for land under different climate policies. Two types of bioenergy technologies are modeled. One for the production of liquid biofuels for transportation and the other for the production of bio-electricity. Forecasts shows that under all analyzed GHG stabilization scenarios, bioenergy production is expected to be higher than under no climate policy. Even without any climate policy bioenergy production is expected to increase substantially. Additionally, as the authors also expect that food demand will grow steadily, they conclude that if bioenergy is going to have a significant role in future energy supply, there will be increased competition for land and major impacts on agricultural markets.

On the macro-econometric side, Piroli et al. (2011) [43] uses a near-VAR approach to investigate empirically land use change caused by corn ethanol in the US using agricultural and energy data from 1950 to 2007. Variables were grouped into 3 categories: exogenous variables, endogenous macroeconomic variables and crop specific endogenous variables. The authors tests the degree of interdependency between oil prices and agricultural land use. Their theoretical model consist of five interdependent markets: agriculture, fossil fuel, biofuel, biofuel by-products and inputs. Two transmissions channels between energy and agricultural markets are identified. The first is a direct channel as higher oil prices should induce more biofuel demand affecting agricultural markets. The second an indirect channel, as energy is an input in agriculture production, thus, higher oil prices increases agricultural costs. The two channels have opposite effects on land use. Higher oil prices stimulate biofuel production and crop expansion, while the increased costs of energy input reduces profitability in the agricultural sector. Their estimates indicate that each additional dollar in crude oil prices increase land use in the United States by 45 to 56 thousand hectares. The conclusion is that biofuels cause significant land use change.

The model has some caveats. There is a problem regarding the definition of direct and indirect land use change. The authors interpret direct land use change as the total variation in land use for 5 crops. This is not accurate as part of expansion is caused by food demand and cannot be attributed to biofuels in its totality. Additionally, crops that are not used for biofuel production cannot enter in the computation of direct land use change caused by biofuels. Indirect land use change is defined as total variation in land use. This measure is also inaccurate because if direct land use change does not displace any cropland or pasture land, then indirect land use change is zero, although total agricultural land use has increased. Additionally, the authors cannot distinguish from estimations, on their own definition, between direct and indirect land use change.

There is a very strong missed link from energy markets to agriculture markets which is independent of biofuels. As corn production increases, so does the demand for basic fertilizers, nitrogen included. And nitrogen fertilizers, mainly ammonia and urea, are produced using natural gas. Thus their price fluctuates with natural gas prices. This additional transmission mechanism may be weaker or stronger than the fuel usage. Whatever is the case the linkage between energy prices and agricultural production costs are actually stronger than depicted in their model if another input is included, the fertilizer channel.

On the micro-econometric side, Hausman (2012) [27] investigates land use in Brazil caused by variations in agricultural commodities prices from 1973 to 2005. Harvested land is regressed on past land use, crops spot and futures prices, past yields per hectare and risk measures. The author finds that soybean land use is responsive to price changes but sugarcane land use is not.

More recently, great effort has been devoted to understanding direct and indirect impacts on land use caused by biofuels production and quantifying future land use change effects of bioenergy production.

The most accurate way to evaluate past land use change, which is employed in the second national communication of Brazil to the UNFCCC [40], is using satellite images and digital processing tools. Satellite images can capture precisely past land use change of any crop over any type of biome and enter as emissions into the Land Use Change and Forestry sector.

But despite the precision provided by satellite images to analyze past data, it is difficult to disentangle land use change caused by biofuels or food production, and to provide forecasts or scenario analysis.

To overcome these caveats, researchers have developed methodologies that are very similar to net present value cash flow analysis, and even the words employed, carbon credit and debt, refer to financial jargon. The difference is that emissions are treated equally over time, i.e. future emission are not discounted by any sort of discount rate.

Fargione et al. (2008) [14] investigates that whenever biofuel production requires native vegetation conversion, there is an initial carbon release, a *"biofuel carbon debt"*, that can have a payback time of more than half a century. This net present value of carbon flows is based on the fact that direct land use change causes a stream of CO_2 emissions that needs to be repaid by the displacement of crude oil products.

The payback time varies with the type of land converted and the food crop used as feedstock for biofuel production. In the Brazilian case, the authors estimate that sugarcane ethanol production converting a Savannah biome, *Cerrado*, has a payback time of 17 years, while soybean oil biodiesel has a payback time of 37 or 319 years, depending whether it is produced converting a Savannah, *Cerrado*, biome or a tropical rainforest, Amazon biome, respectively. Brazilian biofuels, both ethanol and biodiesel, produced in the *Cerrado* exhibits the lowest payback period in their study which includes other 3 analysis for corn ethanol produced in the US and palm oil biodiesel produced in Malaysia on different local biomes.

The payback time of sugarcane ethanol may be even lower as the authors do not consider sugarcane bagasse can be used for bio-electricity production which will further displace fossil fuel power generation.

Due to this biofuel carbon debt the authors advocates that biofuels should be produced using waste biomass or from food crops in abandoned or degraded land in order to yield immediate green house gas emissions reduction. These figures refers exclusively to direct land use change and do not include indirect land use change consequences.

Searchinger et al. (2008) [49] analyze direct and indirect impact on global land use caused by ethanol production in the United States using a global agricultural model over a 30 year framework. Their estimates are based on additional production of 56 billion liters above a baseline scenario with projections for 2016. Calculations represent a flow of carbon where growing biomass results in a benefit, the *"feedstock carbon uptake credit"*, while direct and indirect land use change cause the *"biofuel* *carbon debt"*. It is identified another opportunity cost caused by land use change, the potential foregone carbon sequestration service that maturing forests and grasslands would provide. The authors estimate that, on average, each hectare converted emits 351 tons of CO_2 equivalent. Thus increased corn ethanol production in the United States instead of mitigating greenhouse gases emissions, almost double CO_2 release due to land use change worldwide when compared to using only gasoline over the same period.

Regarding Brazil, the authors estimate that 170 millions hectares can be converted to cropland. They estimate a payback time for sugarcane ethanol that varies from 4 to 45 years depending on the indirect reaction of displaced ranchers from the tropical grazing lands. The authors conclude that biofuels should be pursued only when carbon benefits exceeds carbon costs.

The is still much debate about quantifying the exact net greenhouse gas emissions caused by biofuel production, when direct and indirect land use change are accounted for and thus their real contribution to climate change mitigation in the short and long run. One reason is that models are still much assumption dependent.

Finally, direct and indirect land use change will occur, even in the absence of biofuel production, as agricultural commodities relatives price fluctuates in time with supply and demand dynamics. For example, as world income and population grows or trade and agricultural policies change.

1.4 The Neoclassical Land Use Model

The mathematical model presented in this section is drawn from Beckmann (1972) [3], Reanud (1972) [46], Beckmann (1999) [4] and Fujita and Thisse (2002) [19].

The production function Φ for crop i, has labor (X_i) and land (T_i) as inputs, one output (Y_i). Labor and output markets are competitive. Land market can deviate from perfect competition, thus from marginal product pricing. Transportation market is also perfectly competitive.

Capital is not included in the original model, but Alessio (1973) [2] treats the variable (X) as a bundle of inputs. The parameter A_i is defined by Beckmann (1972) [3] as a *"factor of proportionality"* while Renaud (1972) [46] treats it as a *"technological parameter"*. It represents the natural and different productivity of each crop type. Mathematically, $A_i > 0$ is the output that can be produced when one unit of labor is employed in one hectare of land.

$$Y_i = \Phi(A_i X_i, T_i) \tag{1.1}$$

Equation (1.1) will, hereafter, take the specific Cobb-Douglas form.

$$Y_i = A_i X_i^{\alpha} T_i^{\beta} \tag{1.2}$$

Equation (1.2) has a degree of homogeneity $(n = \alpha + \beta)$ where $0 < \alpha < 1$.

Normalizing it by the amount of land (T_i) in order to have output or yield per hectare (y_i) as a function of labor employed per hectare (x_i), results in:

$$Y_i \left(\frac{1}{T_i}\right)^n = A_i \left(\frac{X_i}{T_i}\right)^{\alpha} \left(\frac{T_i}{T_i}\right)^{\beta}$$

$$\frac{Y_i}{T_i}\left(\frac{1}{T_i^{n-1}}\right) = A_i x_i^{\alpha}$$

$$y_i = T_i^{n-1} A_i x_i^{\alpha} \tag{1.3}$$

Where $y_i = \dfrac{Y_i}{T_i}$, or yield per hectare and $x_i = \dfrac{X_i}{T_i}$, or labor per hectare.

The physical marginal product of labor is given by:

$$\frac{\partial y_i}{\partial x_i} = \alpha T_i^{n-1} A_i x_i^{\alpha-1} > 0$$

The production function exhibits diminishing physical marginal product with respect to labor per hectare:

$$\frac{\partial^2 y_i}{\partial x_i^2} = (\alpha - 1)\alpha T_i^{n-1} A_i x_i^{\alpha-2} < 0$$

The physical marginal product of land is given by the residual of output per hectare, after labor payment. In the possible absence of constant returns to scale, factor payment using marginal products will be different from total output.

$$
\begin{aligned}
y_i - x_i \frac{\partial y_i}{\partial x_i} &= T_i^{n-1} A_i x_i^{\alpha} - x_i \alpha T_i^{n-1} A_i x_i^{\alpha - 1} \\
&= T_i^{n-1} A_i x_i^{\alpha} - \alpha T_i^{n-1} A_i x_i^{\alpha} \\
&= (1 - \alpha) T_i^{n-1} A_i x_i^{\alpha}
\end{aligned}
\tag{1.4}
$$

Besides the distance (r) to the market, other important variables driving the model results are the cost of transportation of crop i per ton per kilometers (t_i) and the prevailing exogenous price of each commodity (p_i) in its final destination.

The agricultural producer receives, for each ton of crop (i) sold in the market, the prevailing market price (p_i) minus the total cost of transportation ($t_i r$). Thus, each producer has, in order to be competitive in the reference market, subtract all costs incurred to place its product there.

The model employs a common feature used by producers in commodity markets, **netback pricing**. Netback pricing is the maximum price a producer can charge, at the production site, taking into account the logistic costs for physical delivery of a commodity in a reference market. For example, crude oil and its products are priced at reference markets such as the United States Gulf Coast, Rotterdam and Singapore. Note that, despite the model assumption of perfect competition in output markets, commodity markets may not be competitive, e.g. energy and metals commodities. Also agricultural commodities can contain mark-up in their prices.

A profit maximizer producer, located at a distance (r) from the market, decides how much labor to hire for each crop $(x_i) \in X = \{x_1, x_2, \ldots, x_k\}$.

Profit per hectare (π) is a function of prices (p_i), output per hectare (y_i), which in turn is only a function of the amount of labor employed in each crop (x_i); distance from the reference market (r), cost of transporting one ton of each crop (t_i) per km, the prevailing wage rate (w) and land rent (R).

$$\max_{x_i \in X} \pi(r, X) = \sum_i \left[(p_i - t_i r) T_i^{n-1} A_i {x_i}^{\alpha} - w x_i\right] - R(r) \tag{1.5}$$

$$\forall \quad x_i \in X = \{x_1, x_2, \ldots, x_k\}$$

Subject to the following non-negativity constraints:

$$\begin{aligned} A_i &> 0 \\ x_i &\geq 0 \\ (p_i - t_i r) &\geq 0 \end{aligned}$$

For any agricultural commodity (i) whenever $r = p_i/t_i$, the netback price is zero, and it is not profitable to produce that crop beyond that distance.

For crops that are actually produced $x_i > 0$, profit maximization with respect to labor per hectare for each crop (x_i) implies in the following first order conditions:

$$\frac{\partial \pi}{\partial x_i} = (p_i - t_i r) T_i^{n-1} A_i \alpha x_i^{\alpha - 1} - w \tag{1.6}$$

As labor is assumed to be ubiquitously supplied at a constant wage rate (w) throughout the territory, optimal labor demand can be found by rearranging equation (1.6).

$$x_i^*(r) = \left[\frac{A_i \alpha (p_i - t_i r) T_i^{n-1}}{w}\right]^{\frac{1}{1-\alpha}} \tag{1.7}$$

From the labor demand equation (1.7), it is possible to verify directly that optimal labor demand is increasing in the product price and decreasing in wage, distance and transportation costs.

The relation between labor and distance can be found by differentiating (1.7) with respect to (r).

$$\frac{dx_i(r)}{dr} = \frac{-t_i [A_i \alpha T_i^{n-1}]^{\frac{1}{1-\alpha}} (p_i - t_i r)^{\frac{\alpha}{1-\alpha}}}{(1-\alpha) w^{\frac{1}{1-\alpha}}} < 0 \tag{1.8}$$

Equation (1.8) demonstrates that optimal labor per hectare for any crop (i) is a decreasing function of distance.

As labor per hectare declines with distance and output is a function of labor per hectare, output per hectare also falls with distance for all crops. Moreover for each crop, both labor per hectare and output per hectare reaches zero when $r = p_i/t_i$.

This implies in result 1 quoted from Fujita and Thisse (2002) [19].

Result 1: *"...for each activity, less and less labor is used as one moves away from the market town so that the equilibrium output is decreasing and continuous in distance to the market town."*

The maximum amount a producer of crop i located at r will be willing to pay for land rent, known as the bid-rent function for each crop $\Psi_i(r)$, is defined when profits are set to zero.

$$\Psi_i(r) = (p_i - t_i r) T_i^{n-1} A_i {x_i}^{\alpha} - w x_i \tag{1.9}$$

Substituting for (x_i) from equation (1.7) into the bid-rent function (1.9) results that the bid-rent is equal to the value of the marginal product of land for each crop at each location.

$$\Psi_i(r) = (1-\alpha)\left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}} \left[A_i T_i^{n-1}(p_i - t_i r)\right]^{\frac{1}{1-\alpha}} \tag{1.10}$$

It follows that the bid-rent function for each crop is decreasing in distance $\Psi_i^{'}(r) < 0$ and convex $\Psi_i^{''}(r) > 0$.

The land owner or producer, located at any distance r, will choose the single crop with the highest bid-rent function. That is, crop i will be produced whenever $\Psi_i(r) > \Psi_j(r)$.

Land rent in each location is determined by competing producers that bid to rent that hectare driving profits to zero. Thus, land rent at each location $R(r)$ is equal to the maximum bid $\Psi_i(r)$ a producer can make to rent that land.

$$R(r) = \max\left[\max_i \Psi_i(r), 0\right] \quad \forall i \tag{1.11}$$

$$R(r) = max\left[(1-\alpha)\left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}} \max_i \left[A_i T_i^{n-1}(p_i - t_i r)\right]^{\frac{1}{1-\alpha}}, 0\right] \quad \forall i \tag{1.12}$$

As the distance from the market increases, land rent decreases proportionately to transportation costs, t_i.

At any given distance (r), it must be that $A_i T_i^{n-1}(p_i - t_i r)$ is greater, less or equal to $A_j T_j^{n-1}(p_j - t_j r)$ for any two crops $i \neq j$. Thus, the land owner or producer will choose the single crop which yields the highest value for the output produced with 1 unit of labor.

It results that only at a particular point in space, if any at all, at some distance (r_{ij}) land owners will be indifferent between producing crop i or j. This occurs because the bid-rent function is equal for the two crops i and j only when:

$$A_i T_i^{n-1}(p_i - t_i r) = A_j T_j^{n-1}(p_j - t_j r)$$

In any other point closer to or farther from the market from (r_{ij}) it will be more profitable to produce only crop i or only crop j, depending on the output per hectare with 1 unit of labor ($A_i T_i^{n-1}$), transportation cost (t_i) of each crop and prevailing market prices (p_i). This leads to result 2 quoted from Beckamnn (1972) [3].

Result 2: *"... at a given distance from the market one and (except at hairline boundaries) only one agricultural product will be produced"*, i.e. there is *"specialization of land use"*.

In equilibrium, the constant wage is equal to the value of the marginal product of labor in each location (r), when optimal quantity labor is employed (x_i^*).

$$w = (p_i - t_i r) T_i^{n-1} A_i \alpha [x_i^*(r)]^{\alpha - 1} \tag{1.13}$$

Taking the land rent-wage ratio for one specific crop i from equations (1.12) and (1.13).

$$\frac{R(r)}{w} = \frac{(1-\alpha) T_i^{n-1} A_i [x_i^*(r)]^{\alpha} (p_i - t_i r)}{(p_i - t_i r) T_i^{n-1} A_i \alpha [x_i^*(r)]^{\alpha - 1}}$$

$$\frac{R(r)}{w} = \frac{(1-\alpha)}{\alpha} x_i^*(r) \tag{1.14}$$

Taking into account that the wage does not vary with distance and that at the transition of one land use zone (i) to the next (j), at r_{ij} the bid-rent functions must be the same and equal to the land rent, $\Psi_i(r_{ij}) = \Psi_j(r_{ij}) = R(r_{ij})$, it must be true that:

$$\frac{\Psi_i(r_{ij})}{w} = \frac{\Psi_j(r_{ij})}{w}$$

$$\frac{(1-\alpha)}{\alpha} x_i^*(r_{ij}) = \frac{(1-\alpha)}{\alpha} x_j^*(r_{ij}) \tag{1.15}$$

It follows that equation (1.15) holds only when $x_i^* = x_j^*$. This leads to model results 3 quoted from Beckamnn (1972) [3].

Result 3: *"The labor/land ratio is a continuous decreasing function of distance from the market."*

Moreover, at the transition from crop i to crop j, at r_{ij}, the bid-rent function of crop i has to be stepper than that of crop j, with respect to distance.

$$\Psi_i'(r_{ij}) < \Psi_j'(r_{ij}) \tag{1.16}$$

$$\Psi_i' = -t_i \left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}} \left[A_i T_i^{n-1}(p_i - t_i r)\right]^{\frac{\alpha}{1-\alpha}} A_i T_i^{n-1} \tag{1.17}$$

But equation (1.17) can be written as a function of (x_i^*) from equation (1.7).

$$\Psi_i' = -t_i \alpha^{\frac{\alpha}{1-\alpha}} [x_i^*]^{\alpha} A_i T_i^{n-1} \tag{1.18}$$

Substituting in equation (1.16) yields:

$$-t_i \alpha^{\frac{\alpha}{1-\alpha}} [x_i^*]^{\alpha} A_i T_i^{n-1} < -t_j \alpha^{\frac{\alpha}{1-\alpha}} [x_j^*]^{\alpha} A_j T_j^{n-1} \tag{1.19}$$

$$-t_i A_i T_i^{n-1} < -t_j A_j T_j^{n-1} \tag{1.20}$$

$$t_i A_i T_i^{n-1} > t_j A_j T_j^{n-1} \tag{1.21}$$

Taking into account that t_i is the cost of transportation of one unit of crop i per ton per km and that $A_i T_i^{n-1}$ is the output of by employing one unit of labor in one unit of land, lead us to result 4 quoted from Beckmann (1972) [3] and Renaud (1972) [46].

Result 4: *"Land use zones are arranged in order of decreasing transport cost of the acre output produced with equal amounts of labor"* and at the transition of land use there is a *"discontinuity of transport cost for the output per unit of land area"*.

1.5 Sample and Dataset

The Brazilian territory has an area of 8,515,767 km^2 divided into five political regions; North, Northeast, Center-West, Southeast and South described in picture (1.1).

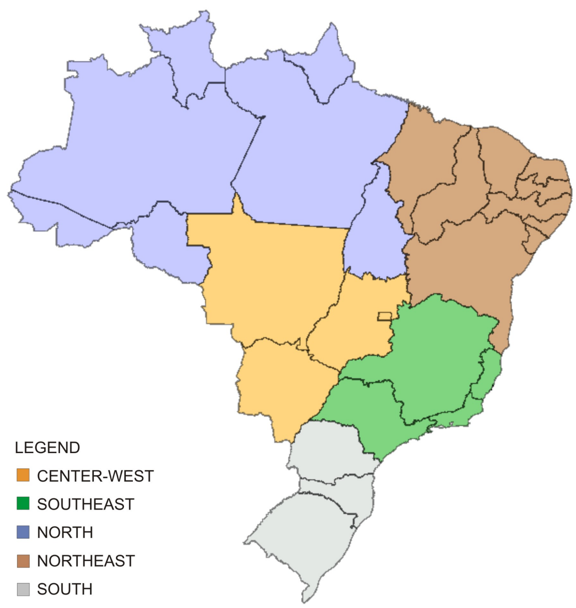

Figure 1.1: Brazilian Territory and Political Division

The sample consists in part of the Brazilian territory, the Center-West, Southeast and South political regions encompassing 306 (out of 558) micro-regions as of the political division of 2010. Two regions are dropped form the analysis, the Northeast and North. The sample, with the code of each state and the location of the two selected ports are depicted in figure (1.2).

The level of analysis chosen is the micro-region, which is composed by a set of

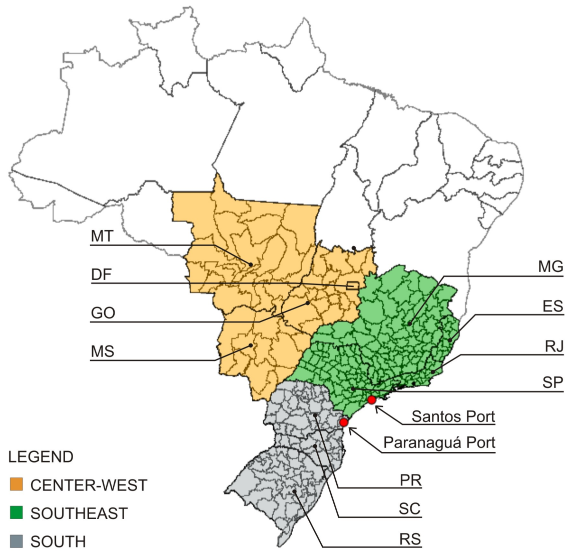

Figure 1.2: Sample Description and Market Centers

municipalities and is similar, but not equivalent, to NUTS 3 classification in Europe or counties in the United States. Table (1.1) presents the political division of the sample into regions, states and number of micro-regions in each state.

Table 1.1: Description of the Sample as of 2010

State Code	State Name	Region	Micro-regions
DF	Distrito Federal	Center-West	1
GO	Goiás	Center-West	18
MT	Mato Grosso	Center-West	22
MS	Mato Grosso do Sul	Center-West	11
MG	Minas Gerais	Southeast	66
SP	São Paulo	Southeast	63
RJ	Rio de Janeiro	Southeast	18
ES	Espírito Santo	Southeast	13
PR	Paraná	South	39
SC	Santa Catarina	South	20
RS	Rio Grande do Sul	South	35
Total	11	3	306

Three connected factors lead to the restriction of the sample to micro-regions that belong to the Center-West, Southeast and South regions.

The first reason is related to characteristics of the Brazilian territory, its biomes in particular depicted in figure (1.3).[7]

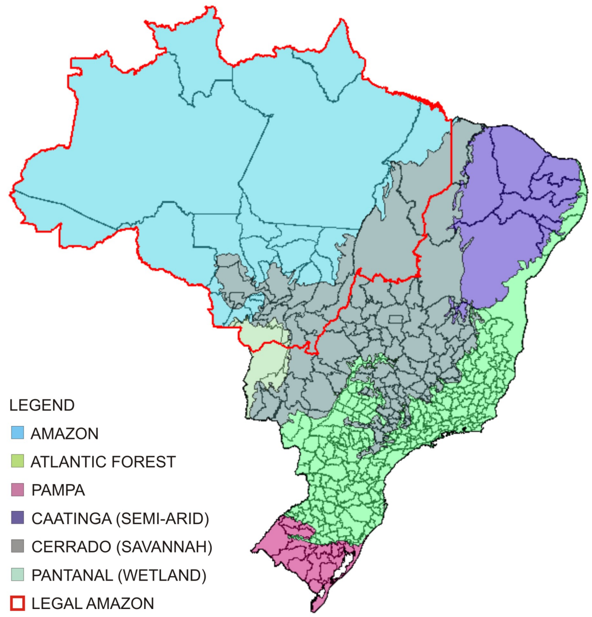

Figure 1.3: Brazilian Biomes and Legal Amazon

The Northeast region main biome is the *Caatinga*, or semi-arid, which is rather unsuited for most types of agriculture, sugarcane and soybean included. It is also covered by the Atlantic Forest biome in its coastline and the *Cerrado* biome in its western part.

The North region is almost entirely covered by the Amazon Biome, i.e. rainforest, with some parts covered by the *Cerrado*, Savannah, biome.

The second reason, which adds to the first, regards land use legislation. Law number 4,771 from 1965, known as the Forest Code, establishes a mandatory preservation area of native vegetation in rural properties, called Legal Reserve, that varies according to regions and biomes.

[7] The Brazilian Government defines biome as *"a set of vegetation types that covers large contiguous areas on a regional scale, with similar flora and fauna, as defined by the physical conditions prevailing in those regions"*. http://www.brasil.gov.br/sobre/environment/geography/

It is complemented by law number 5,173 from 1966 that defines the Legal Amazon, an area which encompasses the North region, the state of Mato Grosso in the Center-West region and part of Maranhão state in the Northeast region. Legal Amazon boundaries are presented in figure (1.3).

This legal framework determines the preservation area, or Legal Reserve, to be 80% for rural properties inside the Legal Amazon and in the Amazon biome, 35 % for properties inside the Legal Amazon and in the *Cerrado* biome and 20% for properties elsewhere in the country. Thus, a rural property inside the Legal Amazon covered by the Amazon biome can use only 20% of its area for agricultural activities. This number increases to 65% for rural properties inside the Legal Amazon covered by the *Cerrado* biome and to 80% in the rest of the country in any biome.

The third reason, regards logistic infrastructure in the North and Northeast regions which is less developed than in the rest of the country. In particular, as the agricultural frontier expands towards north, over the *Cerrado* biome in the North and Northeast regions, between the Amazon and *Caatinga* biomes, major investments will be necessary to improve transportation modes and upgrade ports on the northern coast in order to increase the incipient agriculture production outflow northbound.

These three effects combined results in a spatial concentration of output in the selected regions. The bulk of sugarcane, ethanol, soybean and biodiesel production is concentrated in regions inside the sample. In 2010, sugarcane production inside the sample amounted to 90.1 % of total Brazilian production, ethanol production ammounted to 93.3 %, soybean production inside the sample ammounted to 89.9 % and biodiesel production to 88.3 %.

Thus, including the remainder two regions would add little information and many noises to the sample. The ethanol industry also separates itself into these regional categories, as there is another smaller sugarcane producing region in coastline of the Northeast region. Additionally, 8 out of 10 existing crude oil refineries in the country are inside the selected sample. It results that the area of interest can be narrowed to the Center-West, Southeast and South regions.

The constructed dataset consists of a panel with variables from several sources. All variables sources are described in detail in Appendix B.

Data on sugarcane and soybean production, harvested area, planted area and yield per hectare per micro-region is obtained from the Brazilian Statistical Bureau, IBGE - *Instituto Brasileiro de Geografia e Estatística - Pesquisa Agrícola Municipal*.[8] Data encompass the period from 1990 to 2010, $t = 21$.

Formal employment in sugarcane and soybean production per micro-region in 31.12 of each year, is obtained form the Brazilian Ministry of Labor database, *Ministério do Trabalho e Emprego, RAIS - Relação Anual De Informações Sociais*.[9] Micro-region employment data is available from 1994 to 2010, $t = 17$.

International sugar and soybean prices are obtained from the US Department of Agriculture for the period 2008 to 2010, $t = 3$.

Transportation costs for sugarcane, sugar and soybean are obtained for the period 2008 to 2010, $t = 3$, from several publications from the University of São Paulo, *Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz*, ESALQ-USP.

Distances in km were obtained from the centroid of each micro-region to each port of interest taking a straight line using the software QuantumGis. This is also a simplification as accurate road, rail or river distances and quality of each of these infra-structures is not available.

1.6 Empirical Results and Discussion

1.6.1 The Agricultural Production Function

The aggregate neoclassical production function and its normalized form are estimated for each crop i. As the theoretical model requires that the exponents of the Cobb-Douglas production function be the same for both crops, a dummy variable I_{cane} is employed, where:

$$I_{cane} = \begin{cases} 1 & \text{if} \quad i = \text{sugarcane} \\ 0 & \text{if} \quad i = \text{soybean} \end{cases}$$

[8] www.ibge.gov.br

[9] www.mte.gov.br

Only the intercept, or *"technological parameter"*, may be different for each crop. Micro-regions or years with zero hectares harvested or zero output were dropped. When formal labor was equal to zero and the other variables were positive, observations were also dropped. This can happen for two reasons. First, output can be produced by informal labor even if formal labor is zero. There is no data on informal labor to overcome this hurdle. Second, employment data refer to 31.12 of each year, so it is possible that there was some formal employment during the year that generated output, but employment on the last day of the year was zero.

A Pooled OLS estimation in logs is employed. Equations (1.22) and (1.23) gives the pooled OLS between estimator specification for the aggregate production function and its normalized form.

$$\log \overline{Y}_n = \log A_{soy} + \gamma I_{cane} + \alpha \log \overline{X}_n + \beta \log \overline{T}_n + \epsilon_n \tag{1.22}$$

$$\log \overline{y}_n = \log A_{soy} + \gamma I_{cane} + \alpha \log \overline{x}_n + \theta \log \overline{T_n} + \nu_n \tag{1.23}$$

Equations (1.24) and (1.25) refer to the pooled OLS specification.

$$\log Y_{n,t} = \log A_{soy} + \gamma I_{cane} + \alpha \log X_{n,t} + \beta \log T_{n,t} + \eta_{n,t} \tag{1.24}$$

$$\log y_{n,t} = \log A_{soy} + \gamma I_{cane} + \alpha \log x_{n,t} + \theta \log T_{n,t} + \mu_{n,t} \tag{1.25}$$

In equations (1.23) and (1.25), the parameter θ is defined as $\theta = \alpha + \beta - 1$. Regression results are presented in table (1.2) with robust standard errors.

Table 1.2: Agricultural Production Function Estimation in Logs

	Between Estimator		Pooled OLS	
	Aggregate	Normalized	Aggregate	Normalized
log A_{soy}	0.482****	0.481****	0.339****	0.339****
	(0.088)	(0.088)	(0.034)	(0.034)
$\gamma = I_{cane}$	3.179****	3.179****	3.327****	3.327****
	(0.025)	(0.025)	(0.010)	(0.010)
α (Labor)	0.055****	0.055****	0.033****	0.033****
	(0.008)	(0.008)	(0.003)	(0.003)
β (Land)	1.022****		1.038****	
	(0.011)		(0.004)	
θ (Land)		0.077****		0.072****
		(0.008)		(0.003)
R-squared	0.992	0.972	0.985	0.971
adj. R-squared	0.992	0.972	0.985	0.971
F	15,632	5,006	130,000	68,465
p-value	0.000	0.000	0.000	0.000
AIC	101.8	102.3	1,661.5	1,661.5
BIC	118.1	118.6	1,687.4	1,687.4
N	438	438	4,835	4,835

Note: Robust standard errors in parenthesis.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

Among the four estimated equations the first model is selected, the between estimator of the aggregate production function, as it exhibits the highest adjusted $R^2 = 0.992$, an almost perfect fit. From this estimation, the values of the technological parameters $A_{cane} = e^{(0.481672+3.17899)} = 38.887$ and $A_{soy} = e^{0.481672} = 1.61878$ are calculated. These values represent output of each crop whenever 1 unit of labor is employed in 1 hectare of land. The 95% confidence intenrval for these parameters, with robust standard errors, are $A_{cane} \in [31.111, 48.606]$ and $A_{soy} \in [1.361, 1.925]$.

The estimated Cobb-Douglas production function for sugarcane and soybean are described in equations (1.26) and (1.27), respectively.

$$Y_{cane} = 38.89 X_{cane}^{0.055} T_{cane}^{1.022} \tag{1.26}$$

$$Y_{soy} = 1.61 X_{soy}^{0.055} T_{soy}^{1.022} \tag{1.27}$$

The production function exhibits increasing returns to scale, homogeneous of degree 1.077. This result can be almost fully attributed to the land input as $\beta > 1$. The 95% confidence interval, with robust standard erros, is $\beta \in [0.9997523, 1.044547]$. At 90% confidence level there are increasing returns to scale exclusively due to the land factor as $\beta \in [1.003365, 1.040934]$. A 94% confidence interval already guarantees a confidence interval where $\beta > 1$. Additonaly, the three remainder regressions exhibit increasing returns to scale in the land factor at 95% confidence level.

This result has major implications. First, the marginal product of land is increasing. As the aggregate production function reflects micro founded production functions, this implies that biofuel production or mandates are more feasibly met if there is land concentration. This result unveils that bioenergy production will work better under a plantation structure of land use rather than from small farmers. It requires vast extensions of land and consequently well defined property rights in rural areas.

Moreover, with increasing returns to scale on land factor, land competition becomes even more important to the agricultural and bioenergy sector.

Although this result is valid, some precaution is necessary.

First, regarding the absence of technological change in the model, which could be captured by a varying technological parameter $A_{cane,t}$ and $A_{soy,t}$. Evidence of technological change exists as yield per hectare for both crops do increase over time. But regressions fails to account for it as the technological parameter is treated as a constant.

Nevertheless, from the production function it is possible to compute the impact of technological change on land use. For simplification, if output and total labor are kept constant, each 1% increase in the technological paramenter, A_{cane} or A_{soy}, reduces land demand by 0.98%.

The second reason for precaution regards mechanization as capital is not included in the model. In sugarcane production, harvesting techniques are evolving replacing labor by machines, while soybean harvesting techniques are already highly mechanized.

Both technological change and mechanization contribute to increase yield per hectare while reducing or keeping constant labor input per hectare.

These effects have been more pronounced in the sugarcane industry as production techniques evolved, particularly with restrictions on fields pre-harvesting burning that frequently precedes manual harvesting, leading to substitution of labor for capital.

Other model assumption that soil is supposed to have everywhere the same fertility and there are no restrictions for agriculture are very unlikely to hold. Soil quality varies according to nutrient characteristics, irrigation requirements and declivity, which enables higher or lower degrees of mechanization.

A fixed effect model that could control for soil quality, climate, amount of rainfall in each micro-region cannot be employed because a time invariant dummy variable is required, the exponents of the production function need to be the same for both crops.

There are also some protected areas where agriculture is forbidden on the selected sample, especially in important biomes such as *Pantanal*, a wetland, and the Amazon Forest. Since these biomes are located at the boundary of the sample, far from selected ports, and not in the middle of agricultural land use zones, they do not impact the results.

Tests for autocorrelation of the error term are not presented as they depend on the arbitrary way observations were ordered. A correct measure of autocorrelation would require spatial regressions, but the theoretical land use model does not requires it.

1.6.2 Result 1

Result 1: *"...for each activity, less and less labor is used as one moves away from the market town so that the equilibrium output is decreasing and continuous in distance to the market town."*

Labor per Hectare and Distance

Labor per hectare as a function of distance from the two selected ports is tested using a between estimator specified in equation (1.28) and a pooled OLS specified in equation (1.29). Data is available from 1994 to 2010 and all micro-regions (n) are selected for both crops (i). Observations with zero labor per hectare were dropped as this procedure wipes out non-producing micro-regions.

$$\log \overline{x}_{i,n} = \omega + \phi \log \overline{r}_n + \epsilon_{i,n} \tag{1.28}$$

$$\log x_{i,n,t} = \omega + \phi \log r_n + \nu_{i,n,t} \tag{1.29}$$

Results for the between estimator are presented in table (1.3) and for the pooled OLS in table (1.4).

Table 1.3: Between Estimator of Log of Labor per Hectare and Log of Distance

	Sugarcane		Soybean	
	Santos	Paranagua	Santos	Paranagua
Intercept	-0.375	-1.403	-6.290^{****}	-7.757^{****}
	(1.215)	(1.681)	(1.024)	(0.784)
Log of Distance	-0.771^{****}	-0.594^{**}	-0.036	0.194
	(0.197)	(0.259)	(0.155)	(0.122)
R-squared	0.056	0.034	0.000	2.53
adj. R-squared	0.052	0.030	-0.005	0.003
F	15.29	5.25	0.06	1.546
p-value	0.000	0.023	0.814	0.113
AIC	948.7	953.8	683.0	681.5
BIC	955.5	960.6	689.5	688.0
N	225	225	199	199

Note: Robust standard errors in parenthesis.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

Table 1.4: Pooled OLS estimation of Log of Labor per Hectare and Log of Distance

	Sugarcane		Soybean	
	Santos	Paranagua	Santos	Paranagua
Intercept	-1.031^{***}	-2.145^{****}	-4.234^{****}	-6.812^{****}
	(0.315)	(0.420)	(0.384)	(0.297)
Log of Distance	-0.584^{****}	-0.394^{****}	-0.384^{****}	0.012
	(0.052)	(0.066)	(0.059)	(0.047)
R-squared	0.042	0.019	0.016	0.000
adj. R-squared	0.042	0.019	0.015	-0.000
F	125.8	35.9	42.5	0.060
p-value	0.000	0.000	0.000	0.804
AIC	9,024.0	9,081.1	8,922.1	8,961.4
BIC	9,035.6	9,092.7	8,933.7	8,973.0
N	2,382	2,382	2,453	2,453

Note: Robust standard errors in parenthesis.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

The model that best explain the relation between labor per hectare and distance for sugarcane production is the between estimator with distance from Santos Port as it exhibits the best adjusted R^2, although it is extremely low, and the best AIC and BIC. Results are to be interpreted as elasticities. In the sugarcane regression, each 1 % increase in distance away from Santos port decreases labor per hectare by 0.8 %. The 95% confidence interval for this parameter is $[-1.16, -0.38]$. The pooled OLS estimation is also good.

$$\log \overline{x}_{cane,n} = -0.375 - 0.771 \log r_n \tag{1.30}$$

For the relation between labor per hectare and distance for soybean production results of the pooled OLS for Santos port provide the best model, in terms of statistical significance, expected sign of coefficients and adjusted R^2. All other estimations for soybean do not exhibit statistically significant slopes for the parameter (ϕ).

$$\log \overline{x}_{soy,n} = -4.234 - 0.384 \log \overline{r}_n \tag{1.31}$$

Thus, each 1 % increase in distance away from Santos port causes labor per hectare to decrease 0.4 % in soybean production. The 95% confidence interval for this parameter is $[-0.50, -0.27]$.

