

# Implications of Unleashing Brazilian Ethanol: Trading Off Renewable Fuel for How Much Forest and Savanna Land?

Sriniketh Nagavarapu\*

May 18, 2010

## Abstract

Lowering US import barriers for Brazil's sugarcane ethanol could boost the production of this low-carbon renewable fuel significantly, but also lead to deforestation in Brazil. I examine this trade-off by using annual household- and region-level data from 1995-2005 to estimate a general equilibrium model of Brazil's regional agricultural markets, accounting for competition over land, labor, and capital. Simulations using the estimates indicate Brazil could export more than 5 billion gallons of ethanol with little decline in non-agricultural land. However, higher world prices of oil and agricultural goods lead to dramatic land clearing while raising exports only 10 billion additional gallons.

---

\*Department of Economics, Brown University and Center for Environmental Studies, Brown University. Box B, Brown University, Providence, RI 02912, ssn@brown.edu. I owe a large debt to Thomas MaCurdy, John Pencavel, and Jayanta Bhattacharya for their advice, support, and encouragement. I am also grateful to Luigi Pistaferri and Aprajit Mahajan for their guidance. I benefited greatly from conversations with Orazio Attanasio, Nick Bloom, Marcus Edvall, Andrew Foster, Zephyr Frank, Giacomo De Giorgi, Gopi Shah Goda, Vernon Henderson, Caroline Hoxby, Lovell Jarvis, Seema Jayachandran, Katja Kaufmann, Soohyung Lee, Colleen Manchester, Naercio Menezes-Filho, Pedro Miranda, Jonathan Meer, Marc Muendler, Kevin Mumford, Gerald Nelson, Alex Pfaff, Alejandro Ponce-Rodriguez, Felix Reichling, Chris Timmins, Frank Wolak, Joanne Yoong, attendees of Stanford's Applications Seminar, and members of Stanford's Labor Reading Group. Suggestions received at Brown University, RAND, the USDA, Mathematica Policy Research, the AEA annual meetings, Duke University, North Carolina State University, and RTI International are greatly appreciated. For their immense help with acquiring data used here, I thank Steven Helfand, Frank McIntyre, Marcia Moraes, and the staff at Fundacao Getulio Vargas. I appreciate the financial support provided by SIEPR for data acquisition and by the Taube/SIEPR dissertation fellowship and the NSF Graduate Research Fellowship. I am grateful to Soo-Ryun Kwon, Mohan Nagavarapu, and Kamala Paritala for their constant support. Remaining errors are, of course, my own.

Amid growing economic, security, and environmental concerns related to oil usage, the United States and other countries have shown greater interest in ethanol. The 2007 energy bill passed by the US Congress requires that the total amount of transportation fuels used in the US contain at least a minimum level of renewable fuels, with the minimum growing over time. For instance, renewables must constitute 15.2 billion gallons of total transportation fuel by 2012 and 36 billion gallons by 2022.<sup>1</sup> While bio-diesel may play a role in the coming years, this renewable fuel mandate will be fulfilled primarily through the use of ethanol. Fuel ethanol is of growing importance in a variety of countries, but the US is the world's largest market for ethanol, and the renewable fuels mandate will ensure this is likely to be the case for many years to come.

No country is in a better position to take advantage of this surging interest in ethanol than Brazil. Brazil's sugarcane-based ethanol has three important advantages over US corn-based ethanol and other types of ethanol. First, beginning in the 1970s, Brazil spent a substantial amount of government resources to develop infrastructure for ethanol production and distribution. Second, Brazil's natural endowments are conducive to growing sugarcane at low cost. Third, sugarcane can be converted into ethanol with a smaller energy input than that needed for other conventional feedstocks. Brazil is the most cost- and energy-efficient producer of ethanol in the world, and the country has vast potential for expanding sugarcane cultivation further.

These facts make the US-Brazil interaction a key facet of the ethanol industry going forward. Brazil would be a natural source for the United States' ethanol requirements, except for one fact: The US protects its own ethanol producers with a 2.5% ad valorem tariff and a 54 cent per gallon duty on imports. These recently extended measures have come under increasing criticism by the Brazilian government. In fact, the government has aggressively pitched freer markets for its ethanol as a potential "win-win," a simultaneous spur to rural economic growth and an environmentally friendly product.<sup>2</sup>

Nevertheless, more open markets for Brazilian ethanol generate uncertain implications for the environment in Brazil and elsewhere. There is a crucial tradeoff. On the one hand, Brazil's sugarcane ethanol is a renewable, lower-carbon alternative to petroleum and US corn ethanol. Replacing US consumers' use of petroleum and corn ethanol with Brazilian ethanol could therefore reduce carbon emissions sizeably. On the other hand, this additional sugarcane has to be produced somewhere in Brazil. Brazil's land mass features a wide array

---

<sup>1</sup>The "Energy Independence and Security Act of 2007" became Public Law 110-140 on December 19, 2007. The relevant portion of the law is Title II, Subtitle A, Section 202. The Environmental Protection Agency (EPA) has the authority to implement the volume requirements with appropriate regulations.

<sup>2</sup>For instance, see "Our Biofuels Partnership", by President Luiz Inacio Lula da Silva, in the March 30, 2007 edition of the Washington Post (page A17).

of forest and savanna land, with both carbon sequestration and bio-diversity functions. The expansion in sugarcane production required to produce more ethanol could lead to greater deforestation and other environmentally harmful land clearing.

This paper addresses this tradeoff by answering the research question: Would freely importing Brazilian ethanol into the US lead to enough land clearing to offset the environmental benefits of greater ethanol use?

To answer this question, it is necessary to determine how responsive Brazilian ethanol production is to the opening of the market, and how much land will enter agricultural use as sugarcane cultivation increases. As the US ethanol market opens, sugarcane demand will increase. Sugarcane production in each region will respond to the resulting change in prices. But precisely how much it will respond depends on a large number of factors. For instance, the appropriateness for sugarcane production of land not currently used for sugarcane will help determine the response on the extensive margin. The substitutability between labor and capital will help determine how feasible it is to cultivate existing land more intensively. This substitutability, along with the elasticity of labor supply and capital supply to the sugarcane sectors, will have consequences for wages and the amount of labor used in sugarcane and other sectors, which will in turn impact sugarcane supply. Preferences for other agricultural goods and sugar, and the ability to import these goods to satisfy domestic demand, will also play a role. In summary, answering the question above involves the consideration of the intimate but complicated links between land use decisions, input markets, and product markets.

In view of this, in this paper I develop and estimate a general equilibrium model of regional agricultural markets in Brazil. Producers of sugarcane, other agricultural goods, ethanol, sugar, and non-agricultural goods compete over land, labor and capital inputs. These features are common to computational general equilibrium (CGE) models of land use, with the most relevant to Brazilian biofuels being recent work by Onil Banerjee, Alexander Macpherson & Janaki Alavalapati (Forthcoming). There are also studies of agricultural trade liberalization in Brazil that have used CGE models in combination with micro-simulation models that estimate wage equations and migration equations using a year or two of survey data (see, e.g., Maurizio Bussolo, Jann Lay & Dominique van der Mensbrugghe (2006)).<sup>3</sup> Unlike these models, I pursue a maximum likelihood estimation approach that estimates most parameters of the model with one set of data over multiple years, ensuring that all key elasticities come from one coherent estimation framework. Every error term in the model

---

<sup>3</sup>See Anne-Sophie Robilliard, Francois Bourguignon & Sherman Robinson (2001) for a CGE model including parameters estimated from a self-employment/employment choice in Indonesia. For a different sort of micro-simulation approach in Brazil, see Joaquim Bento de Souza Ferreira Filho & Mark Horridge (2009). Thomas Hertel & Jeffrey J. Reimer (2005) provide a good overview.

has an economic interpretation, and the estimation and simulation approaches both account for the stochastic nature of the error terms.

Heterogeneity of land quality and heterogeneity of labor supply preferences play an important role in the model. On the former count, the model is much less detailed than work by Gerald C. Nelson & Richard D. Robertson (2008), who examine the biofuel issue using a cross-section of highly disaggregated data on land use.<sup>4</sup> I complement this work by examining the influence of labor supply decisions and using annual individual-level and aggregate data from the 1995-2005 period to estimate a model that clarifies underlying mechanisms. Despite the simultaneous use of individual-level and aggregate data, my methodology imposes constraints that ensure the logical consistency and coherence of the empirical model.

Throughout, my focus is on understanding the effects of greater international ethanol demand on changes in land use within Brazil. My treatment of the international trade portion of the question is necessarily simplified. Amani Elobeid & Simla Tokgoz (2008) instead focus on the international ethanol market in their CGE examination of the effect of removing US trade barriers on trade and production of ethanol, and I use two elasticities from these authors in the international component of my model.<sup>5</sup> While labor supply is a critical component of my model, the framework is too highly aggregated and focused on land clearing to provide a detailed understanding of workers' wages. Ekaterina Krivonos & Marcelo Olarreaga (2009) instead focus on the effects of sugarcane expansion on wages.<sup>6</sup>

After estimating the model, I use the parameters to predict the consequences of increasing international ethanol demand. I find that Brazil could increase ethanol exports sharply in response to a removal of US import restrictions with little or no effect on non-agricultural land. Holding production shocks and exogenous variables constant at their 2005 values, a removal of the import tariffs is predicted to lead to an additional 4.8 billion gallons of ethanol exports. To provide an idea of how large this is, consider that total ethanol production in the US in 2007 was 6.5 billion gallons. A diversion of sugarcane from sugar production and a decrease in other agricultural land allows this gain to come at no cost to total non-agricultural land in most of Brazil. In fact, wage spillovers in other agriculture make it less profitable, leading to predicted increases in non-agricultural land in most areas. I examine the sensitivity of this result to uncertainty by drawing vectors of production shocks, computing

---

<sup>4</sup>For other detailed studies of land use, but using a CGE approach, see Angelo Gurgel, John M. Reilly & Sergey Paltsev (2007) and Jerry M. Melillo, John M. Reilly, David W. Kicklighter, Angelo C. Gurgel, Timothy W. Cronin, Sergey Paltsev, Benjamin S. Felzer, Xiaodong Wang, Andrei P. Sokolov & C. Adam Schlosser (2009).

<sup>5</sup>For a detailed empirical analysis of ethanol imports from Brazil, see Barbara Farinelli, Colin A. Carter, C.-Y. Cynthia Lin & Daniel A. Sumner (2009) and Hyunok Lee & Daniel A. Sumner (2010)).

<sup>6</sup>See, also, Alexandre Rands Barros, Tiago V. de V. Cavalcanti, Fernando Mendonca Dias & Andre Matos Magalhaes (2005) and Marcia Azanha Ferraz Dias De Moraes (2007).

equilibria under each one, and then examining the distribution of key quantities across the simulations. The basic result continues to hold. Consequently, Brazil could provide enough ethanol to have a significant impact on the use of renewables in the US with little effect on its most environmentally sensitive areas.

I conduct two additional simulation exercises to address other pertinent concerns. In the first, I not only remove US import barriers, but also increase the price of other agricultural goods by 20%. Increases in soy prices in recent times makes this a relevant exercise. This leads to large simulated declines in non-agricultural land. This is exacerbated in the second exercise, where I additionally increase the world price of petroleum over the 2005 baseline by 20%, bringing the price to around \$60 per barrel. This shock to ethanol demand dramatically increases ethanol production and exports. At 2005 values of the production shocks and exogenous variables, ethanol exports climb to over 16 billion gallons. The median over the set of simulations allowing for random draws of production shocks is just under 15 billion gallons. To put this in perspective, note that with 13 billion gallons of ethanol, the US could have used ten percent ethanol blends in all of its gasoline consumption in 2007.<sup>7</sup> However, the cost in terms of lost non-agricultural land is large, more than 100 million acres relative to the baseline when using 2005 values, with large swathes coming from a region with a substantial amount of Amazon Rainforest. The results suggest that the tradeoff between greater ethanol exports and more land clearing is non-linear and dependent on economic factors relatively external to Brazil, such as petroleum prices.

This paper provides a new perspective on the debate over Brazilian ethanol by accounting for input and output market linkages in an approach that estimates all key elasticities with one set of individual-level and aggregate data, in one coherent framework. But there are important limitations to the results. For instance, limitations to the labor force data prevent me from examining the sparsely populated region of the country with the densest areas of the Amazon Rainforest. As another example, behavioral responses not accounted for here – on the part of government or private actors – could alter the outcomes significantly.<sup>8</sup> I focus on potential declines in non-agricultural public land, since this seems of first-order importance in Brazil. But the degree to which changes in planted and natural forests occur on privately owned land could affect my results in important ways (see, for instance, Andrew Foster & Mark Rosenzweig (2003) on village forest land in India).<sup>9</sup> Additionally, future improvements

---

<sup>7</sup>Most vehicles in the US cannot run entirely on ethanol. For the source of these numbers, see the Department of Energy article at [http://apps1.eere.energy.gov/news/news\\_detail.cfm/news\\_id=11633](http://apps1.eere.energy.gov/news/news_detail.cfm/news_id=11633).

<sup>8</sup>For studies examining the management of forests by communities in developing countries, see, for example, Eric Edmonds (2002) and Jennifer Alix-Garcia (2007).

<sup>9</sup>The most recent Agricultural Census suggests that the majority of privately owned natural forest land in Brazil is used for uses other than sustainable wood products, including the grazing of animals and growing of particular crops. These private forests may therefore be less acceptable from an environmental point of

in infrastructure – such as new road networks or new ports – could have important effects (see, e.g., Alexander Pfaff (1999)).<sup>10</sup> I discuss these and other limitations in the conclusion. The framework presented here could be fruitfully adapted to address many of these points in the future.

The remainder of the paper proceeds as follows. Section 1 describes the data sources. Data availability helps dictate modeling choices later in the paper. Section 2 provides an overview of sugarcane and ethanol production in Brazil, in addition to its geographic relationship to key environmental land. The discussion motivates the need for a model. Section 3 develops the model, while Section 4 describes the structural estimation approach and presents the estimation results. Section 5 discusses the methods used in the policy simulations, as well as the simulation results. Section 6 discusses the most important limitations of the model and concludes.

## 1 Data

Here, I describe the primary data sources briefly. For further details, please see the appendix. The analysis requires information on several economy-wide aggregates. To construct aggregate sugar and ethanol supplies, I use data from the Ministry of Agriculture, Livestock, and Food Supply. The price data come from *Fundacao Getulio Vargas* (FGV), a foundation that compiles a large amount of economic data for Brazil. I also require data on total regional production, national capital investment, government spending, and net exports (overall, agricultural goods, ethanol, and sugar). For this purpose, I use national and regional income accounts data from *Instituto Brasileiro de Geografia* (IBGE), and data on imports and exports from Brazil's *Secretaria de Comercio Exterior* (SECEX).

The empirical analysis also requires an estimate of the total value of production in, and total amount of land devoted to, sugarcane and other agriculture in every region and year. For information on production and land allocations by state, I use the *Producao Agricola Municipal* (PAM). The PAM does not capture four types of land uses: pasture land; land temporarily at rest; forest land; and land used for minor crops (a problem in the 1981-1989 period). To address this problem, I use data from the 1980, 1985, 1995, and 2005-06 Agricultural Censuses, which contain information on pasture, forests, and all other agricultural land (at rest or otherwise). I also use annual data on cattle herd sizes from *Producao Pecuaria Municipal* (PPM), and IBGE data on the total value of forest production.