Output per Hectare and Distance

The driving force of a declining output per hectare with distance, in the normalized production function, is the declining labor per hectare with distance. As labor per hectare is indeed declining with distance in both crops, as shown previously, it must be the case that also output per hectare is declining in distance for each crop. To test this result both a between estimator and a pooled OLS are employed, according to equations (1.32) and (1.33) respectively. Data on output per hectare is available from 1990 to 2010.

$$\log \overline{y}_{i,n} = \omega + \phi \log \overline{r}_n + \epsilon_{i,n} \tag{1.32}$$

$$\log y_{i,n,t} = \omega + \phi \log r_n + \nu_{i,n,t} \tag{1.33}$$

Regression results are presented in tables (1.5) and (1.6). Again, estimates are to be interpreted as elasticities.

Table 1.5: Between Estimator of Log of Output per Hectare and Log of Distance

	Sugarcane		Soybean	
	Santos	Paranagua	Santos	Paranagua
Intercept	3.961****	3.324****	−1.487**	-0.475
	(0.649)	(0.563)	(0.602)	(0.722)
Log of Distance	-0.034	0.066	0.303***	0.147
	(0.100)	(0.085)	(0.091)	(0.113)
R-squared	0.001	0.003	0.046	0.015
Adj. R-squared	-0.002	0.000	0.041	0.010
F	0.12	0.59	11.09	1.7
p-value	0.73	0.44	0.001	0.19
AIC	643.1	642.4	566.1	573.7
BIC	650.5	649.8	573.1	580.6
N	297	297	237	237

Note: Robust standard errors in parenthesis.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

Table 1.6: Pooled OLS Estimation of Log of Output per Hectare and Log of Distance

	Sugarcane		Soybean	
	Santos	Paranagua	Santos	Paranagua
Intercept	5.488****	4.593****	0.912****	0.925****
	(0.069)	(0.074)	(0.050)	(0.045)
Log of Distance	−0.256****	−0.114****	−0.023***	−0.026****
	(0.011)	(0.011)	(0.008)	(0.007)
R-squared	0.090	0.020	0.002	0.002
adj. R-squared	0.090	0.020	0.001	0.002
F	537.8	121.955	8.29	12.50
p-value	0.000	0.000	0.004	0.000
AIC	7,788.3	8,219.5	2,181.7	2,178.5
BIC	7,801.6	8,232.9	2,194.4	2,191.2
N	5,863	5,863	4,215	4,215

Note: Robust standard errors in parenthesis.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

For sugarcane production the best specification is the pooled OLS from Santos port.

$$\log \overline{y}_{cane,n} = 5.488 - 0.256 \log \overline{r}_n \tag{1.34}$$

This means for each 1 % increase in distance away from Santos Port, output per hectare falls by 0.25 %. The 95% confidence interval for this parameter is $[-0.28, -0.23]$.

For soybean production the best result is the between estimator from Santos port. Nevertheless, in this regression the sign of the coefficient is the opposite to that expected, i.e. output per hectare increases in distance. Results from the pooled OLS are also statistically significant and with the expected sign. For both ports, the between estimator gives opposite, unexpected results, compared to the pooled OLS.

To reinforce this point, from the production function, it is clear that a decreasing relation between labor per hectare and distance should imply a decreasing yield per hectare with respect to distance. In fact from the production function itself, each 1% decrease in labor per hectare should decrease yield per hectare by 0.055 %. From previous result, each 1% increase in distance, reduces soybean labor per hectare by 0.4%. It follows that each 1% increase in distance should decrease yield per hectare by 0.022 %, which is very close to the pooled OLS estimation of 0.023 % for Santos port. For this reason, the pooled OLS estimation is chosen. The coefficients are almost equal for both ports but the Paranaguá port regression is overall more significant in terms of adjusted R^2 and also the F test, AIC and BIC. But to be coeherent, the decline of labor per hectare in soyeban production is only statistically significant form Santos port, thus this is the selected equation.

$$\log y_{soy,n} = 0.912 - 0.023 \log r_n \tag{1.35}$$

The production function shows that soybean yield per hectare should decline by 0.022% for each 1% increase in distance. The regression result shows this relation to be 0.023%. The 95% confidence interval for this parameter is $[-0.049, -0.007]$.

1.6.3 Result 2

Result 2: *" . . . at a given distance from the market one and (except at hairline boundaries) only one agricultural product will be produced"*, i.e. there is *"specialization of land use"*.

Full specialization of land use, if any, is expected to be seldom observed. One of the most important techniques to maintain or regenerate soil quality is crop rotation. Also, land use for animal husbandry is expected to be connected to agricultural activities that provides its feed. This does not pose a problem as the analysis is restricted to the competition between soybean and sugarcane.

Land use specialization is first analyzed through the cumulative production of each type of crop as a function of distance from ports. If land use specialization exists, it is expected that the bulk of production of one crop is done up to a certain distance from the market, where transition to the other crop will occur. Consequently, the production of sugarcane and soybean shall not overlap in space.

Data on output per micro-region is available from 1990 to 2010. On figure (1.4) the average quantity produced on the aforementioned period of each crop and their cumulative density are depicted as a function of distance from ports.

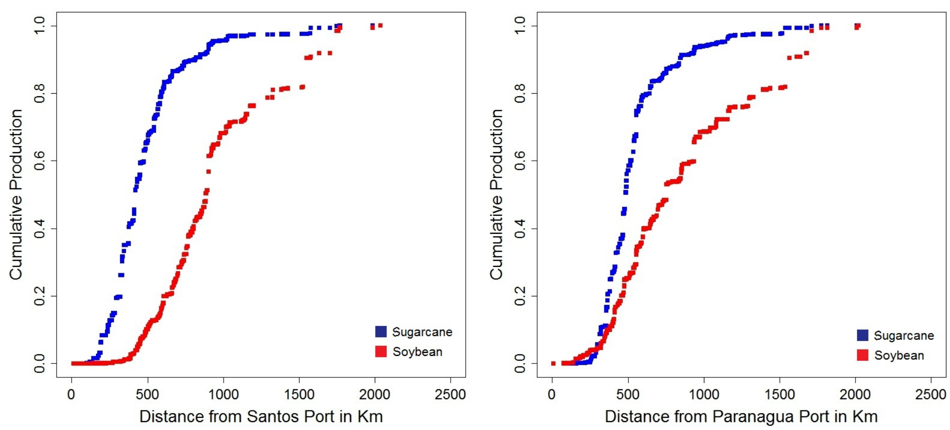

Figure 1.4: Cumulative quantity density and distance from ports

There is a certain overlap in space of production, nevertheless it is clear that sugarcane production is concentrated closer to the market in both cases. In particular, land use pattern from Santos port is better defined. Taking, for example a distance of 415 km from Santos port the cumulative production of sugarcane amounts to approximately 50 % while that of soybean has barely started.

The selection of a precise distance for a specific land use zone would be arbitrary. Nevertheless, if any choice is to be done, it needs to reflect at least where the bulk of production is done. In table 1.7, the cumulative density quantiles are presented.

Table 1.7: Average Production Quantiles from 1990-2010 and Distance in Km

	From Santos		From Paranaguá	
	Sugarcane	Soybean	Sugarcane	Soybean
Start	≈ 75	≈ 120	≈ 7	≈ 76
Q1 = 25%	≈ 320	≈ 685	≈ 390	≈ 490
Q2 = 50 %	≈ 415	≈ 885	≈ 480	≈ 754
Q3 = 75 %	≈ 565	$\approx 1,160$	≈ 570	$\approx 1,170$
End at 99 %	$\approx 1,575$	$\approx 1,770$	$\approx 1,535$	$\approx 1,775$

Based on these quantile results, the beginning and the end of each land use zone can be established, although, as already mentioned, in an arbitrary manner by selecting a cut-off point for cumulative sugarcane production, such as 75 %. Considering that from Santos port separation of land use zones is more evident, 565 km would be an approximate measure of the transition between sugarcane and soybean crops.

Another way to test this result is using the share of planted area for each crop per micro-region over time. Other crops are ignored, although some micro-regions in the sample have orange, coffee, cotton and cattle ranching, to cite some, as their main agricultural activity. Nevertheless, the focus is on the dichotomy between the two bioenergy crops.

Specialization of land use requires that whenever one micro-region produces one type of crop, it shall produce a very small quantity or nothing of the other crop. This effect can be captured by a negative relationship between land use for sugarcane and land use for soybean production in each micro-region.

For the pooled OLS the share of planted land for soybean PL_{soy}, is regressed on the share of planted land for sugarcane PL_{cane}. Observations where neither crop was produced were dropped. Data is available from 1990 to 2010. The between estimator is defined in equation (1.36) and the pooled OLS in equation (1.37).

$$\overline{PL}_{soy,n} = \eta + \theta \overline{PL}_{cane,n} + \epsilon_n \tag{1.36}$$

$$PL_{soy,n,t} = \eta + \theta PL_{cane,n,t} + \nu_{n,t} \quad (1.37)$$

Results are presented in table (1.8).

Table 1.8: Share of Soybean Planted Area as a Function of Sugarcane Planted Area

	Between Estimator	Pooled OLS
Intercept	0.057****	0.059****
	(0.007)	(0.002)
PL_{cane}	-0.015	-0.023
	(0.067)	(0.017)
R-squared	0.000	0.000
adj. R-squared	-0.003	0.000
F	0.05	3.15
p-value	0.825	0.0759
AIC	-433.4	-8,511.4
BIC	-426.0	-8,500.9
N	301	6,153

Note: Robust standard errors in parenthesis.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

Estimations show a negative relation, although the slope coefficient is not statistically significant. The model has an extremely poor fit, measured by the adjusted R^2. In the pooled OLS estimation the F test is statistically significant at 10%, validating the model.

Again, specialization of land use can be confirmed. In a given micro-region, as the share of land area employed in the production of sugarcane increases the share of area employed for soybean production decreases and vice-versa. Figure (1.5) make land use specialization more clear. As it is expected to be a binary phenomenon, either a micro-region shall produce sugarcane or soybean, the bulk of observations are on the axis implying in a clear pattern of land use specialization, but in an extremely poor fit for any linear model.

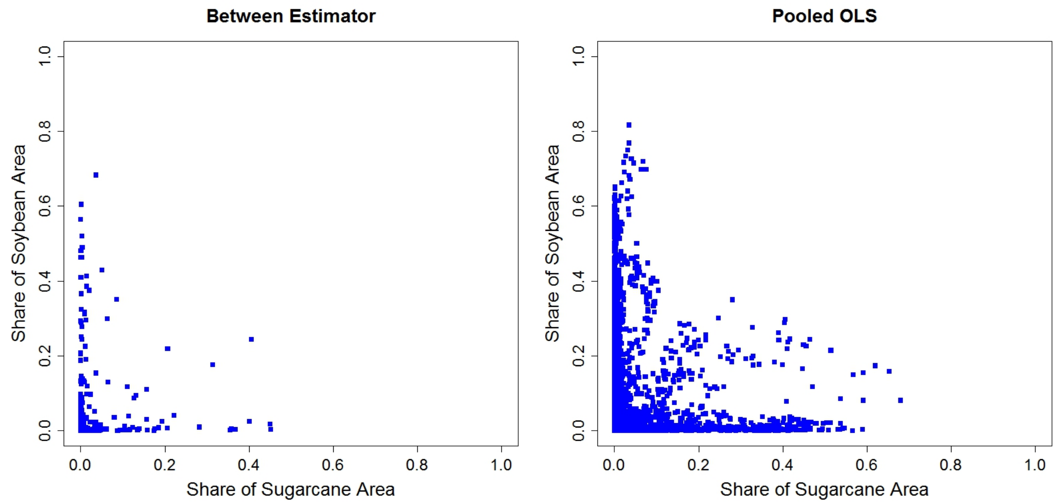

Figure 1.5: Share of Micro-Region Area used for Sugarcane and Soybean Production

1.6.4 Result 3

Result 3: *"The labor/land ratio is a continuous decreasing function of distance from the market."*

If in fact the underlying data generating process is a continuous function, then, an equation can be estimated with these characteristics linking labor per hectare for both crops (x_{cane}) and (x_{soy}) with distance (r). Moreover, the estimated slope coefficient has to be negative.

Since some micro-regions produce both agricultural products, another variable must be created to assess this result. Labor per hectare in each micro-region in both crops is defined by $x_{n,i+j}$. It represents the ratio of employment in both crops ($i+j$) divided by the amount of harvested land for both crops in each micro-region (n). Again, both Santos and Paranaguá ports are used as the reference measures of distance. The between estimator is specified in equation (1.38) and the pooled OLS is defined in equation (1.39).

$$\log \overline{x}_{n,i+j} = \alpha + \beta \log \overline{r}_n + \epsilon_n \tag{1.38}$$

$$\log x_{n,i+j,t} = \alpha + \beta \log r_n + \nu_{n,t} \tag{1.39}$$

Results are presented in table (1.9).

Table 1.9: Test for Result 3

	Between Estimator		Pooled OLS	
	Santos	Paranaguá	Santos	Paranaguá
Intercept	0.628	-3.598^{***}	3.033^{****}	-3.569^{****}
	(0.973)	(1.043)	(0.270)	(0.283)
Log of Labor	-1.005^{****}	-0.338^{**}	-1.376^{****}	-0.338^{****}
	(0.151)	(0.160)	(0.043)	(0.044)
R-squared	0.133	0.017	0.193	0.013
adj. R-squared	0.129	0.013	0.193	0.013
F	44.27	4.45	1,034.1	59.39
p-value	0.000	0.036	0.000	0.000
AIC	1,049.8	1,084.9	13,815.2	14,546.0
BIC	1,051.1	1,092.2	13,827.6	14,568.4
N	281	281	3,638	3,638

Note: Robust standard errors in parenthesis.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

The pooled OLS, having Santos port as the market center, has a better fit in terms of adjusted R^2. Both estimates from Santos port have a better fit in terms of adjusted R^2 compared to Paranaguá. All estimates have negative and statistically significant coefficients as expected. The selected estimated equation is the pooled OLS from Santos port.

$$\log \overline{x}_{n,i+j} = 3.033 - 1.376 \log \overline{r}_n \tag{1.40}$$

This imply that for each 1 % increase in distance, labor per hectare falls by 1.37 %. The 95% confidence interval is $[-1.46, -1.29]$.

Figure (1.6) depicts this relation for the pooled OLS for both ports.

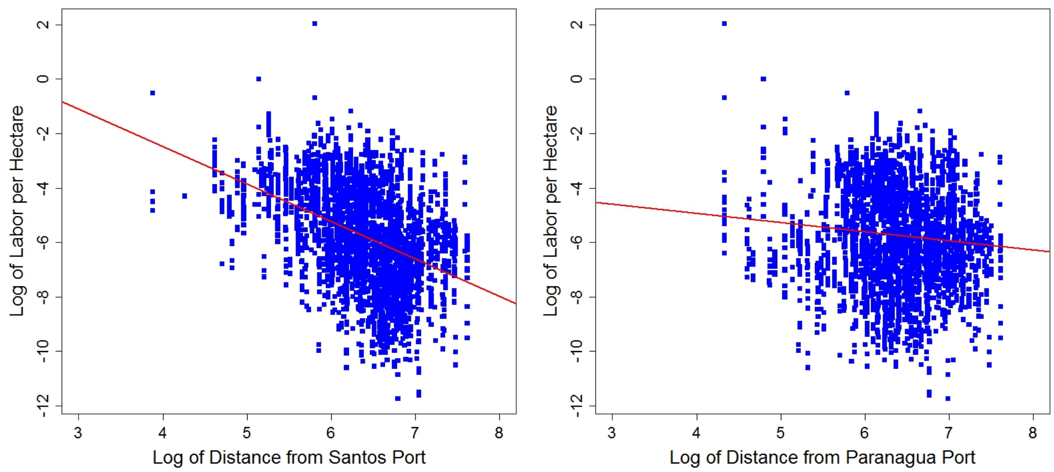

Figure 1.6: Average of Labor per Hectare and Distance

1.6.5 Result 4

Result 4: *"Land use zones are arranged in order of decreasing transport cost of the acre output produced with equal amounts of labor"* and at the transition of land use there is a *"discontinuity of transport cost for the output per unit of land area"*.

Only recent data from 2008 to 2010 is available regarding average road freight costs for sugarcane, soybean and sugar. Costs are presented in table (1.10) in USD/Ton/Km, and consequently are linear in distance and weight and can be interpreted as marginal transportation costs.

Table 1.10: Average Transportation Costs in Nominal USD/Ton/Km

	2008	2009	2010
Sugarcane (a)	0.1163	0.1427	0.1371
Sugar (b)	0.0623	0.0692	0.0852
Soybean (b)	0.0506	0.0585	0.0671
Ratio Sugarcane/Sugar	1.8668	2.0621	1.6097
Ratio Sugarcane/Soybean	2.2984	2.4393	2.0439
Ratio Sugar/Soybean	1.2322	1.1828	1.2697
Exchange Rate (c)	2.0854	1.8690	1.7128

(a) Data from University of São Paulo, Pecege-Esalq-USP detailed in Appendix B.

(b) Data from University of Sao Paulo, Esalq-Log-USP detailed in Appendix B.

(c) Yearly data from IPEA in BRL/USD (www.ipeadata.gov.br), average of bid and ask exchange rates.

In order to demonstrate this result, it is necessary to recall result 3, that labor per hectare is a continuous decreasing function of distance. It implies that on the transition of one land use zone to another, labor per hectare should be almost the same, i.e. $x_{cane}(r) = x_{soy}(r)$.

Suppose there is a point in space (r) where the transition from one crop to the other occurs. At the transition point (r), transportation costs to the market of the output produced with equal amounts of labor (x) in one hectare is given by equation (1.41) for sugarcane and equation (1.42) for soybean.

$$t_{cane} \times r \times A_{cane} \times x^{\alpha} \tag{1.41}$$

$$t_{soy} \times d \times A_{soy} \times x^{\alpha} \tag{1.42}$$

Plugging in the values of A_{cane} and A_{soy} from the production function and transportation costs for 2010:

$$t_{cane} \times A_{cane} > t_{soy} \times A_{soy}$$

$$0.1371 \times 38.887 > 0.0671 \times 1.61878$$

$$5.33 \quad \text{USD/km} > 0.11 \quad \text{USD/km}$$

Transporting the output per hectare at point (r) of sugarcane costs almost 50 times more than transporting soybean. Thus, the sugarcane land use zone has to be closer to ports. Furthermore, transportation costs falls discontinuously on the transition from one land use to the other. This result is invariant to any value of the 95% confidence interval for A_{cane} and A_{soy}.

$$0.1371 \times [31.111, 48.606] > 0.0671 \times [1.361, 1.925]$$

$$[4.27, 6.66] \quad \text{USD/km} > [0.09, 0.13] \quad \text{USD/km}$$

Taking the lowest boundary of A_{cane} and the highest boundary for A_{soy}, the transportation cost of the output per hectare at (r) is 30 times higher for sugarcane than for soybean. This result holds for 2008 and 2009 as well and again if the 95% confidence interval is employed.

If instead the average yield per hectare is used the same conclusions holds.[10]

$$t_{cane} \times d \times y_{cane} > t_{soy} \times d \times y_{soy}$$

$$0.1371 \times 79.044 > 0.0671 \times 2.947$$

$$10.84 \quad \text{USD/km/hectare} > 0.20 \quad \text{USD/km/hectare}$$

Using this alternative approach, logistic costs differ on each land use zone by approximately 55 times. This approach holds fro 2008 and 2009.

As the production of sugar or ethanol consists in a significant weight loosing process, each ton of sugarcane contains around 145 kg of total reducible sugars, yielding

[10] Brazilian average yield per hectare for 2010 in tons/hectare.

approximately 130 kilos of sugar, result 4 is tested as if sugar was directly harvested. Accounting for this, and using sugar transportation costs does not change the general validity of the result with respect to the pattern of land use and the discontinuous fall in transportation costs. The calculation is demonstrated for the year 2010 comparing sugar and soybean, but also holds for 2008 and 2009. Accounting for the 95% confidence interval for A_{cane} and A_{soy} does not alter the result.

$$0.0852 \times 38.887 \times 0.13 \quad \text{USD/km} > 0.11 \quad \text{USD/km}$$

$$0.43 \quad \text{USD/km} > 0.11 \quad \text{USD/km}$$

Transporting sugar output of one hectare costs 4 times more than transporting soybean output of the neighboring hectare with equal amount of labor. Again, transportation costs falls discontinuously on the transition from one land use to the other. Result 4 is validated. In any point in space it costs much more to transport the yield per hectare of sugarcane crops and sugar, than of soybean crops with with the same amount of labor.

This result reinforce the validation result 2, of land use specialization, that sugarcane and sugar production have to be closer to the market, i.e. ports, than soybean production due to their higher costs of transporting output per hectare. Thus, the first land use zone ring should be that of sugarcane, followed by an outer ring of soybean.

1.6.6 Bid-Rent Functions and Elasticities

It is possible to calculate bid-rent functions for each crop from 2008 to 2010, according to equation (1.10) using data presented on table (1.11). Since sugarcane is not directly exported, the bid-rent function as if refined sugar was directly produced from land and labor is calculated, considering that each ton of sugarcane yields 121.97 kg of refined sugar.[11]

[11] This figure is the amount of total reducible sugars used to price sugarcane at the processing mill. The use of raw sugar was also considered. The main change is to shift the bid-rent curve inward as raw sugar prices are lower than refined sugar.

Table 1.11: Parameters for Calculating Bid-Rent Functions

		2008		
		Sugarcane	Refined Sugar	Soybean
A_i	Technological Parameter	38.887	38.887	1.61878
T_i	Harvested Area (Hectares)	6,877,409	6,877,409	19,157,766
p_i	Price (USD/Ton)	17.3	351.9	388.5
t_i	USD/ton/km	0.1163	0.0623	0.0506
w	Wages (USD/year)	5,986.48	7,603.38	5,307.30
		2009		
		Sugarcane	Refined Sugar	Soy
A_i	Technological Parameter	38.887	38.887	1.61878
T_i	Harvested Area (Hectares)	7,385,248	7,385,248	19,612,381
p_i	Price (USD/Ton)	18.5	487.9	367.1
t_i	USD/ton/km	0.1427	0.0692	0.0585
w	Wages (USD/year)	7,250.46	10,000.19	6,646.93
		2010		
		Sugarcane	Refined Sugar	Soy
A_i	Technological Parameter	38.887	38.887	1.61878
T_i	Harvested Area (Hectares)	7,810,665	7,810,665	20,909,370
p_i	Price (USD/Ton)	24	612.4	359.8
t_i	USD/ton/km	0.1371	0.0852	0.0671
w	Wages (USD/year)	9,163.03	11,575.01	8,016.14

From table (1.11), it is possible to observe that one model assumption is violated, average wages are not equal for both crops. Average wage in sugarcane production is higher than in soybean production.

In figures (1.7), (1.8) and (1.9) bid-rent functions are presented for the years 2008, 2009 and 2010 respectively. The farthest distance in the sample from Santos port, is at 2,450 km to the northwest in Mato Grosso state, already inside the Amazon biome. For bid-rent calculations, results are presented up to 2,500 km.

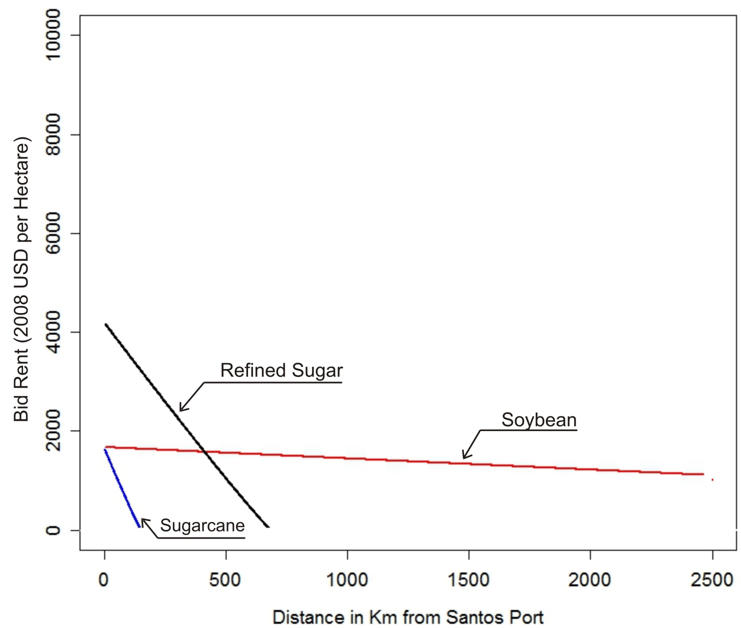

Figure 1.7: Bid-Rent Function for Sugarcane, Refined Sugar and Soybean in 2008

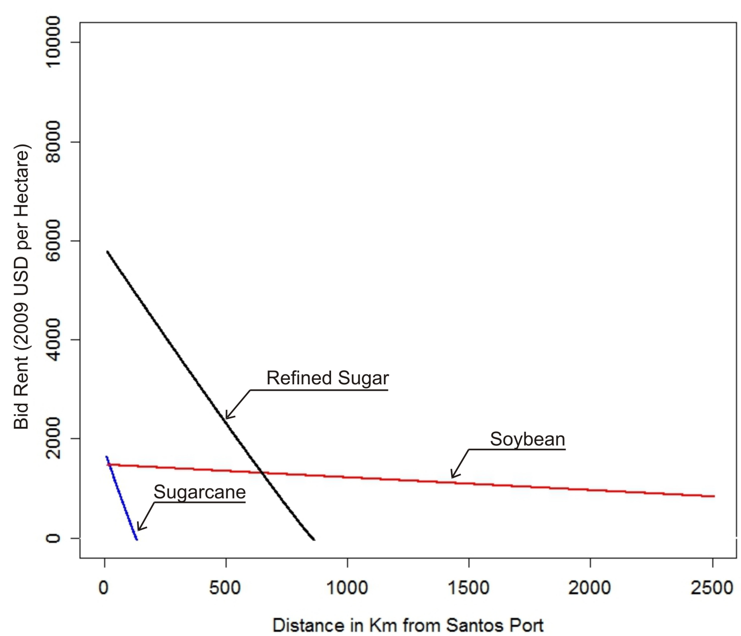

Figure 1.8: Bid-Rent Function for Sugarcane, Refined Sugar and Soybean in 2009

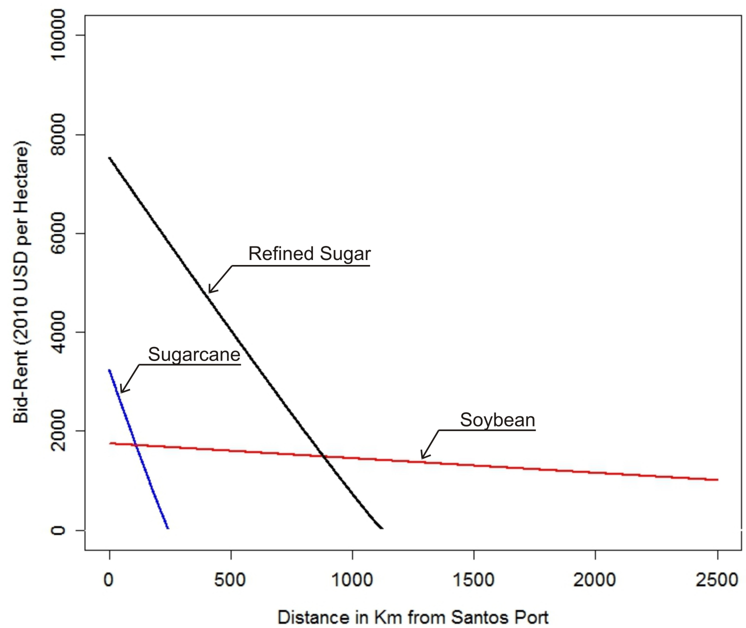

Figure 1.9: Bid-Rent Function for Sugarcane, Refined Sugar and Soybean in 2010

Bid-rent functions allows some conclusions.

Transporting sugarcane is very costly compared to other products. The upper limit for a sugar and/or ethanol mill radius of influence, or raw material procurement, can be estimated at 148 km in 2008, 129 km in 2009 km and 175 km in 2010. These distances measure where the sugarcane bid-rent function becomes zero in each year. This imply that sugar and ethanol mills must be near sugarcane supplies or supply oriented. There is a discrepancy between estimated and real values for the radius of a sugar/ethanol mill presented in table (1.12). Real data regarding average sugarcane procurement, indicates an approximate radius of 25 km for each mill.

Table 1.12: Sugarcane Average Procurement Radius

Year	2008	2009	2010
Average Distance in Km	24.90	21.50	26.00
From Bid-Rent	148	129	175

Note: Data from University of São Paulo.

Thus the area of procurement is determined not only by transportation costs, but by the scale of processing mills. A large sugarcane ethanol mill with an output of 1.000 m^3 of ethanol/day should have a radius of approximately 13.5 km, which is more in line with real figures.[12]

At 418.1 km in 2008, 641.4 km in 2009 and 909.2 km in 2010 land use pattern should change from "refined sugar" to soybean crop.

Bid-rent functions allow to calculate how the distance where transition of land use occurs change as prices and transportation costs varies, i.e. refined sugar own and cross elasticities. This can be interpreted a measure of competition for land between the two crops as market conditions varies.

Calculations are shown in table (1.13). Taking 2010 as the reference year, each 1% increase in refined sugar prices expands the inner ring radius by 1.28 %. Reducing soybean transportation costs is the variable that causes less impact on land use competition. Each 1 % decrease in soybean transportation costs should reduce sugarcane land use by 0.05 %. Sugarcane land use is elastic to its own variables and inelastic to soybean related variables.

In the period analyzed, from 2008 to 2010 both crops expanded their land use, but sugarcane and sugar prices increased significantly while soybean prices decreased. Thus, from elasticities, as sugarcane land use zone expands, it pushes soybean area further inland.

Table 1.13: Land Use Transition Distance Elasticity

Year	2008	2009	2010
Transition Distance in Km	418.1	641.4	909.2
Refined Sugar Price	1.71	1.39	1.28
Sugar Transport Cost	-1.03	-1.03	-1.04
Soybean Price	-0.71	-0.39	-0.28
Soybean Transport Cost	0.04	0.04	0.05

[12]This figure considers that each hectare of sugarcane produces 80 tons/year that can be converted into 6,400 liters of ethanol and the plant operates 365 days per year.

The same sign of elasticities found for refined sugar are expected to be observed for ethanol prices and ethanol transportation costs, but most likely with different values. The same reasoning holds for elasticities of soybean meal, soybean oil and biodiesel prices along with transportation costs of these goods compared to results of soybean price and transportation costs.

The soybean bid-rent function is quite flat. At a distance of 2,500 km from Santos port, already inside the Amazon Forest, soybean bid-rent is still positive. This means that there are still economic incentives to expand crop area over the Von Thünen "uncultivated wilderness". This does not imply that soybean caused, causes or will cause deforestation as there are mechanisms and policies to avoid this from happening. But, the bid-rent shows that there exists economic incentives for doing so, or at least, incentives to displace other activities towards the forest. Any policy to prevent deforestation has to take this opportunity cost into account, which amounted to 1,098.5 USD per hectare in 2008, 905.9 USD per hectare in 2009 and 776.5 USD per hectare in 2010. Even accounting for the Forest Code, that enforces preservation of 80% of land inside the Legal Amazon and Amazon biome, land rents are still positive as described in table (1.14).

Table 1.14: Land Rent of the Outermost Hectare

Year	2008	2009	2010
USD/Hectare	1,098.5	905.9	776.5
With Forest Code	219.7	181.18	155.3

Land rent elasticity of this outermost hectare at 2,500 km distance from Santos port is detailed in table (1.15).

Table 1.15: Land Rent Elasticities for an Hectare at 2,500 km

Year	2008	2009	2010
Soybean Price	1.57	1.76	1.98
Soybean Transport Cost	-0.51	-0.70	-0.92

The outermost hectare land rent is more elastic to price increases of soybean than to transportation costs improvements.

As soybean transportation costs causes the lowest impact on land use competition with sugarcane and the lowest impact on land rent of the outermost hectare, investments or policies to reduce this cost should be pursued.

1.7 Conclusion

This paper investigates empirically land use patterns for liquid biofuels production in Brazil using the neoclassical land use model. Sugarcane and soybean crops are analyzed as these are the main feedstocks employed in country's bioethanol and biodiesel production, respectively.

The estimation of the agricultural production function exhibits increasing returns to scale, mainly due to the land factor. This is the first result of the paper.

This has major consequences for biofuels policy making, to be further explored, mainly the difficulty of making small farmers participate in this industry. The frequent claimed social objective, such as rural development, from biofuels program may come through other channels as income in these micro-regions may increase but not by including small holders in the production process. Additionally, with increasing returns to scale, land competition becomes even more important to the agricultural sector and land concentration should be observed.

The Brazilian experience indicates that these monoculture agricultural rings, structured as plantations, are a *sine qua non* condition for biofuel development.

The absence of capital in the model is a caveat, as it should be an input into the production function. As no data is available to overcome this hurdle, it was not addressed. Another shortcoming is the absence of technological progress that could increase yields per hectare and reduce competition for land.

The role of non-renewable fertilizers in biofuel production, such as Nitrogen (N), Phosphorus (P), Potassium (K) has been quite neglected in biofuel modeling and policy making.

Next, the four results of the neoclassical land use model are tested empirically. Two major export ports, Santos and Paranaguá, are chosen as market centers that generates the spatial pattern of land use for sugarcane and soybean. Santos port is the most important for sugar and ethanol exports while Paranaguá port is the most important for soybean, soybean meal and soybean oil exports. All results of the neoclassical land use model have been strongly validated empirically.

Santos port is validated as the unique market center as all results stemming from the neoclassical land use model holds. The pattern of land use can be understood as circular rings from it. First, closer to the port, land is employed in sugarcane production due to its much higher output per hectare and higher transportation costs followed by the soybean land use zone which has lower output per hectare and lower transportation costs. Labor and output per hectare decline with distance from ports for both crops. There is, to a great extent, specialization of land use. And labor per hectare declines continuously from the market center.

This is the second result of the paper, which shows that one port in Brazil explains land use patterns inside the sample. The neoclassical land use model is able to explain the formation of these monoculture agricultural rings. The production function, prices, transportation costs and distances, generates the patterns of land use. Of special importance are biofuel policies, which affects demand and prices of agricultural products, logistics, which determine transportation costs and technology, in determining zones of land use.

Transporting sugarcane is extremely costly compared to other products. The upper limit for a sugar and/or ethanol mill radius of influence, or raw material procurement, can be estimated at 164 km in 2008, 173 km in 2009 km and 245 km in 2010. These distances measure where the sugarcane bid-rent function becomes zero. The actual average procurement radius is much lower around 25 km. This is not a transportation cost restriction but a technological restriction in the scale of processing mills. It implies that sugar and ethanol mills must be near sugarcane supplies or supply oriented.

The distance where transition of land use occurs is elastic to sugar price and transportation cost and inelastic to soybean price and transportation costs. As land use for both crops is increasing, and so does the transition point, it must be that sugarcane crops are pushing soybean crops further inland.

The soybean bid-rent function is quite flat. At 2,500 km of distance from Santos port, already inside the Amazon Forest, soybean bid-rent is still positive. This means that there are still economic incentives to expand the crop area over the Von Thünen "uncultivated wilderness". This does not imply that soybean caused, causes or will cause deforestation as there are mechanisms and policies to avoid this from happening. But, the bid-rent shows that there exists economic incentives for doing so, or displacing other activities towards the forest. Any policy to prevent deforestation has to take this opportunity cost into account, which amounts to 1,435 USD per hectare in 2008, 1,132 USD per hectare in 2009 and 1,004 USD per hectare in 2010. Thus around 1,000 USD per hectare, throughout the period. Even considering the Forest Code land rents are positive, amounting to 20% of the aforementioned values.

As soybean transportation cost has little impact on land use competition and on the land rent of the outermost hectare, reducing logistic costs for this industry should be pursued. Each 1% decrease in soybean transportation costs, captures 0.05 % of land use for sugarcane and increases land rent of the outermost hectare by 0.72 %, *ceteris pariburs*.

Logistics in general, and transportation and port infra-structure in particular, can be significantly improved in Brazil reducing export bottlenecks.

The formation of the agricultural ring precedes biofuel production. The Brazilian experience shows moreover that the agricultural commodity is mainly exported while the biofuel is sold domestically.

The patterns of land use predicted in the neoclassical land use model, under hybrid land use for food and energy, is unchanged, although the hybrid land use affects significantly the variables inside the model and intensifies fuel on fuel competition for land use.