---

view. To the extent that privately managed production on forest land is instead environmentally sustainable and non-invasive, my results could over-state the threat to environmentally sensitive areas.

<sup>10</sup>I also rule out endogenous improvements in agricultural technology. See, e.g., D.A. Hay (2001) on Brazilian manufacturing during trade liberalization.

With these data sources, I use simple extrapolation procedures to project total agricultural area in the years of my analysis. Finally, the PAM data do not provide the total value of agricultural production for all years. For this purpose, I use regional income accounts information for the 1985-2005 period.

For information on the labor supply decisions of individuals and sectoral wages, I rely on the *Pesquisa Nacional por Amostra de Domicílios* (PNAD). The PNAD is a cross-sectional representative household survey collected by IBGE. On average, the PNAD surveys roughly 100,000 households each year. In the figures below, I will use PNAD data from 1981 through 2005, and in the estimation I use PNAD data from 1995-2005.<sup>11</sup> The only region of the country not represented fully in the PNAD sample is the North Census region.<sup>12</sup> This will prevent me from addressing the sugarcane markets in the Northern portion of the country.

## 2 Brazil's Sugarcane and Ethanol Industries

This section describes aspects of the sugarcane and ethanol industries that are essential to understanding the debate behind the research question and to motivating my model. I begin by discussing the location of forest and savanna land within Brazil and then presenting a regional classification system used in the remainder of the paper.

### 2.1 Areas of Environmental Concern

The areas of primary environmental concern are the Amazon Rainforest, the Atlantic Rainforest, and the savanna land known as the *Cerrado*.<sup>13</sup> Of these, the Amazon has received the greatest attention in popular discourse, both because of its bio-diversity and its carbon storage potential. However, as illustrated in Nelson & Robertson (2008), the Atlantic Rainforest and the *Cerrado* are also environmentally important areas, because of the housing of various species and carbon storage.

Figure 1 illustrates the location of the various biomes within Brazil. Large swathes of the country fall under the Amazon Rainforest (Amazonia), Atlantic Rainforest (Mata Atlântica), and *Cerrado* biomes. Despite the extent of the biome in the map, the Atlantic Rainforest today is often sparse and exists primarily as a thin strip along the length of

---

<sup>11</sup>In 1991 and 2000, the PNAD was not conducted because those were national demographic census years. In 1994, the PNAD was not conducted due to other constraints. Although I have the national census data for 1991 and 2000, these were conducted at different times of year from the PNAD, making comparisons difficult in the agricultural sector.

<sup>12</sup>The rural parts of the six states in the North region are not represented through 2003. In the 2004 and 2005 data, this deficiency has been remedied, and the PNAD is now nationally representative.

<sup>13</sup>The Pantanal, in Mato Grosso and Mato Grosso do Sul, is another ecologically rich area, but the threats to these wetlands may not come from agricultural expansion.

Brazil’s eastern coast. In both the Amazon and Atlantic Rainforests, the holding of a certain percentage of private land as forest is legally mandated and there are also public reserves, though enforcement is minimal.<sup>14</sup> Regarding the *Cerrado* savanna lands, large swathes of this scrubby vegetation area are already under agricultural use, and there is intense concern about further expansion.

Figure 2 presents the regional classification that I will use in the remainder of the paper. Table 1 lists the states composing each region, with the first and second column providing the number and “shorthand reference” that I use to refer to each region in future tables and figures. The table also shows how each of my regions come together to form the government-defined “Census regions.” To avoid confusion, whenever I use the term “region” in the paper, I am referring to my definition of the seven regions and not the Census definition.<sup>15</sup> In the analysis, I am unable to include the sparsely populated North Census region. This is unfortunate, given that the majority of the Amazon Rainforest is located in this region.<sup>16</sup> However, the Mato Grosso region contains a significant acreage of rainforest, and the other regions I consider include the Atlantic Rainforest and the *Cerrado*.

## 2.2 Factors Affecting Sugarcane and Ethanol Production

With this context in mind, we can examine the regional pattern of sugarcane production to better understand the debate over Brazilian ethanol. Four factors help to determine the appropriateness of land for sugarcane production: reliability of precipitation at particular points of the growth cycle; soil composition; land gradient; and distance to the nearest sugar mill or ethanol distillery. The gradient of the land is important because flatter land is easier to harvest, either with manual harvesting or – if the capital is available – mechanical harvesting.<sup>17</sup> Sugar mills and ethanol distilleries are typically near sugarcane fields: sugarcane

---

<sup>14</sup>Note, however, that Pfaff (1999) suggests that compliance is more prevalent than commonly thought.

<sup>15</sup>I choose to group states together into regions based primarily on geographical proximity, but also consider anecdotal evidence about the integration of labor markets. In determining the number of regions, I strike a balance between, on the one hand, using so few regions that important heterogeneity within regions is neglected and, on the other hand, using so many regions that the household survey does not contain a reasonably large number of sugarcane workers in some regions. I sometimes refer to region 5 as the “Center-West,” even though Tocantins is technically a part of the North census region, and not the Center-West. I do so because Tocantins was once connected to Goias.

<sup>16</sup>The North Census region is a small, though non-trivial, part of the economy. In a recent year, about 5% of GDP was produced in the North. Using the household survey data described below, I find that in 2005, approximately 6.3% of the over-15 population lives in the North, and approximately 6.6% of over-15 workers are located in the North. To assess the importance of omitting the North for this analysis, one could look at migration rates into the North using the post-2003 data.

<sup>17</sup>The majority of production shifted from the Northeast to the Center-South of Brazil by the 1950s, partly because of the advantage of flatter lands. The Northeast still produces sugarcane, sugar, and ethanol, due in part to seasonal patterns of production across Brazil. Barbara Numberg (1986) describes this historical shift and Brazil’s sugarcane policies before the 1980s.

is sold almost entirely to either mills or distilleries, and fields are typically located relatively close to these facilities since transportation over long distances is expensive and results in deterioration in the quality of the harvested sugarcane.

At the same time, other factors are at work, factors not considered in the model below. The largest such factor is government policy.<sup>18</sup> Until 1997, the Brazilian government intervened strongly in the sugarcane, sugar, and ethanol markets. The *Instituto do Acucar e do Alcool* (IAA) operated a system of quotas and price controls for sugarcane, sugar, and ethanol over a period of 60 years. As time went on, the system of quotas and administered prices came under increasing pressure. In 1990, the IAA was abolished and sugar exports were privatized. In 1995, sugar quotas were ended, and from 1997-1999, prices for sugar, ethanol, and sugarcane were liberalized. For simplicity, the model assumes a free market for the entire industry during the entire 1995-2005 period. While the adequacy of this assumption for sugarcane and ethanol should be examined in the future, Steven M. Helfand (2003) shows that through the 1990s, domestic sugar prices moved roughly in tandem with international sugar prices.

### 2.3 Patterns of Sugarcane Production Across Space and Time

One line of argument points to the current distribution of sugarcane production across Brazil to suggest that any future sugarcane expansion might take place primarily in Sao Paulo, Pernambuco, and possibly other states in the South-Southeast. This would limit consequences for the Amazon or the *Cerrado*, as long as the migration of displaced other agriculture is not extensive (a point returned to below). To illustrate the regional distribution of production, Table 2 shows land allocations and production quantities in 2005 for all seven regions used in the analysis. The Pernambuco region and the South-Southeast portions of the country dominate sugarcane production. The differences in sugarcane production across regions translate naturally into differences in ethanol and sugar production. Table 3 displays differences across regions in ethanol and sugar production, as well as the value of that production. These numbers suggest that many parts of the country face little threat from expanding sugarcane production, since their current contribution is so marginal.<sup>19</sup>

However, historical land use patterns make clear that technological changes and other

---

<sup>18</sup>For major events in government policy, see Geraldo Sant'ana De Camargo Barros & Marcia Azanha Ferraz Dias De Moraes (2002), Mauro De Rezende Lopes & Iñez Vidigal Lopes (1998), and Brent Borrell, Jose R. Bianco & Malcolm D. Bale (1994).

<sup>19</sup>Figures 6 and 7 in the appendix illustrate the regional distribution of production in 2007. The first figure divides the hectares of cultivated land in Brazil's counties into eight quantiles, and depicts areas in higher quantiles with darker colors. The second figure instead classifies the counties into eight bins of equal size, defined by the absolute level of hectares cultivated with sugarcane. The second figure makes clear that the majority of sugarcane in Brazil still comes from Sao Paulo.

pressures can quickly lead to changing patterns of cultivation. Consider Figure 3, which shows sugarcane land shares (top row) and production totals (bottom row) over time for each region, relative to 1990. The top panel shows the trends for the four regions in the Center-South, and the bottom panel refers to the regions in the Northeast. In all the regions, sugarcane land grows over the 1981-1990 period, though data for the Mato Grosso region are not available in this period. Since 1990, the picture is very different. In the Center-South, all regions except Minas Gerais show growth since 1990, while no regions in the Northeast show growth since 1990. The strongest relative growth is in Mato Grosso.<sup>20</sup> These figures tell a very different story from the cross-sectional distributions, and suggest that growth in sugarcane cultivation may not be confined to Sao Paulo and Pernambuco.

## 2.4 Motivating the Model

The movements over time and the cross-sectional distributions of regional sugarcane production are each consistent with one line of argument in the ethanol debate in Brazil. In reality, both the patterns across space and those across time are the result of a complicated set of interactions among underlying forces, including the demand and supply of inputs, the re-location of non-sugarcane agricultural production amid sugarcane growth, and more.

As an example, consider the roles of labor and capital in regional sugarcane production. The household survey data, in combination with the data on sugarcane production, reveal substantial regional differences in labor productivity (output divided by labor hours). Table 4 shows median hours of work and estimated total number of workers in 2005 in each sector/region combination. By comparing this with the production totals in Table 2, one sees that the Center-South regions have a strikingly higher level of production per hour of labor. Considering the previous figures, this suggests the regions with greatest labor productivity have seen the greatest growth in sugarcane production in the 1990s.

One possible explanation for this is that the difficulty and health consequences of sugarcane work make it hard to attract labor into the sector.<sup>21</sup> Consistent with this, sugarcane wages are generally higher than wages in other agriculture. Figure 11 (appendix) depicts the evolution of median hourly wages in sugarcane, other agriculture and non-agriculture in each of the seven regions.<sup>22</sup> For many of the regions, there is a persistent gap between sugarcane

---

<sup>20</sup>To put these patterns into context, Figures 8 and 9 in the appendix illustrate the changes over time in the prices of sugarcane, sugar, and ethanol (and petroleum, for comparison). Figure 10 in the appendix compares changes in the relative price of sugarcane to the relative price of “other agriculture.”

<sup>21</sup>For instance, manual harvesting of sugarcane involves the pre-harvest burning of fields to ease the harvesting process by eliminating pests, etc. The smoke is an immediate health hazard.

<sup>22</sup>All wages are in 2000 Reais, and the Real/US Dollar exchange rate was approximately 1.84 Reais per US Dollar at the time.

wages and other agricultural wages.<sup>23</sup> Regions that have higher labor productivity may shift into capital more quickly as wages are bid up, enabling them to expand easily in the face of increasing demand for ethanol and sugar. Even these regions will face severe constraints to expansion eventually because of the finite supply of capital and high-quality land.

The merit of this explanation – and its implications for future sugarcane expansion – cannot be deciphered from the figures and tables alone. The model developed below is a vehicle to disentangle these and other underlying forces and predict the response to an opening of the US market to Brazilian ethanol.

### 3 Model

I cover the key aspects of the model in this section, and I leave details to the appendix. The model is static. The economy contains seven regions, detailed in Table 1 and depicted in Figure 2. Wages differ across these regional markets, but I assume all other prices are national. In each region, firms produce one of five goods: sugarcane, another agricultural good (which I will call “other agriculture”), ethanol, sugar, and a composite consumption good. Sugarcane is an intermediate good that is purchased only by ethanol producers or sugar producers. Ethanol, sugar, the other agricultural good, and the composite consumption good are all final goods, either consumed by individuals in the Brazilian economy or exported.

In the following sub-sections, I discuss each of the model’s components separately. I describe the individual utility maximization problem underlying aggregate labor supply and product demand and the parcel-specific profit maximization problem underlying aggregate supply and labor demand in the sugarcane and other agriculture sectors. In the appendix, I use these underlying problems and assumptions about the distribution of heterogeneity to produce expressions for corresponding economy-wide aggregates. All other components of the model deal directly with economy-wide aggregates. The role of the government is left to the appendix.

#### 3.1 Individual Utility Maximization

In period  $t$ , individual  $i$  chooses consumption levels, whether or not to work and – if she works – a region-sector combination  $s$  and hours of work.<sup>24</sup> The individual’s choices happen

---

<sup>23</sup>In Mincer regressions not shown here, I examine the role of crop-specific wage premia for agricultural workers. I regress the logarithm of wages on controls for year, age, education, literacy, and gender, as well as on dummy variables for specific crops and livestock. I find large wage premia in sugarcane. This finding is not driven by the fact that the PNAD takes place during the sugarcane harvest. Using data from the 2000 Census, which takes place at a different time period, I find the same feature.

<sup>24</sup>The vast majority of workers in Brazil report working in only one occupation. Using the PNAD, I find that in 2005, the share of all workers with hours in a second job is less than 5%, and less than 9% for

simultaneously, in a static setting with no uncertainty. I use a Cobb-Douglas utility function because it simplifies aggregation (see, e.g., Angus Deaton & John Muellbauer (1983) and Richard Blundell & Thomas MaCurdy (2008)). The individual's maximization problem is:

$$\max_{\ell_{it}, \bar{y}_{it}, j_{it}} \sum_s 1(j_{it} = s) [\beta \ln(\ell_{it}) + \ln(c_{it}) + \kappa_{st} + \eta_{ist}]$$

subject to  $c_{it} = (y_{eit})^{\gamma_{et}} (y_{sit})^{\gamma_{st}} (y_{ait})^{\gamma_{at}} (y_{oit})^{\gamma_{ot}} (y_{kit})^{\gamma_{kt}}$ ,  $\ell_{it} \leq T$ ,  $\ell_{it} \geq 0$ ,  $y_{zit} \geq 0 \forall z$  and

$$w_{jit} T + r_t \bar{K}_{it} + \pi_{it} - \tau_{it} = w_{jit} \ell_{it} + p_{ot} y_{oit} + p_{2t} y_{at} + p_{et} y_{eit} + p_{st} y_{sit} + r_t y_{kit}$$

The region-sector choice is denoted by  $j_{it}$  and leisure is  $\ell_{it}$ . Meanwhile,  $y_o, y_a, y_e, y_s$ , and  $y_k$  are the amount of the composite good, the other agricultural good, ethanol, sugar, and capital that the individual consumes. The price of good  $x$  is given by  $p_{xt}$ , with  $r_t$  denoting the capital price and  $p_{2t}$  denoting the price for the other agricultural good. The wage that the individual faces in the chosen market-sector,  $w_{jit}$ , is common across all individuals.<sup>25</sup> The time and capital endowments are  $T$  and  $\bar{K}_{it}$ ; the former is common across individuals but the latter is not. Land profits appear as  $\pi_{it}$  and net taxes are  $\tau_{it}$ . Denote total non-labor income as  $M_{it}$ , with  $M_{it} = r_t \bar{K}_{it} + \pi_{it} - \tau_{it}$ .

Having individuals value holding capital is unconventional. However, this is the simplest way to maintain consistency with the national income accounts in a static setting. I treat capital analogously to labor; each individual has a capital endowment and chooses to rent some of it out and hold the remainder.<sup>26</sup> Since capital does not have natural units, I assume that the aggregate capital stock in the economy moves one for one with the total number of individuals. This pins down the units of capital and the rental rate.

The remaining facet of the maximization problem is the set of items describing preferences. To begin with, the additive term  $\kappa_{st}$  gives preferences for a particular region-sector combination  $s$  that are common across individuals. I parameterize  $\kappa_{st}$  as  $\kappa_{st} = \kappa_s + \varphi_{st}$ . The  $\varphi_{st}$  terms are region-sector-year specific shocks to preferences common across individuals. Heterogeneous preferences for a region-sector combination arise in  $\eta_{ist}$ . Unemployment is also a sector, denoted with  $s = 0$ , and I treat it analogously to region-sector combinations for work (subject to the additional assumptions below).<sup>27</sup>

---

self-described employers. While these numbers are higher – and have an increasing trend over time – for agricultural workers, even the percentage for agricultural employers does not increase past about 15%.

<sup>25</sup>In practice, I use the median wage in a sector/region as the wage for that sector/region.

<sup>26</sup>See Lars A. Bovenberg, Larry H. Goulder & Mark R. Jacobsen (2008), who make the same assumption. For convenience, we can think of this as individuals holding goods for future consumption.

<sup>27</sup>For simplicity, I consider the value of unemployment to be the same regardless of where people live.

Regarding items governing the marginal utilities of consumption and leisure, I assume the share parameters  $\beta$ ,  $\gamma_{ot}$ ,  $\gamma_{et}$ ,  $\gamma_{st}$ ,  $\gamma_{at}$ , and  $\gamma_{kt}$  are common across individuals, with  $\beta + \sum_i \gamma_{it} = 1$ . I allow the  $\gamma$  parameters to vary by year, subject to year-specific shocks. In particular, I use the parameterization  $\gamma_{it} = (1 - \beta) \frac{\exp(a_i + \delta_{it})}{1 + \sum_j \exp(a_j + \delta_{jt})}$ , where the  $\delta$  are year-specific shocks to preferences and  $a_o = \delta_o = 0$  is a normalization. Together with the condition that  $0 < \beta < 1$ , this parameterization ensures that the  $\gamma$  parameters remain in the unit interval.

I treat individuals who declare themselves as employers, self-employed, or non-remunerated family workers analogously to wage laborers. Employers and the self-employed work in their chosen sector for the same wage as other hired laborers in that sector, and profits from their enterprises accrue to them as non-labor income. A producer (employer or self-employed) is indifferent as to whether a unit of labor comes from a household member or someone outside of the household. At the same time, a member of a farm household is indifferent between working inside the household or outside the household. These assumptions ensure that producers maximize profits independently of their work decisions.<sup>28</sup>

In the appendix, I present expressions for aggregate product demand based on this individual maximization problem. I also use assumptions about the distribution of heterogeneity to derive expressions for aggregate labor supply to each region and sector.

### 3.2 Parcel-Specific Profit Maximization

This section describes the portion of the model covering agricultural production and land use. Land owners use their land for sugarcane production, agricultural production, or non-agriculture.<sup>29</sup> I assume the production function for sugarcane and other agriculture on a parcel of land takes a CES form, with labor and capital as inputs:<sup>30</sup>

$$y_{jrt}(L_{jrt}, K_{jrt}; u_{jrt}) = e^{\lambda_{jrt} + u_{jrt}} ((1 - \alpha_{jrt}) K_{jrt}^{\rho_{jr}} + \alpha_{jrt} L_{jrt}^{\rho_{jr}})^{\theta_j / \rho_{jr}} \quad \text{for } j = 1, 2$$

where  $\alpha_{jrt} = \frac{1}{1 + e^{\phi_{jrt}}}$ ,  $\rho_{jr} < 1$ , and  $\theta_j < 1$ .  $L_{jrt}$  and  $K_{jrt}$  indicate the amount of annual labor hours and capital used on the parcel of land.

The amount of production a parcel can generate with a given amount of labor and capital depends on the scale term in production, which is  $e^{\lambda_{jrt} + u_{jrt}}$  in use  $j$ , where  $\lambda_{jrt}$  is a region-

<sup>28</sup>See Pranab Bardhan & Christopher Udry (1999) on “separation.” Fixed costs to working outside the household, or other labor market frictions, could undermine this assumption. Calculations with the PNAD data for households with someone working in agriculture show that after 1992, between 20 and 25% of households contain both a self-employed person/employer and a hired laborer.

<sup>29</sup>See Alberto Valdes & Johan A. Mistiaen (2003) on the limited prevalence of renting.

<sup>30</sup>Other inputs do not enter the production function due to data constraints, since information on other inputs is only available during agricultural census years. Differences in parameters across regions will in part reflect differences in usage of other inputs.

time-specific indicator of land quality that is common across all parcels in region  $r$  and  $u_{jrt}$  is the portion of land quality that is heterogeneous across parcels. This term refers to the quality of a parcel of land in a particular use.<sup>31</sup> In sugarcane production, for example, quality differences arise through differences in distance from the closest mill or distillery, availability of irrigation, weather, and land gradient. With these assumptions, it is straightforward to perform the optimization and write expressions for optimal product supply, labor demand, and profits on each parcel as a function of wages, the rental rate, and land quality.

To determine the use to which a parcel of land is allocated, the profits from the two agricultural endeavors - sugarcane ( $j = 1$ ) and other agriculture ( $j = 2$ ) - are compared to a threshold. This threshold indexes two characteristics that are monotonically and positively related to the “quality” of untouched forest land: one, the expected penalty imposed by the government for developing the land; and two, the cost of preparing the land for use in agricultural production. That is, the government will be more likely to penalize the development of the most environmentally important forest areas. These areas are also likely to have the largest cost of development.<sup>32</sup>

Let  $j = 3$  denote the non-agricultural use. The parcel is used for  $j$  if and only if  $\pi_{jrt}^* > \pi_{krt}^*$  for  $k \neq j$ , where  $\pi_{3rt}^* = e^{\lambda_{3rt} + u_{3rt}}$  and

$$\pi_{jrt}^* = \max_{L_{jrt}, K_{jrt}} e^{\kappa_{jrt}} [p_{jt} y_{jrt}(L_{jrt}, K_{jrt}; u_{jrt}) - w_{jrt} L_{jrt} - r_t K_{jrt}] \quad \text{for } j = 1, 2$$

The output price and regional wage are  $p_{jt}$  and  $w_{jrt}$ . The vector  $(\kappa_{1rt}, \kappa_{2rt})$  can be viewed as frictions in the agricultural output market; those who farm land lose a portion of the profits due to frictions. I set  $\kappa_{1rt} = 0$  since it cannot be identified, and hereafter use  $\kappa_{rt}$  to refer to other agriculture.<sup>33</sup>

This set-up necessitates additional assumptions to ensure coherence of the model. First, if the government fines an agricultural producer, I assume this fine is returned to individuals in lump-sum fashion, so that net transfers from the government are unaffected. Second, any fixed costs paid to prepare land for production are assumed to be a transfer from agricultural

---

<sup>31</sup>My approach is closely related to the approach used in Christopher Timmins (2006). The notion of “effective land units” is analogous to a Roy model for labor supply decisions. For a classic discussion, see for instance James J. Heckman & Guilherme Sedlacek (1985).

<sup>32</sup>This approach to non-agricultural land use is not fully satisfactory. A (likely small) portion of land is used for industrial development or residential housing in urban areas. Moreover, landowners sometimes hold land purely for speculative or savings purposes, and do not use it in agriculture for this reason. See, e.g., Juliano Assuncao (2006) and Gervasio Castro de Rezende (2002).

<sup>33</sup>Another way to interpret these parameters is as static optimization error: people incorrectly perceive agricultural returns or are unable to respond completely to current agricultural returns because of dynamic considerations. From a practical point of view, without the vector of  $\kappa$ 's, the model imposed too tight a relationship between shares of land and production in preliminary estimation.

producers to other individuals in the economy. This leaves total non-labor income unchanged. Third, and similarly, if  $\kappa_{rt}$  is viewed as a measure of relative “frictions”, those agricultural profits lost in transactions accrue to some individual in the economy.

Solving the profit maximization problem above indicates the use of each parcel, as a function of parcel-specific unobservables  $(u_{1rt}, u_{2rt}, u_{3rt})$ . In the appendix, I use distributional assumptions on the parcel-specific heterogeneity to aggregate over the parcels in each region and produce expressions for aggregate product supply and aggregate labor demand.

### 3.3 Non-Agricultural Good Production and Input Demand

In the production of sugar and ethanol, substitution possibilities between sugarcane and other inputs are extremely limited; a producer cannot easily economize on sugarcane by using more of other inputs. Therefore, I use Leontief production functions for sugar (s) and ethanol (e):

$$Y_{jrt} = \min \{ \alpha_{jt} Y_{1jrt}, \beta_{jt} K_{jrt} \} \quad \text{for } j = s, e$$

where  $Y_{1jrt}$  and  $K_{jrt}$  are the amounts of sugarcane and capital demanded by producers of good  $j$ .<sup>34</sup> While  $\beta_{st} = \beta_s$  for all  $t$ , I assume that  $\alpha_{st} = e^{\alpha_s + \epsilon_{st}}$ ,  $\alpha_{et} = e^{\alpha_e + \epsilon_{et}}$ , and  $\beta_{et} = e^{\beta_e + \epsilon_{ebt}}$  where  $\epsilon_{st}$ ,  $\epsilon_{et}$ , and  $\epsilon_{ebt}$  are shocks to production.<sup>35</sup>

For simplicity, I assume that production of the composite good is Cobb-Douglas, taking capital and labor as inputs. The production function for each region  $r$  is:

$$Y_{ort} = z_{ort} L_{ort}^{\beta_{ort}} K_{ort}^{1-\beta_{ort}}$$

where  $z_{ort} = e^{\phi_{or} + \epsilon_{ort}^p}$ ,  $\beta_{ort} = \frac{1}{1 + \epsilon_{or}^{\beta_{or}} + \epsilon_{ort}^b}$ , and  $\epsilon_{ort}^p$  and  $\epsilon_{ort}^b$  are region-year-specific shocks.

### 3.4 International Market

I model the interaction between Brazil and the rest of the world in a very simple way. All final goods can be exported and imported. I assume that the prices of the other agricultural good and the composite good are determined outside of Brazil.<sup>36</sup> Net exports of the other agriculture good, ethanol, sugar, and the composite good are denoted by  $Y_{at}^X$ ,  $Y_{et}^X$ ,  $Y_{st}^X$ , and

<sup>34</sup>It is not possible to clearly identify those who work in the ethanol and sugar industries prior to 2004 in the PNAD. Consequently, I omit labor from the ethanol and sugar production functions.

<sup>35</sup>Though I assume independence for simplicity, in practice the shocks to production should be correlated, partly because in some cases sugar and ethanol are produced in the same facility.

<sup>36</sup>This assumption is less controversial for the case of agriculture than for the composite good, which includes services. Brazil’s trade barriers in agriculture fell markedly in the early 1990s. Helfand (2003) illustrates how, once changes in real exchange rates are accounted for, domestic agricultural prices moved closely with international prices during the 1990s.

$Y_{ot}^X$ , respectively. I assume that the total value of net exports is zero.

For the functions determining net exports of ethanol and sugar, I rely heavily on a simplified version of the model used by Elobeid & Tokgoz (2008).<sup>37</sup> For ethanol, I assume:

$$Y_{et}^X = x_e + \beta_D(t_e p_{et} + d_e)(I^*) + \beta_p p_{pt} - \beta_S(t_e p_{et} + d_e)(I^*) + \epsilon_{xet}$$

Here, the intent is to capture net demand for Brazilian ethanol from the US, where  $\beta_D$  and  $\beta_S$  are parameters based on US demand and supply elasticities from Elobeid & Tokgoz (2008),  $t_e$  is the tariff rate of 1.025 and  $d_e$  is the duty of 0.54 US cents per gallon, converted into Brazilian Reais. Unexplained net exports are given by the shock  $\epsilon_{xet}$ .  $I^*$  takes a value of 1 if net exports are positive, and zero otherwise. If  $I^* = 1$ , then the import restriction-based world ethanol price is low enough to be relevant in the US. I estimate  $x_e$  and  $\beta_p$ .

For sugar, I assume:

$$Y_{st}^X = x_s + \alpha_D p_{st} - \alpha_S p_{st} + \epsilon_{xst}$$

where  $\alpha_D$  governs the price response of international demand for sugar and  $\alpha_S$  governs the price response of international supply of sugar. For these values, I used a value in the range of the relevant elasticities in Elobeid & Tokgoz (2008) for the largest sugar producing and consuming countries. Here,  $\epsilon_{xst}$  is a shock and  $x_s$  is a parameter to be estimated.

### 3.5 Equilibrium

These components come together to form a market equilibrium. The only “exogenous” variables are: prices of the other agricultural good, the composite good, and petroleum; total population; total capital stock; government tax revenue; and the total amount of land in each region. All other variables are determined endogenously, based on the exogenous variables, the parameters, and the values of the shocks. I define equilibrium formally in the appendix; it involves market-clearing in goods, labor, and capital markets, zero-profit conditions, an economy-wide budget constraint, and zero total net exports.

## 4 Estimation

Next, I move to estimation. Dale W. Jorgenson (1998a) and Dale W. Jorgenson (1998b) provide an overview of estimation of large-scale macroeconomic models. Here, I use individual and aggregate data to estimate a model derived from underlying heterogeneity of individual preferences and land quality. In this way, this paper is more related to the relatively recent

---

<sup>37</sup>Details about the choices of coefficients are available on request.

work on dynamic general equilibrium models with individual labor supply and human capital decisions (see seminal papers such as James J. Heckman, Lance Lochner & Christopher Taber (1998) and Donghoon Lee & Kenneth I. Wolpin (2006)). I restrict myself to a static setting to explore cross-region patterns and deal with data limitations. In exchange for my simpler setting, I incorporate an intensive margin of work hours, give an economic role to all shocks, and use an estimation approach that guarantees the existence and local uniqueness of a vector of production and preference shocks that produces the equilibrium quantities in each year of the data.

The last feature relies on my use of maximum likelihood. Most existing work on empirical general equilibrium models uses SMM or GMM methods, due to problems with MLE (See, e.g., Heckman, Lochner & Taber (1998), Lee & Wolpin (2006)). However, I pursue the MLE approach for four reasons. First, maximum likelihood ensures that all the constraints implied by the equilibrium system are accounted for. It would be possible to implement all these constraints in a more agnostic method of moments approach, but it would require a great deal of caution. Second, using GMM and SMM does not address the more fundamental economic problem of existence and uniqueness of equilibria. With SMM, for example, we know that at all the error draws used in the calculation of the moments, one has an equilibrium. However, this does not guarantee the existence (or local uniqueness) of an error draw that can result in each equilibrium we see in the data. Third, for the simulations I need to know what distribution to draw the errors from. Even with GMM, I would then be forced to make stringent distributional assumptions. Fourth, MLE dictates exactly what moments to use, while GMM and SMM would force me to choose moments somewhat arbitrarily. To be sure, in small samples it is unclear if the MLE moments are any “better” than the GMM/SMM moments, even assuming the distributional assumptions hold.