1.8 Appendix A

Table 1.16: Summary Statistics for Production Function - Table (1.2)

Variable	Units	Mean	Std. Dev.	Min.	Max.	N
Log Output	Tons	11.027	3.062	-2.140	16.904	438
I_{cane}	Dummy	0.523	0.500	0	1	438
Log Land	Hectares	8.535	2.757	-2.833	13.993	438
Log Labor	Workers	2.871	2.952	-2.833	9.037	438
Log Yield per Hectare	Tons/Hectare	2.491	1.610	0.187	4.631	438
Log Labor per Hectare	Labor/Hectare	-5.664	1.897	-11.421	3.135	438
Log Output	Tons	12.117	2.354	1.386	17.264	4,835
I_{cane}	Dummy	0.493	0.500	0	1	4,835
Log Land	Hectares	9.638	2.157	0	14.469	4,835
Log Labor	Workers	3.930	2.390	0	9.644	4,835
Log Yield per Hectare	Tons/Hectare	2.479	1.676	-1.148	4.807	4,835
Log Labor per Hectare	Labor/Hectare	-5.708	1.889	-11.718	4.384	4,835

Table 1.17: Summary Statistics for Result 1 - Tables (1.3) and (1.4)

Variable	Units	Mean	Std. Dev.	Min.	Max.	N
Log Cane Labor per Hectare	Workers	-5.208	2.037	-11.513	1.553	225
Log Distance Santos Port	Km	6.268	0.626	3.878	7.620	225
Log Distance Paranaguá Port	Km	6.411	0.637	2.001	7.613	225
Log Soy Labor per Hectare	Workers	-6.525	1.336	-9.721	0.134	199
Log Distance Santos Port	Km	6.472	0.545	4.705	7.620	199
Log Distance Paranaguá Port	Km	6.364	0.609	4.333	7.613	199
Log Cane Labor per Hectare	Workers	-4.647	1.643	-10.127	4.385	2,382
Log Distance Santos Port	Km	6.190	0.579	3.878	7.620	2,382
Log Distance Paranaguá Port	Km	6.351	0.576	2.001	7.613	2,382
Log Soy Labor per Hectare	Workers	-6.738	1.502	-11.513	2.426	2,453
Log Distance Santos Port	Km	6.514	0.493	4.705	7.620	2,453
Log Distance Paranaguá Port	Km	6.364	0.585	4.333	7.613	2,453

Table 1.18: Summary Statistics for Result 1 - Tables (1.5) and (1.6)

Variable	Units	Mean	Std. Dev.	Min.	Max.	N
Log Cane Output per Hectare	Tons/year	3.744	0.711	-0.743	4.557	297
Log Distance Santos Port	Km	6.321	0.631	3.878	7.620	297
Log Distance Paranaguá Port	Km	6.376	0.633	2.001	7.613	297
Log Soy Output per Hectare	Tons/year	0.454	0.813	-3.390	1.049	237
Log Distance Santos Port	Km	6.396	0.571	4.615	7.620	237
Log Distance Paranaguá Port	Km	6.322	0.667	2.001	7.613	237
Log Cane Output per Hectare	Tons/year	3.861	0.493	1.281	4.807	5,863
Log Distance Santos Port	Km	6.353	0.577	3.878	7.620	5,863
Log Distance Paranaguá Port	Km	6.404	0.615	2.001	7.613	5,863
Log Soy Output per Hectare	Tons/year	0.762	0.314	-2.120	1.347	4,215
Log Distance Santos Port	Km	6.450	0.530	4.615	7.620	4,215
Log Distance Paranaguá Port	Km	6.352	0.587	2.001	7.613	4,215

Table 1.19: Summary Statistics for Result 2 - Table (1.8)

Variable	Units	Mean	Std. Dev.	Min.	Max.	N
Share of Cane Planted Area	Percentage	0.030	0.074	0	0.452	301
Share of Soy Planted Area	Percentage	0.057	0.117	0	0.683	301
Share of Cane Planted Area	Percentage	0.031	0.080	0	0.681	6,153
Share of Soy Planted Area	Percentage	0.059	0.121	0	0.816	6,153

Table 1.20: Summary Statistics for Result 3 - Table (1.9)

Variable	Units	Mean	Std. Dev.	Min.	Max.	N
Log Cane + Soy Labor per Hect.	Workers	-5.751	1.673	-11.188	-0.725	281
Log Distance Santos Port	Km	6.348	0.606	3.878	7.620	281
Log Distance Paranaguá Port	Km	6.378	0.646	2.001	7.613	281
Log Cane + Soy Labor per Hect.	Workers	-5.720	1.780	-11.731	2.028	3,638
Log Distance Santos Port	Km	6.361	0.574	3.878	7.620	3,638
Log Distance Paranaguá Port	Km	6.357	0.613	2.001	7.613	3,638

1.9 Appendix B

Coordinate Reference System

All shapefiles were projected using SAD69/Brazil Polyconic projection which uses as references, the 0° parallel (Equator line) and the 54° West meridian. All distances were obtained in this projection.

Granularity

As the political division of 2007, Brazil had 5 levels of granularity. Country (1), Regions (5) equivalent to NUTS 1, States (27) equivalent to NUTS 2, Meso-Regions (137), Micro-Regions (558) equivalent to NUTS 3, Municipalities (5564). Shapefiles were obtained from www.ipea.gov.br retrieved in January 2011.

Agricultural Data

Quantity of sugarcane and soybean produced per year, yield per hectare, harvested and planted area according to the Brazilian National Statistics Bureau, *Instituto Brasileiro de Geografia e Estatistica*, www.ibge.gov.br, Pesquisa Agrícola Municipal. Data at micro-region level from 1990 to 2010.

Labor Data

Obtained from Labor Ministry, *Ministério do Trabalho e Emprego*, www.mte.gov.br, RAIS, retrieved in November 2011. Data on formal employment as of 31.12 of each year, from 1994 to 2010 per micro-region.

Prices

Refined sugar and soybean international prices used in the computation of bid-rent functions from USDA for 2008 to 2010. Data on exchange rates from IPEA, Instituto de Pesquisa de Economia Aplicada, www.ipea.gov.br. Bid and ask yearly average. Average wages from from Labor Ministry, *Ministério do Trabalho e Emprego*, www.mte.gov.br, RAIS from 2008 to 2010.

Transport Costs

Data on sugar and soybean transportation costs from, University of São Paulo, ESALQ-LOG. Data for sugarcane transportation costs were estimated using University of São Paulo, Pecege-ESALQ-USP, reports which presents data on production costs for the sugarcane, sugar and ethanol industry for 2008 [54], for 2009 [42] for 2010 [41].

For example, for the year 2008 it includes 4 transportation costs of sugarcane in BRL/ton for the traditional plantation area of 5.82 BRL/Ton for sugarcane stalks and 5.92 BRL/Ton for sugarcane billets and an average mill radius of 27 Km. On the expansion area transport costs are of 6.40 for sugarcane stalks and 5.80 for sugarcane billets with an average mill radius of 22.8 Km. First take the average of both types of sugarcane transportation cost for each region, 5.87 and 6.10 respectively. Then divide by the average radius to obtain the figures in BRL/Ton/Km for the traditional and expansion area. Divide both figures by the corresponding exchange rate and take the average. The same procedure was used in all years.

Chapter 2

Biofuel Mills: Location and Capacity Decisions in Brazil

2.1 Introduction

Ethanol production is expanding in several parts of the world from feedstocks such as maize, sugarcane and sugar beet. In the United States and Brazil, ethanol from maize and sugarcane respectively, already contributes to displace significant amounts of gasoline in Otto cycle engines.

Biodiesel production is also expanding worldwide, from a myriad of oil crops such as soybeans in the Americas, oil palm in Asia, rapeseed in Europe, cottonseed, sunflower, castor beans and jathropa, to cite some and also from animal fats. In 2010, Europe was the world's leading biodiesel producer, with output concentrated mostly in Germany and France.

The biofuel industry stands on the intersection of two mature global markets, agricultural commodities which provide feedstocks for their production and energy commodities, especially crude oil, which are the end use products being displaced.

Agricultural commodities markets are distorted by current international trade framework, mainly allowing for subsidies and barriers. In fact, agricultural commodities trade is one, if not the most, controversial issues on the unsettled Doha Development Round started in 2001. Crude oil and oil products markets are also distorted by the well known oligopolist market structure of the industry.

Considering that the expansion of the biofuel industry is taking place on the intersection of these two global markets, that deviate from perfect competition, understanding how biofuels programs starts and becomes successful is important for industrial organization, for private decisions and public policies.

The goal is to build an understanding about the economic forces at work in the formation of these bioenergy producing regions, drawing upon the Brazilian experience on ethanol and biodiesel, by eliciting regional features that attract investments in biofuels production.

This paper investigates empirically, using regional variables, location and capacity decision drivers for ethanol and biodiesel mills in Brazil as of 2011.

The choice of Brazil follows from the country being in 2010, the second major world producer of ethanol and biodiesel and thus a rare case of producing large quantities of both biofuels to be analyzed. But the choice is not restricted to these figures.

The country expects to be a major crude oil producer and exporter (of crude oil or its products) and also a major ethanol exporter. Brazil expects to increase oil production three fold by 2020, up to 6 million barrels per day, and half of this figure will be beyond its needs and available to international markets. Ethanol production is also expected to increase three fold to approximately 70 billion liters per year, of which 10 % will be exported.[1] Also according to Mckinsey (2007) [52], the country is bound to be a major player in ethanol export markets.

If this scenario is confirmed, the claim that biofuels or renewable energies reduce foreign oil dependency will not hold anymore, at least for some time frame. In fact, this claim cannot hold for any net oil exporting country. The net environmental consequences need also to be reassessed, because as biofuels displace oil products in the domestic market, more oil surplus is generated to be exported and consumed elsewhere.

Other frequent claims about biofuels programs will still hold partially. There will be continued creation of green jobs in the biofuel industry, but for the reasons explained above, there will be also continued job creation in the oil industry. Finally, the claim that biofuels increase energy security also holds, not by energy availability itself, but

[1] Forecasts from Empresa da Pesquisa Energética. www.epe.gov.br, Plano Decenal de Expansão da Energia, 2020.

by the diversification of the energy mix.

Of course these inconsistencies will be only transitory, lasting an unknown number of years. In the long run, as crude oil is produced and depleted, all the claims about biofuels will hold again.

Nevertheless, this awkward situation is worth pointing out because it helps to highlight one important characteristic of the Brazilian biofuel industry, ethanol in particular. It has become a business on its own, where profits can be made. And as long as there are profits in the oil and biofuel industries, employment and output will expand, and this alone, explains exporting both crude oil and ethanol.

The novelties in this paper are five. First, together with the previous chapter, it consists of an integrated microeconomic analysis ranging from the agricultural location of feedstock crops to the location of biofuel mills, from "*fields to fuels*" or from "*soil to oil*".

Second, previous studies do not address determinants of installed capacity. Preceding research employ either binary dependent variables, in probit or logit regressions, or count data regressions to analyze location decisions of agro-industries and corn ethanol mills.

This paper overcomes this aspect and analyzes also installed capacity decisions. For ethanol mills, the number of workers in ethanol manufacturing in each micro-region is used as proxy for capacity, while for biodiesel mills actual installed capacity in each micro-region is available.

Profits are not only determined by siting decisions but also by the choice of the scale of operations, especially if there are economies of scale in biofuel production. The dynamics of installing a small biofuel mill are much different of installing a large one in terms of capital investment, feedstock requirements and market access.

Investigating capacity decisions also helps to explain a frequent feature observed in Brazil of multiple mills in the same micro-region. This approach is expected to yield better results than count regressions for two reasons. Count regressions can overweight the presence of many small mills in detriment of one large mill and capacity expansions are neither captured by binary nor by count variables.

Although variables that affect location decisions may not be the same as those that affect mills capacity, in the case analyzed herein, capacity decisions are driven by almost the same set of variables that drive location decisions.

Third, analyses of ethanol mills locational drivers have been carried out for the US but have never been done for Brazil, which exhibits different market dynamics.

For example, different from corn ethanol, sugarcane processing yields sugarcane bagasse as by-product which is used to produce bio-electricity. Thus, new variables are employed, reflecting local conditions such as access to power transmission grid.

Fourth, there is no similar empirical evidence regarding the location of biodiesel mills. For biodiesel mills, a whole new set of variables are necessary.

Fifth, combining information of location drivers for both types of biofuel mills, it is possible to analyze where first generation bio-refineries could locate.

A Probit regression shows that ethanol mills are located in micro-regions with abundant sugarcane production, low feedstock price, with high river density, near ethanol storage terminals and significant local demand for ethanol measured by the number of automobiles. A Tobit regression shows that the capacity of ethanol mills, measured by the number of workers in ethanol manufacturing in each micro-region, is determined by the same set of variables plus the past number of cattle heads in the micro-region.

Ethanol suppliers are facing problems to keep up with increasing demand, especially during sugarcane off season. Identifying the set of variables that affect location and capacity decisions is important to promote supply investments and reduce excess demand. If Brazil is going to export ethanol, than much more production capacity will be needed along with additional sugarcane supply. Moreover, due to past and current regulatory framework, there is a huge untapped potential for bio-electricity production derived from sugarcane bagasse. In order to benefit from this potential, more investments will be required.

Biodiesel mills are located in abundant soybean producing micro-regions, near soybean crushing mills and with large number of workers in raw vegetable oil production, a proxy for the scale of soybean crushing mills. Biodiesel mills capacity are determined by the same variables.

The biodiesel industry currently has huge excess capacity caused by over-entry. The biodiesel program gives locational incentives, through tax exemptions, to attract producers to poor regions of the country and to use castor beans and oil palm as feedstocks, especially if procured from small farmers. But as the market evolved, the usage of soybean oil as a feedstock revealed to be much more economically attractive. This resulted in an initial movers disadvantage, misled by locational incentives. Thus, not all existing biodiesel mills are optimally positioned to use soybean oil as feedstock and there are still many empty optimal locations. So paradoxically, this is an industry with excess capacity and continued entry that can be explained by mislocation of many mills. To support this idea, there is evidence that entry is occurring near existing soy crushing mills where entrants face lower production costs than mislocated incumbents.

Another minor source of mislocation regards the choice of technological route. Biodiesel production requires an alcohol in its process, either methanol or ethanol. Some millers opted for the ethylic route and placed themselves near ethanol mills, betting on the country's large ethanol availability. But biodiesel production process using ethanol, with existing technology, is more expensive and less efficient than using methanol. Additionally, there is already excess demand for ethanol for direct use as a gasoline substitute and consequently a significant opportunity cost of using it to manufacture biodiesel.

As the ethylic route can be adapted to the methylic route, the bulk of biodiesel production is done using methanol, which is a non-renewable feedstock normally produced from natural gas. This has raised concerns that biodiesel is not a completely renewable fuel. And has also soared methanol imports. But since the technological route can be reverted, the mislocation becomes again a matter of feedstock procurement costs.

Combining information on the location of both types of biofuel mills allows to identify where first generation bio-refineries can be located. Most of these sites coincide with where transition of land use from one crop to the other occurs. But with current technology and the fact that there is just one case of an integrated ethanol-biodiesel mill, it is very difficult to evaluate if there are economies of scope in the production of both biofuels. Additionally, it is the existence of sugarcane production, due to its higher logistics cost and after harvest decay that determines the possibility of co-existence of both biofuel mills.

It is found that the formation of these bioenergy producing regions is endogenous and obeys to great extent the economic forces identified by theoretical and empirical locational models. This endogeneity, also identified by Hausmann and Wagner (2009) [29], means that the formation of these bioenergy producing regions do not depend exclusively on their exogenous natural conditions that determines its agriculture potential, such as land availability, soil quality, water availability and adequate climate conditions. These clusters are formed endogenously and evolve due to economic conditions, markets and market structure, policies, deployed technologies, interaction with other industries, research and development, infrastructure and institutions among others.

The main implication of this study is to understand what are regional features in Brazil that attract investments in biofuels production. Results can help to formulate policies to promote biofuel industry expansion and to foster regional development.

Understanding the Brazilian experience with biofuels can be useful to other countries investing in this industry. For example, Hausmann and Wagner (2009) [29] proposes to replicate the Brazilian experience with ethanol in other tropical countries.

Finally, these regional revealed comparative advantages on first generation biofuels production can be used to understand industry dynamics and help advancing to second generation biofuel production, including not only sugarcane and soybean but also the paper and pulp industry.

2.2 Liquid Biofuels in Brazil

2.2.1 Ethanol

The Brazilian ethanol program was the first successful large scale biofuel implementation. Its success, according to Goldemberg and Moreira (1999) [23], is the result of an intended long-run government policy launched in 1975, the Pró-Alcóol program, as a reaction to the first oil shock in 1973 and falling sugar prices in international markets. The program aimed at improving macroeconomic conditions by reducing the trade deficit, replacing imported fuel, and strengthening the sugar industry by creating another market for sugarcane and sugar producers. On the onset of the ethanol program in 1975, neither environmental nor social objectives were present.

In 1979, the same year of the second oil shock, production of neat hydrated ethanol fueled automobiles started. Licensing of new neat hydrated ethanol vehicles peaked in 1986 with approximately 90 % of maret share.[2]

Towards the end of the 1980s as crude oil prices decreased and sugar prices in the international market increased, the country begun to face ethanol supply shortages, aggravated by a car fleet mainly comprised of neat hydrated ethanol fueled engines. This undermined consumers' confidence in a continuous ethanol supply without price spikes and in 1990 licensing of new gasoline vehicles surpassed that with neat ethanol engines.

To counteract declining ethanol demand, law 8.723 from 1993, established a countrywide mandatory blending of 22% of anhydrous ethanol into gasoline.

In 1996, Ministry of Finance ordinances 59, 242 and 244 liberalized wholesale and retail gasoline and ethanol prices. Another landmark for the industry was the reform of the national oil and gas sector and creation of the National Petroleum Agency in 1997.

In 2003, production of flexible fuel vehicles, that can run with any mixture of hydrated ethanol and gasoline started, giving a new boost to the industry. This technological breakthrough had a very fast market penetration and in 2005 licensing of flexible fuel vehicles surpassed that of gasoline vehicles. Currently, the bulk of light-duty vehicles sold in Brazil are flexible fuel vehicles, although specialists claim that flexible fuel vehicles are less energy efficient than pure gasoline or hydrated ethanol engines.

Pousa et al. (2007) [44] stress that since the 1980s no pure gasoline is used in Brazil, only a blend of gasoline and anhydrous ethanol called gasohol or Gasoline C, that currently according to law 12.490 from 2011 can vary from 18% to 25%, also known as E18 and E25. As blending anhydrous ethanol in petroleum gasoline increases fuel performance acting as an octane enhancer, this allowed Brazil to be the first country to abandon the addition of lead into gasoline.

In the last decade, from 2001-2010, the domestic ethanol industry has undergone a consolidation process with merger and acquisition of mills, foreign direct investment

[2]Data from Anfavea; Associação Nacional dos Fabricantes de Veículos Automotores. www.anfavea.com.br

and capacity expansions. Following the global trend identified by Chen and Reiner (2011) [9], Brazil had not only foreign entrants from the sugar and ethanol sector itself but also from other industries such as food processors, agro-commodity traders, oil and gas companies, engineering and construction companies and investors from the banking industry.

In 2010, Brazil was the second largest world ethanol producer, with an output of 486 thousand barrels per day, using only sugarcane as feedstock.[3]

The world top producer since 2005 is the United States that produced in 2010, 867 thousands barrels per day. Although the ethanol volume produced in the US is almost twice as that in Brazil, gasoline market sizes are also very different. It results that the ethanol market share in Brazil is much higher than in the US. Furthermore, ethanol market share in Brazil is endogenously determined by price competition of hydrated ethanol with gasoline and a blending mandate of anhydrous ethanol into gasoline.

In 2010, sugarcane products, ethanol and bio-electricity from sugarcane bagasse, ranked as the second source of energy supply in Brazil with a share of 19.3 % or 48.9 billion tons of oil equivalent, behind only crude oil with a share of 42.1 % or 106.5 billion tons of oil equiavalent. It is the major source of renewable energy as hidroelectricity, the second largest renewable source, has a share of 13.7 % or 26.8 Gtoe.[4] In fact, Weidenmier et al. (2008) [53] provide evidence of the macroeconomic benefits of the Brazilian ethanol program caused by reduced oil imports and reduced business cycle volatility.

Currently no subsidies are given to ethanol producers in Brazil, but ethanol and gasoline have different taxation schemes, in which gasoline has a higher tax burden. Also flexible fuel vehicles have lower taxation than its gasoline counterpart.

Overall, Brazilian ethanol is very cost competitive. Goldemberg et al. (2004) [22] have shown that current competitiveness of Brazilian ethanol with fossil fuels results from economies of scale, technological progress and learning-by-doing in the sugarcane and ethanol industry.

[3]Data from US Energy Information Administration, www.eia.gov

[4]Data from Brazilian Ministry of Mines and Energy, www.mme.gov.br

2.2.2 Biodiesel

The biodiesel program is more recent, dating back to 2005. To a certain extent, it is linked to the ethanol program, but not only by the experience the country has on producing and marketing a biofuel. Pousa et al. (2007) [44] propose that biodiesel production could benefit from Brazilian large ethanol availability as the transesterification process to produce biodiesel requires an alcohol that can be either methanol or ethanol.

The program, enforced by law 11.097 from 2005, established a countrywide optional blending of 2% of biodiesel (B2) into refinery diesel from 2005 to 2008. In 2008, the optional blending of 2 % became mandatory and was expected to ramp up to 5 % (B5) by 2013. In july, 2008 mandatory blending increased to 3%, and one year after, in July, 2009 to 4%. Since January, 2010 the country's mandate is to blend 5 % of biodiesel (B5) with refinery diesel. Thus, biodiesel market share is exogenously determined by a blending mandate into diesel.

The biodiesel program started with an economic objective, reducing diesel imports[5], an environmental objective, displacing fossil fuel, and according to Pousa et al. (2007) [44] a strong social objective.

The program was designed for the usage of any vegetable oil and tallow but placed strong incentives, through tax exemptions, for the use of certain feedstocks produced by small farmers in less developed regions of the country.

The fiscal regime and tax incentives for biodiesel production were introduced by law 11.116 from 2005 and decrees numbers 5.297 from 2004 and 5.457 from 2005.

This set of legislation equalized biodiesel taxation with refinery diesel for two federal taxes, namely PIS/PASEP[6] and COFINS[7], and allowed for exemptions on the same tributes that vary according to the region of the country where the biodiesel mill procures its feedstock and thus where it is located, the feedstock used and the type of feedstock supplier.

A tax exemption of 30.5 % is granted if the raw material used as feedstock, from

[5]Brazil diesel net imports as of 2010 amounted to 6.3 million tons of oil equivalent.

[6]PIS/PASEP, Programa de Integração Social/Programa de Formação do Patrimônio do Servidor Público.

[7]COFINS, Contribuição para o Financiamento da Seguridade Social.

any type of supplier, is castor beans or oil palm, produced in the poorest regions of the country, the north and northeast regions, which are not in the sample of this study.

Decree number 5.297 from 2004 also introduced the Social Fuel Stamp, a certification given to each biodiesel mill that meets a minimum feedstock purchase requirement from small farmers, which entitles further tax exemptions.

These minimum purchase requirements were set by the Ministry of Agrarian Development Ordinance number 2 from 2005 at 10% for the Center-West and North regions, at 30% for the South and Southeast regions and 50% for the Northeast region.

The Social Fuel Stamp grants full tax exemption if the raw material used as feedstock is castor beans or oil palm, in the north and northeast regions of the country. It also grants a tax exemption of 67.9 % for any feedstock produced in any region of the country. This legal framework represents a clear federal locational incentive for private agents.

Locational incentives can be justified, according to Glaeser (2001) [20], because they stir firm siting decision to regions where greater social surplus and/or greater agglomeration economies can be generated. Although, the author stress that the debate if locational incentives generate or correct spatial distortions is still unsettled. In the case of the Brazilian biodiesel program, these federal locational incentives had a clear objective of creating producer surplus on local input markets of raw material and labor in the poorest regions of the country.

Nevertheless, blending mandates immediately determines the size of the market for biofuels. And as claimed by Lee et al. (2008) [38], the cheapest way to meet volume requirements is to use or expand an already competitive agricultural crop, rather than through small farmers or new crops. In fact, that is exactly what happened in Brazil with biodiesel.

Instead of relying on the production of castor beans or oil palm from small farmers in the Northeast and North regions, biodiesel producers found more economically attractive, despite all government incentives to use these feedstocks, to use soybean oil as the main feedstock and to locate in the Center-West, Southeast and South regions where the bulk of soybean production occurs. Far behind soybean oil, the second most used feedstock is tallow also in the same regions.

There are no direct incentives to use soybean oil or tallow as the main feedstocks, just the standard tax exemption applied if the biodiesel mill holds the Social Fuel Stamp.

In 2010, Brazil was the second largest world biodiesel producer, with an output of 41.3 thousand barrels per day, behind Germany which produced 49 thousand barrels per day.

There is huge excess capacity for biodiesel production in Brazil caused by over-entry. The installed capacity to produce biodiesel in the country (5.8 billion liters) is more than twice as that required to meet the 5% blending mandate (2.4 billions liters) in 2010, according to the National Petroleum Agency. Put differently, at the same time the country has potential to double output and meet a B10 demand and become the world top producer, and more than half of the country's installed capacity is currently idle and the industry will undergo some structural changes.

Over entry probably has been caused by economic and technological uncertainties regarding feedstocks and production techniques coupled with government incentives to locate mills in poor regions of the country to foster small farmers agriculture.

Another factor that may have caused over entry is that biodiesel is sold only in auctions to the Brazilian National Oil Company, Petrobras, at prices above diesel prices.

There has been already cases of market exit due to bankruptcy, mainly of first movers, ownership change and backward integration with agriculture production.

Most likely initial movers have made a siting decision to benefit from government locational and feedstock incentives, which revealed ex-post, to be insufficient to compete with other more economically viable feedstocks, mainly soybean oil.

Thus, it could be that initial movers mills, guided by federal incentives, are not optimally located for the use of soybean oil as feedstock and as the market evolved subsequent location decisions improved. Another set of mills may have opted to use other vegetable oils as feedstock not directly contemplated by federal incentives, and are placed near these sources of vegetable oil, but given current economic conditions are also misplaced. Or alternatively, are not at an optimal location for an unique feedstock but have more input flexibility.

Summing both biofuels Brazil ranks second behind the United States, producing 28 % of world biofuels in 2010, or 527 thousand barrels per day. This number is more than two fold the European figure, that produced 248 thousand barrels per day of biofuels in the same year.

The birth of both Brazilian biofuel programs were decided by the government. But the success of both Brazilian biofuel programs is linked to the previous existence of an agro-industrial complex, such as sugar, soybean and meat processing, that were already competitive in international markets. Moreover, in the Brazilian case, food output is mainly directed for exports while biofuel output is directed to domestic markets.

This conjecture holds also for US maize ethanol. As of 2010 the US was the world top producer and exporter of maize. It also holds for biodiesel production from soybean in the United States (1st producer and 1st exporter) and Argentina (3rd producer and 3rd exporter), from palm oil in Indonesia (1st producer and 1st exporter) and Malaysia (2nd producer 2nd exporter) and from rapeseed in European Union (1st producer and 2nd exporter).[8]

The success of both Brazilian biofuels programs is also linked to major efforts in agricultural R&D and cooperation among producers, researchers, government and the automotive industry.

2.3 Literature Review

This paper is related mainly to previous literature on agroindustry location and agri-food districts, both theoretical and empirical.

Hsu (1997) [33] presents the first theoretical framework for agroindustry location, which defines agroindustry, where biofuels mills can be included, as *"the industry of processing of agricultural products"*. The most important distinction between a non-agroindustrial producer and an agroindustrial producer is that the latter uses an input of the von Thünen type, i.e. ubiquitously supplied that has to be transported to the processing facility. It follows that Hsu's model is directly linked to the neoclassical land use model proposed by Beckmann (1972) [3] and employed in the previous chapter.

[8] Data from FAOSTAT as of 2010.

The model exhibits a single profit maximizer price taker firm choosing optimal location inside an exogenously given land use zone. That is, the agroindustrial firm has to decide where to locate and what input mix to use given already existing market center and agriculture production that is shipped to the market center by farmers.

Hsu demonstrates that the optimal location of an agroindustrial firm is different from its non-agroindustrial counterpart. The author proves that the duality that the profit maximizing site coincides with the total cost of transportation (TTC) minimizing site for non-agroindustrial manufacturing, may not always be true for an agroindustrial firm. Accordingly, the author derives existence conditions of an interior optimal location for both cases, a total cost of transportation minimizing firm and a profit maximizing firm.

For a total cost of transportation minimizing firm, Hsu demonstrates that if output per hectare is *"monotonically decreasing in distance ... it is possible to have an interior optimal location ..."*. This result also holds if output per hectare is monotonically increasing.

It was demonstrated empirically in the previous chapter that output per hectare is declining in distance, in line with the neoclassical land use model. Both sugarcane and soybean output per hectare declines in distance from Santos port, the market center. Thus, the necessary condition that output per hectare is monotonically decreasing on the range of each zone of land use is satisfied and an interior optimal location for a total transportation cost minimizing biofuel mill should also exist.

For a profit maximizing firm, Hsu shows that the existence of an optimal interior solution requires a concave profit function with respect to distance from the market, otherwise boundary locations will be preferred. According to the author, *"A price-taking agroindustrial firm will locate at one of the end-points if the firm's profit function is non-concave over space, that is,* $\frac{\partial^2 \pi}{\partial r^2} \geq 0$*"*.

Results in Hsu's model are hard to test empirically as they are existence conditions or comparative statics conclusions, but they provide the closest theoretical framework that justifies the existence of optimal interior locations for biofuel mills. This mitigates the caveat that regressions employed here are not directly based on the results of a theoretical model.

Kilkenny and Coleman (2006) [36] try to bridge the gap between theory and empirics, and propose the first empirically testable model of agro-industry location derived directly from first order conditions. In their model an initial plant locates on the center of an existing raw material producing region and this site becomes the market center. Subsequent plants decide not only their location but also the size of their input supplying area to maximize profits which are a function of the input-output ratio, fixed and variable costs, input and output prices and transportation costs. The authors find that any site, including intermediate locations, can be optimal because the total cost function is strictly convex in distance and the profit function is concave in distance. They propose a three category typology for agroindutries: i) concentrated, when a new plant locates on the same site of an existing plant; ii) co-located, when new plants locate within the input market area of an incumbent plant and iii) dispersed, when new plants locate outside the input market area of incumbent plants. This typology is tested empirically for 7 agro-industrial sectors in the United States. The dependent variable is the minimum distance between a pair of plants on the same industry transformed into a discrete variable according to the aforementioned typology. The input-ouput ratio, relative transport costs ratio, and the endogenous radius of each plant input area are used as explanatory variables. The authors validate their typology, finding for example that soybean processing is dispersed, and propose that all *"agro-industry is input oriented and city located"*. One caveat of the model, recognized by the authors, is that the price of raw material inputs for entrant plants and competition for it among plants is not endogeneized. As showed by Jones and Krummel (1987) [34] and Hsu (1997) [33] the location decision of an agro-industry modifies the price schedule and bid-rents over space.

These theoretical models do not include technological progress that can increase yields per hectare and/or modify the processing input/output ratio. Increased yields per hectare can lead to the concentration of several plants in the same site, which can be considered equivalent (except maybe for input competition) as an expansion of an already existing plant.

These agricultural raw material processing, value added activities, share also many characteristics with the concept of agri-food districts put forward in Brasili and Fanfani

(2006) [7], which in turn is related to the concept of industrial districts (IDs) or local production systems. The authors defines IDs as *" . . . systems of enterprises and institutions which interface in a specific geographic area to produce specialized and specific types of products".* These industrial districts are characterized by agglomeration of small and medium enterprises in a region to benefit from positive externalities created by the spatial concentration and specialization of production. Hausmann and Wagner (2009) [29], mention that agglomeration externalities occur in ethanol production.

Markusen (1996) [39] investigates the characteristics of regions that have been successful in attracting and maintaining value-added activities and creates a typology for these *"sticky places"* comprised of four categories of industrial districts. Industrial districts are defined as *"a sizable and spatially defined area of trade-oriented economic activity which has a distinctive economic specialization, be it resource related, manufacturing, or services."* The author refutes the idea that Marshallian districts are the unique form of industrial agglomeration and unique solution for regional development. The research based on regions that exhibited above average growth in United States, Japan, South Korea and Brazil leads to the proposal of three additional forms of industrial agglomerations; the hub-and-spoke district, the satellite platform and state-anchored districts. In reality, industrial districts may exhibit hybrid features of the pure types of districts. The four types of industrial districts are distinguished by characteristics such as market structure, labor markets, regional growth dynamics, income distribution, responses to changes in markets and technologies, to cite some. Thus each type of industrial district yields a different welfare outcome for regional economies.

Brasili and Fanfani (2006) [7] highlights that the main difference between industrial and agri-food districts regards their geographical coverage. Agri-food districts can encompass a much larger area due to the characteristics of its inputs. This is the case analyzed here where production of sugarcane, sugar and ethanol and of soybean, soybean meal, soybean oil and biodiesel cover significant land area. Another feature of these agri-food districts highlighted by Brasili and Fanfani is the importance of exports on industry revenues.

The cases analyzed here of biomass energy production in Brazil resembles more the hub-and-spoke district. Their market structure is characterized by few vertically integrated firms, economies of scale and scope are important, there are economic links with other firms outside the district, district firms have influence on local and even national politics and trade associations play an important role in industry coordination.

Furthermore, these are natural resource and technology anchored districts. As natural resources cannot be reallocated, together with exploitation technologies, it gives rise to natural resource anchored industrial districts, such as biomass energy districts.

A model for vertically integrated industries is proposed by Venables (1996) [51]. In this model, the existence of an upstream industry that supplies an intermediate good for a downstream industry, generates *"demand and cost linkages"* between them, that depending on transport costs, can lead to agglomeration of both industries in a single location.

Regarding spatial competition, comes into play the economic forces first described by Hotelling (1929) [32], but in a reversed form. Millers have a quasi-monopsonistic power due to spatial limitations. Agriculture producers will be willing to sell to a closer mill even if it is not the highest prevailing price to reduce logistic costs or the deterioration of the raw material. Another aspect pointed out by Hotelling, is that one company can control more than one mill and thus the price of raw material in more than one point of the market. In fact, several prices can exist for the same homogeneous commodity at the same time. Hotelling also stress that there exists a strong tendency for suppliers of the same commodity to cluster in nearby points of space.

Spatial competition turns the cluster into a gravitational pull for new entrants leading to excessive concentration of suppliers if compared to a centrally planned solution that minimizes transportation costs.

The agglomeration of producers in one region is driven also by other economic benefits such as knowledge spillovers of production techniques, usage of shared infrastructure and concentration of specialized labour supply and suppliers in general, to cite some.

Unfortunately the link between agricultural and agro-industry location theories is not so direct. There are two aspects not very well resolved due to their high complexity. The first regards the fact that the decision of locating an agriculture processing plant drives the price of raw materials in the surroundings of the plant, increasing it, and consequently increasing land rents. Agents should anticipate this effect in their decision process to locate a biofuel mill. Second, the spatial competition for agricultural inputs among plants and their eventual dynamic entry or exit is not theoretically modeled.

On the pure empirical side, Henderson and Macnamara (2000) [30] investigate using negative binomial regressions locational factors in 936 counties, in 10 states in the US corn belt that attracted investments in new food manufacturing plants from 1987 to 1995. Fourteen explanatory variables were classified into 5 categories: input and output markets, labor market, infrastructure, agglomeration economies and fiscal policy. The dependent variable, number of new plants in each county, was classified into supply oriented, demand oriented and footloose firms. In all regressions, at least one variable from each category was relevant to location decisions but the set of relevant variables varied with the type of firm considered. Results showed also that agglomeration economies, measured by the number of existing food processing plants in a county and if the county was a regional economic node, contributed positively to location decisions. Although this evidence does not capture the effects of competition among plants that produce the same product. The authors conclude that supply oriented firms seek sites with favorable input market conditions while demand oriented firms seek sites with favorable output market conditions.

Sarmiento et al. (2012) [48], employed logistic regression and spatial correlation to elicit locational factors that drives siting decision of ethanol mills in the contiguous United States, 48 states. The authors used as explanatory variables agricultural features of each county, state subsidies to ethanol producers and two forms of spatial interaction. The first spatial interaction reflects the fact that access to raw material, measured in acreage planted to corn, is very important in locational decisions not only at the county level but also from neighboring counties. The second spatial interaction regards the competition among ethanol mills. An incumbent ethanol mill repels investment in new plants in its surroundings, or put differently, plants avoid competition for raw materials among them. State subsidies are also important in locational decisions. Finally, the presence of livestock in the county favors locational decisions as one by-product of ethanol production, dry distillers grains, can be sold as feed. Although the causality in this last result can be self-reinforcing, or eventually reversed. That is, the availability of a low cost feed for livestock attracts cattle ranching activities, after the ethanol mill is already installed.

Lambert et al. (2008) [37], analyzes using probit regressions ethanol mills location in the contiguous United States from 2000 to 2007. The authors use 19 explanatory variables grouped into 5 categories: input and output markets, labor market, infrastructure, local fiscal policies and industry incentives. They find that feedstock access is the most relevant variable in determining siting decisions. Access to by-product markets is also important. The authors employ exploratory spatial data analysis to identify clusters of counties that exhibit high probability, assigned from the probit regressions, of hosting an ethanol mill surrounded by other counties with high probabilities, high-high clusters.

Haddad et al. (2010) [25], investigated using probit regressions the locational factors that drives siting decision of ethanol mills in the US Midwest corn belt. More precisely, for 381 counties in Iowa, Illinois, Minnesota and Nebraska. Explanatory variables were separated into 5 categories at the county-level: input availability, infrastructure, education of labor force, market and community concern. The authors validated the hypothesis that decision making of plant location is made in two-steps, with increasing granularity, and that the set of meaningful variables that determines location changes in each step. For the first step the authors found that access to the feedstock is the determinant factor in plant siting decisions, as ethanol plants locate in abundant corn producing municipalities. In the second step, other variables comes into play such as access to markets and to infrastructure. They conclude that corn bioethanol production is a supply-oriented industry.

Kilkenny and Coleman (2006) [36], criticize this empirical literature as *"ad-hoc"* models, not coming explicitly from a profit maximizer decision maker. These procedures can be regarded as unveiling a multicriteria analysis of decision makers for plant location.

Another problem mentioned in Henderson and Macnamara (2000) [30], is that binary variables do not capture that existing plants may invest in expansions. In fact, the expansion of planted area over other crop for biofuel production, or increases in yield per hectare may justify investments in plant expansions, or investments in additional plants in the same region.

All these models do not capture the fact that plants do vary in their installed capacity. To illustrate this problem, the smallest biodiesel plant in the dataset has an installed capacity of 864 $m^3/year$, located in Varginha, and the largest 486,720 $m^3/year$, located in Rondonópolis, a difference of 563 fold. Certainly, investors have preferences not only about optimal location but also about optimal capacity. If there exists any possibility of increasing returns to scale or economies of scale in biofuel production, addressing this issue becomes extremely important.

Additionally, a firm that has many biofuel plants may have preferences regarding the location of its portfolio. Last, the time line of investments is also an important variable to understand the dynamics of sequential locational decisions of incumbents and new entrants. These last two issues are not addressed in this paper.

2.4 Conceptual Framework

Agriculture and energy markets share an important feature, which is the spatial mismatch between supply and demand. Agriculture output has to be harvested and transported to processing facilities, warehouses, cities or export ports. Crude oil has to be transported from oil fields to oil refineries by pipelines or vessels, coal has to be transported from coal mines to power plants and industries, natural gas has to be transported by pipelines or liquefied, shipped and regasified and electricity has to be transported in power transmission and distribution grids in order to reach final demand. Therefore, significant transportation, processing and ancillary infrastructure is necessary in order to connect supply and demand, and bring these markets to equilibrium. Biofuel markets are no different as they inherit characteristics from both agricultural and energy markets.

Accordingly, besides standard supply and demand dynamics, this additional infrastructure dimension should be taken into account when dealing with these markets. More specifically, to deal with this spatial detachment between supply and demand, a profit maximizing firm or central planner need to include these transportation, processing and ancillary costs into their decision making process.

As the goal is to investigate location and capacity decision drivers of profit maximizing biofuel mills, location theory provides an adequate framework, because it addresses exactly problems where supply and demand do not necessarily coincide in space and transportation costs are relevant. In particular, the theoretical model of agroindustry location developed by Hsu (1997) [33] is presented in more detail, with focus on how the author has modeled the firm profit maximization problem. Thus, all the equations presented in this section are drawn from this study.

The objective of analyzing this model is to understand what economic forces are at work and what variables are necessary to approximate empirically a biofuel mill profit function at the micro-region level. Even though a micro-region can host more than one biofuel mill, the representative biofuel firm refers to or is aggregated at the micro-region level.

Hsu (1997) [33] models the profit maximization problem of an agroindustrial firm with a neoclassical production function exhibited in equation (2.1). The firm uses as inputs raw materials (M), for example sugarcane or soybean produced according to the neoclassical land use model, and labor (L) to produce an output (Q) that can be either an upgraded agriculture commodity, a biofuel, or both.

$$Q = F(L, M) \tag{2.1}$$

Capital is not employed in the model. The amount of raw material produced at a certain point in space at distance (x) from the market center, yield per hectare, is defined as $\rho(x)$. Raw material (M) has to be shipped from fields to the agroindustrial plant only in the market direction, without any reverse hauls. A plant sited at a distance (r) from the market center uses input (M) as defined in equation (2.2).

$$M = \int_{r}^{r1(r)} \rho(x)dx \tag{2.2}$$

Where ($r_1(r)$) is the limit where feedstock is procured. Thus, the agroindustrial firm employs all raw material produced from (r) to ($r_1(r)$). Hsu assumes that the agroindustry is a price taker on input markets and pays all farmers the price ($p_m(r)$) which is the prevailing price of raw material (M) at site (r), where the plant is located.

Additionally, the costs of transporting raw material to site (r) are incurred by the firm according to a transportation cost function (t_m).

The cost of feedstock procurement, $C(M)$ is given in equation (2.3). The first term on the right hand side refer to the cost of raw material and the second term refer to the cost of transporting raw material to site (r).

$$C(M) = p_m(r)\int_r^{r_1(r)} \rho(x)dx + \int_r^{r_1(r)} t_m(x-r)\rho(x)dx \tag{2.3}$$

Hsu assumes that workers have to commute from the market center to the plant site. Here the assumption is that labor is readily available at any site without any modification in results. The agroindustrial firm profit function is presented in equation (2.4).

$$\pi = (p_q - t_q(r))Q - wL - C(M) \tag{2.4}$$

Where (p_q) is the price of output at the market center, (t_q) is the output transport cost function and (w) is wage. Plugging in the expression for $C(M)$ from equation (2.3), yields:

$$\pi = (p_q - t_q(r))Q - wL - p_m(r)\int_r^{r_1(r)} \rho(x)dx - \int_r^{r_1(r)} t_m(x-r)\rho(x)dx \tag{2.5}$$

Profit maximization requires choosing optimal combination of inputs (L^*) and (M^*) and the optimal location (r^*), for given prices and transportation cost functions.

From equation (2.5) it is possible to observe that the agroindustrial firm profit function depends on variables referring to total revenues, output price and quantity, or demand related, variables referring to total production costs, input prices and quantities, or supply related and variables referring to input and output transportation costs, or infrastructure related.

Therefore, a biofuel mill profit function for micro-region (i), can be approximated empirically by local supply (S), infrastructure (I) and demand (D) conditions. Moreover, the biofuel mill profit maximization problem can be regarded as the choice of an optimal location and an optimal installed capacity for that location.

It results that, for a profit maximizing biofuel mill, location and capacity decisions can be captured by regional supply (S), infrastructure (I) and demand (D) variables. This is the empirical approach employed here.

2.5 Sample, Dataset and Empirical Model

2.5.1 Sample

The sample consists in part of the Brazilian territory, the Center-West, Southeast and South regions or 306 (out of 558) micro-regions as of the political division of 2010. Two regions are dropped, the Northeast and North.

The reason for that is because the bulk of sugarcane, ethanol, soybean and biodiesel production is done in the regions inside the sample. In 2010, sugarcane production inside the sample amounted to 90.1 %, ethanol production ammounted to 93.3 %, soybean production inside the sample ammounted to 89.9 % and biodiesel to 88.3 % of total Brazilian production. Including these two regions would add little information to address this investigation and lots of noises to the sample. The ethanol industry separates itself into these regional categories, as there is another much smaller sugarcane cluster in the coastline of the Northeast region.[9]

The level of analysis chosen is the micro-region, which is composed by a set of municipalities and is similar to NUTS 3 classification in Europe or counties in the United States. This choice is done for two reasons. The first is to obtain a more balanced sample with micro-regions that host a biofuel mill and those who does not. The second is because there is evidence that feedstock procurement is done not only inside the boundaries of the hosting municipality, but also on neighboring municipalities.

The selected sample, with the code of each state and the location of the main export port, used as market center, are depicted in figure (2.1).

The sample description with Brazilian political regions, states, number of micro-regions in each state and the number of micro-regions with one or more biofuel mill in each state is presented in table (2.1). Data on existing biofuel mills refer to 2011.

The usage of tallow as a feedstock for biodiesel production is excluded from the

[9] A detailed motivation of sample selection is presented in Chapter 1.

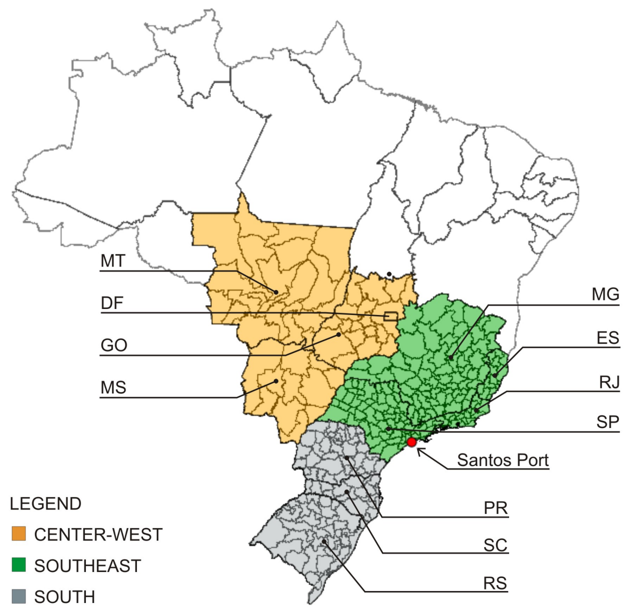

Figure 2.1: Sample Description and Market Center

present analysis, as well as those biodiesel mills running on it. This led to the exclusion of 5 mills, but only 3 micro-regions. Mills running on tallow had an installed capacity of 292,042.8 $m^3/year$, representing 5% of the Brazilian total capacity of 5,837,929.2 $m^3/year$ at the end of 2010.

The reason for excluding it, is because tallow is a secondary feedstock in biodiesel production and directly linked to livestock processing industries, which exhibits other non-agricultural dynamics. Of the remaining biodiesel mills, almost all, process soybean oil. So the focus can be restricted on two specific crops or feedstocks, sugarcane used for ethanol production and soybean used for biodiesel production.

Table 2.1: Sample Description as of 2011

Code	State Name	Region	Number of Micro-Regions	With Ethanol Mills	With Biodiesel Mills*
DF	Distrito Federal	Center-West	1	0	0 (0)
GO	Goiás	Center-West	18	10	6 (7)
MT	Mato Grosso	Center-West	22	7	11 (11)
MS	Mato Grosso do Sul	Center-West	11	8	2 (3)
MG	Minas Gerais	Southeast	66	16	5 (5)
SP	São Paulo	Southeast	63	42	7 (7)
RJ	Rio de Janeiro	Southeast	18	2	1 (1)
ES	Espírito Santo	Southeast	13	4	0 (0)
PR	Paraná	South	39	17	2 (4)
SC	Santa Catarina	South	20	0	0 (1)
RS	Rio Grande do Sul	South	35	2	6 (8)
Total	11	3	306	108	40 (47)

* Number in parenthesis include already authorized entries.

The starting year of the analysis is 2005. In this year, licensing of flexible fuel vehicles surpassed that of gasoline vehicles and the biodiesel program started with an optional blending target up to 2 %. This is the best year to start analyzing both industries recent dynamics.

2.5.2 Dataset

The constructed dataset is in the form of a cross-section, though some variables embed information averaged over a 6 years time frame.

In particular, all non spatial variables were averaged from 2005 to 2010 to smooth eventual fluctuations in agriculture output, employment, prices, livestock and road transportation fleet. For all spatial variables only recent data is available and the evolving dynamic of distances among them cannot be captured.

The same spatial coordinates of each micro-region centroid were assigned to every facility inside a micro-region, with the exception of Santos port, even if they actually are located somewhere else in space inside the micro-region. This is a simplification, because not all facilities have their exact spatial coordinate and to avoid that one facility in a micro-region was actually closer to another micro-region centroid.

For spatial variables, distances in km were obtained from the centroid of each micro-region to each point of interest taking a straight line using the software QuantumGis. This is also a simplification as accurate road, rail or river distances and quality of each of these infrastructures is not available. All variables sources are described in the Appendix.

According to the empirical approach, explanatory variables are separated into 3 categories: supply (S), infrastructure (I) and demand (D). Demand variables can be further divided into fuel demand (D_1) and feed demand (D_2).

Next, all explanatory variables employed in Probit and Tobit regressions are described along with their expected contribution to plant location and capacity decisions. The expected contribution of each explanatory variable is based on previous empirical work and microeconomic theory. *A priori*, it is assumed that variables that contribute positively to locational decisions should exert the same influence on installed capacity.

Supply Variables

The variable distance from Santos port (Dist Port) is used as the measure of distance from the market center for biofuel mills siting decisions. In the previous chapter, it is shown that this variable captures land use zones as sugarcane occupies the inner ring around Santos Port and soybean occupies the outer ring.

Thus, it is expected that location decisions for ethanol mills are negatively related with distance from Santos port because, moving away from it, the sugarcane land use zone ends. For the same reason, this variable is expected to be positively related to location decisions for biodiesel mills, as moving away from Santos port the soybean land use zone begins.

Next, the positive contribution of access to raw materials has already been identified in empirical literature by Henderson and Macnamara (2000) [30], Lambert et al. (2008) [37], Haddad et al. (2010) [25] and Sarmiento et al. (2012) [48].

The average quantity of sugarcane and soybean produced from 2005 to 2010 in each micro-region, (Cane Quantity) and (Soy Quantity) respectively, is expected to contribute positively to locational decisions.

Feedstock prices are expected to contribute negatively to location decisions, as higher prices should drive investors off. Real average prices of sugarcane and soybean from 2005 to 2010 in each micro-region is employed, (Cane Price) and (Soy Price). Some micro-regions did not produce any of these two crops during this period. When this occurred, the regional price of each crop was attributed to that observation. Previous studies cited herein have not employed the price of feedstock into their analysis. Only Henderson and Macnamara (2000) [30] include the value of all crops produced in a county, but as a measure of access to feedstock and thus with the opposite sign from that expected here.

The area of micro-regions (Area) is expected to be positively related to location and capacity decisions. Area is related to availability of land input. Micro-regions with larger areas can produce more feedstock, ceteris paribus. Henderson and Macnamara (2000) [30] use the size of farmland in each US county as an infrastructure variable that gives the potential area for crop expansion, with a positive contribution. There is a correlation of 0.68 between the distance from Santos port and the area of micro-regions. Moving away from the port, micro-regions become larger.

Specifically for biodiesel mills, two additional variables are employed. Soybean after harvested can be sold directly in international markets or processed at a soybean crushing mill where soybean meal and soybean oil are produced. Only soybean oil is used as input for biodiesel production. Thus, for biodiesel mills, the minimum distance to the nearest soybean crushing mill in 2010 (Soy Crush Dist) is employed as a measure of access to soybean oil. It is expected that as distance increases the likelihood of observing a biodiesel mill decreases and their capacity should decrease as well. Thus this variable is expected to have a negative sign.

Unfortunately, there is no data available on the capacity of all soybean crushing mills, which is expected to influence positively location and capacity decisions of biodiesel mills. The closest proxy for it, is the number of workers in the production of raw vegetable oil in each micro-region (Veg Oil Labor). More workers in raw vegetable oil production, soybean oil included, implies that more feedstock is potentially available. This variable has another benefit. It can help to capture partially, one or two mills that use cottonseed oil as feedstock.

Summing up, it is expected that biodiesel mills locate in abundant soybean producing micro-regions, expected plus sign, and near soybean crushing mills. As the distance from a soybean crushing mill increases, the likelihood of observing a biodiesel plant should decrease, expected minus sign, while being near to soy crushing mills with large capacities should increase the likelihood of plant location, expected plus sign for the number of workers in raw vegetable oil production.

Exceptions may happen because the biodiesel mill can either procure soybean oil from the soy crushing mill or crush soybean directly. But since the oil content in soybean is low, compared to other oil crops, it is not reasonable to crush soybean only for biodiesel production without having a market for soybean meal.

Infrastructure Variables

The role of infrastructure has been identified in previous studies as always contributing positively to location decisions. Henderson and Macnamara (2000) [30] use miles of roads divided by county area, Lambert et al. (2008) [37] use road density, rail density and a dummy for river adjacency while Haddad et al. (2010) [25] employ railroad miles per county as infrastructure related explanatory variables.

In this study four infrastructure variables are employed. Two variables measure transportation related infra-structure. Highway density (Highway Density) and railway density (Railway Density) measures the kilometers of each of these modes of transportation divided by the area of each micro-region. It is expected that they contribute positively to location decisions of both types of mills.

Historically in Brazil, transportation infrastructure such as roads and rails were, to a great extent, built from areas with abundant raw materials to export ports. This indicate that micro-regions with roads, and especially railways, are on the path between regions with abundant natural resources and main export ports.

Specifically for ethanol mills two other variables are considered, river density and power transmission grid density measured in kilometers divided by the area of the micro-region.

River density (River Density) is not a measure of access to navigable rivers as in Lambert et al. (2008) [37]. This variable captures access to water that is intensively used in sugarcane ethanol manufacturing process. This is expected to be positively related to location decisions.

Power transmission grid density (Grid Density) is expected to be positively related with ethanol mills location decisions. One of the by-products of sugar and ethanol production, sugarcane bagasse, is used to generate combined heat and power for the mill and any excess electricity can be sold to the grid.

Demand Variables

There are two types of demand to consider. The first, is directly related to biofuel end use and to the displacement of crude oil products. The other is related to the co-products of the biofuel production chain, mainly used as feed for livestock. Both types of demand are considered here.

The price of the fossil fuel competitor at the state level is employed. More precisely the price of gasohol, the gasoline sold in Brazil that already contains ethanol (Gasohol Price) and the diesel price (Diesel Price). It is expected that the higher is the price of fossil fuels the more likely will be to observe a biofuel mill, as biofuels become more competitive in that state.

For example, to be competitive, 1 liter of anhydrous ethanol has to be priced around 70 % of the gasoline price. According to Goldemberg and Moreira (1999) [23], ethanol has lower and higher heating values of 21.2 MJ/liter and 23.4 MJ/liter while gasoline has 30.1 MJ/liter and 34.9 MJ/liter. This means that anyhydrous ethanol has an energy content ranging from 67 % to 70.4 % compared to 1 liter of gasoline.

The average number of automobiles in each micro-region (Automobiles) is employed as a measure of local retail market for gasoline and ethanol. Brazil forbids diesel engines in light-duty vehicles as opposed to Europe. This variable is expected to contribute positively to ethanol mills location decisions.

Also in the case of ethanol, the minimum distance to the nearest ethanol storage terminal (Et Storage Dist) was calculated for each micro-region. On the regions inside the sample, sugarcane harvesting is concentrated from April to November and ethanol has to be stored to meet demand spread throughout the year. It is expected that as the distance from an ethanol storage terminal increases, the likelihood of observing an ethanol mill decreases.

The average number of trucks plus buses (Trucks plus Buses) in each micro-region is used as a measure of local retail diesel demand. It is expected that this variable contributes positively to location decisions of biodiesel mills.

The minimum distance to the nearest fuel terminal (Fuel Term Dist) was calculated for each micro-region. This variable is expected to be negatively related to the location of a biofuel mill. As the distance from a fuel terminal increases, the likelihood of making a positive location decision decreases. Haddad et al. (2010) [25] use distance to ethanol blending terminals as a market variable with negative impact on location decisions. Fuel terminals gives indication where fuel demand is located. Additionally, it captures the position of all crude oil refineries as all of them host fuel terminals. This is a wholesale measure of fuel market demand.

Regressions were run also using the average population density from 2000 and 2010, as a proxy for the population density in 2005. It was expected that population density would contribute positively to location and capacity decisions as a supply variable, as more populated areas have more dynamic labor markets and concentrated stocks of human capital.

This variable captures, indirectly, also the fact that population agglomerations are associated with supply of other services, such as schools, banking and hospitals, to cite some which should contribute positively to locational decisions of biofuel mills. In fact, Henderson and Macnamara (2000) [30] showed that agglomeration economies, measured by the number of existing food processing plants in a county and if the county was a regional economic node, contributed positively to location decisions.

A problem of imperfect multicollinearity emerged when this variable was utilized. Population density is highly correlated with energy demand. In particular the correlation is 0.90 with the number of automobiles and 0.88 with trucks plus buses. If the population in 2005 is used this correlation rises to 0.94 for automobiles and 0.93 for trucks plus buses.

Despite that population density can capture to some extent labor supply or agglomeration effects, clearly these effects are dominated by the energy demand effect. Having the exact number of automobiles, buses and trucks the variable population density or population became redundant, thus discarded.

Haddad et al. (2010) [25] on the other hand, use population density as a proxy for community concern capturing the *"not in my backyard"* effect, which should be negative related to plant location. The *"not in my backyard"* effect is very unlikely to hold for Brazil in the biofuel industry as it is perceived as a value added activity that creates jobs, income, local economic development and is strongly supported by local politicians. And again, the correlation with energy demand dominates all other possible effects.

Other variables referring to local labor market conditions such as unemployment rates or average wage rates are not available at the micro-region level. Thus, the number of automobiles and trucks plus buses, being durable goods, can also be understood as a proxy for the general level of welfare in each micro-region.

Finally, the University of São Paulo, Pecege-ESALQ-USP [41], using a sample of mills in the Center-West, Southeast and South regions, find that in 2010, anhydrous ethanol manufacturing labor costs ranged from 8.1% to 10.1% of total costs, while sugarcane costs represented 60.6% to 64.7% of total costs. For hydrated ethanol manufacturing, labor costs ranged from 8.1% to 10.2% of total costs, while sugarcane costs respresented 60.1% to 64.4% of total costs. These figures also help to mitigate the absence of variables directly related to local labor market conditions.

Regarding co-products demand, three variables are employed.

In the US, ethanol production from corn generates the by-product dry distiller grain which is used as feed for livestock. Accordingly, Lambert et al. (2008) [37], Haddad et al. (2010) [25] and Sarmiento et al. (2012) [48] uses the number of cattle heads as an explanatory variable that contributes positively to locational decision. Sarmiento et al. (2012) [48] use also the quantity of hogs as an explanatory variable that contributes positively to locational decisions.

Sugarcane and soybean have different degrees of integration with livestock activities. According to the Brazilian Development Bank (2008) [12], sugarcane bagasse, if added to some nitrogen source such as urea, can be fed to cattle but cannot be used to feed poultry or hogs. Soybean is not used directly as livestock feed, but after crushed, soybean meal can be fed to all livestock types. The presence of livestock in a micro-region is an indication of potential demand for soybean meal.

The variable cattle (Cattle) measures the average number of cattle heads per micro-region from 2005 to 2010. It is expected to be positively correlated with both types of biofuel plants. The variables hogs (Hogs) and poultry (Poultry) measures the average number of hogs and poultry in each micro-region from 2005 to 2010 and is expected to contribute positively only to biodiesel mills decisions.

The demand for another main by-product of biodiesel production, glycerol, is not analyzed. First because glycerol is used in many industries making it difficult to find one or two variables that could represent its market and second because it is not yet a relevant source of revenue for biodiesel producers.

The expected contribution of each explanatory variable for each type of biofuel, their classification as supply (S), infrastructure (I) or demand (D), which is further divided into fuel demand (D_1) and feed demand (D_2) is summarized in table (2.2).

Table 2.2: Explanatory Variables and Expected Contribution

Variable	Group	Ethanol	Biodiesel
Distance to Santos Port	S	-	+
Feedstock Quantity	S	+	+
Feedstock Price	S	-	-
Area	S	+	+
Distance to Soy Crushing Mill	S	n/a	-
Raw Vegetable Oil Employment	S	n/a	+
Highway Density	I	+	+
Railway Density	I	+	+
Power Transmission Grid Density	I	+	n/a
River Density	I	+	n/a
State Fossil Fuel Substitutes Price	D_1	+	+
Automobiles	D_1	+	n/a
Trucks plus Buses	D_1	n/a	+
Distance to Ethanol Storage Terminal	D_1	-	n/a
Distance to Fuel Terminal	D_1	-	-
Cattle	D_2	+	+
Poultry	D_2	n/a	+
Hogs	D_2	n/a	+

Note: n/a stands for not applied.

2.5.3 Empirical Model

Probit and Tobit regressions are employed with the same set of regressors $X_b = [S, I, D_1, D_2]$. The set of regressors inside vector X_b varies depending on the type of biofuel (b) considered, according to table (2.2). The same procedure is employed for each biofuel separately.

The Probit regression indicates if a micro-region hosts at least one ethanol or biodiesel mill. It captures location decisions by eliciting micro-region features that attract investment in each type of biofuel mill.

The binary dependent variable, ethanol mill, is defined as (EM_i).

$$EM_i = \begin{cases} 1 & \text{if} \quad \text{Micro-region i hosts at least one ethanol mill.} \\ 0 & \text{if} \quad \text{Micro-region i does not host an ethanol mill.} \end{cases}$$

The binary dependent variable, biodiesel mill, is defined as (BM_i).

$$BM_i = \begin{cases} 1 & \text{if} \quad \text{Micro - region i hosts at least one biodiesel mill.} \\ 0 & \text{if} \quad \text{Micro - region i does not host a biodiesel mill.} \end{cases}$$

The specification of Probit regressions are presented in equations (2.6) and (2.7). The probability of observing at least one ethanol or biodiesel mill in each micro-region $(p_{i,b})$ is determined by supply (S), infrastructure (I) and demand (D) explanatory variables.

$$p_{i,b} = \text{Prob}[EM_i = 1|S, I, D_1, D_2] = X_b\beta + \epsilon_i \qquad \epsilon_i \sim \mathcal{N}(0, \sigma_b^2) \tag{2.6}$$

$$p_{i,b} = \text{Prob}[BM_i = 1|S, I, D_1, D_2] = X_b\beta + \mu_i \qquad \mu_i \sim \mathcal{N}(0, \sigma_b^2) \tag{2.7}$$

The interpretation of probabilities follows Bresnahan and Reiss (1991) [8], with some modifications. First, the cut-off point that maximizes model accuracy, below 50 %, is selected, $\overline{p} \leq 0.5$. This cut-off point is interpreted as an indication of profitability in that micro-region. If $p_i \geq \overline{p}$, then profits at micro-region (i) are expected to be greater or equal to zero. If $p_i < \overline{p}$, profits in that location are expected to be negative.

Marginal effects at the sample average of each variable are calculated for both Probit regressions.

The Tobit regression estimates micro-region biofuel output capacity, conditional on the existence of at least one biofuel mill. The observed installed capacity $(C_{i,b})$ in each micro-region (i) for biofuel (b) is censored according to equation (2.8), where $(C^*_{i,b})$ is the latent variable.

$$C_{i,b} = \begin{cases} C^*_{i,b} & \text{if} \quad C^*_{i,b} > 0 \\ 0 & \text{if} \quad C^*_{i,b} \leq 0 \end{cases} \tag{2.8}$$

A Tobit regression is estimated for the latent dependent variable $C^*_{i,b}$ specified in equation (2.9).

$$C^*_{i,b} = X_b\beta + \nu_{i,b} \qquad \nu_{i,b} \sim \mathcal{N}(0, \sigma^2_b) \tag{2.9}$$

Out of the four possible marginal effects on the Tobit regression the focus is on the effect on the probability that the variable becomes uncensored and on uncensored capacity. The marginal effects on the probability that the variable becomes uncensored indicate potential market entries or exits and are compared to Probit marginal effects.

The marginal effects on uncensored capacity indicates the impact of changes in explanatory variables in output expansion or contraction of existing mills.

2.6 Empirical Results and Discussion

2.6.1 Ethanol Mills Location

As of September, 2011 there were 108 micro-regions with at least one ethanol mill, yielding a ratio of 108/306 = 35.3% of the sample. Thirteen explanatory variables are employed, four related to supply, four to infrastructure and five to demand according to table (2.2). Probit regression estimates are presented in table (2.3), alongside with Probit and Tobit marginal effects.

Table 2.3: Regression for Ethanol Mills Location and Marginal Effects

	Probit		Tobit	
	Coefficients	Marg. Effects	Marg. Effects	Units
Intercept	-3.151			
	(3.758)			
Dist Port	1.887***	0.085	0.361***	10^3 Km
	(0.733)	(0.108)	(0.126)	1 km = + 0.0361 %
Cane Quantity	2.286****	0.103	0.027****	10^6 tons/year
	(0.367)	(0.106)	(0.008)	10^3 tons = + 0.0027%
Cane Price	−19.742**	-0.886	−9.176****	BRL/kg
	(8.278)	(1.240)	(0.932)	1 BRL/ton = - 0.918%
Area	−36.007**	-1.616	−4.978**	10^6 km^2
	(14.353)	(2.077)	(2.484)	10^3 km^2 = - 0.4978%
Highway Density	-9.341	-0.420	0.021	Km/km^2
	(7.072)	(0.582)	(0.732)	
Railway Density	12.909	0.580	0.531	Km/km^2
	(11.160)	(0.745)	(1.151)	
Grid Density	−20.405*	-0.916	-1.388	Km/km^2
	(10.523)	(1.160)	(1.146)	m/km^2 = - 0.1388%
River Density	19.158**	0.860	2.193**	Km/km^2
	(8.255)	(1.115)	(0.922)	m/km^2 = + 0.2193%
Gasohol Price	1.145	0.051	-0.377	BRL/liter
	(1.853)	(0.102)	(0.383)	
Automobiles	1.701*	0.076	0.088****	10^6 cars
	(1.033)	(0.102)	(0.025)	10^3 cars = + 0.0088%
Et Storage Dist	−2.285***	-0.103	−0.352***	10^3 Km
	(0.800)	(0.133)	(0.128)	1 km = - 0.0352%
Fuel Term Dist	-0.972	-0.044	0.259	10^3 Km
	(1.575)	(0.087)	(0.290)	
Cattle	0.243	0.011	0.173***	10^6 heads
	(0.253)	(0.019)	(0.057)	10^3 heads = + 0.0173%
McFadden R-sq.	0.6983			
Wald-test	72.59			
p-value	0.0000			
Log-Likelihood	-59.947			
AIC	147.90			
BIC	200.02			
N	306			

Note: Robust standard errors in parenthesis for column 2 and delta-method for columns 3 and 4.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

The model exhibits a fit of 0.7 measured in terms of pseudo R-squared and is statistically significant from the Wald test.

From Probit estimates, the variables distance from Santos port (Dist Port), average produced sugarcane quantity from 2005 to 2010 (Cane Quantity), average real price of sugarcane from 2005 to 2010 (Cane Price), micro-region area (Area), power transmission grid density (Grid Density), river density (River Density), average number of automobiles from 2005 to 2010 (Automobiles) and the distance to the closest ethanol storage terminal (Et Storage Dist) are statistically significant. Probit marginal effects were calculated at the sample average and no one was statistically significant.

Tobit marginal effects on the probability that the variable becomes uncensored, provides two different results compared to the Probit regression. The variable (Grid Density) is not statistically significant and the variable (Cattle), measuring the average number of cattle heads in each micro-region from 2005 to 2010, becomes statistically significant. The remainder variables behave as in the Probit regression. As Probit and Tobit marginal effects do not show any statistically significant contradictory effect, the model can be validated. Due to their statistical significance Probit coefficients and Tobit marginal effects are analyzed.

Distance from Santos port (Dist Port) has a positive sign and is statistically significant in both cases. In the Probit regression, this unexpected result can be attributed to the usage of a binary response variable, that gives equal importance to micro-regions hosting one or many mills. For example, the micro-region of Ribeirão Preto, at 322 km from Santos port, has 33 authorized ethanol suppliers that were collapsed into a single observation and exerts the same influence as the micro-region of Arinos, at 1,770 kilometers from Santos port with 1 authorized ethanol supplier. There is also evidence of non-linearities from the spinogram. Employment in the ethanol sector exhibits the same pattern, first it increases in distance, then it decreases. This non-linearity was not introduced in the empirical model as there is no theoretical basis for doing so and to avoid an additional explanatory variable.

Marginal effect shows that each 1,000 km further away from the average distance from Santos port, increases the probability of a positive ethanol mill location decision by 36.1 %. Each additional kilometer has a very small impact, increasing location probability by 0.0361 %.

The statistical significance and the positive sign of average produced sugarcane quantity (Cane Quantity) are as expected and implies that sugarcane ethanol mills need to be located inside or near abundant sugarcane supplying regions, thus, inside the sugarcane land use zone. At the sample average, each additional million tons of sugarcane produced increases the probability of having an ethanol mill by 2.7 %. Therefore, each additional thousand tons of sugarcane produced increases location probability by 0.0027 %. One additional standard deviation increases probabilities by 10.61 %.

The next statistically significant variable, average real sugarcane price (Cane Price) in each micro-region has a negative sign as expected. Ethanol mills locate where feedstock procurement exhibits lower costs. From marginal effect, each additional unit in the price of sugarcane reduces location probability by 0.918 %. One additonal standard deviation decreases probability by 32.36 %.

The variable area (Area) is statistically significant but has an unexpected negative sign. The explanation is that this variable captures almost the same spatial dimension as the distance from Santos port (Dist Port). The correlation between (Area) and (Dist Port) amounts to 0.68. Moving away from Santos port and out of the sugarcane land use zone, micro-regions are larger. This is another indication that moving away from the sugarcane land use zone the likelihood of observing an ethanol mill decreases. Each additional thousand square kilometers of a micro-region area reduces probability of a positive location by 0.50 %.

Although the effects of the variables (Dist Port) and (Area) are canceling out each other, the area of each micro-region has to be controlled for, as agricultural output depends to a great extent on land availability.

All supply explanatory variables are statistically significant in both regressions.

The density of highways in each micro-region (Highway Density) has an unexpected negative sign but is not statistically significant.

Railway density (Railway Density) exhibits the expected sign but is not statistically significant for location decisions.

Power transmission grid density (Grid Density) is statistically significant in the Probit regression at 10% level, but not in the Tobit regression, and exhibits an unexpected negative sign. Another variable was tested to evaluate potential acessibility of ethanol mills to the power transmission grid, the distance from each micro-region centroid to the nearest electrical substation. This variable exhibited the expected sign but was not statistically significant in any regression or marginal effects.

Although the contribution of this variable is not as expected, its statistical significance is also questionable for the aforementioned reasons.

The explanation for this result is that, according to UNICA (2010) [35], until the mid 1990s when the Brazilian power sector went through major reforms, ethanol mills were forbidden to sell any excess electricity to the grid. For this reason, in order to avoid storing sugarcane bagasse, very inefficient co-generation technologies were deployed. The objective was not to maximize electricity production, but maximize the amount of sugarcane bagasse burned subject to meeting only energy self-sufficiency of the mill. New regulation of the sector allowed independent electricity producers, ethanol mills included, to sell their surplus to the grid. Older ethanol mills might have neglected access to the transmission grid in their siting decision process and deployed inefficient co-generation technologies. More recent or new mills, are more likely to have more efficient technologies for combined heat and power and to be located closer to power transmission grid accesses.

UNICA (2010) [35] estimates a considerable untapped potential for bio-electricity production, currently around 10 Gigawatts on average. It is very likely that access to the power transmission grid will become increasingly important.

River density (River Density) is statistically significant with the expected sign. Ethanol production requires significant amounts of water and mills have to be located near abundant water sources. Marginal effect indicates that each additional meter of river per kilometer squared of area, increases the probability of an ethanol mill location by 0.2193 %. Only this infrastructure variable is significant in both regressions.

Gasohol price (Gasohol Price) is not statistically significant in both regressions.

The average number of automobiles in each micro-region has the expected sign and is statistically significant. As the number of automobiles increase, there is more local gasoline and ethanol demand, increasing attractiveness of that location. The marginal effect of automobiles is to increase the probability of existence of an ethanol mill by 0.0088 % for each 1,000 additional cars. An additional standard deviation increases probabilities by 2.90 %.

As already mentioned, the number of automobiles is highly correlated with population density as one would expect. As sugarcane land use zone overlaps with smaller, more populated and more developed micro-regions, mainly in the State of São Paulo, the wealthiest of the country, the variable (Automobiles) captures also a spatial aspect that reinforces the supply orientation of sugarcane ethanol mills. As one moves away from Santos port, and out of the sugarcane land use zone, micro-regions gets larger, with lower population density and lower number of automobiles.

The results found up to here are completely in line with Kilkenny and Coleman (2006) [36] that the *"agroindustry is input oriented and city located"*.

The minimum distance to an ethanol storage terminal (Et Storage Dist) has the expected sign and is statistically significant. Micro-regions farther from any ethanol storage terminal, have a lower likelihood of hosting an ethanol mill. Currently, there are eight existing ethanol storage terminals that can store up to 90 million liters of ethanol. Distance from an ethanol storage terminal is important and likely to become even more important as the industry is struggling to keep up with increasing demand, particularly outside sugarcane harvesting season. The variable (Et Storage Dist) has a correlation of 0.76 with the distance form Santos Port (Dist Port). Ethanol storage terminals are inside or near sugarcane land use and as the distance from Santos port increases so does the distance to the nearest ethanol storage terminal.

The marginal effect of ethanol storage terminals tells that each 1,000 km away from it, the likelihood of a positive ethanol plant location decreases by 35.2 %. So each kilometer has a negative impact of 0.0352 %. One additional standard deviation decrases probabilities by 9.83 %.

The distance to the closest fuel terminal (Fuel Term Dist) has also the expected sign in the probit regression, but without statistical significance.

The average number of cattle heads in each micro-region (CattleAv) captures the possible integration of sugar and ethanol production with cattle ranching, as sugarcane bagasse can be used as feed. It exhibits a positive sign as expected but not statistically significant in the probit regression. There is no evidence of integration between these activities.

In the Tobit regression the variable (Cattle) becomes statistically significant, but not for the expected reason. It results that cattle ranching is one of the most easily displaced activities by sugarcane cropland expansion. Thus, the past and current presence of large quantity of cattle gives a good indication of land availability to increase sugarcane production. This variable captures part of the direct land use change caused by the expansion of sugarcane production. One thousand additional cattle heads increases the probability of observing an ethanol mill by 0.0173 %. One additional standard deviation increases proabibilities by 10.22 %.

All categories of variables are jointly significant to explain the locational pattern of ethanol mills in the probit regression. Results of joint significance tests are presented in table (2.4).

Table 2.4: Test for Joint Significance of Variables by Category on Probit

	Supply	Infrastructure	Demand	D_1	D_2
Number of Variables	4	4	5	4	1
χ^2	59.50	12.45	11.77	10.80	n/a
p-value	0.0000	0.0143	0.0381	0.0289	n/a

The model reaches the maximum accuracy of 92.16% at a cut-off point of 48%.[10] It misclassifies 24 micro-regions, detailed in table (2.5). Retriving probabilities from Tobit estimates yield worse results. Using Tobit estimates achieves a maximum accuracy of 86.27 % at a lower cut-off threshold of 39 %.

[10]The same accuracy is reached at a cut-off of 32 %, but 48 % is chosen because it is closer to the rule of thumb of a 50% cut-off.