MLE has its own obstacles.<sup>38</sup> The small number of observations force me to make stringent assumptions about the independence of unobservables to secure identification.<sup>39</sup> Another problem is that there is no guarantee that every possible combination of the error terms allowed by the normality assumptions has an equilibrium for the endogenous variables associated with it. Even if a solution exists, there is no guarantee that it must be unique. (For a brief discussion of these issues, see for instance Takeshi Amemiya (1985)). Therefore, the normality assumptions do not capture restrictions on the possible distribution of the

---

<sup>38</sup>Ray Fair & William Parke (1980) compare MLE to non-linear 3SLS and 2SLS. See also Ray Fair & John Taylor (1983).

<sup>39</sup>In a linear FIML model without constraints on the contemporaneous correlation of the error terms, identification is impossible if the number of observations is smaller than the sum of the number of endogenous and exogenous variables in the model (see J.D. Sargan (1975)). The independence assumptions below help substantially, turning portions of the problem into simple OLS.

error terms that are implicit in the model.

Keeping these limitations in mind, the following sub-sections describe my MLE approach, and present estimates of the model.

## 4.1 Estimation Procedure

Here, I outline the estimation procedure. Theoretical identification of the labor supply parameters using only the aggregate data, if even possible, would rely heavily on the nonlinearities of the labor supply functions. Given the small sample size, identification using only the aggregate data would be practically difficult regardless. Therefore, I proceed with a two-step estimation procedure. First, I estimate the labor supply parameters using an individual-based likelihood function. Second, I maximize a likelihood function that comes from writing down the density of the aggregate endogenous variables. I outline these steps below. In the appendix, I specify my distributional assumptions, explicitly state the parameters to be estimated, and present the components of the likelihood functions in detail.

### 4.1.1 Likelihood of Individual Labor Supply Decisions

For each individual, we potentially observe four choices: whether or not they work, the region of work, the sector of work, and hours. The last three are observed only for those individuals who work. The likelihood function conditions on the vector of region-sector wages  $\vec{w}_{jr}$  and national-level non-labor income  $M_t$ . Let  $d_{it} = 1$  if individual  $i$  works, and equal zero otherwise. The likelihood function is:

$$L = \prod_t \prod_{i=1}^{N_t} [Pr(d_{it} = 0 | \vec{w}_{jr}, M_t)]^{1-d_{it}} \prod_{r,j} [L_{it}(d_{it} = 1, h_{ijrt}, s_{it} = (r, j) | \vec{w}_{jr}, M_t)]^{d_{it} 1(s_{it}=(r,j))}$$

where  $h_{ijrt}$  is the hours choice and  $s_{it}$  is the region-sector choice, with  $s_{it}$  taking a value  $(r, j)$ , for  $r = 1, \dots, 7$  and  $j = 1, 2, 3$ . Here,  $1(s_{it} = (r, j))$  is an indicator function taking a value of one if the chosen region-sector is  $(r, j)$ . The expression for  $L_{it}(d_{it} = 1, h_{ijrt}, s_{it} = (r, j) | \vec{w}_{jr}, M_t)$  and details of the likelihood function appear in the appendix.

I take two steps to ensure the coherence of the individual-level estimates with the aggregate economy.<sup>40</sup> First, I ensure that in any given year, the expected value of non-labor income will equal per-capita non-labor income in the aggregate data. Second, I constrain

---

<sup>40</sup>There is an alternative approach to estimation that can avoid the separation of the individual estimation from the aggregate estimation. The procedure is related to the ‘‘Micro-BLP’’ approach, presented in Steven Berry, James Levinsohn & Ariel Pakes (2004). Unlike in the Micro-BLP case, there is no clear guarantee that one can find a parameter vector such that all the non-linear constraints are satisfied.

the  $\varphi$  preference shifters to approximately ensure that the analytical expression for aggregate labor supply to any region-sector combination is equal to aggregate labor hours in that region-sector in the data. In the constraints, I use a first-order Taylor expansion of the labor supply expressions as a function of wages, non-labor income, and the shocks  $\varphi_{jrt}$ , expanded about region-sector-specific average wages, average per capita non-labor income over the ten year period, and a value of zero for all the shocks. From these expressions, I can solve for the  $\varphi_{jrt}$  and substitute these back into the likelihood function. <sup>41</sup>

I maximize the resulting likelihood function using simplifications for computational purposes. First, the probability of being unemployed is an integral over non-labor income. I numerically approximate this integral using Gauss-Hermite quadrature with 20 nodes (See, e.g., Ken Judd (1998)). Second, for certain values of non-labor income, the probability of being in particular region/sector combinations is zero (due to the zero hours constraint). This leads to a discontinuity in the likelihood function. To smooth the likelihood function, I multiply the value of being in each region-sector by  $\Phi(h/0.1)$ , where  $h$  gives desired hours of work.  $\Phi(h/0.1)$  is close to 1 when  $h > 0$ , and close to zero when  $h < 0$ .<sup>42</sup>

#### 4.1.2 Likelihood Function for Aggregate Quantities

Let the notation  $\vec{Z}_r = (Z_1, Z_2, \dots)$  represent the vector of variables  $Z$  for all seven regions, and leave the subscripts “t” implicit. Let  $K^{DL}$  be the product of the rental rate and capital demanded by consumers. Define  $\vec{V}_{2t} \equiv (\vec{L}_{1r}, \vec{L}_{2r}, \vec{L}_{3r})$ ,  $\psi_{2t} \equiv (\vec{\varphi}_{1r}, \vec{\varphi}_{2r}, \vec{\varphi}_{3r})$  and

$$\begin{aligned}\vec{V}_{1t} &\equiv (\vec{Y}_{1r}, \vec{Y}_{2r}, \vec{S}_{1r}, \vec{S}_{2r}, \vec{Y}_{or}, Y_e, M, p_1, p_e, Y_s^X, Y_a^X, K^{DL}, \vec{w}_{1r}, \vec{w}_{2r}, \vec{w}_{3r}, Y_e^X, p_s) \\ \psi_{1t} &\equiv (\vec{\epsilon}_{1r}, \vec{\nu}_{1r}, \vec{\epsilon}_{2r}, \vec{\nu}_{2r}, \vec{\nu}_{3r}, \vec{\nu}_{kr}, \vec{\epsilon}_{or}^b, \vec{\epsilon}_{or}^p, \epsilon_s, \epsilon_{e1}, \epsilon_{e2}, \delta_a, \delta_e, \delta_s, \delta_k, \epsilon_{xe}, \epsilon_{xs})\end{aligned}$$

Note that the vector of endogenous variables  $\vec{V}_t = (\vec{V}_1, \vec{V}_2)$  has the same length (86) as the vector of errors  $\vec{\psi}_t = (\vec{\psi}_{1t}, \vec{\psi}_{2t})$ .

The joint density of the endogenous variables  $\vec{V}_t$  is:

$$f(\vec{V}_{1t}, \vec{V}_{2t}) = g(\vec{\psi}_{1t}(\vec{V}_{1t}, \vec{V}_{2t}))g(\vec{\psi}_{2t}(\vec{V}_{1t}, \vec{V}_{2t}))\det \begin{pmatrix} \frac{\partial \psi_1}{\partial V_1} & \frac{\partial \psi_1}{\partial V_2} \\ \frac{\partial \psi_2}{\partial V_1} & \frac{\partial \psi_2}{\partial V_2} \end{pmatrix}$$

since  $\vec{\psi}_1$  and  $\vec{\psi}_2$  are independent. In fact, the errors within the vectors  $\vec{\psi}_1$  and  $\vec{\psi}_2$  are

<sup>41</sup>During estimation, I constrain the average of the  $\varphi_{jrt}$  to be close to zero (within 0.1 of zero) for each of the three sectors. This ensures that the linearization around shock values of zero is an adequate approximation.

<sup>42</sup>Finally, in order to ease identification in practice, I fix  $\kappa_{02} = 6$ , a value chosen based on preliminary estimation. This parameter is explained in the appendix.

independent as well, which breaks apart this density even further.

The next step in constructing the likelihood function involves finding expressions for the error vectors  $\vec{\psi}_1$  and  $\vec{\psi}_2$  in terms of the endogenous variables. These expressions can then be substituted into the densities above in the change of variables formula. There are convenient expressions for  $\vec{\psi}_1$ , and I present these in the appendix.

Unfortunately, there are not analogous, simple analytical expressions for  $\vec{\psi}_2$ . Instead, recalling that the elements of  $\vec{\psi}_2$  will appear in the preference shifters  $\kappa$  (as described in the appendix), I linearize the labor supply functions, using the same expressions as those used in the constraints for the individual-based estimation above. Using this linearization, I can solve for  $\vec{\psi}_2$  and finish the construction of the likelihood.

I impose one additional constraint motivated by preliminary estimation. I constrain the elasticity of substitution for sugarcane and other agriculture to be no higher than 2. In preliminary estimation, I allowed the elasticity of substitution to be unconstrained, and found that the maximum typically occurred with values on the order of 15 or more for sugarcane and other agriculture. These estimates seemed implausibly high, necessitating a constraint. The choice of an upper bound of 2 was designed to allow the production function to take a Cobb-Douglas form; however, the maximization procedure suggests the data are more consistent with an elasticity of substitution greater than 1, as the estimates hit the upper bound for both sugarcane and other agriculture.

## 4.2 Estimates of the Model

Tables 5-8 show the estimates of the model. I do not show standard errors; while asymptotic standard errors would be meaningful for the estimates of the labor supply parameters – which rely on a large number of individual-level observations from the PNAD data – they would be very misleading for the majority of parameter estimates. These latter parameters are estimated from only 10 years of data, necessitating small sample standard errors not presented here.

To begin with, Table 5 displays estimates of the parameters affecting labor supply, except for the region-sector preference shifters. The estimates of the non-labor income distribution parameters refer to non-labor income for one week, rather than one year. About 59% of people fall into the class of people with  $\theta = \theta_1$ . Given the estimate of  $\sigma_1$ , the average weekly non-labor income in this class is about 240 Reais (less than US \$150). The non-labor income in the other category is much lower, about 26 Reais. The optimization procedure results in estimates of  $d_{1|l}$  and  $d_{1|h}$  that hit their bounds. Therefore, conditional on having  $\theta = \theta_2$ , a person is never in the work force. Conversely, conditional on being a high income type, the probability of being in the work force is 1. This is plausible, considering that all workers 15

and older are included in the sample. A large fraction may be non-working students with low non-labor income.

Table 6 presents the remaining parameters determining labor supply choices. The table contains estimates of the sum of  $\kappa_{jr}$  and  $z_{jr}$ ,<sup>43</sup> which indicate the non-time varying portion of preferences for a region-sector, as well as the variances of the  $\varphi_{jrt}$  shocks. Conditional on identical wages in all sectors, the sugarcane sector is much less desirable than the other sectors. Non-agriculture is the most preferred sector within the Center-South regions, while other agriculture is more preferred in the more rural Northeast. Using the linearized expression for aggregate labor supply into each region-sector, at 2005 values of labor hours and wages in Pernambuco, Sao Paulo, Parana, and Minas Gerais the own wage elasticity of labor hours is higher in sugarcane than in other agriculture, ranging from 2.2 to 5.1. In every region but Maranhao, the own wage elasticity for sugarcane is higher than that in non-agriculture. Of course, the linearization is in terms of wage levels and non-labor income levels, so this will change quickly as labor hours are drawn into sugarcane during an expansion. The preference parameters – in combination with the parameters describing the non-labor income distribution – do, however, suggest that in every region except Bahia and Maranhao, the elasticity of labor hours with respect to weekly non-labor income is more negative in sugarcane than in non-agriculture.<sup>44</sup>

Estimates of the parameters common to all regions and using the aggregate data appear in Table 7. As noted before, the elasticity of substitution in both sugarcane and agriculture hits the upper bound of the parameter space considered.<sup>45</sup> This is in contrast to the case of Reinaldo Alcantara & Anthony Prato (1973), who assume generalized Cobb-Douglas production and use farm-level data in Sao Paulo to estimate returns to scale and input elasticities.<sup>46</sup> Given the increase in mechanization over time, it is plausible that substitution possibilities are now greater than those reflected in the authors' pre-1973 data. The estimates of  $\theta_j$  suggest that a slightly larger share of revenue goes to profit in other agriculture than in sugarcane. Here,  $\theta_1$  helps determine the elasticity of production with respect to labor use on any given parcel. In fact, it is an upper bound on this elasticity. This value of 0.746 is

---

<sup>43</sup>Note that without the linearization, it is not possible to distinguish  $z_{jr}$  from  $\kappa_{jr}$ . The linearization allows us to obtain estimates of the  $z_{jr}$ , which can be thought of as a measure of the inaccuracy of the linear approximations.

<sup>44</sup>These elasticities vary from a low of -4.5 to a high of -0.81 in sugarcane. In non-agriculture, the highest elasticity is in Sao Paulo, the only positive value. This is possible because, based on the estimates, some individuals will always work and this region-sector is the most desirable to move into. This effect is enough to counteract the negative effect coming through the intensive margin.

<sup>45</sup>Preliminary estimation indicated that the same would be true if the elasticity of substitution was permitted to differ across regions for sugarcane.

<sup>46</sup>The returns to scale estimates do not directly speak to the ones here, given that the authors treat land as a chosen input.

higher than the upper end of the elasticity estimated by Alcantara & Prato (1973), 0.513. The value of  $\gamma$  indicates a large amount of variance in land quality within each region. The remaining agricultural parameters describe the variance of shocks to land or productivity.

The next rows of estimates deal with non-agricultural production. For ethanol and sugar, the estimates of  $\alpha_e$  and  $\alpha_s$  suggest that each metric ton of sugarcane can produce slightly more metric tons of sugar than cubic meters of ethanol at a production shock of zero for both parameters. The estimates of  $\beta_e$  and  $\beta_s$  imply that each unit of capital can produce more sugar than ethanol. As might be expected, the ratios of  $\alpha$  to  $\beta$  are similar across ethanol and sugar. Finally, the last rows relate to aggregate consumption, and show that people on average spend their budget shares in highest to lowest order on capital, the other agricultural good, ethanol, and sugar.

Table 8 shows the estimates of the remaining parameters, which vary by region. The first two rows indicate parameters governing the marginal product of labor and capital, the “share” parameters in the CES functions. Transforming the parameters using the expressions for the shares, and setting production shocks to zero, suggests that for sugarcane, the multiplier on labor in production varies from a low of 0.045 to a high of 0.110. The third row, for  $\lambda_{1r}$ , shows that Sao Paulo and Pernambuco have the highest estimates for sugarcane land quality, followed by the remaining regions of the South-Southeast, Parana and Minas Gerais. Examining the variance of the land quality parameters,  $\sigma_{v_1}^2$  indicates relatively high variance to land quality. This is particularly true of Maranhao, where the variance is high enough to create questions about how well the model fits the data for this region. The estimates of  $\kappa$  suggest that all regions have a bias toward other agricultural land. In regards to non-agricultural land, the highest estimated values are found in Maranhao, Bahia, and Mato Grosso (row 5). The first and third regions contain portions of the Amazon Rainforest.