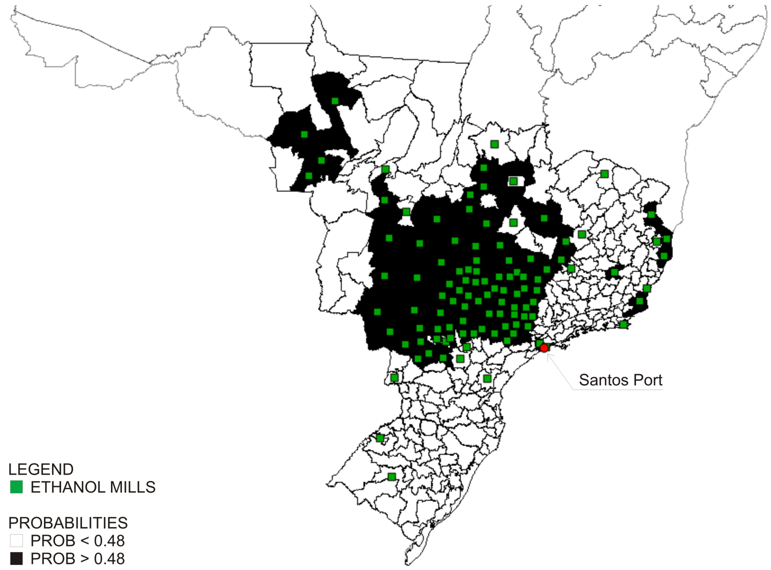

Figure 2.2: Probit Regression for Ethanol Mills

Table 2.5: Accuracy of Probit Regression For Ethanol Mills

	$< 48\%$	$\geq 48\%$	Total
0	191	7	198
1	17	91	108
Total	208	98	306

In figure (2.2), estimated probabilities are plotted against the true existing ethanol mills. Micro-regions in white are expected to yield negative profits for ethanol mills as ($p_i < 0.48$) and micro-regions in black are expected to yield positive profits as ($p_i \geq 0.48$).

2.6.2 Ethanol Mills Capacity

Unfortunately, there is no data available on installed capacity of each ethanol mill. Installed capacity is measured in terms of sugarcane crushing capacity per year. The only available data on this variable refers to the distribution by size of ethanol mills in each state, and not to single observations.

To overcome this problem, the number of employees working on ethanol production in each micro-region at the end of 2010 is used as a proxy for installed capacity in the micro-region in the same year. A Tobit regression is employed with the same explanatory variables used for eliciting locational decisions.

A caveat of this approach is that employment is expected to increase at a lower pace than capacity increases.

Another caveat refers to the usage of two different sources of data. There are 24 cases of ethanol mills with zero employment and 23 cases of positive employment without an ethanol mill. To avoid incoherences, priority was given to the dataset from the National Petroleum Agency with existing mills used in the Probit regression. Existing mills with zero employment were kept. Cases of positive employment without mills were treated as if censored observations. A Tobit regression was run with all observations from the Ministry of Labor dataset. The only changes are that the variables (Area) and (River Density) are not statistically significant. But the model performs worse in all aspects, McFadden R-squared, log-likelihood, AIC and BIC and the F-test statistics. Thus, there is evidence to treat the cases of mismatch of positive employment in micro-regions without mills as censored.

Results of the Tobit estimation are presented in table (2.6). Estimates are presented for the latent variable, for the marginal effects of independent variables on capacity of existing mills or uncensored dependent variable, and for the marginal effects on censored and uncensored dependent variable.

Table 2.6: Tobit Regression for Ethanol Mills Employment

	Coefficients	$\partial E[C\|X, C \geq 0]$	$\partial E[C\|X]$	Units
Intercept	3,862.113			Number of Employees
	(3,600.636)			
Dist. Port	1.653****	0.310****	0.224***	Km
	(0.447)	(0.083)	(0.080)	
Cane Quantity	0.122****	0.023****	0.016**	10^3/tons
	(0.027)	(0.006)	(0.006)	
Cane Price	−42.079****	−7.892****	−5.689****	BRL/ton
	(11.022)	(1.388)	(0.872)	
Area	−0.023**	−0.004**	−0.003*	km^2
	(0.010)	0.002	(0.002)	
Highway Density	96.632	18.124	13.065	Km/km^2
	(3,362)	630.424	(454.106)	
Railway Density	2,436	456.956	329.402	Km/km^2
	(5,125)	(972.934)	(720.303)	
Grid Density	-6,366	-1,194.06	-860.750	Km/km^2
	(4,956)	(945.778)	(732.289)	
River Density	10,056**	1,886.217**	1,359.698**	Km/km^2
	(4,557)	(827.843)	(650.897)	
Gasohol Price	-1,728	-324.107	-233.635	BRL/liter
	(1,701)	(318.066)	(234.987)	
Automobilies	0.405****	0.076****	0.055***	10^3 cars
	(0.082)	(0.016)	(0.018)	
Et Storage Dist	−1.614***	−0.303***	−0.218**	Km
	(0.525)	(0.099)	(0.089)	
Fuel Term Dist	1.189	0.223	0.161	Km
	(1.342)	(0.252)	(0.186)	
Cattle	0.793***	0.149***	0.107**	10^3 heads
	(0.293)	(0.054)	(0.046)	
McFadden R-sq.	0.0923			
F-test	6.23			
p-value	0.0000			
Log-Likelihood	-947.17			
AIC	1,924			
BIC	1,980			
N	306			

Note: Robust standard errors in parenthesis in column 2 and delta-method in columns 3 and 4.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

The model fit, measured in terms of adjusted R-squared, is low but the model is statistically significant from the F-test.

From the Tobit regression and marginal effects, the average quantity of sugarcane produced in a micro-region (Cane Quantity) also affects the size of the ethanol mills measured by the number of employees. At the sample average, each additional 1,000 tons of sugarcane produced in a micro-region that already produces ethanol, increases the number of workers by 0.023 units. One additional standard deviation should increase employment by 90.4 employees.

One unit increase in the price of sugarcane should decrease employment by 7.89 units. One additional standard deviation should reduce employment by 278.8 units.

One thousand units increase in the number of automobiles should increase employment by 0.076 units. One additional standard deviation increases employment by 25 units.

If a new ethanol storage terminal is built, each kilometer closer to it should increase employment by 0.3 units. One additional standard deviation increases employment by 84.8 units.

Each additional thousand cattle heads in the micro-region in the past increased employment by 0.15 units. One additional standard deviation increases employment by 88.1 units.

Tests for joint significance of explanatory variables for the Tobit regression are presented in table (2.7). Supply and demand variables are jointly significant, while infrastructure variables are not jointly significant anymore.

Table 2.7: Tobit Joint Significance Test by Category

	Supply	Infrastructure	Demand	D_1	D_2
Number of Variables	4	4	5	4	1
F-test	12.13	1.76	8.73	10.15	n/a
p-value	0.0000	0.1367	0.0000	0.0000	n/a

2.6.3 Biodiesel Mills Location

Probit and Tobit regressions are estimated to assess the probability that a certain micro-region hosts a biodiesel mill. The analysis is simplified to mills that use vegetable oil as feedstock, the vast majority being soybean oil. As of 2011, there were 43 micro-regions in the sample with at least one biodiesel mill. Three micro-regions were excluded as they host only mills running on animal fat. This is done to focus on the soybean industry dynamics. After removing mills running on tallow, there is still one or two at most, running on cottonseed oil that cannot be pinpointed with certainty.

Selecting only mills running on vegetable oil, leaves 40 micro-regions with one or more biodiesel mills installed, out of 306 micro-regions in the Center-West-Southeast-South regions of the country, giving a ratio of $P = 40/306 = 13.07\%$. The micro-region of Rondonópolis, located in Mato Grosso state, has 4 biodiesel mills and also the largest installed capacity.

Regressions include 14 explanatory variables. Results for the Probit model and Probit and Tobit marginal effects are presented in table (2.8).

Table 2.8: Regression for Biodiesel Mills Location and Marginal Effects

	Probit		Tobit	
	Coefficients	Marg. Effects	Marg. Effects	Units
Intercept	-5.483			
	(4.024)			
Dist Santos Port	0.317	0.054	0.068	10^3 Km
	(0.499)	(0.084)	(0.079)	
Soy Quantity	0.445*	0.076*	0.059**	10^6 tons/year
	(0.233)	(0.041)	(0.024)	10^3 tons/year = + 0.0076%
Soy Price	0.652	0.112	0.218	BRL/kg
	(2.535)	(0.435)	(0.381)	
Dist Crush Mill	−1.018*	−0.175*	−0.225**	10^3 Km
	(0.558)	(0.091)	(0.093)	1 km = - 0.0175%
Veg Oil Labor	0.855	0.147	0.181	10^3 Workers
	(0.761)	(0.132)	(0.127)	
Area	5.696	0.977	0.405	10^6 km^2
	(11.293)	(1.950)	(1.520)	
Highway Density	-0.580	-0.099	0.094	Km/km^2
	(3.493)	(0.599)	(0.571)	
Railway Density	9.830	1.685	1.123	Km/km^2
	(8.167)	(1.374)	(1.310)	
Diesel Price	2.145	0.368	0.343	BRL/liter
	(2.186)	(0.377)	(0.369)	
Trucks + Buses	0.740	0.127	-0.117	10^6 units
	(6.893)	(1.182)	(1.130)	
Fuel Term Dist	-0.547	-0.094	-0.059	10^3 Km
	(1.713)	(0.294)	(0.241)	
Cattle	0.133	0.023	0.026	10^6 heads
	(0.235)	(0.041)	0.030	
Hogs	-1.415	-0.243	−0.293*	10^6 heads
	(1.046)	(0.177)	(0.160)	10^3 heads = - 0.0293 %
Poultry	0.028	0.005	0.006	10^6 heads
	(0.022)	(0.004)	(0.004)	
McFadden R-sq.	0.1713			
Wald Test	34.47			
p-value	0.0018			
Log-Likelihood	-98.332			
AIC	226.66			
BIC	282.52			
N	306			

Note: Robust standard errors in parenthesis for column 2 and delta-method for columns 3 and 4.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

The Probit model exhibits a low fit of 0.171 measured in terms of pseudo R-squared but is statistically significant. The biodiesel Probit regression is, not surprisingly, worse than the ethanol Probit. The ethanol industry started in 1975 while the biodiesel industry started in 2005. Furthermore, there are 108 positive observations for ethanol mills and only 40 for biodiesel mills. Thus, ethanol is a mature industry with a consolidated location pattern, while biodiesel is an infant industry with an emerging location pattern.

Out of the 14 explanatory variables, only two are statistically significant. Marginal effects of statistically significant variables are also statistically significant.

The average quantity of soybean produced in each micro-region from 2005 to 2010 (Soy Quantity) has the expected sign and is statistically significant. The more soybean is produced in a micro-region the more likely it becomes to observe a biodiesel mill. Marginal effects show that each additional 1,000 tons of soybean produced in the micro-region increases the probability of hosting a biodiesel mill by 0.0076 % or by 0.0059 %. One additional standard deviation increases probabilities by 3.66 % or 2.84 %.

Distance to the nearest soy crushing mill (Soy Crush Dist) exhibits the expected sign and is statistically significant. Moving away from any soybean crushing mill, the likelihood of observing a biodiesel mill falls. Marginal effects indicate that each kilometer away from a soy crushing mill decreases the probability of observing a biodiesel mill by 0.0175 % or by 0.0225 %. These figures have to be considered with caution as all spatial variables were collapsed into the micro-regions centroids. But, the main result that moving away from a soybean crushing mill reduces the likelihood of observing a biodiesel mill holds. One additional standard deviation decreases probabilities by 4.04 % or by 5.19 %.

Distance from Santos port (Dist Port) has the expected sign but is not statistically significant. Moving away from this port into the soybean land use zone, the probability of observing a biodiesel mill increases.

Average real soybean prices in each micro-region from 2005 to 2010 (Soy Price) has an unexpected positive sign but is not statistically significant.

The average number of workers in raw vegetable oil production from 2005 to 2010 (Veg Oil Labor) contributes positively but is not statistically significant. This is a proxy for vegetable oil output capacity in each micro-region. The bulk consists of employment in soybean crushing mills with some residual employment in cottonseed, sunflower seed and rapeseed crushing mills.

Area (Area) is positive as expected, but not statistically significant.

Highway density (Highway Density) exhibit mixed signs but is not statistically significant. Railway density (Rail Density) gives a positive contribution as expected but is not statistically significant.

Diesel price (Diesel Price) is positive as expected but not statistically significant.

Local retail fuel demand, measured by the average number of trucks plus buses (Trucks + Buses), has mixed signs but is not statistically significant.

The coefficient for the distance to the closest fuel terminal (Fuel Term Dist) is negative as expected but not statistically significant. As distance from a fuel terminal increases the likelihood of observing a biodiesel mill decreases.

Livestocks variables, cattle (Cattle) and poultry (Poultry) have the positive expected sign but are not statistically significant. The variable hogs (Hogs) has an unexpected sign and is statistically significant at the 10 % level in the Tobit marginal effect. It provides indication that existing biodiesel mills are not yet fully integrated with the livestock sector. And some micro-regions can still ramp up capacity to reach a proportional scale to the livestock sector.

The Probit model shows that biodiesel mills are oriented exclusively towards supply abundant regions. There are no statistically significant variables on the infrastructure and demand categories.

In fact, testing for joint significance of groups of variables shows that only supply variables are jointly significant. Results are presented in table (2.9).

Table 2.9: Probit Joint Significance of Variables by Category

	Supply	Infrastructure	Demand	D_1	D_2
Number of Variables	6	2	6	3	3
χ^2	12.61	1.45	3.71	1.01	2.19
p-value	0.0496	0.4846	0.7155	0.7992	0.5336

The Probit and Tobit models have subtle yet very important differences. The Probit model gives relatively more importance to access to soybeans while the Tobit model gives relatively more importance to proximity to soy crushing mills. The Probit model reaches a maximum accuracy of 88.56 % at a cut-off level of 43 %, while the Tobit model reaches a maximum accuracy of 89.22 % at a cut-off level of 47 %. The area under the ROC (Receiver Operating Characteristic) curve is slighlty higher for Probit than Tobit. As the Tobit model has also an unexpected statistically significant explanatory varaible, the Probit model is selected to be analyzed in further detail. The difference betwewn the two models is that Probit misclassifies 2 micro-regions more than Tobit.

Despite the overall good accuracy, the Probit model fails to assign correctly high probabilities to many existing biodiesel mills. It misclassifies 35 micro-regions. Table 2.10 shows the model accuracy in terms of predictive power.

Table 2.10: Accuracy of Probit Regression For Biodiesel Mills

	$< 43\%$	$\geq 43\%$	Total
0	262	4	266
1	31	9	40
Total	293	13	306

In figure (2.3), estimated probabilities are plotted against the true existing biodiesel mills. If $p_i \geq \overline{p}$, then profits at that location are expected to be greater or equal to zero. If $p_i < \overline{p}$, profits at that location are expected to be negative. Negative profits are expressed in white, and positive profits are in black.

Remains to explain altogether the model's poor performance, failing to assign correctly 31 micro-regions with existing biodiesel mills, and how an industry with excess installed capacity can observe continued entry.

The answer to both issues is that many of the existing biodiesel mills are mislocated for the use of soybean oil as feedstock.

Mislocation of many existing biodiesel mills can explain why the model has a relatively poor performance, taking into consideration that soybean oil has accounted for

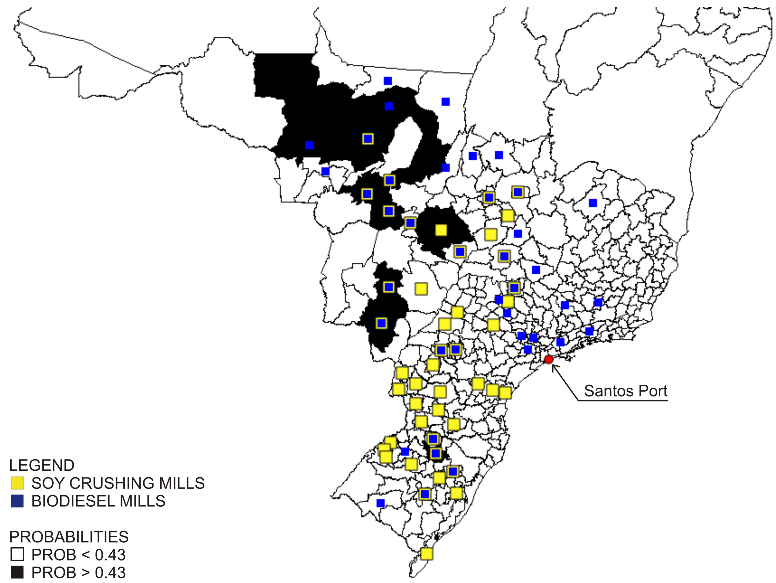

Figure 2.3: Probit Regression for Biodiesel Mills

the bulk of feedstock usage and accordingly, selected supply explanatory variables refer to the soybean industry.

Mislocation can also explain why new mills still find profitable to enter in this market, in spite of excess capacity, as entrants will have lower costs than many of misplaced incumbents for soybean oil procurement.

Analyzing the model's results in table (2.10) quadrant-wise, in counter clockwise manner starting from the bottom right, there are 9 existing micro-regions with biodiesel mills correctly captured by the model. Seven of them are in micro-regions that have also a soy crushing mill, while two, both in Mato Grosso state do not. These are micro-regions with biodiesel mills that the model correctly captures.

Moving one quadrant up, there are 4 micro-regions classified as very good sites for biodiesel mills but that in reality do not have one. Three of them are in Mato Grosso state and do not have a soybean crushing mill, but have access to soybean. The remainder is in the State of Goiás and has a soy crushing mill.

There are 262 micro-regions classified as not good sites for biodiesel mills and that do not host one. These micro-regions are also correctly classified.

The more problematic quadrant is the bottom left, where 31 micro-regions with existing mills are misclassified, that is, assigned a probability lower than 43 %. For these 31 micro-regions a simple criteria is established. Assuming all biodiesel mills attached to a soy crushing mill should be captured by the model, it fails to allocate 11 micro-regions that host both a biodiesel mill and a soybean crushing mill.

This procedure leaves 20 true observations, or half of the total, that cannot be explained. This group can be considered mislocated for the usage of soybean oil. The degree of mislocation of each mill varies and can be assessed by the probability assigned to the hosting micro-region. Of these 20 not captured observations, out of the top ten existing mills with highest assigned probabilities, 5 are in the Center-West, 2 in the South and 3 in the Southeast region. The bottom ten is comprised by 8 biodiesel mills in the Southeast region and 2 in the Center-West region. This is an indication that biodiesel mills in the Southeast region are in general more mislocated than others.

Nevertheless, the most surprising number is that there are 27 micro-regions with soy crushing mills which could potentially host biodiesel mills. This is the reason why, paradoxically, it is being observed a market with excess installed capacity and continued entry.

This mislocation can be explained because biodiesel still is an infant industry in the country and also worldwide and consequently there were, and there are still, many uncertainties regarding the economic attractiveness of feedstocks and technological route to be deployed.

On top of these uncertainties, the government program gives locational incentives through tax exemptions to attract investments to poor regions of the country, foster rural development and generate income for small farmers. This strategy to pick up winning feedstocks and locations, namely castor beans in the Northeast region and palm oil in the North region, in fact attracted investors on the onset of the program. But despite all incentives, time revealed to be more economically attractive the usage of soybean oil, an already abundantly available raw material from an existing, mature and competitive in international markets industry.

It is fair to state that there was an initial movers disadvantage in the biodiesel industry.

More evidence to support this conjecture is found observing already authorized capacity expansions or new entries by the National Petroleum Agency as of 22.03.2012. Considering entries or capacity expansions only inside sample regions, which consists of the bulk of new capacity, there are additional 2,157,404.4 $m^3/year$ already authorized to be constructed. That is 40 % more of the exisiting capacity.

There are 15 expansion or entry announcements, 6 are in micro-regions that already host a biodiesel mill and a soy crushing mill, 5 are entering in micro-regions that already host a soy crushing mill. These 11 announcements account for 70 % of new capacity. Of the 4 remainder entrants, 3 are in micro-regions that neighbors micro-regions with a soy crushing mill and only one is not. Figure (2.4) depicts where entry is occurring, existing soy crushing mills and biodiesel mills and estimated probabilities.

Therefore, the explanation provided here justifies continued entry.

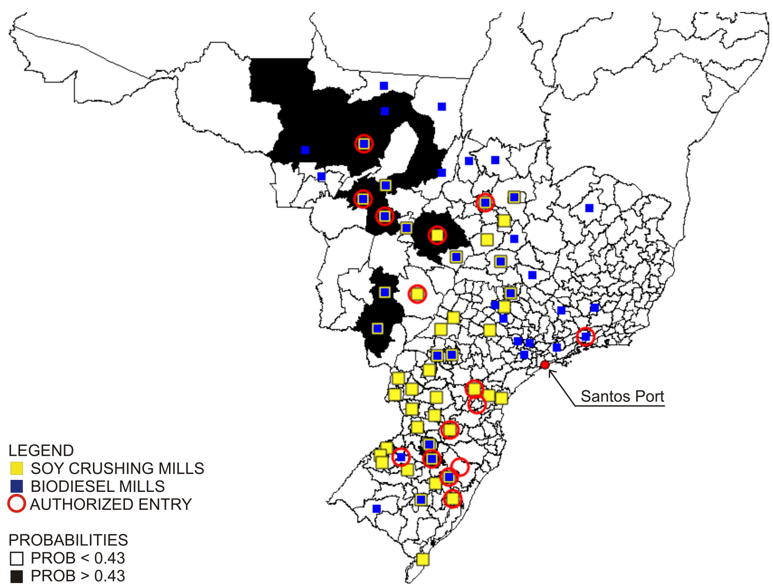

Figure 2.4: Authorized Expansions or Entry in Biodiesel Production

Another source of uncertainty that could help explaining mislocation, but to a much lower extent, refers to the choice of technological route, methylic versus ethylic.

Biodiesel mills that opted for the ethylic route are expected to be close to ethanol supply, thus close to ethanol mills and far from soybean supply. The correlation between micro-regions with ethanol and biodiesel mills is positive but small, 0.14.

Evidence from the National Petroleum Agency shows that out of the 50 mills in the sample, only 4 opted for the ethylic route, 34 for the methlyic route and 12 for a flexible route where both types of alcohols can be used.

Only 3 micro-regions have mills exclusively utilizing ethanolysis. All of them are in the Southeast region, where the model fails to assign high probabilities but where the bulk of ethanol production is located. The fourth micro-region, in the South region, hosts a biodiesel mill using ethanol and a biodiesel mill running on animal fat and using methanol.

Table (2.11) provides evidence that, proportionately, biodiesel mills located in the Southeast region have opted more for the ethanolysis or hybrid route than in other regions.

Table 2.11: Percentage of Technological Route in Each Region

	Methanolosys	Ethanolosys	Flexible
Center-West	78%	0%	22%
Southeast	43%	21%	36%
South	78%	11%	11%

But informal evidence from industry specialists, press and companies websites, shows that currently only two mills use anhydrous ethanol as feedstock in the biodiesel production process, the company Fertibom in the micro-region of Catanduva in São Paulo state and the company Barrálcool in the micro-region of Tangará da Serra in Mato Grosso state, which is classified by the National Petroleum Agency as using the methylic route. Barrálcool is the unique integrated first generation biorefinery producing ethanol, bioelectricity and biodiesel using ethanol.

With existing technology, biodiesel production using anhydrous ethanol is less efficient and more expensive than using methanol. Additionally, the ethylic route can be adapted to the usage of methanol. Consequently, methanol has been the feedstock of choice in biodiesel production, despite being a non-renewable product, normally produced from natural gas. It has also caused a soar in methanol imports.

Moreover, using anhydrous ethanol in the production process of biodiesel has an opportunity cost of not using it directly to displace gasoline. On the other hand, biodiesel production using ethanol consists in a more renewable fuel.

Overall, the choice of technological route plays a minor role in locational decisions of biodiesel mills, as the methylic route is more efficient. All authorized expansions and entries opted for the methylic or the flexible route.

As the ethylic route can be adapted to the methylic route, the only cause of mislocation becomes again a problem of access to feedstock.

2.6.4 Biodiesel Mills Capacity

A Tobit regression is estimated with the same 14 explanatory variables, to investigate determinants of biodiesel mills installed capacity in each micro-region. Results are presented in table (2.12).

Table 2.12: Tobit Regression for Biodiesel Mills Installed Capacity

	Coefficients	$\partial E[C\|X, C > 0]$	$\partial E[C\|X]$	Units
Intercept	-1,204,916			m^3/year
	(887,823.6)			
Dist Santos Port	86.838	14.390	8.149	Km
	(102.255)	(16.752)	(9.346)	
Soy Quantity	75.392***	12.493***	7.075**	10^3 tons/year
	(29.030)	(4.823)	(2.987)	
Soy Price	278.375	46.129	26.122	BRL/Ton
	(484.692)	(80.552)	(46.102)	
Dist Crush Mill	−287.432**	−47.630**	−26.972**	Km
	(133.873)	(21.201)	(11.614)	
Veg Oil Labor	230.854	38.255	21.663	Workers
	(160.392)	(26.432)	(15.219)	
Area	0.518	0.086	0.049	km^2
	(1.934)	(0.321)	(0.182)	
Highway Density	120,523.1	19,971.78	11,309.72	Km/km^2
	(732,679.6)	(121,446)	(68,840.25)	
Railway Density	1,435,664	237,902.6	134,720.7	Km/km^2
	(1,687,423)	(275,748.4)	(152,991.8)	
Diesel Price	438,171.1	72,608.94	41,117.35	BRL/liter
	(477,087.1)	(79,012.68)	(45,206.46)	
Trucks + Buses	-149.158	-24.717	-13.997	10^3 units
	(1,443.998)	(239.236)	(135.438)	
Fuel Term Dist	-75.339	-12.484	-7.070	Km
	(305.404)	(50.698)	(28.842)	
Cattle	33.626	5.572	3.155	10^3 heads
	(38.653)	(6.431)	(3.709)	
Hogs	−373.896*	−61.958*	−35.086*	10^3 heads
	(208.903)	(34.079)	(19.517)	
Poultry	7.302	1.210	0.685	10^3 heads
	(5.054)	(0.830)	(0.475)	
McFadden R-sq.	0.035			
F-Test	2.54			
p-value	0.0019			
Log-Likelihood	-607.4			
AIC	1,246.8			
BIC	1,306.4			
N	306			

Note: Robust standard errors in parenthesis in column 2, and delta-method in columns 3 and 4.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

The model exhibits a low fit in terms of pseudo R-squared but is statistically significant from the F-test.

The average quantity of soybean produced (Soy Quantity) is positive and statistically significant. At the sample average, each additional thousand tons of soybean produced in a micro-region, should allow an increase in capacity of an existing biodiesel mill of 12.5 m^3/year. One additional standard deviation should increase capacity by 6,023.72 m^3/year.

Distance to the closest soybean crushing mill (Dist Crush Mill) is negative as expected and also statistically significant. The distance measure has to be carefully interpreted as all facilities were collapsed into the micro-region centroids. Although the regression tells that each kilometer away from a soybean crushing mill decreases the capacity of the biodiesel mill by 47.63 m^3/year, a better interpretation is that capacity of biodiesel plants is decreasing in distance from the nearest soy crushing mill. One additional standard deviation should decrease biodiesel output capacity by 10,991.58 m^3/year.

This result is quite intuitive. Production of biodiesel requires access to raw material, of which soybean oil is currently by far the most important. Soybean when crushed yields around 80% of soyeban meal and 20% of soybean oil. It follows that biodiesel mills away from soy crushing mills have to crush their own soy to produce soybean oil, but have the additonal burden of finding market for soybean meal. But if it was attractive to produce a large quantity of soybean meal in the first place, then a soy crushing mill could profitably establish there. This explains why capacity decreases in distance from soybean crushing mills. As a stand alone biodiesel mill increases its capacity its portfolio of products becomes similar to the soy crushing mill. Therefore, biodiesel mills are a complementary capital and poor substitutes to soy crushing mills.

Moreover, the capacity of the biodiesel mill should be proportional to the capacity of the soy crushing mill. The closest proxy for this measure, employment in raw vegetable oil manufacturing, is positive but not statistically significant. But the accurate measure could yield better results.

The variable (Hogs) is statistically significant at the 10 % level and exhibits an unexpected negative sign. The explanations for this result are two. First, biodiesel mills are mislocated for the usage of soybean oil, and consequently without an adequate integration with the livestock sector. The second, is that even well located mills can ramp up their capacity to align with operating scales in the livetock sector.

Only supply variables are jointly significant. Tests for joint significance are presented in table (2.13).

Table 2.13: Tobit Joint Significance F-Test by Variable Category

	Supply	Infrastructure	Demand	D_1	D_2
Number of Variables	6	2	6	3	3
F-test	3.07	0.45	0.92	0.29	1.29
p-value	0.0063	0.6381	0.4787	0.8346	0.2784

There are two economic forces at work. The first is proximity to a soy crushing mill, which allows for access to abundant soybean oil supply. Biodiesel mills near soy crushing mills can have a larger installed capacity as soybean meal is not part of their product portfolio.

The second, in the absence of a nearby soy crushing mill, biodiesel mills locate in regions with access to abundant soybean supply as the biodiesel mill can have its own crushing facility. In this case, the biodiesel mill has to be smaller because there will be necessarily soybean meal output to deal with.

This explanation fits into the dynamics of the Brazilian soybean industry identified in Goldsmith and Hirsch (2006) [24].

According to the authors there are two different soybean producing regions to consider in Brazil. The first can be considered the traditional soybean producing region, the South, comprised of the 3 Brazilian southern states, Rio Grande do Sul, Santa Catarina and Paraná where soybean production was introduced in Brazil in the 1960s.

The second can be considered the new or expansion soybean producing region in the Center-West.

Goldsmith and Hirsch (2006) [24] advocates that there is an unbalance between the two regions, Center-West versus South, measured by the quantity of soybean produced and soy crushing installed capacity in each region, created by historical events, distance to export ports, quality of infra-structure and the fact that existing international trade rules favors exports of raw soybean rather than processed products. This unbalance still exists as detailed in table (2.14) with an extended analysis to include the Southeast region, biodiesel, crude oil refineries and diesel demand.

Table 2.14: Soybean Complex and Diesel Unbalance in 2010

	Center-West	Southeast	South
Soy Quantity	51.05%	6.98%	41.97%
(tons)	31,558,236	4,315,398	25,950,387
Crushing Capacity	41.29%	15.83%	42.88 %
(tons/year)	67,775	25,980	70,379
Biodiesel Mills	54.0%	28.0%	18.0%
Number	27	14	9
Biodiesel Capacity	49.3%	18.6%	32.1%
(m^3/year)	2,376,487.9	896,025.6	1,544,518.8
Biodiesel Production	50.0%	15.5%	34.5%
(m^3/year)	977,951.36	303,148.96	675,609.60
Crude Oil Refinery	0	76 %	24 %
(m^3/day)	0 %	206,200	65,000
Diesel Sales	15.3%	58.8%	25.8%
(m^3/year)	5,623.53	21,567.54	9,467.07

The Southeast and South regions produce less soybean than their installed capacity for processing it. The Center-West region which is farther from Santos port, produces more soybean than its processing capacity. It follows, according to Goldsmith and Hirsch (2006) [24] that *"current crushing infrastructure is old, small, and out of position"*.

From the perspective of crude oil refining and diesel demand this unbalance is even more pronounced. The Southeast-South regions have all refineries and the bulk of diesel demand.

Investment in biodiesel mills is helping to correct this regional unbalance. The relative installed capacity of biodiesel in the Center-West region is much higher than its diesel demand. Soybean is cheaper in the Center-West region but looses part of its competitiveness when transported. As biodiesel aims the domestic market, it can profitably establish there. But the biodiesel mills have to be smaller when compared with their counterparts in the South. The reason is the market for soybean meal that makes unattractive for soy crushing mills to establish there. From table (2.14), it is observable that there are much more mills in the Center-West region that in the South region, but the average installed capacity is much higher in the South.

In fact, as there is much idle capacity, who actually produced the bulk of biodiesel output in 2010 were three states, Rio Grande do Sul (RS) with 31% in the South region and Mato Grosso (MT) and Goiás, in the Center-West region with 27% and 22.6%, respectively, not considering mills running on tallow.

2.6.5 Entry and Capacity Expansion in Biodiesel

Including already authorized entries and capacity expansions into the analysis, allows to observe that the disagreement between the data and the model is being reduced. In table (2.15) results of the Probit regression and Probit and Tobit marginal effects are presented for existing and authorized to be built biodiesel mills location.

Table 2.15: Regressions for Biodiesel Mills Location with Entry and Expansion

	Probit		Tobit	Units
	Coefficients	Marg. Effects	Marg. Effects	
Intercept	-5.576			
	(4.016)			
Dist Santos Port	-0.053	-0.010	0.034	10^3 Km
	(0.493)	(0.092)	(0.085)	
Soy Quantity	0.596**	0.111**	0.050*	10^6 tons/year
	(0.262)	(0.052)	(0.028)	10^3 tons/year = + 0.011%
Soy Price	0.686	0.128	0.231	BRL/kg
	(2.527)	(0.472)	(0.392)	
Dist Crush Mill	−1.163*	−0.217**	−0.310***	10^3 Km
	(0.639)	(0.109)	(0.113)	1 km = - 0.0217%
Veg Oil Labor	1.716**	0.320**	0.258*	10^3 Workers
	(0.820)	(0.160)	(0.136)	1 Worker = + 0.032 %
Area	7.233	1.347	0.823	10^6 km^2
	(11.591)	(2.170)	(1.645)	
Highway Density	-1.298	-0.242	0.039	Km/km^2
	(3.620)	(0.673)	(0.648)	
Railway Density	14.030*	2.612*	2.000	Km/km^2
	(8.086)	(1.442)	(1.327)	m/km^2 = + 0.2612 %
Trucks + Buses	-3.476	-0.647	-0.227	10^6 units
	(8.142)	(1.519)	(1.340)	
Diesel Price	2.241	0.417	0.486	BRL/liter
	(2.163)	(0.407)	(0.382)	
Fuel Term Dist	-0.116	-0.022	0.016	10^3 Km
	(1.700)	(0.317)	(0.255)	
Cattle	0.307	0.057	0.033	10^6 heads
	(0.241)	(0.046)	(0.033)	
Hogs	-0.634	-0.118	-0.165	10^6 heads
	(0.665)	(0.125)	(0.127)	
Poultry	0.027	0.005	0.005	10^6 heads
	(0.019)	(0.004)	(0.004)	
McFadden R-sq.	0.2303			
Wald Test	49.42			
p-value	0.0000			
Log-Likelihood	-101.022			
AIC	232.04			
BIC	287.90			
N	306			

Note: Robust standard errors in parenthesis for column 2 and delta-method for columns 3 and 4.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

The model including authorized entries and capacity expansions yields a better adjusted R-squared and is statistically significant. But more important is that the model unveils almost all expected results.

The average quantity of soybean produced (Soy Quantity) remains positive and statistically significant. At the sample average, each additional thousand tons of soybean produced in a micro-region increases probability of observing a biodiesel mill by 0.011 % or 0.005 %. Each additional standard deviation increases proabibilities by 5.3 % or by 2.41 %.

Distance to the nearest soybean crushing mill (Dist Crush Mill) remains statistically significant, with increased importance both in the significance level and in the point estimate coefficient. Each Km away from a soybean crushing mill reduces probabilities by 0.0217 % or by 0.031 %. One additional standard deviation reduces probabilities by 5 % or by 7.15 %.

The average number of workers in raw vegetable oil production (Veg Oil Labor) is positive and becomes statistically significant in both regressions. Each additional worker in this sector increases probabilities by 0.032 % or 0.0258 %. One additional standard deviation increases probabilities by 4.80 % or 3.87 %.

Railway density (Railway Density) becomes statistically significant and with the expected positive sign, but only in the Probit regression at the 10% significance level. Each additional meter per squared kilometer increases probabilities by 0.2612 %. One additional standard deviation increases probabilities by 3.78 %.

Only supply variables are jointly significant as described in table (2.16).

Table 2.16: Probit Joint Significance by Variable Category

	Supply	Infrastructure	Demand	D_1	D_2
Number of Variables	6	2	6	3	3
χ^2	20.16	3.01	5.49	1.28	3.72
p-value	0.0026	0.2218	0.4826	0.7351	0.2937

The Probit model reaches an accuracy 86.93 % at a 47 % cut-off level. Retrieved probabilities from the Tobit model have the same maximum accuracy at a 49 % cut-off level, but it attributes less positive observations to existing mills. Results are detailed in table (2.17), and the model misclassifies 40 micro-regions.

Table 2.17: Accuracy of Probit Regression For Biodiesel Mills

	< 47%	≥ 47%	Total
0	252	7	259
1	33	14	47
Total	285	21	306

In figure (2.5), estimated probabilities are plotted against the existing and authorized biodiesel mills.

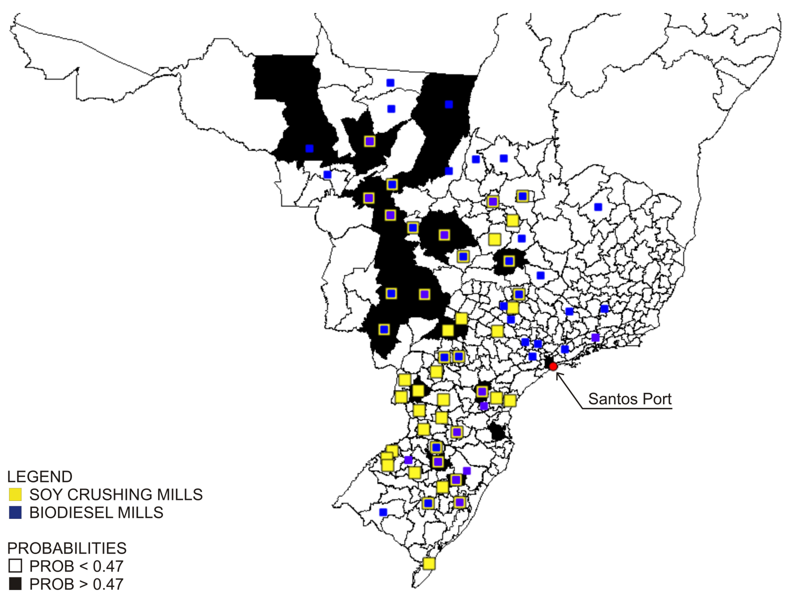

Figure 2.5: Probit Regression for Biodiesel Mills with Entry and Expansion

Considering authorized entries and capacity expansions also changes Tobit estimates as presented in table (2.18).

Table 2.18: Regressions for Biodiesel Mills Installed Capacity with Entry and Expansion

	Coefficients	$\partial E[C\|X, C > 0]$	$\partial E[C\|X]$	Units
Intercept	$-1,562,672$*			$m^3/year$
	(914,129.7)			
Dist Santos Port	44.316	7.637	4.691	Km
	(112.691)	(19.327)	(11.784)	
Soy Quantity	65.974*	11.369*	6.983*	10^3 tons/year
	(33.723)	(5.937)	(3.987)	
Soy Price	302.692	52.160	32.038	BRL/ton
	(505.936)	(87.522)	(54.461)	
Dist Crush Mill	-405.288**	-69.839**	-42.897***	Km
	(173.077)	(27.367)	(15.198)	
Veg Oil Labor	337.367*	58.135*	35.708*	Workers
	(172.259)	(29.761)	(19.392)	
Area	1.076	0.185	0.114	km^2
	(2.151)	(0.370)	(0.228)	
Highway Density	51,121.19	8,809.152	5,410.9	Km/km^2
	(848,281.3)	(146,193.8)	(89,825.78)	
Railway Density	2,615,767	450,746.4	276,864.7	Km/km^2
	(1,789,148)	(298,328.3)	(17,6861.5)	
Trucks + Buses	-297.021	-51.182	-31.438	10^3 units
	(1,827.37)	(314.960)	(193.623)	
Diesel Price	635,527.4	109,513.4	67,267.12	BRL/liter
	(493,972.3)	(85,078.21)	(53,493.2)	
Fuel Term Dist	20.627	3.555	2.183	Km
	(333.686)	(57.459)	(35.244)	
Cattle	42.539	7.330	4.503	10^3 heads
	(41.911)	(7.273)	(4.597)	
Hogs	-215.707	-37.170	-22.831	10^3 heads
	(165.588)	(28.552)	(18.002)	
Poultry	6.998	1.206	0.741	10^3 heads
	(5.044)	(0.871)	(0.553)	
McFadden R-sq.	0.040			
F-Test	3.21			
p-value	0.0001			
Log-Likelihood	-710.391			
AIC	1,452.782			
BIC	1,512.359			
N	306			

Note: Robust standard errors in parenthesis in column 2 and delta-method in columns 3 and 4.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

The model exhibits a poor fit measured in terms of pseudo R-squared but is statistically significant.

The average quantity of soybean produced in a micro-region (Soy Quantity) has a positive contribution to capacity and is statistically significant. Each additional thousand tons of soybean produced increases plant capacity by 11.4 $m^3/year$. One additional standard deviation should increase capacity by 5,481.76 $m^3/year$.

Distance to the nearest soy crushing mill (Dist Crush Mill) is negative as expected and statistically significant. Each kilometer away from the nearest soy crushing mill decreases capacity by 69.8 $m^3/year$. One additional standard deviation should decrease capacity by 16,127.35 $m^3/year$.

The average number of workers in the raw vegetable oil industry (Veg Oil Labor) is positive and statistically significant. This is a proxy for the capacity of the soybean crushing mill in the micro-region. Each additional worker in raw vegetable oil production should increase biodiesel output capacity by 58.14 $m^3/year$. Each additional standard deviation should increase biodiesel mills capacity by 8,714.62 $m^3/year$.

The variable (Hogs) loses its statistical significance.

Tests for joint significance presented in table (2.19) for the Tobit regression reveals that only supply variables are jointly significant.

Table 2.19: Tobit Joint Significance by Variable Category

	Supply	Infrastructure	Demand	D_1	D_2
Number of Variables	6	2	6	3	3
F-test	3.35	1.21	0.98	0.58	1.09
p-value	0.0033	0.3011	0.4381	0.6300	0.3523

The emerging pattern is very clear. Biodiesel mills capacity depend exclusively on feedstocks. It depends on soybean availability, proximity to a soy crushing mill or both. Moreover, capacity of the biodiesel mill is proportional to the capacity of the soybean crushing mill.

2.7 Towards First Generation Bio-refineries

There are currently 21 micro-regions in the sample that already hosts both types of biofuels mills.

To understand what economic forces may allow the flourishment of first generation biorefineries, such as the one aforementioned, probit regressions are plotted in figure (2.6), identifying micro-regions with high probability of hosting both type of biofuel mills.

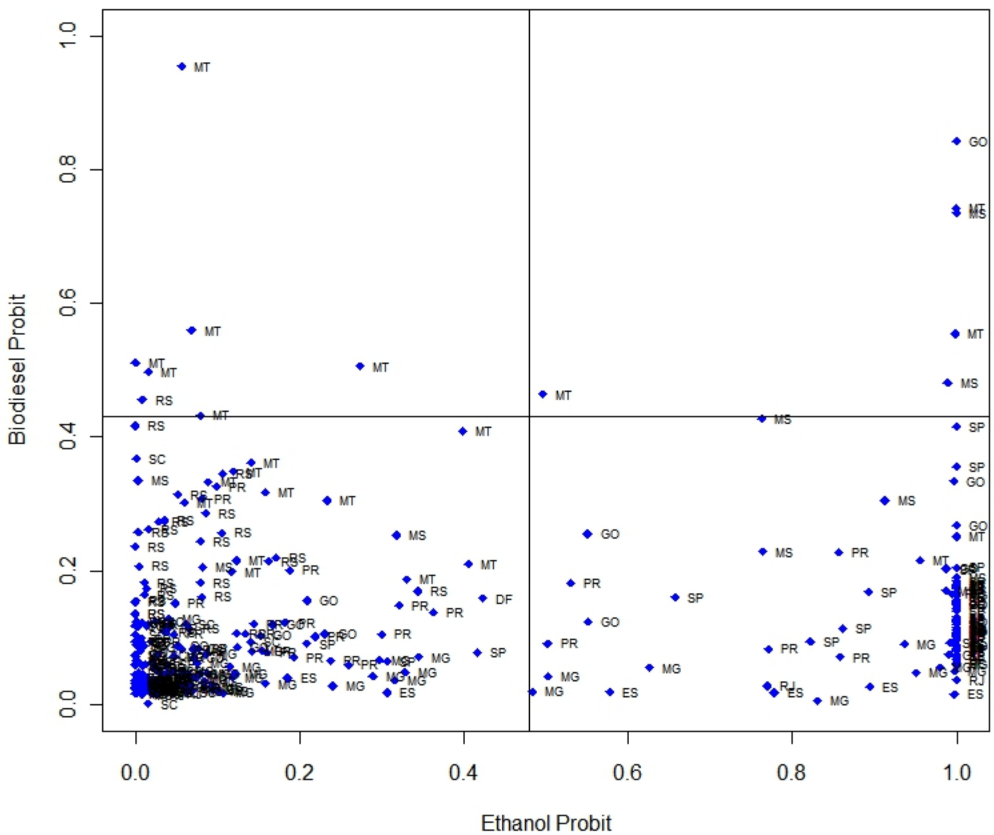

Figure 2.6: Biorefineries

Figure (2.6) reveals 6 micro-regions, all in the Center-West region, that are suitable for hosting both types of biofuel mills. Although the unique existing biorefinery is not included in these 6, this procedure captures one micro-region adjacent to it.

But the most important result is that together with the economic forces identified by isolated probit regressions, in particular to be near raw material supply for both types of biofuels, there is evidence that biorefineries should locate more or less on the transition of land use from sugarcane to soybean. Although, as sugarcane logistics is more costly and difficult, it becomes the determinant factor in the co-existence of both biofuel mills.

Figure (2.6) shows the 21 existing micro-regions that already hosts both types of mills, the 6 micro-regions pinpointed by probit regressions, and where they coincide together with transition of land use in 2010 calculated in the previous chapter.

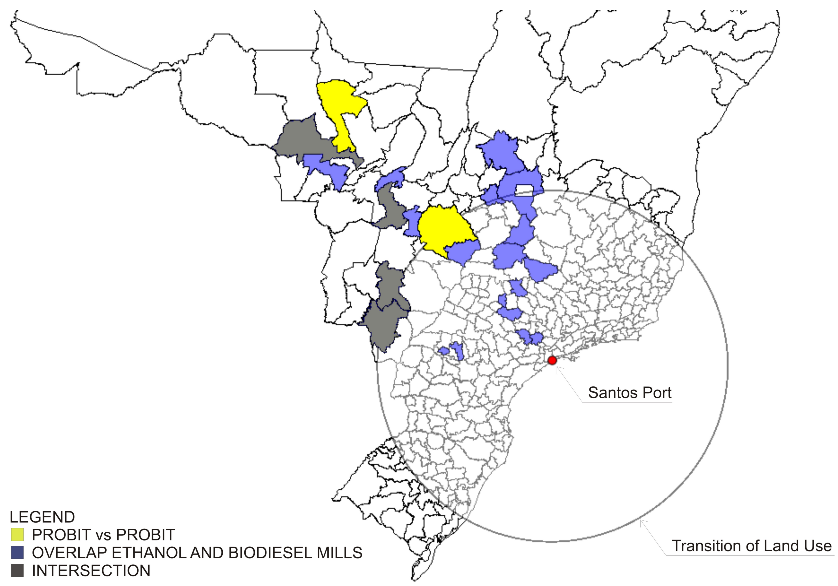

Figure 2.7: Potential Biorefineries Sites and Transistion of Land Use

The reason why this result works only in the northwest direction and not in the northeast or southwest direction are due to climatic conditions. Going southeast the sugarcane land use zone is interrupted more or less around the tropic of Capricorn when climate changes from tropical to temperate. Yet, taking into consideration biodiesel entry from figure (2.4), this will probably happen soon.

To the northeast the estimated point of land use transition coincides also with the beginning of a semi-arid biome, unsuited for most types of agriculture.

Including entry and expansion of biodiesel mills the number of micro-regions with both types of mills increase to 23.

The sort and size of economies of scope in the production of both biofuels is still to be further understood and explored.

2.8 Conclusion

This paper investigates empirically, using Probit and Tobit regressions, location and capacity decisions drivers for ethanol and biodiesel mills in Brazil using regional variables.

The Probit regression shows that ethanol mills locate in micro-regions with abundant and low-priced sugarcane supply, high river density, high number of automobiles and near ethanol storage terminals. Estimated probabilities provide evidence that ethanol production agglomerates in São Paulo state and surroundings.

The Tobit regression employed to analyze ethanol mills capacity decisions, measured by the number of workers in ethanol manufacturing in each micro-region in 2010, yields almost the same set of statistically significant variables as the Probit regression. Ethanol output capacity is influenced by sugarcane availability at low prices, access to water sources measured by river density, high number of automobiles and proximity to ethanol storage terminals. It also indicates land availability for sugarcane production expansion, captured by the number of cattle heads in the micro-region, as a statistically significant variable.

Ethanol Probit and Tobit regressions do not exhibit conflicting results. This assertive stems from the absence of statistically significant variables with contradictory signs in the two regressions. In the Probit regression all categories of variables are jointly significant, while in the Tobit regression only supply and demand variables are jointly significant.

The biodiesel program is particularly interesting because it gives clear location incentives. It is designed to promote biodiesel production from castor beans in the Northeast region and from oil palm in the North region, especially if produced by small farmers.

In fact, these government incentives given for selected feedstocks and locations steered decisions and attracted investors on the onset of the biodiesel program. But despite all incentives, time revealed, ex-post, to be more economically attractive the use of soybean oil in biodiesel manufacturing. Soybean oil is a feedstock mainly available in the Center-West, Southeast and South regions, produced by an already existing, mature, at the technological frontier and competitive in international markets agroindustry.

The soybean industry is the major agroindustrial activity and the top exporter of agricultural commodities in Brazil and consequently could immediately provide large volumes of vegetable oil to biodiesel production.

This resulted in an initial movers disadvantage. More precisely, ex-ante uncertainties regarding the economic attractiveness of feedstocks and technological process to be deployed coupled with government feedstock and location incentives steered initial entrants to sites unsuited for the usage of soybean oil. To highlight this point, biodiesel produced from castor beans does not meet the standards required by the National Petroleum Agency as due to its high viscosity.

The biodiesel industry has huge excess capacity but paradoxically, continues to observe entry. This can be almost fully explained by the mislocation of some biodiesel mills for the usage of soybean oil, especially initial movers. There are still 27 micro-regions with unattended soy crushing mills that could potentially host biodiesel mills. It is expected to observe new entrants, better placed to use soybean oil, simply because they will have lower feedstock procurement costs than mislocated incumbents.

In fact, when accounting for already authorized entries and capacity expansions by the National Petroleum Agency, there is strong evidence that these dynamics are already being observed and eventually, in the long run, mislocated mills should be driven out of the market. Including authorized entries and expansions, the Probit regression shows that biodiesel mills locate in micro-regions with abundant soybean supply, close to soybean crushing mills and with high number of workers in raw vegetable oil production.

Biodiesel capacity decisions, explored in the Tobit regression, indicates the influence of the same variables. Therefore, biodiesel production capacity in a micro-region is proportional to the soybean crushing capacity in the same micro-region, measured by the number of workers in raw vegetable oil production. These results are intuitive. As the vegetable oil content in soybean is low, a stand alone biodiesel mill has to find markets for the soybean meal it will produce. Whereas a biodiesel mill coupled with a soy crushing mill does not inccur in this additional burden.

Biodiesel Probit and Tobit estimates are similar.

For each biofuel, it was identified that almost the same set of statistically significant variables drives location and capacity decisions. It follows that the influence these variables exert on biofuel industry dynamics is much more pronounced than if location or capacity decisions were considered in isolation. This happens because as explanatory variables are shocked, their effect is transmitted through two channels, the location decision of entry or exit, and the capacity decision of expansion or contraction. From Tobit estimates these two effects are very clear. Marginal effects on the probability that the dependent variable becomes uncensored can be interpreted as the location decision, while marginal effects on the dependent variable, conditional on being uncensored, can be interpreted as the capacity decision.

The categories of jointly statistically significant variables for each biofuel reflect the way competition with fossil fuels occurs and how market sizes are determined. For ethanol, considering both Probit and Tobit regressions, supply and demand variables are jointly statistically significant. This occurs because the market size for ethanol is endogenously determined by competition with gasoline. Thus, demand variables are also important. For biodiesel, on the other hand, also considering Probit and Tobit regressions, only supply variables are jointly statistically significant. This occurs because the market size for biodiesel is exogenously determined by a blending mandate into diesel. Therefore, demand variables are not as important.

Regarding the production of both biofuels, there is clear evidence of the uttermost importance of access to agricultural feedstock in large scales. In all regressions supply variables are jointly significant. In particular, raw material variables coefficients, sugarcane and soybean, are always positive and statistically significant. Thus, biofuel production is ultimately linked to the underlying industrial organization of the feedstock provider agricultural sector and depends on its dynamics and bottlenecks. This link is forged by technological factors and cost structure derived from the bioenergy production function.

The model can be applied to make inferences about entry or exit and capacity expansions or contractions by shocking explanatory variables. For example, analyzing the optimal micro-region to build a new ethanol storage terminal or a new soybean crushing mill that will induce greater entry and capacity expansions in biofuel production. Results may also be used to identify which variables or policies can affect both industries.

In terms of policy implications, it is necessary to look primarily for actions that may have a widespread impact on the industry.

As access to raw materials is very important, any policy that can improve efficiency, competitiveness, profitability and resilience in the agricultural sector should be pursued, e.g. foreign trade policy, improving infrastructure and research and development.

Specifically for ethanol, improving storage infrastructure is important for industry expansion as it can potentially affect entry and capacity in many micro-regions. During sugarcane off season, ethanol production slumps, but supply has to be smoothed throughout the year to avoid major price spikes. Besides, ethanol storage represents an additional cost for producers or wholesalers. Policy makers could foster investments in ethanol storage terminals and take measures to relief the additional costs of storing it.

There is a large untapped potential in the sugar and ethanol industry to produce bio-electricity from sugarcane bagasse, caused by past and current regulatory framework. Policies can be implemented to make bio-electricity more competitive with other sources of power generation. In particular, current regulation does not take into account positive externalities provided by sugarcane bagasse based bio-electricty production. First, it is a renewable energy source. Second, sugarcane harvesting season coincides with the winter, a dry period when hydroelectric reservoirs are not being replenished. Thereby, it is complementary to the main power generation source in Brazil. Third, it is mainly located inside the most economically important state in Brazil, near power demand.

A policy to promote bio-electricity should also contemplate upgrading old and inefficient co-generation technologies for new ones and providing better access of sugar and ethanol mills to the power transmission grid. In the future, sugar and ethanol millers may also opt to use sugarcane bagasse to produce second generation biofuels.

Any active policy for the biodiesel industry has to take into account that the use of diesel in Brazil is very inefficient. The country has a transport mix mainly based on road transportation which is very diesel-intensive per ton of goods carried compared to rail and water transportation. Thus, any investment or subsidies to this sector should be compared to other means of reducing diesel demand by directly investing in changing the transport mix. Diesel consumption could be reduced by improving transportation infrastructure in the country connecting cities and ports and inside cities, moving to less diesel intensive transports.

To deal with the problem of excess capacity, mislocation and continued entry many policies can be implemented.

If government opts for a do nothing policy, continued entry will be observed as there are still many empty locations with lower productions costs and in the long run mislocated mills will be driven out of the market.

Alternatively, government could increase the biodiesel blending mandate. The biodiesel producers trade union proposes to reach B7 by 2013 and B10 by 2014. This policy could also involve different regional blending mandates to adjust for local idle capacity.

Excess capacity could also be used to export biodiesel while domestic diesel demand does not catch up with installed biodiesel capacity. But this is also very difficult as biodiesel is not cost competitive with regular diesel.

It results that excess capacity also makes it unlikely that in the short and medium term, other oil crop will be employed as feedstock in large scale.

Combining results from Probit regressions, it is possible to identify micro-regions that could potentially host both types of biofuel mills. These sites are more or less where transition of land use from sugarcane to soybean occurs because it is necessary to have access to both feedstocks, sugarcane and soybean or soybean oil. But it is actually the presence of sugarcane production that determines the possibility of co-existence of both biofuel mills, due to its higher logistic complexity and its after harvest deterioration. Yet, the sort and size of economies of scale and scope that may exist between the two biofuels industries is to be further analyzed and understood. This study represents one step in analyzing the economics of future bio-refineries.

The results found here consent to speculate on some further implications. In particular, there seems to be a pattern for a first generation biofuel program to be successful, or alternatively, to have the minimum subsidies requirements.

Biofuel production starts in general as a spin-off of an already existing, mature, at the technological frontier and competitive in international markets agroindustry. Main examples are the production of ethanol in the United States from corn and in Brazil from sugarcane. For biodiesel production, examples are soybean in the United States, Brazil and Argentina, palm oil in Indonesia and Malaysia and rapeseed in the European Union.

The statement that bioenergy should derive from consolidated exporting agroindustries provides an explanation of how this source of energy can actually contribute to energy security, in contrast with production in small scale from small farmers.

Finally, existing agro-industrial complexes, due to their economic importance, are also more likely and more experienced in influencing policies to their benefit.

All these considerations are much in line with the concept of agro-industrial districts. Moreover, the formation of these bioenergy producing regions is endogenous and obeys to great extent, the economic forces identified on theoretical and empirical locational models. This endogeneity means that exogenously given natural resources endowment adequate for agriculture are necessary, but not sufficient conditions for the emergence of bioenergy production. They need to be combined with endogenous factors such as economic conditions, markets and market structure, policies, deployed technologies, interaction with other industries, research and development, infrastructure and institutions among others.

2.9 Appendix A

Table 2.20: Ethanol Summary statistics

Variable	Units	Mean	Std. Dev.	Min.	Max.	N
Ethanol Mill	Binary Variable	0.353	0.479	0	1	306
Employment	Workers	271.003	722.004	-1 *	6,771	306
Dist Port	Km	648.563	362.882	12.91	2,037.949	306
Cane Quantity	10^3 tons	1,682.546	3,930.231	0	27,595.504	306
Cane Price	BRL/ton	53.977	35.326	24.806	237.384	306
Area	Km^2	10,113.613	14,308.643	601.1	123,950	306
Highway Density	Km/Km^2	0.068	0.034	0.008	0.176	306
Rail Density	Km/Km^2	0.012	0.014	0	0.091	306
Grid Density	Km/Km^2	0.018	0.017	0	0.082	306
River Density	Km/Km^2	0.034	0.016	0	0.1	306
Gasohol Price	BRL/liter	2.238	0.105	2.136	2.514	306
Automobiles	10^3 cars	88.878	329.613	0.65	4,949	306
Et Storage Dist	Km	400.465	279.782	0	1,400.035	306
Fuel Term Dist	Km	106.53	74.759	0	431.802	306
Cattle	10^3 Heads	445.765	590.956	0	3,833.458	306

* As there are observations with zero employment, censored observations were given a value of -1.

Table 2.21: Biodiesel Summary statistics

Variable	Units	Mean	Std. Dev.	Min.	Max.	N
Biodiesel Mill	Binary Variable	0.131	0.338	0	1	306
Biodiesel Mill 2	Binary Variable	0.154	0.361	0	1	306
Capacity	$m^3/year$	16,844.289	61,478.518	0	528,120	306
Capacity 2	$m^3/year$	24,185.224	78,023.831	0	544,320	306
Soy Quantity	10^3 tons	170.604	482.167	0	5,578.208	306
Soy Price	BRL/ton	469.798	46.904	320.172	708.586	306
Soy Crush Dist	Km	243.559	230.779	0	880.647	306
Veg Oil Labor	Workers	46.288	149.903	0	1,323.333	306
Diesel Price	BRL/liter	1.726	0.072	1.674	1.925	306
Trucks + Buses	10^3 units	6.031	13.956	0.166	200.706	306
Hogs	10^3 heads	91.602	194.463	0	1,988.459	306
Poultry	10^3 heads	3,182.355	5,852.943	0	35,421.836	306

2.10 Appendix B

Coordinate Reference System

All shapefiles were projected using SAD69/Brazil Polyconic projection which uses as references, the 0° parallel (Equator line) and the 54° West meridian. All distances were obtained in this projection.

Granularity

As the political division of 2007, Brazil had 5 levels of granularity. Country (1), Regions (5) equivalent to NUTS 1, States (27) equivalent to NUTS 2, Meso-Regions (137) , Micro-Regions (558) equivalent to NUTS 3, Municipalities (5564). Shapefiles were obtained from www.ipea.gov.br retrieved in January 2011.

Agricultural Data

Quantity of sugarcane and soybean produced per year, yield per hectare, harvested and planted area according to the Brazilian National Statistics Bureau, Instituto Brasileiro de Geografia e Estatística, www.ibge.gov.br, Pesquisa Agrícola Municipal. Data at micro-region level from 1990 to 2010.

Labor Data

Obtained from Labor Ministry, Ministério do Trabalho e Emprego, www.mte.gov.br, RAIS, retrieved in November 2011. Data on formal employment as of 31.12 of each year, from 1994 to 2010 per micro-region.

Ethanol Mills

Authorized ethanol suppliers by municipality informed by the Brazilian National Petroleum Agency, Agência Nacional do Petróleo, www.anp.gov.br updated in 27.09.2011.

Biodiesel Mills

Authorized plants to operate by the Brazilian National Petroleum Agency, Agência Nacional do Petróleo, www.anp.gov.br, located in the Center-Southeast-South political regions of Brazil. Data accessed on October, 2011. Data refers to each plant, capacity expressed in $m^3/year$ and municipality where it is installed. Totally, 43 micro-regions had one or more biodiesel plants with a summed installed capacity of $15,131m^3/day$.

Ethanol Storage Terminals

Existing Ethanol Storage Terminals according to the Brazilian Petroleum Institute, Instituto Brasileiro de Petróleo, www.ibp.org.br, retrieved in September 2011.

Fuel Distribution Terminals

Existing fuel distribution terminals according to the Sindicato Nacional das Empresas Distribuidoras de Combustíveis e Lubrificantes, www.sindicom.com.br. Distribution terminals are categorized into primary and secondary, yet this distinction is not relevant here. On September 2011, there were 45 fuel distribution terminals in the Center-Southeast-South political region of the country.

Soybean Crushing Mills

Existing and operating soy crushing mills according to ABIOVE - Associação Brasileira das Indústrias de Óleos Vegetais, Brazilian Association of Vegetable Oil Industries, www.abiove.com.br in 2010. There were 45 micro-regions with one or more soy crushing mills, in the Center-Southeast-South political region of the country.

Highways

Shapefile from Ministry of Transportation, Ministério dos Transportes, www.transportes.gov.br retrieved in September, 2011.

Railways

Shapefile from Ministry of Transportation, Ministério dos Transportes, www.transportes.gov.br retrieved in September, 2011.

Electricity Transmission Grid

Shapefile from the Ministry of Enviornment, Ministério do Meio Ambiente, www.mma.gov.br retrieved in September 2011. The original data source is from the National Electricity Authority. Only operating transmission grid of different voltages were included.

Rivers

Shapefile from the National Electricity Authority, www.aneel.gov.br retrieved in September 2011. The original data source is from the Brazilian National Statistics Bureau, Instituto Brasileiro de Geografia e Estatística, www.ibge.gov.br on a 1:2,500,000 scale.

Livestock Data

Data on cattle, poultry and hogs from the National Statistics Bureau, Instituto Brasileiro de Geografica e Estatística, www.ibge.gov.br, Pesquisa Pecuária Municipal.

Fossil Fuel Prices

Data from the Brazilian National Petroleum Agency, Agência Nacional do Petróleo, www.anp.gov.br. Average state level prices for gahosol and diesel.

Area

Data from the National Statistics Bureau, Instituto Brasileiro de Geografica e Estatística, www.ibge.gov.br, 2010 Census.

Automobiles, Trucks and Buses

Data on number of automobiles, trucks and buses from DENATRAN - Departamento Nacional do Trânsito. www.denatran.gov.br

Chapter 3

Bioenergy Clusters: A Model of Bioenergy Production

3.1 Introduction

Climate change mitigation efforts will require that regions and countries deploy a portfolio of low-carbon energy technologies that better adapts to their local conditions such as natural resources endowments, human capital, technological expertise, economic structure and energy demand, among others.

Due to these diverse regional characteristics, some are more prone to benefit from different types of low-carbon energy solutions such as wind power, solar energy, hydroelectricity, geothermal, bioenergy and may require more or less investments in carbon capture and storage.

Bioenergy in particular is expected to contribute more to the energy mix in countries with vast arable land, low population density, well defined land property rights and some agricultural sectors at the technological frontier.

Among existing and commercial low-carbon energy technologies, the production of biofuels from biomass, bioenergy, is one of the most cost competitive with fossil fuels. Additionally, these technologies are already well known and current energy distribution infra-structure and end use applications can be adapted for bioenergy products.

According to Best (2003) [5], agro-energy can contribute significantly to climate change mitigation goals, energy security and rural development. The author claims that currently there is a large untapped potential for agro-energy production if already existing state-of-the-art technologies were deployed in less developed countries.

Reilly and Paltsev (2008) [45] forecast that whether or not climate mitigation policies are in place, bioenergy production will expand. In fact, bioenergy production is also being fostered to meet other policy goals. Frayssignes (2011) [18] advocates that agro-energy production can be a promising development strategy for agricultural and rural areas.

Additionally, future demand for biomass based products is probably underestimated because hydrocarbons are also used for non-energy purposes such as petrochemical applications, asphalts, lubricants and fertilizers. Eventually these products also will have to be displaced by renewable sources. Thus, either more bioenergy will be required to displace fossil fuels or more biomaterials will be required to substitute for non-energy use of hydrocarbons.

This paper proposes a microeconomic framework for modeling several types of bioenergy production. Drawing upon the Brazilian experience, three bioenergy chains are considered in detail; (1) sugarcane-ethanol, (2) soybean-biodiesel and (3) forestry-firewood-charcoal-paper and pulp-black liquor. Currently, these are the main sources of biomass energy used in Brazil.

This single framework accommodates all these production possibilities with a nested structure for biomass output in the lower nest and biomass upgrading in the upper nest. It uses a multiple output production function to allow for the production of a vector $\vec{F}$ with five elements which are food, feed, fuels, fibers and forestry products. It can also be employed for modeling bioenergy production from any biomass feedstocks in other regions of the world as well as for second generation technologies of biofuels.

The idea that these microeconomic production structures forms bioenergy clusters from a statistical standpoint is put forward here. A bioenergy cluster consists in a set of similar micro-regions using the same technology, which employs labor, capital and land to produce a primary biomass feedstock that is further processed with labor and capital to generate the output vector $\vec{F}$.

To test this concept, four k-means cluster analyses are performed using a dataset comprised of 306 micro-regions in the Brazilian Center-West, Southeast and South regions, with 15 variables referring exclusively to the nested multiple output production function averaged from 2006 to 2010.

According to Hair et al. (2006) [26], *"The resulting clusters should exhibit high internal (within-cluster) homogeneity and high external (between-cluster) heterogeneity"*.

Results allow to assign each micro-region to one of the three types of bioenergy clusters considered plus one additional cluster unsuited for bioenergy production.

Plotting results of cluster membership in a Geographic Information System to further understand the underlying spatial context of each cluster, reveals that there is geographic proximity among micro-regions belonging to the same cluster. It also allows to determine each cluster location, extension and boundaries in space.

Furthermore, the forestry cluster shows that more than one agglomeration of bioenergy producing micro-regions, employing the same feedstock, can co-exist separated in space in line with the definition of bioenergy clusters presented here.

Micro-regions assigned to one of the bioenergy clusters are expected to be carbon sinks while micro-regions unsuited for bioenergy production are expected to be carbon sources.

Results can be used to identify similar clusters that could benefit from targeted economic, agricultural, energy, environment or climate policies. For example, bioenergy clusters could be entitled to receive payments for environmental services or subsidies derived from revenues of a carbon tax.

3.2 Background

In 2011, primary energy production in Brazil amounted to 256.8 billion tons of oil equivalent.[1] Of this total, 117.7 billion tons of oil equivalent, around 45.8%, were produced from renewable sources. In order of importance, the contribution to renewable energy production comes from sugarcane products (43.3 Gtoe), hydroelectricity (36.9 Gtoe), firewood (26.3 Gtoe) and other renewables (11.2 Gtoe) which includes biodiesel, other biomass, wind energy and biogas. Primary energy production evolution, detailed by source, is presented in figure (3.1).

[1] Data from Brazilian Ministry of Mines and Energy.

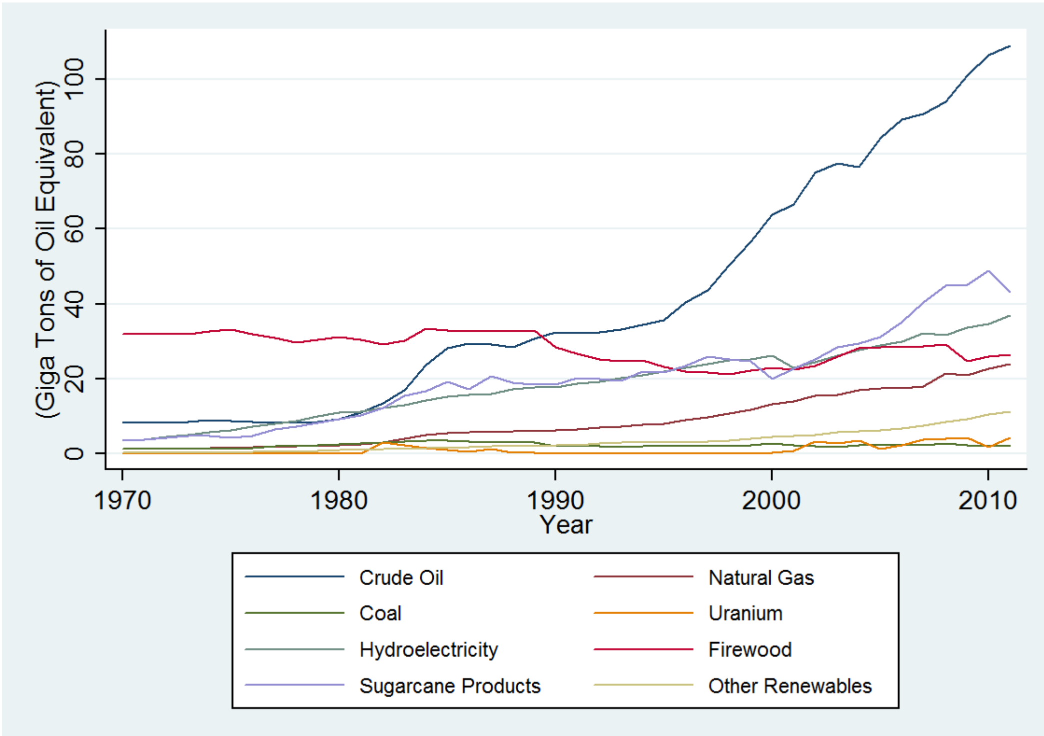

Figure 3.1: Primary Energy Production in Brazil from 1970 to 2011

Also in 2011, total biomass consumption for energy purposes amounted to 65.5 giga tons of oil equivalent, which represents 28.6% of total energy consumption.[2] Sugarcane products amounted to 38.1 Gtoe divided into sugarcane bagasse (27.3 Gtoe) and ethylic alcohol (10.8 Gtoe). Forestry products amounted to 26.0 Gtoe, divided into firewood (16.3 Gtoe), charcoal (5 Gtoe) and black liquor (4.7 Gtoe). Other wastes amounted to 1.4 Gtoe.

Thus, the contribution of bioenergy to the energy mix is quite significant in Brazil. It is particularly important to some specific sectors such as road transportation, power generation and production processes in food and beverage, ceramics, paper and pulp, pig-iron and steel and iron alloys industries, to cite the most relevant.

Firewood and charcoal are also used by low income households for cooking and heating, but according to Cropper and Griffiths (1994) [10] and Goldemberg and Coelho (2004) [21] in an unsustainable and unhealthy way.

[2]Biodiesel is not included in this figure as it is classified into the other renewable sources category, together with wind energy. The difference of this figure with primary energy production, of 4.1 Gtoe, is due to non-energetic uses of biomass.

In the next section, a description of the production chain and output portfolio of the three bioenergy clusters analyzed here is presented as a step in the construction of the formal model.

3.3 Examples of Bioenergy Clusters

Energy production from modern biomass exhibit a common feature independent of the feedstock being used. It is possible to divide the production chain into two distinct activities; the agricultural or upstream phase and the industrial or downstream phase. The agricultural phase refers to crop plantation and harvesting, done outside the mill. The industrial phase refers to activities done inside the mill, which consists in upgrading biomass into a portfolio of products and by-products such as food, feed, fuels, fibers and forestry products.

3.3.1 Sugarcane and Ethanol

Bioethanol production is defined by the International Energy Agency (2007) [1] as a *"...well known process based on enzymatic conversion of starchy biomass into sugar, and/or fermentation of 6-carbon sugars with final distillation of ethanol to fuel grade."*

Output portfolio in the sugarcane-ethanol industry comprises sugar of different quality types, sugar molasses, food grade ethanol for the food and beverage industry, industrial grade ethanol, fuel grade ethanol, i.e. hydrated and anhydrous ethanol, sugarcane bagasse and bio-electricity from sugarcane bagasse.[3]

Figure (3.2) provides a simplified scheme of the sugarcane-ethanol industry production chain. The elasticity of substitution between one output and another is given by the parameter (σ). When the elasticity of substitution is low or inexistent, $\sigma = 0$. When the elasticity of substitution is high or very high, $\sigma = \infty$. For example, when sugarcane is crushed, sugarcane juice and molasses and sugarcane bagasse are produced in almost fixed quantities, thus $\sigma = 0$. On the other hand, substitutability between sugar and ethanol is high, i.e. sugar production competes with ethanol production.

[3] There are basically two types of traded sugars based on their quality: raw, non-food grade sugar, and refined, food grade, sugar.

The miller can, to a certain extent, choose the quantity of ethanol and sugar to be produced depending on prevailing or expected sugar and gasoline prices. In this case, $\sigma = \infty$. The same holds for the decision to produce electricity or cellulosic ethanol from sugarcane bagasse.

According to the Food and Agriculture Organization (2008) [17], using Brazilian figures, one hectare can produce 73.5 tons of sugarcane and each ton of sugarcane can be converted into 74.5 liters of ethanol. Thus, one hectare can produce up to 5,476 liters of ethanol per year.

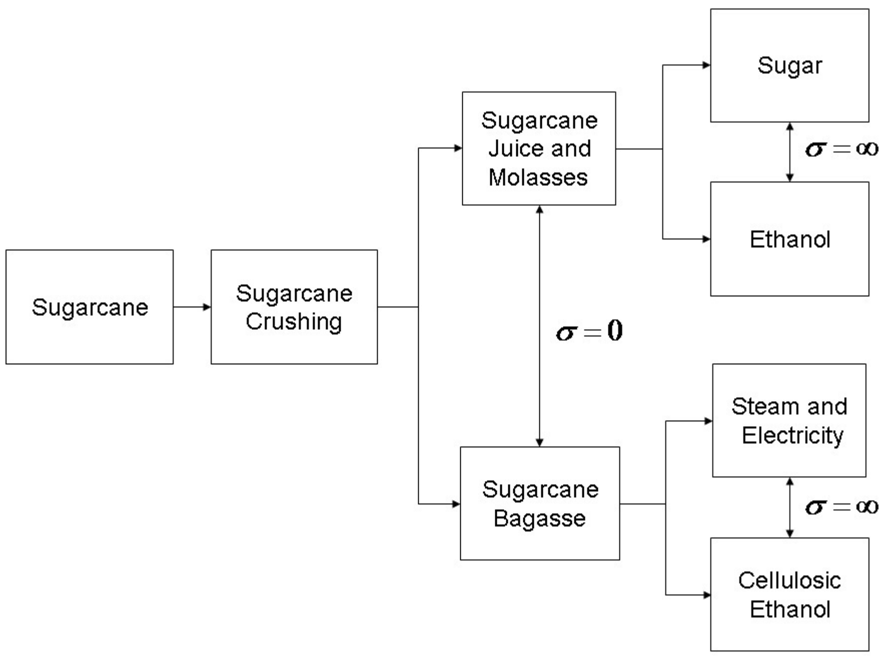

Figure 3.2: Sugarcane Production Chain

In figure (3.3) the demand for sugarcane bagasse and ethanol in Brazil from 1970 to 2011 is depicted. During this period, these two energy products exhibit a correlation of 0.93.