To gain an idea of the implications of these estimates, one can examine key elasticities in sugarcane. The elasticity of sugarcane production in each region with respect to the price, holding wages and other prices constant and holding the value of all production shocks constant at 2005 values, is just under 2.99 for all regions. There are slight differences across regions, but this effect is dominated by the estimated values for  $\theta_1$ , which is common across regions. The same is true for the elasticity of aggregate labor demand with respect to the wage, holding all else constant; this elasticity is approximately 3.86 for all regions.

## 5 Simulations

Using these estimates, I simulate the effect of three changes. In the first, I eliminate the tariff and duty. In the second, I eliminate the restrictions and also increase the other agricultural

price by 20%. In the third, I do the same as the second simulation, and also increase the world petroleum price by 20% (this corresponds to an oil price in 2000 dollars of about \$60 per barrel). During these simulations, I hold all exogenous variables constant at their 2005 values. The first sub-section discusses further details behind the simulation methods. The second sub-section covers the simulation results.

## 5.1 Simulation Methods

The simulations require special considerations that are not conventional in structural models. Conditional on the exogenous variables, any vector of equilibrium quantities is a function of the random vector of production and preference shocks. Therefore, the impact of any policy change will depend on the random vector of shocks. I approach this issue in two ways.

First, I examine the impacts on equilibrium outcomes when holding the production shocks and preference shocks constant at their 2005 values. Equilibria exist for all three policy simulations, and can be compared with the baseline (the actual 2005 realizations). This simulation approach allows us to focus on the mechanisms underlying the simulated changes without being confused by the randomness associated with economic shocks.

Second, I allow for uncertainty in policy predictions by examining the distribution of equilibrium outcomes for ethanol production, sugarcane production, ethanol exports, and regional non-agricultural land allocations. FIML guarantees the existence of a unique equilibrium in the neighborhood of the 2005 values for the error terms. While it is not certain, a small disturbance in the system – through a change in import barriers, agricultural prices, or petroleum prices – may preserve the existence of equilibrium. While there is no guarantee that a unique equilibrium exists for any given random draw of shocks, the closer the error terms are to their 2005 values, the better should be the chance of finding an equilibrium.

Accordingly, I fix the error terms for composite good production, ethanol and sugar production, ethanol and sugar exports, and aggregate good consumption at their 2005 values. I take 300 random error draws for all the agricultural shocks and all the labor supply region-sector preference shocks. I leave these random draws constant across the four cases – i.e., the baseline (no policy or price change) and the three simulated environments. In each case, for each error draw vector, I find the associated equilibrium if it exists. In practice, an equilibrium existed for all error draws for all four cases. The result is 300 vectors of equilibrium outcomes for each case. For four key sets of variables – ethanol production, sugarcane production, ethanol exports, and regional non-agricultural land allocations – I then illustrate the distribution of outcomes using box plots.<sup>47</sup>

---

<sup>47</sup>Throughout all the simulations, I confine myself to equilibria that have wages bounded above by the value of 6 Reais and that have land shares bounded below by 0.0001.

## 5.2 Simulations of Alternative Policy Regimes

Tables 9 to 12 show the results for key quantities when all error draws are fixed at 2005 values, while Figures 4 and 5 illustrate the distribution of outcomes when a subset of the error draws are randomly drawn. Throughout, I refer to the first set of simulations (removing the tariff and duty) as “Sim 1,” the second set (removing barriers and increasing the other agricultural price by 20%) as “Sim 2,” and the third set (doing the same as Sim 2, but also increasing the petroleum price by 20%) as “Sim 3.” I refer to the baseline (no policy or price change) simulation results as “2005” or “Base,” depending on whether the error draws in agricultural and labor preferences are fixed at 2005 values or randomly drawn.

### 5.2.1 Ethanol Production and Exports Increase Substantially

Turning first to Tables 9 and 10, we see that in all three alternative regimes, there is substantial expansion of ethanol production and exports relative to the 2005 baseline. The removal of the tariff and duty leads to a sizeable increase in the amount of ethanol produced, which more than doubles to 8.8 billion gallons. This allows ethanol exports to jump from a baseline level of 685 million gallons to over 5.5 billion gallons. To gauge the size of this number, note that in 2007, total US production of ethanol was 6.5 billion gallons. Correspondingly, the ethanol price increases by about 19%. These results differ from those of Elobeid & Tokgoz (2008), who predict a world ethanol price change of about 24%, and an increase in ethanol exports of only 730 million gallons.

Figure 4 shows these increases are not due to the particular realization of production shocks in 2005. The upper and lower values of each box show the 25th and 75th percentiles, the line inside each box indicates the median, the whiskers outside of the box indicate the adjacent values, and dots indicate values outside the adjacent values. Looking at the third panel of the figure, we see that the median level of exports from “Sim 1” is between three and four billion gallons of ethanol, with the upper adjacent value above five billion.

The results for “Sim 2” and “Sim 3” show ethanol production and exports moving in the expected directions. The projected increases are similar, though slightly smaller, for the regime that additionally includes an increase in the price of the other agricultural product. When international ethanol demand is spurred by the 20% increase in the world petroleum price, ethanol production and exports jump substantially. Examining the fourth row of Tables 9-10, we see that ethanol production and exports are predicted to rise to 19 billion and 16 billion gallons, respectively. An export total of 16 billion gallons would make a sizeable impact on the US fuel mix. As noted in the introduction, with 13 billion gallons of ethanol the US could have used ten percent ethanol blends in all of its 2007 gasoline

consumption.

The box plots in Figure 4 show that there is a substantial amount of uncertainty about the size of the production and export increases. Still, the minimum value of ethanol production is around 13 billion gallons, and the minimum value of ethanol exports is about 10 billion gallons. The brunt of the ethanol export distribution is in the 14-16 billion gallon range. These predictions suggest that, regardless of the exact size of the export increase, Brazil could have a significant impact on the shift to low-carbon fuels in the US.

### **5.2.2 Less Sugar and More Sugarcane Enable Greater Ethanol Exports**

This substantial increase in ethanol exports comes from the movement of sugarcane away from sugar production and into ethanol production, as well as an increase in total sugarcane production. The third column of Table 9 illustrates the massive decline in sugar production upon removal of the barriers. Sugar production falls from 26.6 million metric tons (MT) in 2005 to 12.62 million MT in “Sim 1,” and even lower in the other simulation regimes. Net exports of sugar also fall dramatically. As we would expect, the price of sugar rises, specifically around 21% in moving to “Sim 1.” This is a much larger rise in world sugar prices than seen in Elobeid & Tokgoz (2008), who predict a rise of about 1.8%. While the simplicity of my model of international sugar markets could explain part of this, it is also the case that the larger ethanol export boom I predict necessitates more sugarcane, and this can be easily drawn from sugar when sugarcane production is somewhat inelastic.

The increase in ethanol exports also comes from an increase in total sugarcane production. The first column of Table 9 shows that sugarcane production jumps from 422 million MT in the baseline to 573, 546 and 1073 million MT in the alternative regimes. In the first panel of Figure 4, we see a more nuanced story. In “Sim 1” and “Sim 2,” just a portion of the distribution lies above the level of 2005 production. In the case with petroleum price increases, the entire distribution of sugarcane outcomes moves dramatically higher.

The results in the first panel of Table 11 suggest the importance of extensification in sugarcane production. Table 11 shows the millions of acres allocated to each use under each regime, for all seven regions. In moving from 2005 to the removal of import barriers, sugarcane acreages increase in all regions except Maranhao. Total sugarcane land increases more than 11 million acres to 26 million acres. The largest absolute changes occur in Sao Paulo. The increases in Mato Grosso – despite the low land quality in sugarcane, and the strong preference not to work in sugarcane there – are also substantial. As we move to the third regime with higher oil prices, the sugarcane land allocation changes continue to be largest in Sao Paulo. There appear to be important non-linearities in the yields on sugarcane land. While the first 14.3 million acres are associated with production of 422 million MT,

the next 11 million acres are only associated with production increases of about 150 million MT. The move from there to “Sim 3” brings in 10 million more acres and produces 500 million MT more sugarcane. Thus, in “Sim 3” sugarcane yields again resemble yields in the 2005 baseline.

These non-linearities suggest that either the average quality of land used in sugarcane production or the intensity of input use in sugarcane production are changing across the simulated regimes. Table 16 in the appendix speaks to this. The first panel of the table shows predicted changes in annual labor hours in each region across the alternative regimes. Going from the 2005 baseline to “Sim 1,” the intensity of labor use on land intensifies in Bahia. However, in all other cases, the intensity of labor use actually falls. Looking at the change from the 2005 baseline to “Sim 3,” in every region labor intensity falls. These declines are sometimes too large to be plausible (new values of labor per unit land are sometimes one-tenth of the original value, for instance), suggesting that the model should be re-visited in the future. In any case, the results suggest a shift to more capital-intensive agriculture, and this is not always enough to keep yields stable across regimes.

### **5.2.3 Consequences for Land Clearing are Non-Linear**

What are the ultimate implications of the sugarcane expansions for the clearing of forest and savanna land? As shown in Table 12 and Figure 5 for “Sim 1,” the initial burst of additional ethanol exports come at a cost of little or no non-agricultural land. Columns 5 and 6 of Table 11 show that land for other agriculture not only absorbs the sugarcane land increase, but falls quite steeply. The two areas with the greatest amount of Amazon Rainforest in the regions I consider, Mato Grosso and Maranhao, experience declines of 20 million acres and 55 million acres, respectively. Much of this land gets translated into greater non-agricultural land. The results in Figure 5 suggest that – while the direction and magnitudes show a substantial amount of uncertainty – Mato Grosso and Maranhao are unlikely to see large decreases in non-agricultural land.

The reason for this is apparent in columns 5 and 6 of Table 15 (appendix): In several of the regions, wages in other agriculture increase as sugarcane wages go up. While other agricultural wages in Mato Grosso actually fall slightly, we should keep in mind that increased demand for capital stoked by the sugarcane expansion results in an increased price of capital. This is especially relevant to Mato Grosso, where the ratio of capital to labor is estimated to be relatively high. We should be wary about the size of the projected changes in Table 11 and Table 15, but these changes do suggest an important point to keep in mind: Changes in input prices spurred by sugarcane growth can limit the profitability of other agriculture in areas where it was originally profitable.

Of course, prices of key agricultural products in Brazil have been trending upwards, and one may be wary of relying too heavily on simulations using 2005 prices. Therefore, we turn to “Sim 2,” which removes ethanol trade barriers and increases the price of other agriculture by 20%. This particular percentage is meant to be illustrative. It is also in line with the 25% increase in the price of maize assumed by Nelson & Robertson (2008). Turning first to the 2005 results, Table 9 shows that other agricultural production rebounds, settling at a level slightly above the 2005 baseline. This increase comes from an increase in other agricultural land in every region, as seen in Table 11, and a corresponding decrease in non-agricultural land from the 2005 baseline. The large decreases of 13 million acres and 8 million acres in Mato Grosso and Maranhao, respectively, should be cause for concern; in addition, there are decreases in regions containing remnants of the Atlantic Rainforest, with the steepest decrease occurring in Bahia. Figure 5 corroborates these results. In all cases, the 25th, 50th, and 75th percentiles of the non-agricultural land distribution move down when going from the baseline to “Sim 2.”

Finally, it is important to examine the effect of increased petroleum prices, which will lead to an outward shift of international ethanol demand. When the production and preference shocks are held constant at 2005 values, this change leads an increase of about 40 million acres of total land for other agriculture, relative to “Sim 2” values. Nearly 70 million acres are added in Mato Grosso, with this being offset by declines in other regions. Table 12 shows non-agricultural land falls sharply, by 116 million acres from the baseline and 57 million acres from “Sim 2.” By far, the largest declines from the 2005 baseline are in Mato Grosso and Maranhao. Parana and Sao Paulo see predicted declines of 7-9 million acres, and Bahia sees one of 16 million acres. Figure 5 shows the distribution of outcomes in Mato Grosso and Maranhao is quite wide. The “Sim 3” distributions exhibit the greatest uncertainty in the bottom panel, for the Bahia, Pernambuco, and Maranhao regions.

These results suggest that the growth in sugarcane and other agricultural land in “Sim 3” is likely to come at the expense of significant non-agricultural land. It is crucial to understand the reasons for the great expansion in other agricultural land, particularly in Mato Grosso. Table 16 (appendix) shows the increase in Mato Grosso is driven in part by a large increase in labor supply, which pushes down wages in the region relative to “Sim 2.” Where is this expansion in labor supply coming from? Using the table on annual hours, it appears that there are massive changes in hours predicted for non-agriculture in Minas Gerais and Maranhao. The size of these changes is very large. While the model is not designed to capture changes in hours and wages in non-agriculture, this warrants further investigation in the future.

Overall, the evidence suggests there are important non-linearities in the trade-off between

ethanol exports and non-agricultural land. Brazil could increase exports in response to the policy change by over 5 billion gallons, with little or no cost in terms of non-agricultural land. The additional exports are supported by diversion of sugarcane into sugar and moderate growth in sugarcane land, with input price increases throughout agriculture helping to limit agricultural expansion. Much more non-agricultural land could be displaced by further increases in world ethanol demand in the free trade environment. At 2005 values, the next 10.5 billion gallons of exports comes at the cost of over 100 million acres of non-agricultural land, with some of the largest changes in those regions containing the Amazon Rainforest.

## 6 Conclusion

The possibility of lowering US import barriers for fuel ethanol presents a crucial tradeoff for the environment. Freeing the US ethanol market could encourage the replacement of petroleum and corn ethanol with low-cost and relatively low-carbon sugarcane ethanol. Yet the changes caused by increased demand for Brazilian ethanol could lead to increased conversion of Brazil's forest and savanna land into agriculture, affecting carbon sequestration, bio-diversity, and eco-system functioning. This paper addresses this tradeoff with an estimable general equilibrium model of regional agricultural markets, one accounting for the linkages between sugarcane and other agricultural production decisions through labor, land, and capital inputs. In three simulation exercises, I use the parameters of the model to predict the consequences of removing US import restrictions. The removal is predicted to lead to a large increase in Brazilian exports to the US at little or no cost to non-agricultural land in Brazil. The need to pull more factors of production into sugarcane and wage spillovers into other agriculture serve as natural brakes on the pace of agricultural land expansion. The situation is much more worrisome in the final simulation exercise, where the price of other agriculture and the price of petroleum are both increased by 20% relative to their 2005 values. In this scenario, I predict a massive, additional increase in Brazilian exports, one that leads to significant loss of non-agricultural land throughout Brazil.

While the framework presented here provides a new perspective on the debate over Brazilian ethanol, it is limited in four ways. First, the model defines regions very broadly. Consequently, a predicted decline in non-agricultural land in the Mato Grosso region may or may not come from the Amazon, just as a predicted decline in Parana may or may not come from the Atlantic Rainforest. A more detailed and disaggregated model of land use is required to make more accurate predictions.