Fuel grade ethanol is used almost exclusively in road transportation in spark ignition (Otto cycle) engines. There is minor use in motorcycles and as aviation fuel. Consequently, it displaces gasoline.

Sugarcane bagasse, which results from the sugarcane crushing process is, with current technology, better used either as fuel for power generation or as fuel for industrial

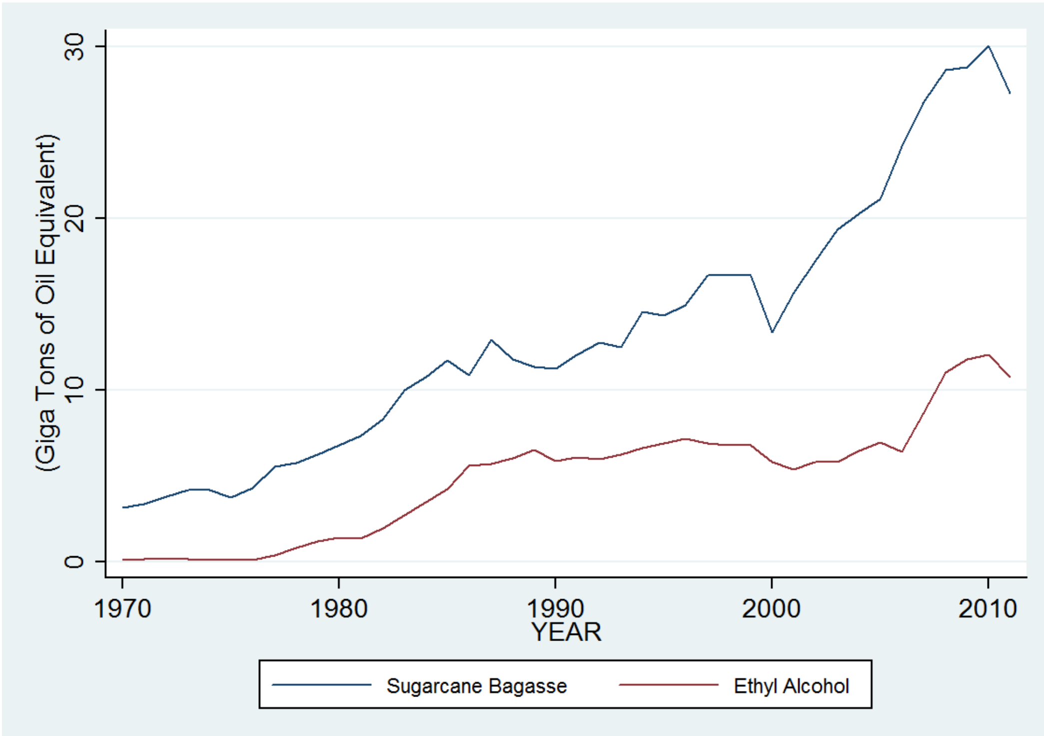

Figure 3.3: Sugarcane Bagasse and Ethanol Demand for Energy

processes heat, mainly in the food and beverages sector.

But as it consists of a cellulose rich biomass, it can be used in the future to produce second generation ethanol, which is defined by the International Energy Agency (2007) [1] as *"Ethanol production from ligno-cellulosic feedstock . . . "* which *". . . includes biomass pre-treatment to release cellulose and hemicellulose, hydrolysis to release 5- and 6-carbon sugars, separation of solid residue and non-hydrolyzed cellulose, and distillation to fuel grade"*.

When sugarcane bagasse is used for bio-electricity production, it is harder to identify with precision which fossil fuels are being displaced. As it is not a base load power generation, it most likely displaces coal, fuel oil, natural gas or even spare hydroelectric power plants reservoirs.

In industrial applications in the food and beverages industry it has displaced mainly firewood and fuel oil. In 2011, sugarcane bagasse accounted for 73.2% of energy use in this industrial sector.

3.3.2 Soybean and Biodiesel

According to the International Energy Agency (2007) [1], *"Biodiesel production is based on trans-esterification of vegetable oils and fats through the addition of methanol (or other alcohols) and a catalyst, giving glycerol as a co-product."*

Figure (3.4) provides a simplified scheme of the soybean-biodiesel industry production chain. Soybean is separated into soybean meal and soybean oil in almost fixed proportions, either through a mechanical or chemical process. Soybean meal can be used as food or feed. Soybean oil can be used as cooking oil or to produce biodiesel.

According to the Food and Agriculture Organization (2008) [17], using Brazilian figures, one hectare can produce 2.4 tons of soybeans and each ton of soybean can be converted into 205 liters of biodiesel. Thus, one hectare can produce up to 491 liters of biodiesel per year.

Biodiesel production yields glycerin as a by-product, used in chemical or pharmaceutical industries.

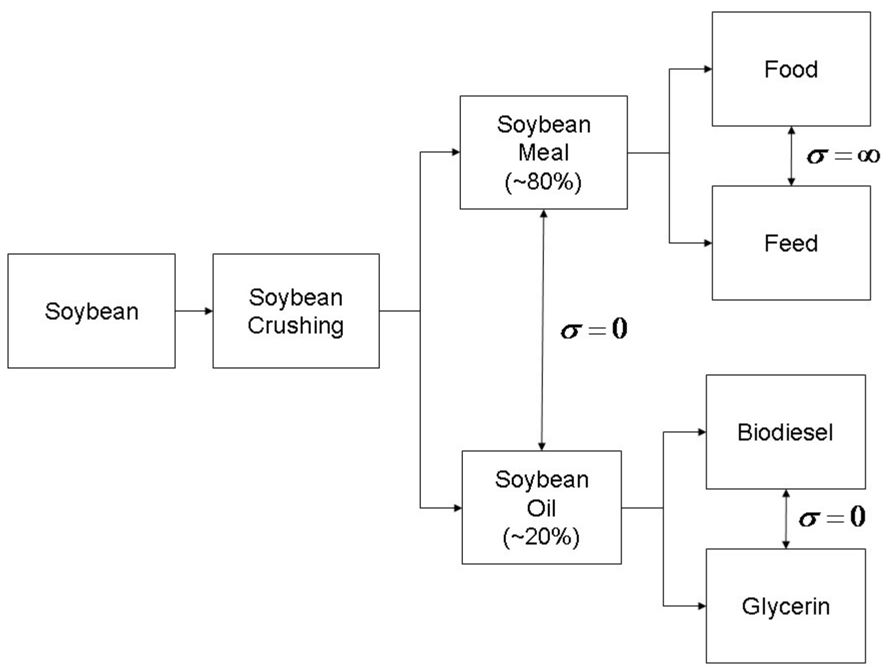

Figure 3.4: Soybean Production Chain

Alternatively, vegetable oils can be blended with refinery diesel and processed at existing crude oil refineries in hydrotreatment units. According to the International

Energy Agency (2007) [1], *"Hydrogenation of oils and fats is a new process that is entering the market"*. This renewable diesel is called hydrotretated vegetable oil, HVO.

Biodiesel is blended into refinery diesel and has the same end uses as diesel itself. This includes mainly road transportation, followed at distance by railway transportation, power generation and some industrial applications, in particular in the mining and pelletization sector.

In figure (3.5) the demand for biodiesel in Brazil from 1970 to 2011 is depicted.

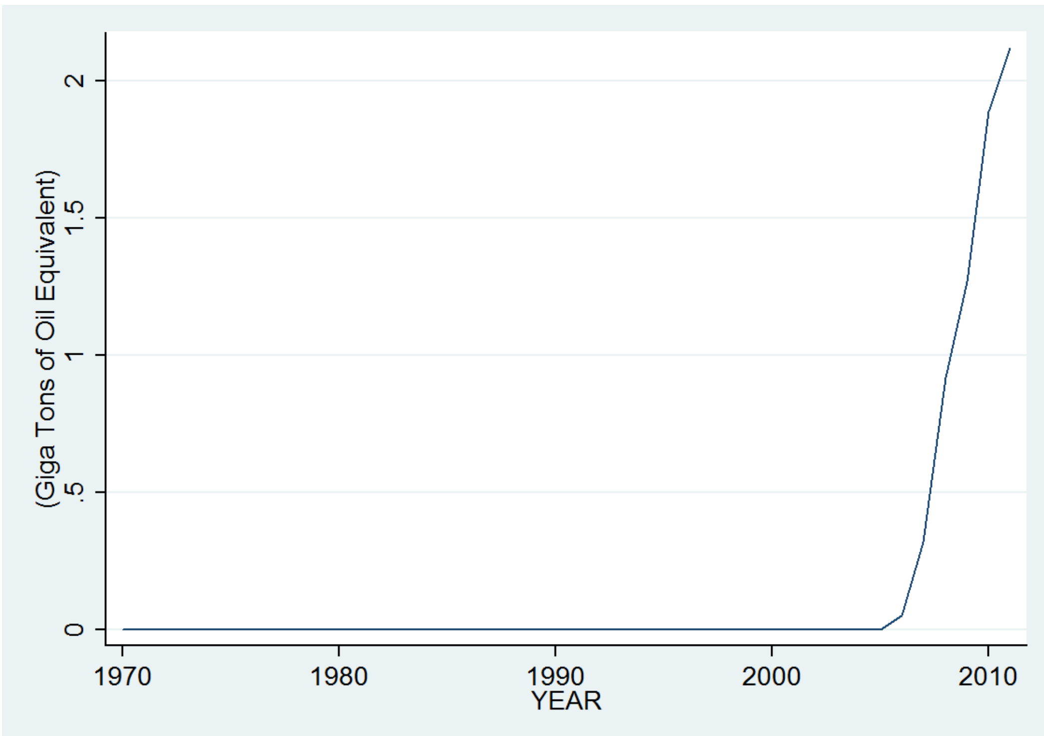

Figure 3.5: Biodiesel Demand for Energy

3.3.3 Forestry, Firewood, Charcoal, Paper and Pulp and Black Liquor

The forestry sector encompass a more complex production chain depicted in figure (3.6). Forestry output can be divided into non-wood and wood products, which include firewood, timber, charcoal and paper and pulp manufacturing.

Fuelwood or firewood is defined by the Food and Agriculture Organization (2008) [16] as *"Wood in the rough (such as chips, sawdust and pellets) used for energy generation"*. In Brazil, firewood is used to produce charcoal, by households for cooking, in

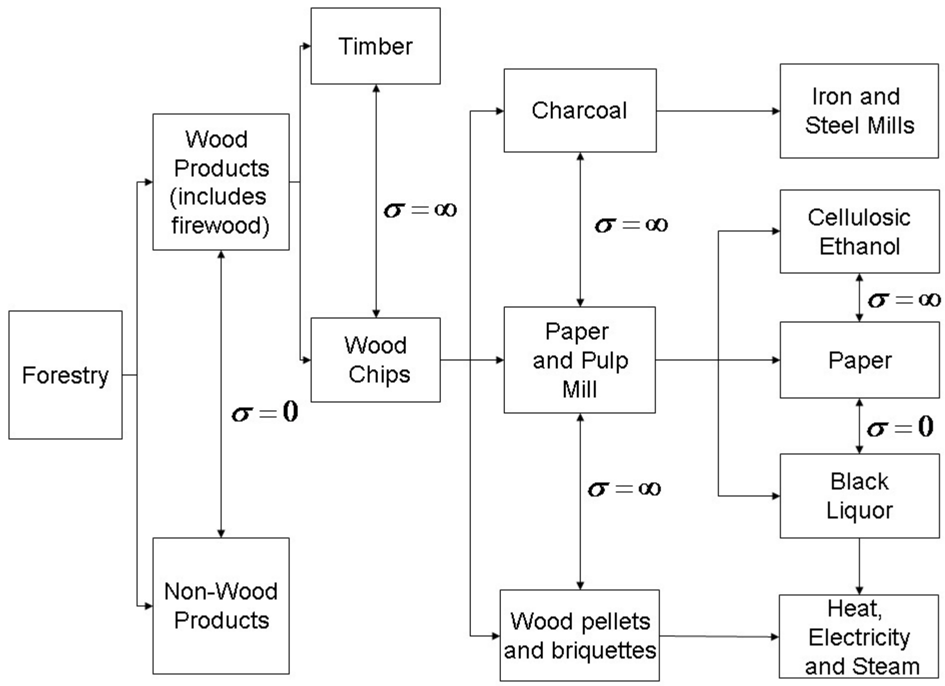

Figure 3.6: Forestry Production Chain

agriculture and in industrial applications, mainly in the ceramics, food and beverages and paper and pulp sectors.

Charcoal is defined by the Food and Agriculture Organization (2004) [15], as a *"solid residue derived from carbonization, distillation, pyrolysis and torrefaction of fuelwood"*. In Brazil, it is mainly used in the industrial sector in the production of pig-iron, steel and iron alloys and also by households.

Black liquor, according to the Food and Agriculture Organization (2004) [15], is an *"alkaline spent liquor obtained from digesters in the production of sulphate or soda pulp during the process of paper production, in which the energy content is mainly originating from the content of lignin removed from the wood in the pulping process"*. In Brazil, black liquor is used to generate electricity and in industrial processes for paper and pulp manufacturing.

The Food and Agriculture Organization (2008) [17] asserts that *"black liquor (a by product of pulp mills) is a major source for bioelectricity generation in countries such as Brazil, Canada, Finland, Sweden, and the United States of America"*.

In figure (3.7) the demand for firewood, charcoal and black liquor in Brazil from

1970 to 2011 is depicted.

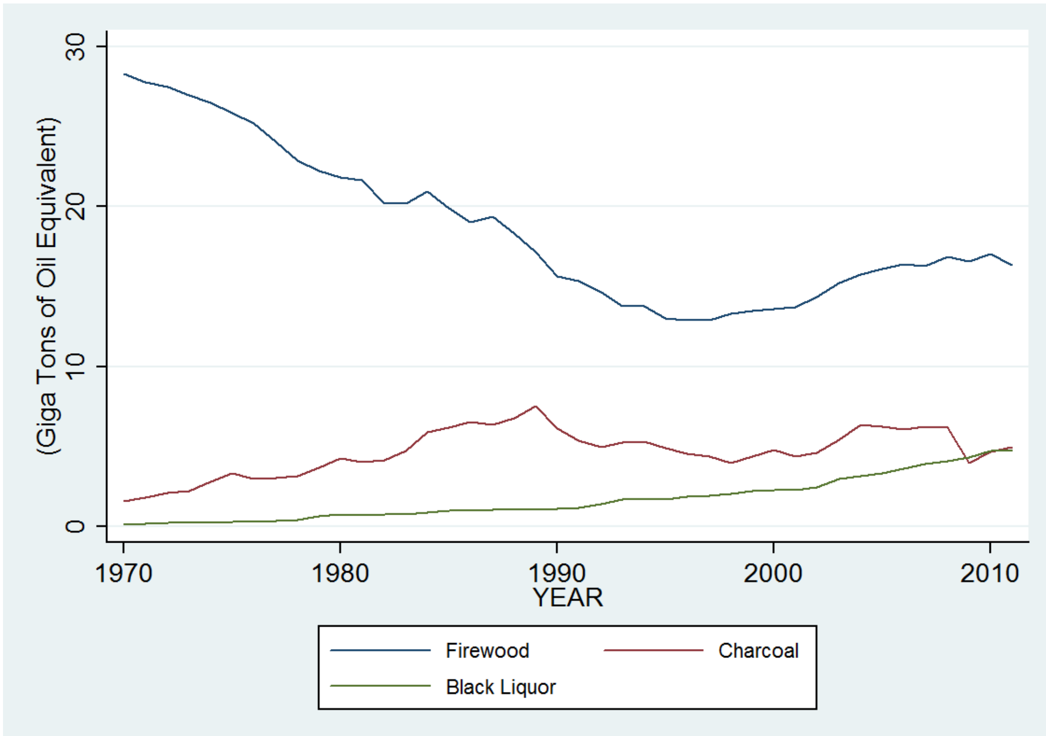

Figure 3.7: Firewood, Charcoal and Black Liquor Demand for Energy

In this period, the correlation between firewood and charcoal is - 0.65, between firewood and black liquor is -0.64 and between charcoal and black liquor is 0.45.

Paper and pulp mills in particular, are expected to be upgraded into future biorefineries, producing 2nd generation biofuels and other biomaterials. Biorefineries are defined by the Food and Agriculture Organization (2008) [16] as *"A new generation of refineries expected to produce not only power and heat, but also transportation fuels and industrial products."*

Finally, biomass to liquids technology can produce a wide range of synthetic hydrocarbon chains such as gasoline, kerosene and diesel. According to the International Energy Agency (2007) [1], this technology can be defined as *"Synthetic biofuel production via biomass gasification and catalytic conversion to liquid using Fischer-Tropsch process (biomass conversion to liquids, BTL) . . . "*.

3.4 Bioenergy Clusters

Definition 1. *Bioenergy Clusters*

A Bioenergy Cluster (C_i) *is an area in space, continuous or not, comprised of a finite set of similar micro-regions using the same technology, which employs labor* (L)*, capital* (K) *and land* (T)*, to produce an output vector* ($\vec{F}$) *from a principal biomass feedstock* (B)*. The vector* ($\vec{F}$) *is composed of* (f_n) *elements, with* $n = 1, \ldots, 5$*:* ***f****ood,* ***f****eed,* ***f****uels,* ***f****ibers,* ***f****orestry products.*

The difference among bioenergy clusters regards the type of technology deployed, the principal biomass feedstock produced (B_i) and the composition of the output vector ($\vec{F_i}$).

3.4.1 The Bioenergy Cluster Technology

The bioenergy cluster technology is represented by a nested multiple output production function. The nested structure is necessary to represent two distinct phases of bioenergy production, the first done outside the mill and the second done inside the mill.

In the lower nest, which represents the agricultural phase done outside the mill, biomass (B) is produced using capital (K_1), labor (L_1) and land (T). Variables have a subscript 1 to denote the first phase when distinction is necessary. The parameter (A_1) is labor augmenting technical change in the agriculture sector. It captures increasing yield per hectare in each crop over time.

Biomass production (B) is a single output from the agriculture production function (Φ) presented in equation (3.1). Biomass (B) can come from any type of agriculture or managed forests.

$$B(t) = \Phi(A_1(t)L_1(t), K_1(t), T(t)) \tag{3.1}$$

In the second stage, done inside the mill, biomass (B) is combined with capital (K_2) and labor (L_2) to produce the scalar output (F) as presented by production function (Λ) in equation (3.2). Mathematically, the scalar output (F) is to be interpreted as the norm of the multiple output vector ($\vec{F}$) in $\mathbb{R}^5$.[4] As several different vectors can have

[4]For example, in a two output food-fuel framework it is possible to measure the norm F in Kcal/year.

the same norm, it is necessary to define other parameters that will indicate an unique vector. This problem can be overcome by dividing the scalar (F) into five shares that sum up to 1. Alternatively, the problem can be overcome by selecting four angles that defines the vector coordinates.

The parameter (A_2) is labor augmenting technical change in the biomass processing sector. It captures increased efficiency in the transformation process of biomass into final products.

$$F(t) = \Lambda(A_2(t)L_2(t), K_2(t), B(t)) \tag{3.2}$$

The capital used in bioenergy clusters can be compared to, or has the same purpose as, carbon abatement capital. Capital employed in agriculture activities (K_1) represents capital responsible for carbon uptake in the production of biomass, while capital employed in bioenergy production (K_2) is necessary to generate outputs that will displace fossil fuels.

A graphical representation of the bioenergy cluster technology is depicted in figure (3.8).

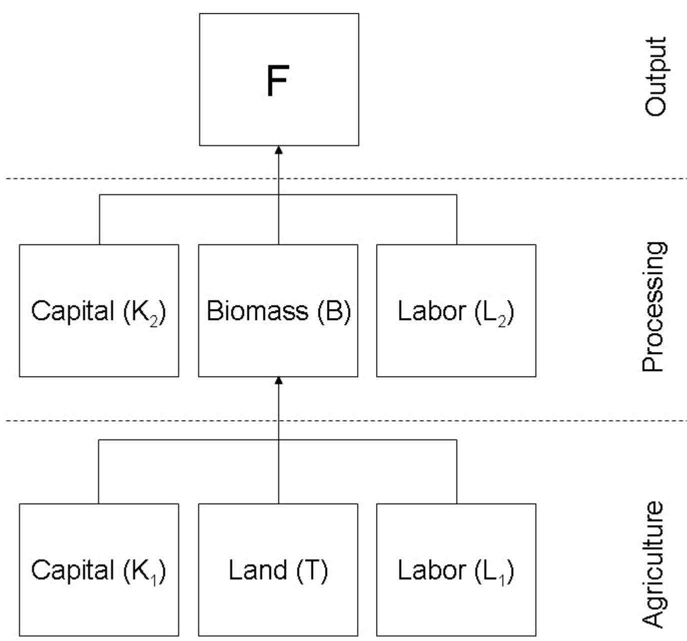

Figure 3.8: Nested Production Structure

The exact parameterization of the nested bioenergy production function depends on the nature of the underlying technology which varies according to the production chain being considered. Thus, each nest may be better characterized by Leontieff, Cobb-Douglas or CES technologies depending on the elasticity of substitution among inputs.

Technology determines also if each possible output will be single or jointly produced, i.e. the flexibility of output vector $\vec{F}$ presented in equation (3.3).

$$\vec{F} = \begin{bmatrix} f_1 & = & \text{food} \\ f_2 & = & \text{feed} \\ f_3 & = & \text{fuels} \\ f_4 & = & \text{fibers} \\ f_5 & = & \text{forestry products} \end{bmatrix} \quad (3.3)$$

The three examples given before shows for instance that some outputs are substitutes, i.e. the production of one excludes the production of the other. Sugar (f_1 = food) and ethanol (f_3 = fuel) production from sugarcane, soybean oil (f_1 = food) and biodiesel (f_3 = fuel), timber (f_5 = forestry products) and firewood (f_3 = fuel) and charcoal (f_3 = fuel) fits into this category.

Other products are technologically limited to be complements. Examples include soybean meal (f_2 = feed) and biodiesel (f_3 = fuel) and paper (f_5 = forestry product) and black liquor (f_3 = fuel).

3.5 Sample and Dataset

The sample consists of 306 micro-regions in the Brazilian Center-West, Southeast and South regions depicted in figure (3.9).

The constructed dataset is comprised of 15 variables referring exclusively to the nested multiple output production function as described in table (3.1). Variables were averaged from 2006 to 2010 to smooth eventual shocks to employment, agricultural and bioenergy production. The time frame is limited to these 5 years because variable

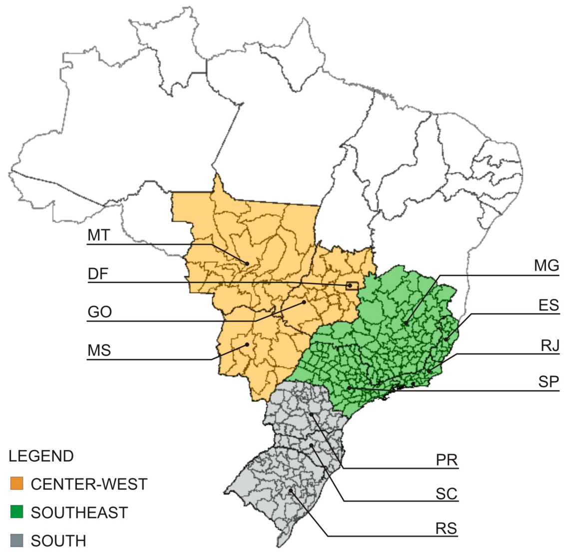

Figure 3.9: Sample Description

number 10, biofuel labor, is available only from 2006 onwards.[5]

For each bioenergy cluster, (1) sugarcane-ethanol, (2) soybean-biodiesel and (3) forestry-firewood-charcoal-paper and pulp-black liquor, 5 variables are employed in order to have the most possible accurate description of the nested multiple output production function. The category column in table (3.1) indicates the position of each variable in the bioenergy technology depicted in figure (3.8). Due to data availability, not all inputs or outputs are included.

[5]This variable refers to biofuel labor, except ethanol labor. As the biodiesel program started in 2005, this variable is available only from 2006 onwards.

Table 3.1: Selected Variables for Cluster Analysis

Number	Name	Category	Var. Name	Cluster
1	Sugarcane Harvested Area	T	Cane Area	Sugarcane
2	Sugarcane Labor	L_1	Cane Labor	Sugarcane
3	Sugarcane Quantity	B	Cane Quantity	Sugarcane
4	Raw Sugar Labor	L_2	Sugar Labor	Sugarcane
5	Ethanol Labor	L_2	Ethanol Labor	Sugarcane
6	Soy Harvested Area	T	Soy Area	Soybean
7	Soy Labor	L_1	Soy Labor	Soybean
8	Soy Quantity	B	Soy Quantity	Soybean
9	Raw Vegetable Oil Labor	L_2	Veg. Oil Labor	Soybean
10	Biofuel Labor	L_2	Biofuel Labor	Soybean
11	Forestry Labor	L_1	Forestry Labor	Forestry
12	Wood Quantity	B	Wood	Forestry
13	Cellulose Labor	L_2	Cellulose Labor	Forestry
14	Firewood Quantity	F	Firewood	Forestry
15	Charcoal Quantity	F	Charcoal	Forestry

To obtain variables in the same dimensional scale for cluster analysis, or unitless, it is necessary to transform them according to equation (3.4).

$$Z_k = \frac{X_k - \overline{X}}{\sigma_k} \tag{3.4}$$

Where Z_k is the transformed variable for micro-region k, X_k is the observed variable for micro-region k, $\overline{X}$ is the sample average and σ_k is the variable standard deviation.

Summary statistics for all variables, in their observed measurement units and after transformation, are presented in Appendix A in tables (3.14) and (3.15), respectively.

3.6 Cluster Analysis Setup

Before performing the k-means cluster analyses, a Pearson correlation matrix for selected variables is presented in tables (3.2), (3.3) and (3.4).

Table 3.2: Pearson Correlation for Sugarcane Cluster Variables

Variable	Cane Area	Cane Labor	Cane Quantity	Sugar Labor	Ethanol Labor
Cane Area	1.000				
Cane Labor	0.813****	1.000	0.810****		
	0.000	0.000	0.000		
Cane Quantity	0.997****	0.810****	1.000		
	0.000	0.000			
Sugar Labor	0.857****	0.653****	0.852****	1.000	
	0.000	0.0000	0.000		
Ethanol Labor	0.545****	0.384****	0.545****	0.350****	1.000
	0.000	0.000	0.000	0.000	
Soy Area	0.030	0.017	0.034	0.001	0.203****
	0.606	0.771	0.552	0.980	0.000
Soy Labor	0.045	0.038	0.045	0.003	0.181****
	0.430	0.507	0.430	0.963	0.001
Soy Quantity	0.027	0.014	0.031	-0.002	0.194****
	0.636	0.807	0.588	0.974	0.001
Veg. Oil Labor	0.091	0.080	0.093	0.036	0.237****
	0.111	0.161	0.106	0.527	0.000
Biofuel Labor	-0.024	-0.033	-0.019	-0.008	0.065
	0.680	0.565	0.743	0.890	0.256
Forestry Labor	-0.077	-0.057	-0.080	-0.058	-0.082
	0.182	0.317	0.163	0.312	0.155
Wood Quantity	-0.029	0.008	-0.033	-0.031	-0.05
	0.610	0.883	0.568	0.586	0.380
Cellulose Labor	0.043	0.042	0.037	0.093	0.033
	0.453	0.468	0.520	0.104	0.564
Firewood Quantity	-0.108*	-0.092	-0.110*	-0.071	-0.096*
	0.060	0.110	0.054	0.218	0.093
Charcoal Quantity	-0.061	-0.063	-0.065	-0.051	-0.035
	0.284	0.275	0.261	0.372	0.540

Note: p-values are presented below correlation coeffcients.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

Table 3.3: Pearson Correlation with Soybean Cluster Variables

Variable	Soy Area	Soy Labor	Soy Quantity	Veg. Oil Labor	Biofuel Labor
Soy Area	1.000				
Soy Labor	0.933****	1.000			
	0.000				
Soy Quantity	0.995****	0.944****	1.000		
	0.000	0.000			
Veg. Oil Labor	0.240****	0.267****	0.238****	1.000	
	0.000	0.000	0.000		
Biofuel Labor	0.489****	0.468****	0.502****	0.227****	1.000
	0.000	0.000	0.000	0.000	
Forestry Labor	-0.066	-0.042	-0.058	0.024	-0.011
	0.251	0.467	0.314	0.676	0.846
Wood Quantity	-0.062	-0.060	-0.056	-0.035	-0.081
	0.278	0.292	0.328	0.537	0.159
Cellulose Labor	-0.047	-0.038	-0.043	0.179***	0.015
	0.415	0.503	0.454	0.002	0.800
Firewood Quantity	-0.020	-0.044	-0.022	0.098*	-0.024
	0.721	0.447	0.701	0.088	0.673
Charcoal Quantity	-0.057	-0.029	-0.053	-0.032	0.056
	0.323	0.618	0.356	0.572	0.325

Note: p-values are presented below correlation coeffcients.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

Table 3.4: Pearson Correlation with Forestry Cluster Variables

Variable	Forestry Labor	Wood	Cellulose Labor	Firewood	Charcoal
Forestry Labor	1.000				
Wood Quantity	0.534****	1.000			
	0.000				
Cellulose Labor	0.275****	0.319****	1.000		
	0.000	0.000			
Firewood Quantity	0.316****	0.412****	0.027	1.000	
	0.000	0.000	0.636		
Charcoal Quantity	0.563****	-0.016	-0.018	0.030	1.000
	0.000	0.787	0.758	0.598	

Note: p-values are presented below correlation coeffcients.

**** Significant at 0.1% level.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

A clear pattern emerges from the analysis of the Pearson correlation matrix. Variables that belong to the same production function or bioenergy cluster exhibit positive and statistically significant correlation coefficients. Moreover, there are very few statistically significant correlations among variables that belong to different bioenergy clusters.

All variables belonging to the sugarcane-ethanol cluster; (Cane Area), (Cane Labor), (Cane Quantity), (Sugar Labor) and (Ethanol Labor) exhibit positive and statistically significant at the 0.1% level correlation coefficients among them.

For the soybean-biodiesel cluster, the same result holds. All variables belonging to the cluster; (Soy Area), (Soy Labor), (Soy Quantity), (Veg. Oil Labor) and (Biofuel Labor) exhibit positive and statistically significant at the 0.1% level correlation coefficients among them.

For the forestry-firewood-charcoal-paper and pulp-black liquor cluster, the variable (Forestry Labor) is positively correlated with all other cluster variables; (Wood Quantity), (Cellulose Labor), (Firewood Quantity) and (Charcoal Quantity) and the coefficients are statistically significant at the 0.1% level. The variable (Wood) also exhibit positive and statistically significant at the 0.1% level correlation coefficients with variables (Cellulose Labor) and (Firewood).

These results suggest that, at the micro-region level, the bioenergy cluster technology consists of an integrated production chain. In all three clusters, it is possible to observe positive and statistically significant correlation at the 0.1% level among all available variables that belong to the agriculture production function or the lower nest, and variables that belong to the upper nest or the output vector of the bioenergy production function.

The few cases of statistically significant correlation among variables from different clusters also suggests that micro-regions tend to specialize in only one bioenergy production technology, in a principal biomass (B_i) and in an output vector ($\vec{F_i}$).

Based upon the bioenergy production function, and the results of the Pearson correlation matrix, the theoretical initial cluster centroids matrix is constructed. The initial cluster centroids matrix provides a first approximation of a representative element, a micro-region, of each bioenergy cluster according to definition (1).

As each bioenergy cluster is comprised of a set of similar micro-regions expected to have an unique and integrated bioenergy production chain, the representative element of each cluster should exhibit high values for all variables belonging to that particular cluster and low values for all remainder variables.

Accordingly, the centroid coordinates of each bioenergy cluster are set at the sample median for all variables belonging to the cluster and at the first quantile for variables that do not belong to the cluster.

Besides the three bioenergy clusters analyzed here, a fourth cluster unsuited for bioenergy production is expected to exist. This no bioenergy cluster should exhibit low values for all variables, and its initial centroid coordinates are set at the sample first quantile for all variables. The initial theoretical cluster centroid matrix is presented in table (3.5).

Table 3.5: Initial Theoretical Cluster Centroids Matrix

Variable	Sugarcane	Soybean	Forestry	No Bioenergy
Cane Area	M	Q1	Q1	Q1
Cane Labor	M	Q1	Q1	Q1
Cane Quantity	M	Q1	Q1	Q1
Sugar Labor	M	Q1	Q1	Q1
Ethanol Labor	M	Q1	Q1	Q1
Soy Area	Q1	M	Q1	Q1
Soy Labor	Q1	M	Q1	Q1
Soy Quantity	Q1	M	Q1	Q1
Veg. Oil Labor	Q1	M	Q1	Q1
Biofuel Labor	Q1	M	Q1	Q1
Forestry Labor	Q1	Q1	M	Q1
Wood	Q1	Q1	M	Q1
Cellulose Labor	Q1	Q1	M	Q1
Firewood	Q1	Q1	M	Q1
Charcoal	Q1	Q1	M	Q1

M: Median of the sample.

Q1: First Quantile of the sample.

Plugging in sample values into table (3.5) yields table (3.6), with the initial centroid coordinates for each cluster.

Table 3.6: Initial Cluster Centroids Matrix

Variable	Sugarcane	Soybean	Forestry	No Bioenergy
Cane Area	-0.417	-0.440	-0.440	-0.440
Cane Labor	-0.353	-0.355	-0.355	-0.355
Cane Quantity	-0.416	-0.428	-0.428	-0.428
Sugar Labor	-0.312	-0.312	-0.312	-0.312
Ethanol Labor	-0.412	-0.412	-0.412	-0.412
Soy Area	-0.382	-0.363	-0.382	-0.382
Soy Labor	-0.337	-0.326	-0.337	-0.337
Soy Quantity	-0.360	-0.343	-0.360	-0.360
Veg. Oil Labor	-0.311	-0.311	-0.311	-0.311
Biofuel Labor	-0.284	-0.284	-0.284	-0.284
Forestry Labor	-0.519	-0.519	-0.429	-0.519
Wood Quantity	-0.383	-0.383	-0.367	-0.383
Cellulose Labor	-0.236	-0.236	-0.236	-0.236
Firewood Quantity	-0.482	-0.482	-0.392	-0.482
Charcoal Quantity	-0.221	-0.221	-0.220	-0.221

In the next section, four k-means cluster analyses are performed based on the initial theoretical centroid matrix. Results of cluster membership are presented in a geographic information system to further elicit spatial features of each bioenergy cluster.

The first two k-means cluster analyses explore agricultural clusters. The analyses employ exclusively variables belonging to the agriculture production function presented in equation (3.1). The first analysis includes only three clusters; (1) sugarcane, (2) soybean and (3) no bioenergy. The second analysis includes four cluster by adding the forestry cluster to the previous three.

Next, bioenergy clusters are examined employing variables belonging to the whole bioenergy production function described in equations (3.1), (3.2) and output vector ($\vec{F}$). Firstly, three clusters are analyzed; (1) sugarcane-ethanol, (2) soybean-biodiesel and (3) no bioenergy. The second analysis includes the forestry-firewood-charcoal-paper and pulp-black liquor cluster, totalizing four clusters.

The motivation to analyze initially three clusters and add subsequently the forestry cluster is to have results comparable with previous chapters. The same reasoning applies to analyzing first agricultural clusters, followed by bioenergy clusters.

3.7 Agriculture Clusters

The first analysis comprises six variables (Cane Area), (Cane Labor), (Cane Quantity), (Soy Area), (Soy Labor) and (Soy Quantity) to form three agriculture clusters; (1) sugarcane, (2) soybean and (3) no bioenergy. The second analysis comprises eight variables, the previous six plus (Forestry Labor) and (Wood Quantity) to form four agriculture clusters; (1) sugarcane, (2) soybean, (3) forestry and (4) no bioenergy. Cluster membership for both analyses are presented in table (3.7).

Table 3.7: Agriculture Cluster Membership

Variable	Sugarcane	Soybean	Forestry	No Bioenergy	Total
Three Clusters	33	4	n/a	269	306
Four Clusters	31	4	28	243	306

Results are robust to changes in the initial centroid matrix specification. In particular, replacing the first quantile (Q1) by the sample minimum do not alter any result, i.e. final clusters centroids matrix, cluster membership, distances among final cluster centroids and distance of each observation to the assigned cluster centroid.

Replacing the median (M) with the third quantile (Q3) does not alter the results either. Thus for four different initial centroid matrix, namely minimum - first quantile (Q1), first quantile (Q1) - median (M), first quantile (Q1) - third quantile (Q3), minimum - third quantile (Q3), the same results hold.

If the sample minimum is utilized with the first quantile then, it is not possible to form three or four cluster because the initial soybean cluster centroid becomes identical to the no-bioenergy cluster centroid.

Analyzing the final three cluster centroid matrix presented in table (3.8), it is possible to verify that both the sugarcane and no bioenergy clusters are well characterized. The sugarcane cluster final centroid has positive values for all variables belonging to the cluster and negative values for remainder variables. The no bioenergy cluster final centroid has negative values for all variables. The soybean cluster has positive values for all variables. In particular, for variables belonging to the cluster the value is too high, which explains why only 4 micro-regions were assigned to this cluster. It captures only micro-regions with very high soybean production.

Table 3.8: Three Agriculture Cluster Final Centroids Matrix

	Clusters		
Variable	Sugarcane	Soybean	No Bioenergy
Cane Area	2.354	0.741	-0.300
Cane Labor	1.991	0.272	-0.248
Cane Quantity	2.317	0.803	-0.296
Soy Area	-0.187	7.035	-0.082
Soy Labor	-0.139	6.659	-0.082
Soy Quantity	-0.182	7.106	-0.083

Figure (3.10) depicts the three agriculture clusters in space. The sugarcane cluster, comprised of 33 micro-regions, displays a clear agglomeration pattern. Thirty micro-regions belonging to the sugarcane cluster neighbor at least one micro-region that also belongs to the cluster.