Second, agricultural decision-making is dynamic, and this is particularly true with sugarcane. Sugarcane has a long growth cycle, and mature sugarcane can be harvested multiple

times before re-planting is necessary. In pursuit of tractability, the model uses short-term changes in prices to say something about production responses without adequately accounting for production response lags, fixed costs of converting land back and forth between sugarcane use, and uncertainty about future prices. For instance, OLS regressions of regional sugarcane land shares on prices, wages and lagged shares (not shown here) suggest a substantial amount of inertia in land allocations over time.

Third, behavioral responses not modeled here could matter. For instance, the Brazilian government could use policy to ensure that any declines in non-agricultural land occurs in the least environmentally sensitive portions of each region. Moreover, the compliance decision among private landowners is a potentially significant contributor to the total amount of forest and savanna land in Brazil. As noted in the introduction, I have essentially ignored private forest land and instead focused on land that is not used for any sort of agricultural production. It could be the case that the forest land set aside by private landowners and used for casual grazing, planted forests, etc., preserves many of the ecosystem services expected of virgin forest land. In this case, it would be important to consider whether changes in existing private land enhance or compensate for declines in public non-agricultural land.

Fourth, data limitations may affect the results. On the labor supply side, information for the North Census region is not available for the entire period of analysis. Moreover, aggregate hours in sugarcane – and median wages in the sector – are estimated using household survey data. Given the relatively small number of people in the sector, sampling variation may lead to variation in these quantities that is not real variation, which would affect the estimation of the model’s parameters. In terms of land use, measurement error may also affect the values for non-sugarcane agricultural land, which are currently imputed in between Agricultural Census years. The availability of repeated cross-sections of satellite data in the future may remedy this difficulty.

In future work, it will be important to take these considerations into account, and the answers presented here should be viewed as preliminary. In considering these factors, though, future models could build on the results here and incorporate the linkages between land use decisions, input markets and product markets, in a setting that allows for estimation of key parameters in a coherent, unified framework.

## References

- Alcantara, Reinaldo, and Anthony Prato.** 1973. “Returns to Scale and Input Elasticities for Sugarcane: The Case of Sao Paulo, Brazil.” *American Journal of Agricultural Economics*, 55: 577–583.

- Alix-Garcia, Jennifer.** 2007. "A Spatial Analysis of Common Property Deforestation." *Journal of Environmental Economics and Management*, 53(2): 141–157.
- Amemiya, Takeshi.** 1985. *Advanced Econometrics*. Cambridge, Massachusetts:Harvard University Press.
- Assuncao, Juliano.** 2006. "Land Reform and Landholdings in Brazil." *UNU-WIDER Research Paper*, 2006/137.
- Banerjee, Onil, Alexander Macpherson, and Janaki Alavalapati.** Forthcoming. "Socioeconomic and Land Use Trade-Offs of Ethanol Expansion in Brazil: A Computable General Equilibrium Model." *Journal of Sustainable Forestry*.
- Bardhan, Pranab, and Christopher Udry.** 1999. *Development Microeconomics*. New York:Oxford University Press.
- Barros, Alexandre Rands, Tiago V. de V. Cavalcanti, Fernando Mendonca Dias, and Andre Matos Magalhaes.** 2005. "The Impact of Sugar's World Trade Liberalization on Poverty in Brazil." Unpublished manuscript.
- Barros, Geraldo Sant'ana De Camargo, and Marcia Azanha Ferraz Dias De Moraes.** 2002. "A Desregulamentacao do Setor Sucroalcooleiro." *Revista de Economica Politica*, 22(2-86): 156–173.
- Berry, Steven, James Levinsohn, and Ariel Pakes.** 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy*, 112(1): 68–105.
- Blundell, Richard, and Thomas MaCurdy.** 2008. "Labour Supply." In *The New Palgrave Dictionary of Economics*, ed. Steven M. Durlauf and Lawrence E. Blume. Palgrave Macmillan.
- Borrell, Brent, Jose R. Bianco, and Malcolm D. Bale.** 1994. "Brazil's Sugarcane Sector: A Case of Lost Opportunity." *World Bank Policy Research Working Paper*, 1363.
- Bovenberg, Lars A., Larry H. Goulder, and Mark R. Jacobsen.** 2008. "Costs of Alternative Environmental Policy Instruments in the Presence of Industry Compensation Requirements." *Journal of Public Economics*, 92: 1236–1253. Working Paper.

- Bussolo, Maurizio, Jann Lay, and Dominique van der Mensbrugge.** 2006. "Structural Change and Poverty Reduction in Brazil: The Impact of the Doha Round." World Bank Policy Research Working Paper 3833.
- Deaton, Angus, and John Muellbauer.** 1983. *Economics and Consumer Behavior*. New York:Cambridge University Press.
- de Rezende, Gervasio Castro.** 2002. "Ocupacao Agricola e Estrutura Agraria no Cerrado: O Papel do Preco da Terra, dos Recursos Naturais e da Tecnologia." *IPEA Texto Para Discussao*, 913.
- Eckstein, Zvi, and Kenneth I. Wolpin.** 1999. "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities." *Econometrica*, 67(6): 1295–1339.
- Edmonds, Eric.** 2002. "Government Initiated Community Resource Management and Local Resource Extraction from Nepal's Forests." *Journal of Development Economics*, 68(1): 89–115.
- Elobeid, Amani, and Simla Tokgoz.** 2008. "Removing Distortions in the U.S. Ethanol Market: What Does it Imply for the United States and Brazil?" *American Journal of Agricultural Economics*, 90(4): 918–932.
- Fair, Ray, and John Taylor.** 1983. "Solution and Maximum Likelihood Estimation of Dynamic Nonlinear Rational Expectations Models." *Econometrica*, 51(4): 1169–1185.
- Fair, Ray, and William Parke.** 1980. "Full-Information Estimates of a Nonlinear Macroeconometric Model." *Journal of Econometrics*, 13: 269–291.
- Farinelli, Barbara, Colin A. Carter, C.-Y. Cynthia Lin, and Daniel A. Sumner.** 2009. "Import demand for Brazilian ethanol: a cross-country analysis." *Journal of Cleaner Production*, 17: S9–S17.
- Filho, Joaquim Bento de Souza Ferreira, and Mark Horridge.** 2009. "Would Trade Liberalization Help the Poor of Brazil?" Agricultural Distortions Working Paper 106.
- Foster, Andrew, and Mark Rosenzweig.** 2003. "Economic Growth and the Rise of Forests." *The Quarterly Journal of Economics*, 601–637.
- Gurgel, Angelo, John M. Reilly, and Sergey Paltsev.** 2007. "Potential Land Use Implications of a Global Biofuels Industry." *Journal of Agricultural and Food Industrial Organization*, 5(2-9).

- Hay, D.A.** 2001. "The Post-1990 Brazilian Trade Liberalization and the Performance of Large Manufacturing Firms: Productivity, Market Share, and Profits." *Economic Journal*, 111: 620–641.
- Heckman, James, and Burton Singer.** 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica*, 52(2): 271–320.
- Heckman, James J., and Guilherme Sedlacek.** 1985. "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market." *The Journal of Political Economy*, 93(6): 1077–1125.
- Heckman, James J., Lance Lochner, and Christopher Taber.** 1998. "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents." *Review of Economic Dynamics*, 1: 1–58.
- Helfand, Steven M.** 2003. "The Impact of Agricultural Policy Reforms on the Agricultural Sector in Brazil in the 1990s: Implications for Pro-Poor Agricultural Policies." Paper Prepared for OECD Global Forum on Agriculture: Designing and Implementing Pro-Poor Agricultural Policies.
- Hertel, Thomas, and Jeffrey J. Reimer.** 2005. "Predicting the poverty impacts of trade reform." *The Journal of International Trade and Economic Development*, 14: 377–405.
- Jorgenson, Dale W.** 1998a. *Growth, Vol. 1: Econometric General Equilibrium Modeling*. The MIT Press.
- Jorgenson, Dale W.** 1998b. *Growth, Vol. 2: Energy, the environment, and economic growth*. The MIT Press.
- Judd, Ken.** 1998. *Numerical Methods in Economics*. Cambridge, Massachusetts: The MIT Press.
- Krivosos, Ekaterina, and Marcelo Olarreaga.** 2009. "Sugar Prices, Labor Income, and Poverty in Brazil." *Economia*, 9(2): 95–128.
- Lee, Donghoon, and Kenneth I. Wolpin.** 2006. "Intersectoral Labor Mobility and the Growth of the Service Sector." *Econometrica*, 74(1): 1–46.
- Lee, Hyunok, and Daniel A. Sumner.** 2010. "International Trade Patterns and Policy for Ethanol in the United States." In *Handbook of Bioenergy Economics and Policy*.

- , ed. Madhu Khanna, Jurgen Scheffran and David Zilberman, Chapter 19, 327–345. Springer.
- Lopes, Mauro De Rezende, and Ignez Vidigal Lopes.** 1998. “Fatores Que Afetam a Competitividade da Cadeia da Cana.” *Centro de Estudos Agrícolas, FGV and IBRE*.
- Melillo, Jerry M., John M. Reilly, David W. Kicklighter, Angelo C. Gurgel, Timothy W. Cronin, Sergey Paltsev, Benjamin S. Felzer, Xiaodong Wang, Andrei P. Sokolov, and C. Adam Schlosser.** 2009. “Indirect Emissions from Biofuels: How Important?” *Science*, 326: 1397–1399.
- Moraes, Marcia Azanha Ferraz Dias De.** 2007. “The Sugarcane Agro-System in Brazil: Labor Market Indicators and Wage Formation Model.” *Mimeo*.
- Nelson, Gerald C., and Richard D. Robertson.** 2008. “Green Gold or Green Wash: Environmental Consequences of Biofuels in the Developing World.” *Review of Agricultural Economics*, 30(3): 517–529.
- Nunberg, Barbara.** 1986. “Structural Change and State Policy: The Politics of Sugar in Brazil Since 1964.” *Latin American Research Review*, 21(2): 53–92.
- Pfaff, Alexander.** 1999. “What Drives Deforestation in the Brazilian Amazon?” *Journal of Environmental Economics and Management*, 37: 26–43.
- Robilliard, Anne-Sophie, Francois Bourguignon, and Sherman Robinson.** 2001. “Crisis and Income Distribution: A Micro-Macro Model for Indonesia.” Working Paper.
- Sargan, J.D.** 1975. “Asymptotic Theory and Large Models.” *International Economic Review*, 16(1): 75–91.
- Timmins, Christopher.** 2006. “Endogenous Land Use and the Ricardian Valuation of Climate Change.” *Environmental and Resource Economics*, 33: 119–142.
- Valdes, Alberto, and Johan A. Mistiaen,** ed. 2003. *Rural Poverty Alleviation in Brazil: Toward an Integrated Strategy*. Vol. World Bank Country Study, The International Bank for Reconstruction and Development / The World Bank.

Table 1: Composition of Regions

| Region Number | Shorthand Reference | States  | Census Region |
|---------------|---------------------|---|---------------|
| 1             | Parana              | Parana<br>Santa Catarina<br>Rio Grande do Sul                               | South         |
| 2             | Sao Paulo           | Sao Paulo   | Southeast     |
| 3             | Minas               | Minas Gerais<br>Rio de Janeiro<br>Espirito Santo                            | Southeast     |
| 4             | Bahia               | Bahia<br>Sergipe  | Northeast     |
| 5             | Mato Grosso         | Mato Grosso<br>Mato Grosso do Sul<br>Goias<br>Tocantins<br>Federal District | Center-West   |
| 6             | Pernambuco          | Pernambuco<br>Alagoas<br>Paraiba<br>Rio Grande do Norte                     | Northeast     |
| 7             | Maranhao            | Maranhao<br>Piaui<br>Ceara  | Northeast     |

Table 2: Land Usage and Production, 2005

| Region       | Land (Mill. Hectares) |           |        | Production |           |
|--------------|-----------------------|-----------|--------|------------|-----------|
|              | Sugarcane             | Other Ag. | Other  | Sugarcane  | Other Ag. |
| Parana       | 0.454                 | 46.733    | 9.154  | 31.228     | 35348.059 |
| Sao Paulo    | 3.085                 | 14.705    | 7.028  | 254.810    | 12482.325 |
| Minas Gerais | 0.582                 | 39.517    | 27.541 | 37.181     | 19310.408 |
| Bahia        | 0.118                 | 29.430    | 29.075 | 7.370      | 6737.708  |
| Mato Grosso  | 0.543                 | 122.708   | 65.124 | 37.939     | 27539.441 |
| Pernambuco   | 0.933                 | 12.625    | 10.018 | 49.101     | 3884.236  |
| Maranhao     | 0.077                 | 30.846    | 41.971 | 4.403      | 6307.208  |

Note: Cane production in millions of metric tons;  
other agriculture in millions of 2000 Reais

Table 3: Ethanol and Sugar Production, 2005

| Region       | Sugarcane | Eth. Quant. | Eth. Value | Sug. Quant. | Sug. Value |
|--------------|-----------|-------------|------------|-------------|------------|
| Parana       | 31.228    | 1.079       | 923.332    | 1.582       | 563.626    |
| Sao Paulo    | 254.810   | 9.654       | 8259.898   | 16.920      | 6028.224   |
| Minas Gerais | 37.181    | 1.241       | 1061.645   | 2.089       | 744.248    |
| Bahia        | 7.370     | 0.135       | 115.753    | 0.224       | 79.764     |
| Mato Grosso  | 37.939    | 2.022       | 1730.309   | 1.712       | 610.121    |
| Pernambuco   | 49.101    | 1.424       | 1218.738   | 4.034       | 1437.392   |
| Maranhao     | 4.403     | 0.108       | 92.198     | 0.019       | 6.743      |

Note: Sugarcane and sugar quantities in metric tons,  
ethanol quantity in cubic meters. Ethanol and sugar values  
in 2000 Reais. All values in millions of respective units.

Table 4: Median Hours and Number of Workers by Sector, 2005

| Region       | Cane  |        | Other Ag. |         | Non Ag. |          |
|--------------|-------|--------|-----------|---------|---------|----------|
|              | Hours | Number | Hours     | Number  | Hours   | Number   |
| Parana       | 46    | 30173  | 40        | 2720925 | 44      | 9567751  |
| Sao Paulo    | 48    | 151129 | 44        | 820975  | 42      | 15725917 |
| Minas Gerais | 48    | 55817  | 40        | 2296550 | 42      | 13301037 |
| Bahia        | 48    | 26977  | 40        | 2172985 | 40      | 3783111  |
| Mato Grosso  | 48    | 23542  | 44        | 1163488 | 44      | 4665797  |
| Pernambuco   | 48    | 223194 | 35        | 1702326 | 40      | 4452067  |
| Maranhao     | 40    | 35155  | 30        | 2499108 | 40      | 4070798  |

Table 5: Utility Function Parameter Estimates I

| $\beta$ | $d_{1 h}$ | $d_{1 l}$ | $\theta_1$ | $\theta_2$ | $\pi_1$ | $\sigma_1$ | $\sigma_2$ |
|---------|-----------|-----------|------------|------------|---------|------------|------------|
| 0.4150  | 1         | 0         | 5.3737     | 2.865      | 0.5862  | 0.4465     | 0.8730     |

Table 6: Utility Function Parameter Estimates II: Region-Sector Preference Shifters

| Parameter              | Parana | Sao Paulo | Minas Gerais | Bahia  | Mato Grosso | Pernambuco | Maranhao |
|------------------------|--------|-----------|--------------|--------|-------------|------------|----------|
| $\kappa_{1r} + z_{1r}$ | -3.920 | -3.225    | -3.496       | -4.368 | -5.323      | -1.825     | -3.235   |
| $\sigma_{z1r}^2$       | 0.124  | 0.085     | 0.134        | 0.266  | 0.211       | 0.066      | 0.250    |
| $\kappa_{2r} + z_{2r}$ | 0.346  | -1.249    | 0.293        | 0.673  | -0.741      | 0.486      | 1.212    |
| $\sigma_{z2r}^2$       | 0.067  | 0.045     | 0.034        | 0.053  | 0.048       | 0.086      | 0.037    |
| $\kappa_{3r} + z_{3r}$ | 0.534  | 0.892     | 0.903        | -0.014 | -0.230      | 0.205      | 0.212    |
| $\sigma_{z3r}^2$       | 0.087  | 0.040     | 0.140        | 0.166  | 0.150       | 0.148      | 0.176    |

Table 7: Estimates of Parameters Common to Regions

| Category      | Parameters            |                         |                       |                       |
|---------------|-----------------------|-------------------------|-----------------------|-----------------------|
| Agriculture   | $\frac{1}{1-\rho_1}$  | $\frac{1}{1-\rho_2}$    | $\theta_1$            | $\theta_2$            |
|               | 2.000                 | 2.000                   | 0.746                 | 0.712                 |
|               | $\gamma$              | $\sigma_{\epsilon 2}^2$ | $\sigma_{\nu 2}^2$    | $\sigma_{\nu 3}^2$    |
|               | 0.994                 | 0.026                   | 0.259                 | 0.013                 |
| Ethanol/Sugar | $\alpha_e$            | $\alpha_s$              | $\beta_e$             | $\beta_s$             |
|               | -2.752                | -1.981                  | -1.477                | 0.487                 |
|               | $\sigma_{e1}^2$       | $\sigma_s^2$            | $\sigma_{e2}^2$       | $x_e$                 |
|               | 0.009                 | 0.065                   | 0.066                 | -4.543                |
|               | $x_s$                 | $\beta_p$               | $\sigma_{xe}^2$       | $\sigma_{xs}^2$       |
|               | 7.071                 | 3.079                   | 0.151                 | 0.008                 |
| Consumption   | $a_e$                 | $a_s$                   | $a_a$                 | $a_k$                 |
|               | -4.382                | -5.870                  | -2.033                | -1.073                |
|               | $\sigma_{\delta e}^2$ | $\sigma_{\delta s}^2$   | $\sigma_{\delta a}^2$ | $\sigma_{\delta k}^2$ |
|               | 0.019                 | 0.362                   | 0.019                 | 0.018                 |

Table 8: Estimates of Region-Specific Parameters

| Parameter                | Parana | Sao Paulo | Minas  | Bahia  | Mato Grosso | Pernambuco | Maranhao |
|--------------------------|--------|-----------|--------|--------|-------------|------------|----------|
| $\phi_{1r}$              | 2.957  | 2.767     | 2.706  | 2.294  | 3.059       | 2.272      | 2.094    |
| $\phi_{2r}$              | 2.883  | 2.918     | 2.725  | 2.208  | 2.937       | 2.290      | 2.163    |
| $\lambda_{1r}$           | -1.234 | 1.785     | -1.615 | -3.866 | -2.074      | -1.016     | -6.015   |
| $\lambda_{2r}$           | -3.421 | -3.197    | -4.516 | -6.508 | -5.151      | -5.914     | -7.323   |
| $\lambda_{3r}$           | -0.815 | -0.272    | -0.317 | -0.051 | -0.089      | -0.298     | 0.037    |
| $\kappa$                 | 0.611  | 0.256     | 0.895  | 1.521  | 2.109       | 0.961      | 1.652    |
| $\phi_{or}$              | -0.571 | -0.500    | -0.583 | -0.599 | -0.552      | -0.920     | -0.999   |
| $\beta_{or}$             | 1.050  | 1.072     | 1.074  | 1.196  | 1.151       | 0.711      | 0.664    |
| $\sigma_{\epsilon 1}^2$  | 0.219  | 0.048     | 0.085  | 0.115  | 0.030       | 0.042      | 0.448    |
| $\sigma_{\nu 1}^2$       | 0.498  | 0.186     | 0.457  | 0.366  | 0.222       | 0.274      | 6.043    |
| $\sigma_{\nu k}^2$       | 0.013  | 0.063     | 0.008  | 0.006  | 0.068       | 0.069      | 0.024    |
| $\sigma_{\epsilon op}^2$ | 0.037  | 0.033     | 0.030  | 0.039  | 0.029       | 0.042      | 0.039    |
| $\sigma_{\epsilon ob}^2$ | 0.035  | 0.043     | 0.020  | 0.028  | 0.018       | 0.019      | 0.017    |

Table 9: Aggregates in 2005 and Simulated Environments - Production

|       | Production (Millions of Units*) |          |       |          |           |
|-------|---------------------------------|----------|-------|----------|-----------|
|       | Sugarcane                       | Ethanol  | Sugar | Other Ag | Composite |
| 2005  | 422.03                          | 4137.87  | 26.58 | 937.06   | 10180.08  |
| Sim 1 | 572.50                          | 8778.04  | 12.62 | 688.91   | 11770.82  |
| Sim 2 | 546.44                          | 8374.49  | 12.07 | 965.68   | 11110.77  |
| Sim 3 | 1072.63                         | 18995.51 | 4.76  | 1299.42  | 10033.41  |

\*Note: Sim 1 removes the duty and tariff. Sim 2 does the same, and increases the other agricultural price by 20%. Sim 3 adds to the Sim 2 changes by increasing the world petroleum price by 20%. Composite quantities are in billions of units. Units of ethanol are US gallons, while units of sugarcane and sugar are metric tons.

Table 10: Aggregates in 2005 and Simulated Environments - Exports and Prices

|       | Net Exports (Millions of Units*) |       |          | Prices (2000 Reais) |         |        |             |
|-------|----------------------------------|-------|----------|---------------------|---------|--------|-------------|
|       | Ethanol                          | Sugar | Other Ag | Sugarcane           | Ethanol | Sugar  | Rental Rate |
| 2005  | 684.75                           | 18.16 | 92.86    | 20.38               | 855.61  | 356.28 | 100.35      |
| Sim 1 | 5516.48                          | 4.79  | -261.58  | 27.73               | 1019.92 | 431.87 | 110.74      |
| Sim 2 | 5159.16                          | 4.35  | 179.69   | 27.75               | 1026.64 | 434.37 | 111.90      |
| Sim 3 | 16011.63                         | -2.27 | 522.12   | 33.86               | 1094.05 | 471.82 | 108.18      |

Table 11: Land Allocations in 2005 and Simulated Environments (Millions of Acres)

|              | Sugarcane |       |       |       | Other Ag. |        |        |        |
|--------------|-----------|-------|-------|-------|-----------|--------|--------|--------|
|              | 2005      | Sim 1 | Sim 2 | Sim 3 | 2005      | Sim 1  | Sim 2  | Sim 3  |
| Parana       | 1.12      | 3.01  | 2.38  | 3.16  | 115.48    | 102.21 | 117.82 | 120.21 |
| Sao Paulo    | 7.62      | 10.27 | 10.84 | 19.34 | 36.34     | 38.71  | 37.33  | 33.25  |
| Minas Gerais | 1.44      | 3.08  | 2.08  | 5.31  | 97.65     | 85.93  | 100.41 | 89.55  |
| Bahia        | 0.29      | 2.18  | 1.02  | 0.18  | 72.72     | 41.43  | 91.06  | 89.29  |
| Mato Grosso  | 1.34      | 3.40  | 2.14  | 3.24  | 303.21    | 283.74 | 315.93 | 383.72 |
| Pernambuco   | 2.31      | 3.95  | 0.97  | 4.73  | 31.20     | 8.02   | 39.97  | 6.29   |
| Maranhao     | 0.19      | 0.04  | 0.02  | 0.03  | 76.22     | 21.44  | 84.17  | 104.15 |
| Total        | 14.31     | 25.93 | 19.45 | 35.99 | 732.81    | 581.48 | 786.68 | 826.47 |

\*Note: Sim 1 removes the duty and tariff. Sim 2 does the same, and increases the other agricultural price by 20%. Sim 3 adds to the Sim 2 changes by increasing the world petroleum price by 20%. Composite quantities are in billions of units. Units of ethanol are US gallons, while units of sugarcane and sugar are metric tons.

Table 12: Land Allocations in 2005 and Simulated Environments (Millions of Acres)

|              | Non-Agriculture |        |        |        |
|--------------|-----------------|--------|--------|--------|
|              | 2005            | Sim 1  | Sim 2  | Sim 3  |
| Parana       | 22.62           | 33.99  | 19.02  | 15.84  |
| Sao Paulo    | 17.37           | 12.34  | 13.16  | 8.73   |
| Minas Gerais | 68.05           | 78.13  | 64.65  | 72.28  |
| Bahia        | 71.84           | 101.26 | 52.78  | 55.39  |
| Mato Grosso  | 160.92          | 178.33 | 147.41 | 78.52  |
| Pernambuco   | 24.75           | 46.29  | 17.32  | 47.24  |
| Maranhao     | 103.71          | 158.64 | 95.93  | 75.94  |
| Total        | 469.27          | 608.98 | 410.25 | 353.93 |

Figure 1: Brazil's Biomes (Source: IBGE)

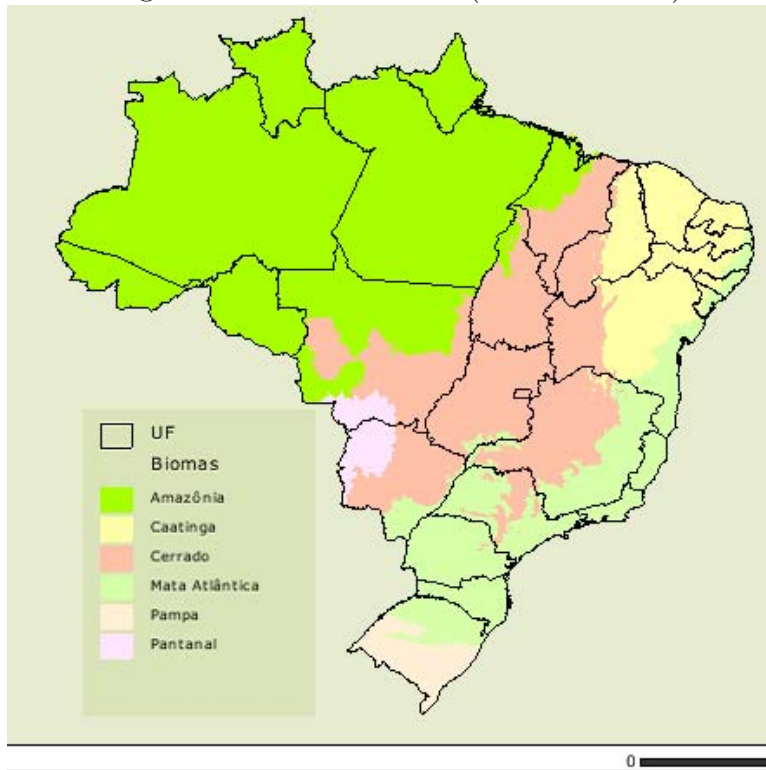


Figure 2: Regional Classification



Figure 3: Regional Sugarcane Land Shares and Production Relative to 1990

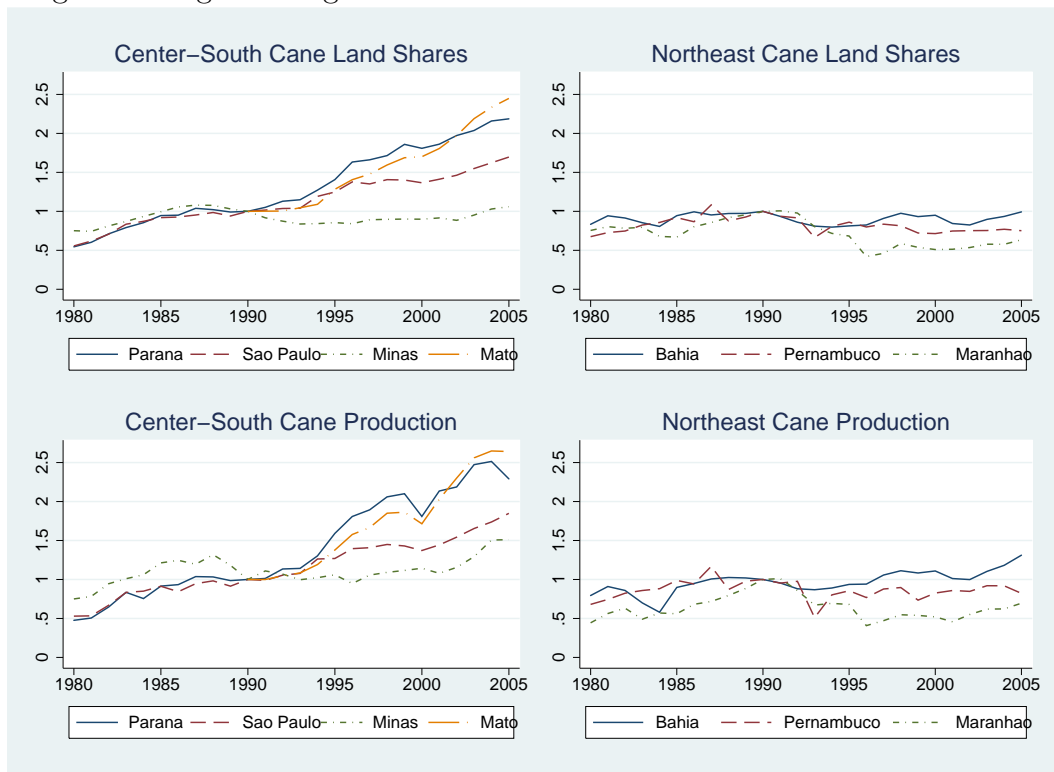


Figure 4: Box Plots for Simulated Values of Sugarcane, Ethanol, and Ethanol Exports (Base, Sim 1, Sim 2, and Sim 3 are scenarios defined in the text)