Next, the forestry cluster is included into the analysis. The final cluster centroid matrix for the four agricultural clusters is presented in table (3.9).

Out of the 28 micro-regions assigned to the forestry cluster, 27 were previously classified in the no bioenergy cluster and 1 in the sugarcane cluster. Another former member of the sugarcane cluster migrates to the no bioenergy cluster.

The forestry cluster is also well characterized. Its final centroid has positive values for all variables belonging to the cluster and negative values for remainder variables. Results for the other clusters are qualitatively unchanged. The sugarcane and no

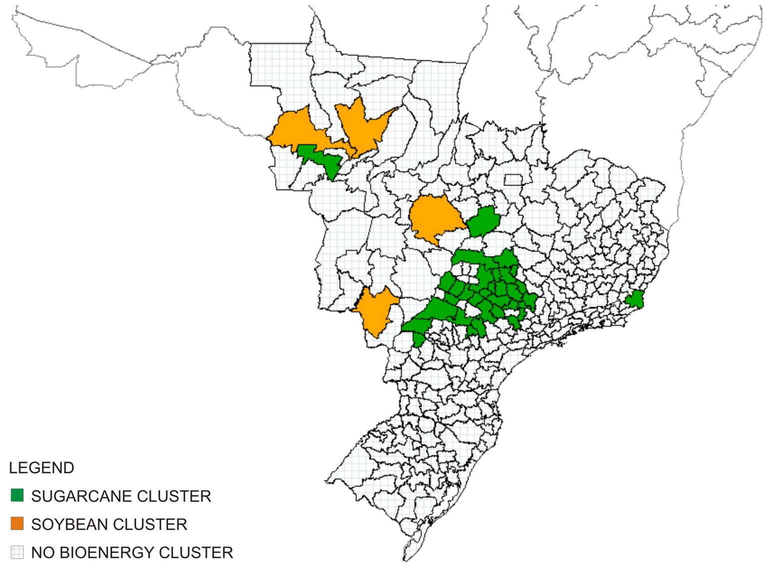

Figure 3.10: Three Agricultural Clusters

bioenergy clusters remain well characterized and the soybean cluster captures the same micro-regions as before.

Table 3.9: Four Agriculture Cluster Final Centroids Matrix

	Clusters			
Variable	Sugarcane	Soybean	Forestry	No Bioenergy
Cane Area	2.422	0.741	-0.215	-0.296
Cane Labor	2.003	0.272	-0.110	-0.247
Cane Quantity	2.384	0.803	-0.225	-0.291
Soy Area	-0.180	7.035	-0.213	-0.068
Soy Labor	-0.131	6.659	-0.184	-0.072
Soy Quantity	-0.175	7.106	-0.193	-0.072
Forestry Labor	-0.310	-0.213	2.267	-0.218
Wood Quantity	-0.239	-0.379	2.425	-0.243

Figure (3.11) depicts the four agriculture clusters in space. The sugarcane cluster displays almost the same agglomeration pattern as before, but with 27 neighboring micro-regions. The forestry cluster exhibit two very close agglomerations. One with 10 neighboring micro-regions, that also borders the sugarcane agglomeration, and another with 6 neighboring micro-regions.

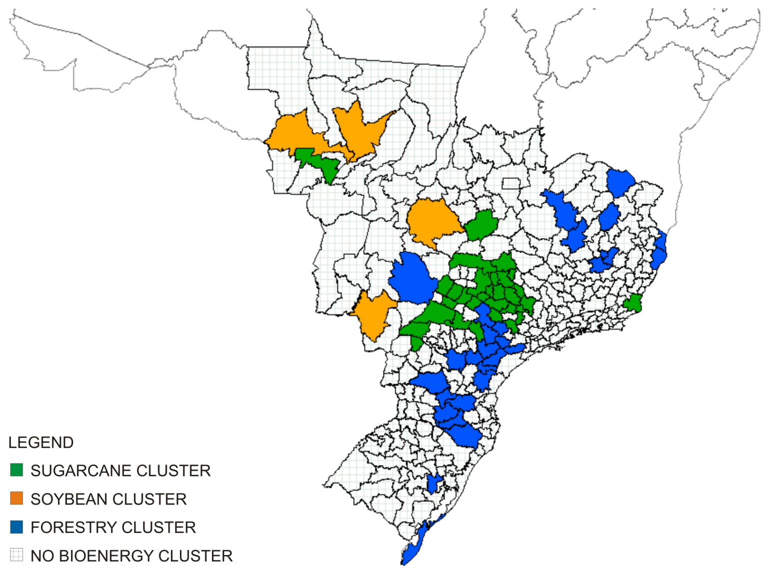

Figure 3.11: Four Agricultural Clusters

3.8 Bioenergy Clusters

The first analysis comprises ten variables (Cane Area), (Cane Labor), (Cane Quantity), (Sugar Labor), (Ethanol Labor), (Soy Area), (Soy Labor), (Soy Quantity), (Veg. Oil Labor) and (Biofuel Labor) to form three bioenergy clusters; (1) sugarcane-ethanol, (2) soybean-biodiesel and (3) no bioenergy.

The second analysis comprises fifteen variables, the previous ten plus (Forestry Labor), (Wood Quantity), (Cellulose Labor), (Firewood Quantity) and (Charcoal Quantity) to form four bioenergy clusters; (1) sugarcane-ethanol, (2) soybean-biodiesel, (3) forestry-firewood-charcoal-paper and pulp-black liquor and (4) no bioenergy.

Cluster membership for both analyses are presented in table (3.10).

Table 3.10: Bioenergy Cluster Membership

Variable	Sugarcane	Soybean	Forestry	No Bioenergy	Total
Three Clusters	25	18	n/a	263	306
Four Clusters	25	7	40	234	306

When only three clusters are considered results are robust to changes in the initial centroid matrix as described in the previous section.

In the four clusters case, there are minor changes in results when the initial centroid matrix is modified from the default first quantile (Q1) - median (M) setting. In particular, when the median (M) is replaced by the third quantile (Q3), 4 micro-regions change cluster membership. All of them migrate from the forestry cluster to the no bioenergy cluster.

It is possible to pinpoint these four micro-regions and according to the values of their variables, there is no evidence against their membership to the forestry cluster. They exhibit at least two, out of five, positive values for variables belonging to the forestry cluster, i.e. above the sample average. The micro-regions are Mogi das Cruzes - SP, Blumenau - SC, Vacaria - RS and Litoral Lagunar - RS.

Moreover, replacing the median (M) with the third quantile (Q3) in the initial centroid matrix does not alter the final centroids of the sugarcane-ethanol and soybean-biodiesel clusters. While, the final cluster centroids of the forestry cluster and the no bioenergy undergo small changes.

Thus, the initial clusters centroid matrix selected using the first quantile (Q1) and the median (M) is quite robust to variations even in the four bioenergy clusters case.

Analyzing the final three bioenergy cluster centroid matrix presented in table (3.11), it is possible to verify that all clusters are well characterized.

Table 3.11: Three Bioenergy Cluster Final Centroids Matrix

	Clusters		
Variable	Sugarcane	Soybean	No Bioenergy
Cane Area	2.680	0.069	-0.259
Cane Labor	2.339	-0.091	-0.216
Cane Quantity	2.663	0.077	-0.258
Sugar Labor	2.414	-0.026	-0.228
Ethanol Labor	1.650	0.706	-0.205
Soy Area	-0.200	2.566	-0.157
Soy Labor	-0.148	2.717	-0.172
Soy Quantity	-0.193	2.546	-0.156
Veg. Oil Labor	0.098	2.304	-0.167
Biofuel Labor	-0.167	1.729	-0.102

The sugarcane-ethanol cluster final centroid has positive values for all variables belonging to the cluster and negative values for almost all remainder variables, except (Veg. Oil Labor).

The soybean-biodiesel cluster becomes better characterized. In the final centroid coordinates, variables belonging to the cluster are not as high as in the agriculture cluster analysis performed in the previous section, and the algorithm assigns eighteen micro-regions to it. Additionally, two variables belonging to the sugarcane-ethanol cluster exhibit negative values. The no bioenergy cluster final centroid has negative values for all variables.

Figure (3.12) depicts the three bioenergy clusters in space. The sugarcane-ethanol cluster, comprised of 25 micro-regions, displays a clear agglomeration pattern. Twenty three micro-regions belonging to the sugarcane-ethanol cluster neighbor at least one micro-region that also belongs to the cluster. The soybean-biodiesel cluster also exhibits an incipient agglomeration pattern. Out of its 18 members, 7 neighbors another cluster member.

Next, the forestry-firewood-charcoal-paper and pulp-black liquor cluster is included

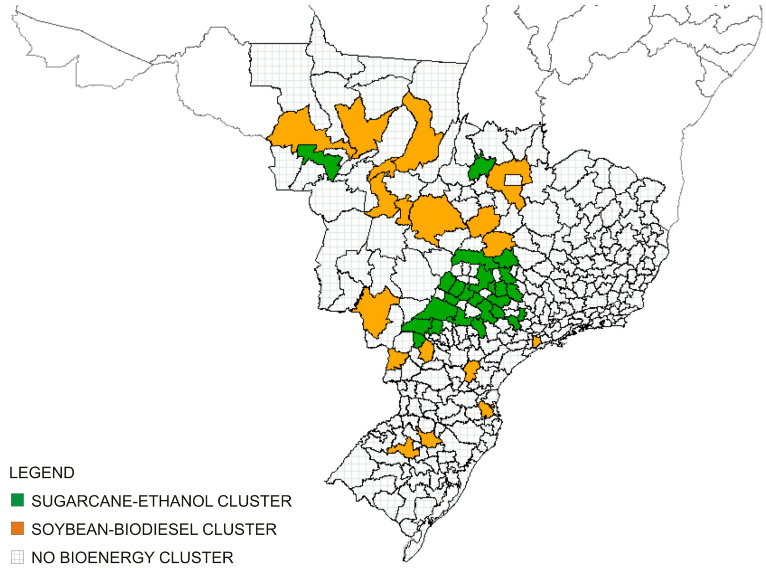

Figure 3.12: Three Bioenergy Clusters

into the analysis. The final cluster centroid matrix for the four bioenergy clusters is presented in table (3.12).

Table 3.12: Four Bioenergy Cluster Final Centroids Matrix

	Clusters			
Variable	Sugarcane	Soybean	Forestry	No Bioenergy
Cane Area	2.656	0.321	-0.265	-0.248
Cane Labor	2.281	0.056	-0.196	-0.212
Cane Quantity	2.643	0.338	-0.273	-0.246
Sugar Labor	2.424	0.188	-0.249	-0.222
Ethanol Labor	1.695	1.270	-0.184	-0.188
Soy Area	-0.122	4.745	-0.190	-0.096
Soy Labor	-0.078	5.129	-0.152	-0.119
Soy Quantity	-0.122	4.787	-0.177	-0.100
Veg. Oil Labor	0.137	1.770	0.561	-0.163
Biofuel Labor	-0.167	3.792	0.011	-0.097
Forestry Labor	-0.273	-0.288	1.851	-0.279
Wood Quantity	-0.261	-0.372	1.608	-0.236
Cellulose Labor	-0.048	-0.236	1.180	-0.189
Firewood Quantity	-0.292	-0.118	1.319	-0.191
Charcoal Quantity	-0.214	-0.220	1.066	-0.153

Out of the 40 micro-regions assigned to the forestry-firewood-charcoal-paper and pulp-black liquor cluster, 36 were previously classified in the no bioenergy cluster, 1 in the sugarcane-ethanol cluster and 3 in the soybean-biodiesel cluster.

Although the sugarcane-ethanol cluster remains with the same number of members, its composition changes by one member. It remains well characterized. The final sugarcane-ethanol cluster centroid coordinates has positive values for all variables that belong to the cluster, and negative values for almost all remainder variables, except one as before (Veg. Oil Labor).

The soybean-biodiesel cluster looses 11 members; 1 to the sugarcane-ethanol cluster, 3 to the forestry-firewood-charcoal-paper and pulp-black liquor cluster and 7 to the no bioenergy cluster. The reason is that the final cluster centroid coordinates exhibits too high values for variables inside the cluster. Additionally, all variables that should belong to the sugarcane-ethanol cluster exhibit positive values. Thus, also in this case, the soybean-biodiesel cluster is not accurately characterized.

The forestry cluster is also well characterized. Its final centroid has positive values for all variables belonging to the cluster and negative values for almost all remainder variables, except two (Veg. Oil Labor) and (Biofuel Labor). The no bioenergy cluster is also well characterized as all its centroid coordinates are negative.

In table (3.13) the distance among the four final bioenergy clusters centroids is presented. The soybean-biodiesel final cluster centroid is the farthest from all others, providing further evidence that its coordinates overshoot.

Table 3.13: Distance Among Final Clusters Centroids

	Sugarcane	Soybean	Forestry	No Bioenergy
Sugarcane	0.000	10.688	6.916	5.803
Soybean	10.688	0.000	10.486	9.839
Forestry	6.916	10.486	0.000	3.759
No Bioenergy	5.803	9.839	3.759	0.000

Figure (3.13) depicts the four bioenergy clusters in space. The sugarcane-ethanol cluster, comprised of 25 micro-regions, displays again a clear agglomeration pattern. Twenty-two micro-regions belonging to the sugarcane-ethanol cluster neighbor at least one micro-region that also belongs to the cluster. The soybean-biodiesel cluster does not exhibit a clear agglomeration pattern, although there is a pair of two neighboring micro-regions that belong to the cluster. The forestry-firewood-charcoal-paper and pulp-black liquor cluster has four agglomerations. One is comprised of 8 micro-regions, other two are are comprised of 7 micro-regions, with one also bordering the sugarcane-ethanol cluster. There is another agglomeration to the northeast of the sample, comprised of 5 micro-regions.

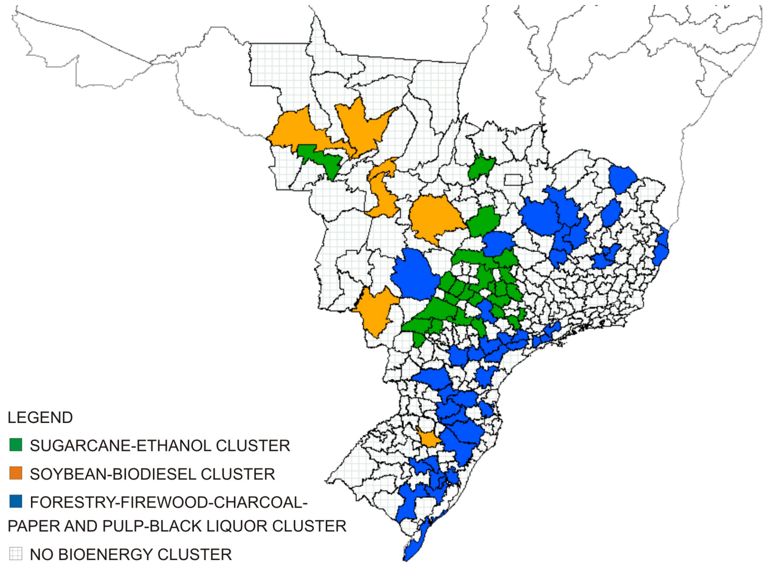

Figure 3.13: Four Bioenergy Clusters

3.9 Conclusions

This paper investigates bioenergy clusters from a statistical standpoint. Three types of biomass based energy clusters are investigated; (1) sugarcane-ethanol, (2) soybean-biodiesel and (3) forestry-firewood-charcoal-paper and pulp-black liquor. These are the main sources of biomass based energy currently produced and consumed in Brazil.

A microeconomic framework employing a nested multiple output production function is proposed to model bioenergy production. The lower nest represents an agriculture production function, where capital, labor and land are combined to produce biomass. In the upper nest, biomass is combined with capital and labor to produce multiple products, described by an output vector ($\vec{F}$). This vector contains five elements; food, feed, fuels, fibers and forestry products.

Based on this technology, described by a nested multiple output production function, the definition of bioenergy clusters is put forward.

The Pearson correlation matrix reveals statistically significant positive correlation among variables that represent the same biomass derived energy productive structure. Moreover, with few exceptions, it does not reveal statistically significant correlations among variables that belong to different bioenergy production chains. Thus, the Pearson correlation matrix provides first evidence that, at the micro-region level, bioenergy clusters are specialized and integrated production chains.

In order to test the bioenergy cluster definition, four k-means cluster analyses are performed. The first two cluster analyses, with 3 and 4-means, aims at classifying micro-regions into agriculture clusters, based only on variables referring to the first nest of the bioenergy production technology, i.e. the agriculture production function.

The last two cluster analyses, with 3 and 4-means, aims at classifying micro-regions into bioenergy clusters employing variables referring to the entire bioenergy production function.

A representative element of each of these agriculture and bioenergy clusters is defined by setting an initial centroid matrix in line with the definition of bioenergy clusters and results from the Pearson correlation matrix. Results are quite robust to different specifications of the initial centroid matrix.

The first analysis considers three agricultural clusters; (1) sugarcane, (2) soybean and (3) no bioenergy. The sugarcane and the no bioenergy clusters are well characterized based on their final centroids coordinates. The soybean cluster is not so accurately depicted because it captures only four micro-regions with very high soybean output. Plotting results of cluster membership in a Geographic Information System reveals that the sugarcane cluster exhibits an agglomeration pattern in space.

The second analysis considers four agricultural clusters; (1) sugarcane, (2) soybean, (3) forestry and (4) no bioenergy. The forestry cluster, comprised of 28 micro-regions, captures mainly former members of the no bioenergy cluster and is well characterized based on its final centroid coordinates. Results for the other three agricultural clusters do not change qualitatively. In particular, the soybean cluster remains not so accurately depicted, consisting of the same 4 micro-regions as before.

The sugarcane cluster exhibits almost the same agglomeration pattern as before, while the forestry cluster displays two separate but close agglomerations. One of the forestry agglomerations, comprised of 10 micro-regions, borders the sugarcane agglomeration.

The third analysis considers three bioenergy clusters; (1) sugarcane-ethanol, (2) soybean-biodiesel and (3) no bioenergy. All clusters are well defined based on their final centroids coordinates.

Twenty-five micro-regions are assigned to the sugarcane-ethanol cluster and eighteen to the soybean-biodiesel cluster. In fact, more micro-regions are assigned to the soybean-biodiesel cluster than in the two previous soybean clusters. This occurs mainly because its final centroid coordinates have lower values for comparable variables belonging to the cluster.

Both the sugarcane-ethanol and the soybean-biodiesel clusters exhibit an agglomeration pattern in space. Moreover, the two agglomerations have bordering micro-regions.

The fourth analysis considers four bioenergy clusters; (1) sugarcane-ethanol, (2) soybean-biodiesel, (3) forestry-firewood-charcoal-paper and pulp-black liquor and (4) no bioenergy. The sugarcane-ethanol, forestry-firewood-charcoal-paper and pulp-black liquor and no bioenergy clusters are well characterized based on their final centroids coordinates. The soybean-biodiesel cluster is not completely accurately characterized. It looses eleven members when compared to the previous analysis with three bioenergy clusters. Twenty-five micro-regions are assigned to the sugarcane-ethanol cluster, 7 to the soybean-biodiesel cluster, 40 to the forestry-firewood-charcoal-paper and pulp-black liquor cluster and 234 to the no bioenergy cluster.

The sugarcane-ethanol cluster agglomerates in space with 22 interconnected micro-regions. The soybean-biodiesel cluster is more spatially dispersed, but there is a pair of two bordering micro-regions belonging to the cluster.

The forestry-firewood-charcoal-paper and pulp-black liquor has four agglomerations. It shows that more than one agglomeration of bioenergy producing micro-regions, employing the same feedstock, can co-exist separated in space.

The main conclusion is that it is possible to aggregate micro-regions based solely on the bioenergy cluster definition presented here, which is in turn based on a microeconomic bioenergy production technology, or production function. These microeconomic structures are specialized and integrated bioenergy production chains.

Results found here can help policy makers to tailor and target economic, agricultural, energy, environment or climate policies to these identified bioenergy clusters.

3.10 Appendix A

Table 3.14: Summary Statistics in Observed Measurement Units

Variable	Units	Mean	Std. Dev.	Min.	Max.	N
Cane Harvested Area	Hectares	21,671.612	48,545.177	0	320,731	306
Cane Labor	Workers	453.754	1,276.553	0	13,385.4	306
Cane Quantity	Tons/year	1,783,109.675	4,130,265.361	0	28,578,518	306
Ethanol Labor	Workers	250.863	608.613	0	4,772.2	306
Raw Sugar Labor	Workers	501.143	1,605.168	0	17,539.801	306
Soy Harvested Area	Hectares	64,306.423	168,524.585	0	1,822,965.6	306
Soy Labor	Workers	233.101	691.759	0	6,766	306
Soy Quantity	Tons/year	174,765.239	485,333.368	0	5,567,910.5	306
Vegetable Oil Labor	Workers	48.604	156.488	0	1,409.2	306
Biofuel Labor	Workers	3.177	11.171	0	106.6	306
Forestry Labor	Workers	157.935	285.097	0	1,911.2	306
Wood Quantity	m^3/year	290,816.627	757,127.666	0	5,855,828.2	306
Cellulose Labor	Workers	33.124	140.114	0	1,102.2	306
Firewood Quantity	m^3/year	131,469.916	264,038.895	0	2,209,766.25	306
Charcoal Quantity	Tons/year	9,815.630	44,423.299	0	394,489.594	306

Table 3.15: Summary Statistics after Transformation

Variable	Mean	Std. Dev.	Min.	Q1	Median	Q3	Max.	N
Cane Harvested Area	0.000	1.000	-0.446	-0.440	-0.417	-0.110	6.160	306
Cane Labor	0.000	1.000	-0.355	-0.355	-0.353	-0.254	10.130	306
Cane Quantity	0.000	1.000	-0.432	-0.428	-0.416	-0.121	6.488	306
Ethanol Labor	0.000	1.000	-0.412	-0.412	-0.412	-0.259	7.429	306
Raw Sugar Labor	0.000	1.000	-0.312	-0.312	-0.312	-0.310	10.615	306
Soy Harvested Area	0.000	1.000	-0.382	-0.382	-0.363	-0.027	10.436	306
Soy Labor	0.000	1.000	-0.337	-0.337	-0.326	-0.134	9.444	306
Soy Quantity	0.000	1.000	-0.360	-0.360	-0.343	-0.034	11.112	306
Vegetable Oil Labor	0.000	1.000	-0.311	-0.311	-0.311	-0.275	8.695	306
Biofuel Labor	0.000	1.000	-0.284	-0.284	-0.284	-0.284	9.258	306
Forestry Labor	0.000	1.000	-0.554	-0.519	-0.429	0.018	6.150	306
Wood Quantity	0.000	1.000	-0.384	-0.383	-0.367	-0.175	7.350	306
Cellulose Labor	0.000	1.000	-0.236	-0.236	-0.236	-0.236	7.630	306
Firewood Quantity	0.000	1.000	-0.498	-0.482	-0.392	0.019	7.871	306
Charcoal Quantity	0.000	1.000	-0.221	-0.221	-0.220	-0.194	8.660	306

3.11 Appendix B

Coordinate Reference System

All shapefiles were projected using SAD69/Brazil Polyconic projection which uses as references, the 0° parallel (Equator line) and the 54° West meridian.

Granularity

As the political division of 2007, Brazil had 5 levels of granularity. Country (1), Regions (5) equivalent to NUTS 1, States (27) equivalent to NUTS 2, Meso-Regions (137) , Micro-Regions (558) equivalent to NUTS 3, Municipalities (5564). Shapefiles were obtained from www.ipea.gov.br retrieved in January 2011.

Agricultural Data

Quantity of sugarcane and soybean produced and harvested area per year according to the Brazilian National Statistics Bureau, *Instituto Brasileiro de Geografia e Estatistica*, www.ibge.gov.br, Pesquisa Agrícola Municipal. Data at micro-region level from 2006 to 2010.

Forestry Data

Quantity of wood, firewood and charcoal produced, according to the Brazilian National Statistics Bureau, *Instituto Brasileiro de Geografia e Estatistica*, www.ibge.gov.br, Produção da Extração Vegetal e da Silvicultura. Data at micro-region level from 2006 to 2010.

Labor Data

Sugarcane Labor, Soybean Labor, Raw Sugar Labor, Ethanol Labor, Raw Vegetable Oil Labor, Biofuel Labor, Forestry Labor, Cellulose Labor obtained from Labor Ministry, *Ministério do Trabalho e Emprego*, www.mte.gov.br, RAIS. Data on formal employment as of 31.12 of each year, from 2006 to 2010 per micro-region.

Bibliography

[1] International Energy Agency. IEA Energy Technology Essentials Biofuel Production. Technical report, Internatioanl Energy Agency, January 2007.

[2] Frank J. Alessio. A Neo-Classical Land Use Model: The Influence of Externalities. *The Swedish Journal of Economics*, 75(4):414–419, December 1973.

[3] Martin J. Beckmann. Von Thunen Revisited: A Neoclassical Land Use Model. *The Swedish Journal of Economics*, 74(1):1–7, March 1972.

[4] Martin J. Beckmann. *Lectures on Location Theory*, chapter 11, pages 121–141. Springer-Verlag, 1999.

[5] Gustavo Best. Agro-Energy: A New Function of Agriculture. *Latin America Thematic Network on Bioenergy*, (3rd Issue):7, June 2003.

[6] International Energy Agency Bioenergy. Bionergy, Land Use Change and Climate Change Mitigation. Technical report, International Energy Agency Bioenergy, 2011. Task 29.

[7] Cristina Brasili and Roberto Fanfani. *The new European rurality: strategies for small firms*, chapter 5, pages 61 – 86. Ashgate, 2006.

[8] Timothy F. Bresnahan and Peter C. Reiss. Entry and Competition in Concentrated Markets. *Journal of Political Economy*, 99(5):pp. 977–1009, 1991.

[9] Jin Hooi Chan and David Reiner. Dynamics of Evolution in the Global Ethanol Fuel-Ethanol Industry. Cambridge Working Paper in Economics CWPE 1129, EPRG Working Paper 1111, University of Cambridge, Electricity Policy Research Group, March 2011.

[10] Maureen Cropper and Charles Griffiths. The Interaction of Population Growth and Environmental Quality. *The American Economic Review*, 84(2):pp. 250–254, 1994.

[11] Associacao Brasileira das Industrias de Oleos Vegetais. Understanding the Soy Moratorium: Responsible Production. Technical report, Soybean Working Group, 2007.

[12] Banco Nacional de Desenvolvimento Economico e Social. *Sugarcane-based bioethanol: energy for sustainable development*. Banco Nacional de Desenvolvimento Economico e Social, Rio de Janeiro, 1st edition, November 2008.

[13] Empresa Brasileira de Pesquisa Agropecuaria EMBRAPA. Sugar Cane Agroecological Zoning: to expand production, preserve life, and ensure a future. Technical report, EMBRAPA Solos, Rio de Janeiro, 2008.

[14] Joseph Fargione, Jason Hill, David Tilman, Stephen Polasky, and Peter Hawthorne. Land Clearing and the Biofuel Carbon Debt. *Science*, 319(5867):1235–1238, 2008.

[15] Food and Agriculture Organization of the United Nations. Unified Bioenergy Terminology. Technical report, FAO Forestry Department, Wood Energy Programme, December 2004.

[16] Food and Agriculture Organization of the United Nations. Forestry and Energy. Key Issues. FAO Forestry Paper 154, Food and Agriculture Organization, 2008.

[17] Food and Agriculture Organization of the United Nations. The State of Food and Agriculture. Biofuels: prospects, risks and opportunities. Biofuels and Agriculture - A technical overview. Technical report, Food and Agriculture Organization, 2008.

[18] Julien Frayssignes. The concept of 'agroenergy district': a relevant tool for sustainable development in rural areas. ERSA conference papers ersa11p318, European Regional Science Association, September 2011.

[19] Masahisa Fujita and Jacques-Francois Thisse. *Economics of agglomeration: cities, industrial location, and regional growth*, chapter 3, pages 62–92. Cambridge University Press, New York, 2002.

[20] Edward L. Glaeser. The Economics of Location-Based Tax Incentives. Discussion Paper 1932, Harvard Institute of Economic Research, November 2001.

[21] Jose Goldemberg and Suani Teixeira Coelho. Renewable Energy - Traditional Biomass vs. Modern Biomass. *Energy Policy*, 32(6):711 – 714, 2004.

[22] Jose Goldemberg, Suani Teixeira Coelho, Plinio Mario Nastari, and Oswaldo Lucon. Ethanol learning curve - the Brazilian experience. *Biomass and Bioenergy*, 26(3):301 – 304, 2004.

[23] Jose Goldemberg and Jose R. Moreira. The Alcohol Program. *Energy Policy*, 27(4):229 – 245, 1999.

[24] Peter Goldsmith and Rodolfo Hirsch. The Brazilian Soybean Complex. *Choices*, 21:97–104, 2nd Quarter 2006.

[25] Monica A. Haddad, Gary Taylor, and Francis Owusu. Locational Choices of the Ethanol Industry in the Midwest Corn Belt. *Economic Development Quarterly*, 24(1):74–86, 2010.

[26] Joseph F. Hair, Bill Black, Barry Babin, Rolph E. Anderson, and Ronald L. Tatham. *Multivariate Data Analysis*. 6th edition, 2006.

[27] Catherine Hausman. Biofuels and Land Use Change: Sugarcane and Soybean Acreage Response in Brazil. *Environmental and Resource Economics*, 51:163–187, 2012.

[28] Ricardo Hausmann. Biofuels Can Match Oil Production. *The Financial Times*, November, 6th, 2007.

[29] Ricardo Hausmann and Rodrigo Wagner. Certification Strategies, Industrial Development and a Global Market for Biofuels. *CID Working Paper 192. Center for International Development at Harvard University. Jointly published as Discussion Paper 2009-15, Belfer Center for Science*, October 2009.

[30] Jason R. Henderson and Kevin T. McNamara. The Location of Food Manufacturing Plant Investment in Corn Belt Counties. *Journal of Agricultural and Resource Economics*, 25(02), 2000.

[31] M. Herrero, P. K. Thornton, A. M. Notenbaert, S. Wood, S. Msangi, H. A. Freeman, D. Bossio, J. Dixon, M. Peters, J. van de Steeg, J. Lynam, P. Parthasarathy Rao, S. Macmillan, B. Gerard, J. McDermott, C. Ser, and M. Rosegrant. Smart Investments in Sustainable Food Production: Revisiting Mixed Crop-Livestock Systems. *Science*, 327(5967):822–825, 2010.

[32] Harold Hotelling. Stability in Competition. *The Economic Journal*, 39(153):pp. 41–57, 1929.

[33] Song-ken Hsu. The Agroindustry: A Neglected Aspect of the Location Theory of Manufacturing. *Journal of Regional Science*, 37(2):259–274, 1997.

[34] Donald W. Jones and John R. Krummel. The Location Theory of the Plantation. *Journal of Regional Science*, 27(2):157–182, 1987.

[35] Nivalde Jose de Castro, Roberto Brandao, and Guilherme de A. Dantas. *Etanol e Bioeletricidade. A cana-de-acucar no futuro da matriz energetica*, chapter 5, pages 136–153. UNICA. Uniao da Industria de Cana-de-acucar, 2010.

[36] Maureen Kilkenny and Scott Coleman. Agro-industry Location: Theory and Test. Draft of work in progress. Unpublished working paper, 2006. Retrieved on 04/09/2011 at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.5989&rep=rep1&type=pdf.

[37] Dayton M. Lambert, Michael D. Wilcox, Alicia English, and Lance A. Stewart. Ethanol Plant Location Determinants and County Comparative Advantage. *Journal of Agricultural and Applied Economics*, 40(01), April 2008.

[38] Henry Lee, William C. Clark, and Charan Devereaux. Biofuels and Sustainable Development: Report of an Executive Session on the Grand Challenges of a Sustainability Transition. San Servolo Island, Venice, Italy: May 19-20, 2008. *CID Working Paper No. 174*, July 2008.

[39] Ann Markusen. Sticky Places in Slippery Space: A Typology of Industrial Districts. *Economic Geography*, 72(3):293–313, July 1996.

[40] Federal Republic of Brazil. Second National Communication of Brazil to the United Nations Framework Convention on Climate Change. Technical report, Ministry of Science and Technology, Brasilia, October 2010.

[41] PECEGE. *Custos de producao de cana-de-acucar, acucar e etanol no Brasil: Fechamento da safra 2010/2011.* Universidade de Sao Paulo, Escola Superior de Agricultura Luiz de Queiroz, Programa de Educacao Continuada em Economia e Gestao de Empresas/Departamento de Economia, Administracao e Sociologia, Piracicaba, Sao Paulo, Brazil, 1st edition, September 2010. Relatorio apresentado a Confederacao da Agricultura e Pecuaria do Brasil.

[42] PECEGE. *Custos de producao de cana-de-acucar, acucar e etanol no Brasil: safra 2009/2010.* Universidade de Sao Paulo, Escola Superior de Agricultura Luiz de Queiroz, Programa de Educacao Continuada em Economia e Gestao de Empresas/Departamento de Economia, Administracao e Sociologia, Piracicaba, Sao Paulo, Brazil, 2nd edition, September 2010. Relatorio apresentado a Confederacao da Agricultura e Pecuaria do Brasil.

[43] Giuseppe Piroli, Pavel Ciaian, and Kancs d'Artis. Land Use Change Impacts of Biofuels: Near-VAR Evidence from the US. Research Paper Series 11, Economics and Econometrics Research Institute, Belgium, 2011.

[44] Gabriella P.A.G. Pousa, Andre L.F. Santos, and Paulo A.Z. Suarez. History and policy of biodiesel in Brazil. *Energy Policy*, 35(11):5393 – 5398, 2007.

[45] John Reilly and Sergey Paltsev. Biomass Energy and Competition for Land. GTAP Working Paper 46, Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, Cambridge, USA, 2008.

[46] Bertrand M. Renaud. On a Neoclassical Model of Land Use. *The Swedish Journal of Economics*, 74(3):400–404, September 1972.

[47] Paul A. Samuelson. Thunen at Two Hundred. *Journal of Economic Literature*, 21(4):pp. 1468–1488, 1983.

[48] Camilo Sarmiento, William W. Wilson, and Bruce Dahl. Spatial Competition and Ethanol Plant Location Decisions. *Agribusiness*, 28(3):260–273, 2012.

[49] Timothy Searchinger, Ralph Heimlich, R. A. Houghton, Fengxia Dong, Amani Elobeid, Jacinto Fabiosa, Simla Tokgoz, Dermot Hayes, and Tun-Hsiang Yu. Use of U.S. Croplands for Biofuels Increases Greenhouse Gases Through Emissions from Land-Use Change. *Science*, 319(5867):1238–1240, 2008.

[50] Giovanni Sorda, Martin Banse, and Claudia Kemfert. An overview of biofuel policies across the world. *Energy Policy*, 38(11):6977 – 6988, 2010.

[51] Anthony J. Venables. Equilibrium Locations of Vertically Linked Industries. *International Economic Review*, 37(2):341–359, May 1996.

[52] Assis Vicente, Elstrodt Heinz-Peter, and Claudio F. C. Silva. Positioning Brazil for Biofuels Sucess. *The Mckinsey Quarterly*, Special Edition: Shaping a new agenda for Latin America, March 2007.

[53] Marc D. Weidenmier, Joseph H. Davis, and Roger Aliaga-Diaz. Is Sugar Sweeter at the Pump? The Macroeconomic Impact of Brazil's Alternative Energy Program. NBER Working Papers 14362, National Bureau of Economic Research, October 2008.

[54] C.E.O. Xavier, L.B. Zilio, D.Y. Sonoda, and P.V. Marques. *Custos de producao de cana-de-acucar, acucar e etanol no Brasil: safra 2008/2009.* Universidade de Sao Paulo, Escola Superior de Agricultura Luiz de Queiroz, Programa de Educacao Continuada em Economia e Gestao de Empresas/Departamento de Economia, Administracao e Sociologia, Piracicaba, Sao Paulo, Brazil, December 2009. Relatorio apresentado a Confederacao da Agricultura e Pecuaria do Brasil.

Estratto per riassunto della tesi di dottorato

Studente:	Breno Pietracci
Matricola:	955551
Dottorato:	Economia
Ciclo:	XXIV

Titolo della tesi: Essays on the Microeconomics of Bioenergy

Abstract: This thesis explores the microeconomics of bioenergy with focus on Brazil. The three chapters are interconnected by this thread. The first chapter investigates empirically land use patterns for liquid biofuels production in Brazil using the neoclassical land use model. It analyzes the determinants of sugarcane and soybean production location. The second chapter inquire empirically, using regional variables, location and capacity decision drivers for ethanol and biodiesel mills in Brazil. The third chapter proposes a microeconomic framework for modeling several types of bioenergy production. It provides a definition of bioenergy clusters and tests it employing statistical cluster analysis considering three different bioenergy production chains.

Estratto: Questa tesi esplora la microeconomia della bioenergia con un focus nel Brasile. I tre capitoli sono interconnessi da questo filo. Il primo capitolo analizza empiricamente i modelli di uso del suolo per la produzione di biocarburanti liquidi in Brasile utilizzando il modello neoclassico d'uso del suolo, analizzando le determinanti di localizzazione della produzione di canna da zucchero e soia. Il secondo capitolo investiga empiricamente i meccanismi alla base delle decisioni di localizzazione e di capacità per impianti d'etanolo e biodiesel in Brasile utilizzando variabili regionali. Il terzo capitolo propone un quadro microeconomico per modellare diversi tipi di produzione di bioenergia e fornisce una definizione di cluster bioenergetico validata usando l'analisi statistica di cluster considerando tre differenti catene di produzione di bioenergia.