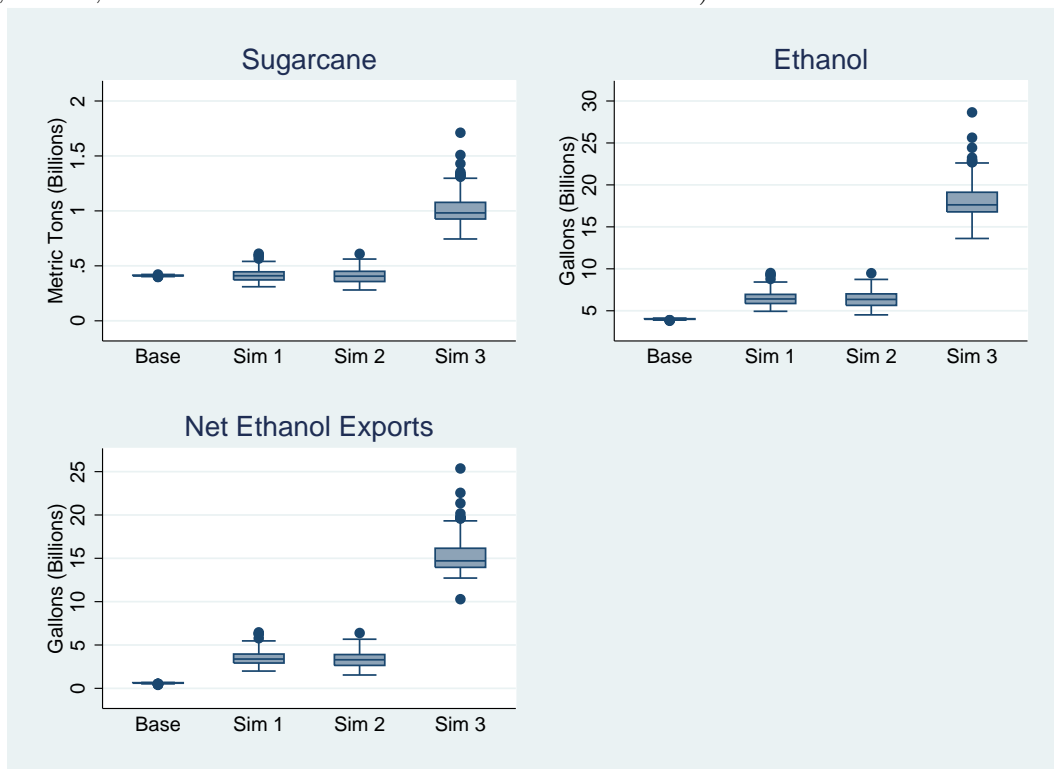
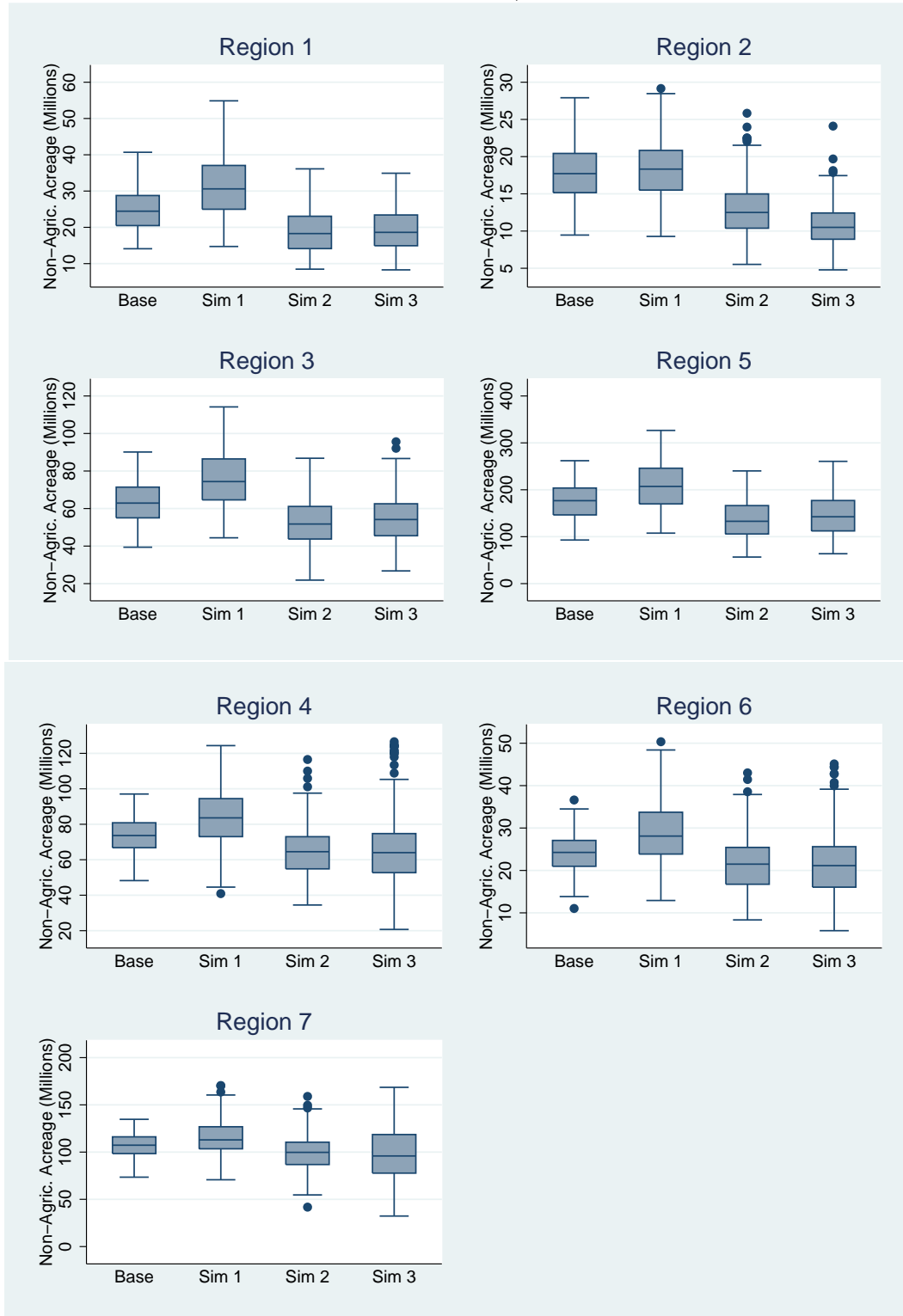


Figure 5: Box Plots for Simulated Values of Non-Agricultural Land (As defined in the text: “Base” refers to the baseline simulation. “Sim 1” removes trade barriers. “Sim 2” additionally increases the price of other agriculture output by 20%. “Sim 3” is the same, but adds a 20% increase in the price of petroleum)



# APPENDIX: Not for Print Publication

## A Data Details

I begin by providing additional detail on the data, dividing the variables into various categories. I then describe the data and assumptions used to create some of the key aggregate quantities used in estimation.

### A.1 Prices

- **Price of Sugarcane:** I take the price of sugarcane directly from the FGV-provided prices, which cover 17 states during the period of study. FGV provides prices on a month-by-month basis, where the prices are collected using farm surveys in the relevant geographical regions. I use an average over these monthly prices. In a case in which a region includes more than one state with a sugarcane price provided, I use the average of the prices of all the states in that region. In a case in which no states within a region have a sugarcane price, I use the sugarcane price of the closest region geographically. For the estimation and simulations, I assume one sugarcane price exists for all of Brazil. For this purpose, I use the median price of sugarcane across the regions.
- **Price of Sugar:** I use the FGV-provided price of a pound of sugar from the New York market. To translate this into Reais from US dollars, I use contemporaneous exchange rate data from the Central Bank of Brazil.
- **Price of Ethanol:** The FGV provides a price index for ethanol. I benchmark this index using the implied price of a cubic meter of exported ethanol in a particular year. I obtained this implied price using data from the organization SECEX, and their online AliceWeb tool.
- **Price of Other Agricultural Good:** The FGV provides a price index for all of agriculture. Using sugarcane prices and data on the share of agricultural production resulting from sugarcane, I calculate a price index for “other agriculture.” Since such a price has no natural units, I simply set the 2000 value to 100 Reais. This defines a “unit” of non-sugarcane agricultural production.
- **Petroleum Prices:** The FGV provides historical series of the month-by-month price of a barrel of oil in US dollars at various points of origin. I use the prices for Nigeria, relying on Central Bank of Brazil exchange rate series to convert these into Reais.
- **Price of Composite Non-Agricultural Good:** I construct relative changes in the nominal price for the composite good by using changes in nominal prices of all

other goods, the overall consumer price index, and the shares of all other goods in the “consumption basket” (from the national income accounts). Then I normalize the price since the composite good has no natural units.

## A.2 Labor Supply

- **Labor Hours:** The PNAD data contain: whether or not the member is employed (with September being the reference month); if she is employed, then her monthly income and usual weekly work hours from her primary job; the sector of her primary job, with sector being narrowly defined enough to allow for the identification of specific crops; and whether the individual is hired labor, a non-remunerated family worker, self-employed, or an employer. For region of work, there are a limited number of cases in which the survey respondent was not present in the current state as of the last week of September. In these cases, I simply drop the respondent. In order to construct aggregate hours worked in a particular region/sector, I multiply weekly hours by 52 and take the weighted sum over all individuals in that region/sector, using the PNAD-provided survey weights. In particular, I include all working individuals – not just hired labor – in these calculations. If “52” is the incorrect factor to multiply weekly hours by, then in general all the estimates of the paper will be affected in uncertain ways that deserve further inquiry in the future.
- **Sectoral Hourly Wages:** To construct the hourly wage, I multiply reported monthly earnings by  $\frac{7}{30}$ , and then divide the result by reported hours of work. In the analysis, I use the unconditional median hourly wages in the sector-region combinations as the regional wages for the three sectors. In calculating the wage, I use only the earnings of hired workers in the relevant sector. In particular, I do not use self-employment earnings. This procedure yielded one wage that appeared to be an outlier, driven by a very small sample of sugarcane workers in that region-year combination (the Maranhao region in 1999). I replace this wage with a predicted wage. To construct this predicted wage, I run a quantile regression for the median of Maranhao wages in other years, as a function of individual characteristics and wages in all the other regions and sectors.

## A.3 Land Use

- **Land Cultivated for Sugarcane:** The *Producao Agricola Municipal* (PAM) data provide information on the number of acres devoted to sugarcane in a given state and year, in the 1981-2005 time period. The PAM data refer to the calendar year, rather than the sugarcane harvesting year. There are data on the acreage planted and the

acreage harvested. These numbers are generally quite close together, and I use the acreage harvested throughout the paper.

- **Land Cultivated for Other Agricultural Uses:** The PAM provides annual information on total area harvested for major crops in each state. Data are available for all permanent and temporary crops in the 1990-2005 time period, and for ten of the most important crops in the 1981-1989 period.<sup>48</sup> Moreover, no years contain data on pasture land. This necessitates an imputation procedure for total agricultural land. For this purpose, I use the PAM data in combination with data from the agricultural censuses of 1980, 1985, 1995, and 2005-06, which provide more complete information. In particular, I use a scale factor defined by the ratio of the PAM acreage to the agricultural census acreage in a census year to estimate total planted acreages for intervening non-census years. The agricultural censuses also provide information on total land used for pasture and privately owned forests. To interpolate pasture values for a region between census years, I first run a regression on time trends and total cattle raised in a region in a year (available in PPM data), allowing for region-specific intercepts. The resulting standard errors are large, but as the purpose is only to obtain reasonable predictions, I use the regression equation to predict values for pasture land for each non-census year in each region. In the small number of cases with negative predicted values, I set pasture land to zero. I use IBGE data on the value of forest production in each year to conduct a similar imputation procedure. I then sum the values for crop land, pasture land, and private forest land across all states within a region. Further details are available upon request.
- **Non-Agricultural Land:** Non-agricultural land in a region is simply the residual resulting from subtracting sugarcane and other agricultural land from the total land area of the region. Data on total land area is obtained from IBGE. A given region may become slightly larger or smaller over time due to changes in method of measurement, changes in state boundaries, etc. I use the given land area for each year.

#### A.4 Production

- **Sugarcane Production:** I use data from PAM on sugarcane production in each state and year for the 1981-2005 period. I aggregate across states to produce regional totals. These data are for the calendar year, rather than the harvest year.

---

<sup>48</sup>The quality of the data varies over time, generally being better near agricultural census years. I obtained the 1990-2005 data from SIDRA, an online database operated by the Brazilian government. The earlier PAM data come from IPEA, an institute affiliated with the government, with a website at [www.ipeadata.com.br](http://www.ipeadata.com.br).

- **Non-Sugarcane Agricultural Production:** I obtain total agricultural production values for each state from state income accounts data. The value of agricultural production from the income accounts includes the value of forest production, though this is typically a small percentage of the total. To construct the value of “other agricultural” production for each state, I use sugarcane prices and sugarcane production to calculate the total value of sugarcane production, and then subtract that from the value of total agricultural production. To form the regional measure, I aggregate the result across the states within each region.
- **Ethanol and Sugar Production:** I obtain state-by-state ethanol and sugar production levels for each harvest year over the 1994/1995-2005/2006 period from the Ministry of Agriculture, Livestock, and Food Supply. These data are for local harvesting years. I impute a measure of the ethanol and sugar quantities for calendar year  $t$  by taking a weighted average of the surrounding harvest years, where the weights differ between the Northeast regions and the Center-South regions. The harvest year definition runs from May to April in the Center-South, and from September to August in the Northeast. The data on sugar and ethanol production also contain the amount of sugarcane crushed by mills and distilleries in each harvest year. In this paper, I use data on sugarcane production from PAM, but the crushed/harvested data from this alternative source suggests that the vast majority of harvested sugarcane is crushed by mills or distilleries, though gaps between the two exist.

## A.5 Construction of Key Quantities

It is useful to clearly delineate what quantities are available in the data sources and what can be constructed from these data sources, with the help of supplementary assumptions. Table 13 provides the quantities available from the raw data, categorized according to the labels in the first column. An asterisk in a superscript indicates that the item is referring to Brazil as a whole, rather than just the non-North regions that are the subject of this paper. A superscript “N” indicates aggregates specific to the North Census region, while aggregates without a superscript refer to the non-North portions of the country. When a quantity appears with the subscript “r”, it should be understood to represent the vector of that quantity for all regions. For instance, the item  $G_t^*$  refers to total government spending. The quantity  $VP_{rt}$  denotes the total value of production in each region.

I use the data from Table 13, along with supplementary assumptions, to construct empirical analogues of key quantities in the model. Table 14 summarizes this construction process. Each row in the first column provides the raw data variables and assumptions that I use to

construct the item in the second column of that row.

The first set of assumptions is necessary to divide up national quantities between the North Census region and the non-North Census regions that are the subject of my analysis. I make several assumptions to use the asterisked national-level items to construct the corresponding quantities for only the non-North regions. These assumptions are:

- Net exports from the North Census region are zero for the intermediate good of sugarcane and for each final good
- Agricultural profits, net transfers from the government, and capital do not flow in or out of the North Census region
- Net taxes in each super-region (the North, and the rest of Brazil) are proportional to GDP in that super-region

Together, these assumptions ensure both the North economy and the non-North economy are coherent, in the sense that the spending, production, and income methods of income accounting all yield the same values for GDP in the North Census region and in the total of the non-North Census regions. The assumptions also ensure that we can construct the amount of capital demanded and the amount of non-labor income in each of the two areas.

As a final note on Table 14, two new expressions appear there. First,  $WB_t$  is defined as  $WB_t = \sum_{r,j} w_{rjt} L_{rjt}$  (total wage bill for all regions and sectors). Second, the table lists the quantity  $VP_{rt}$ . This piece of data from the regional income accounts represents the total value of production in a region. It does not have a directly analogous item in the model, since it incorporates the value of taxes and some fraction of the value of newly produced capital goods.

Under one additional assumption, I can use  $VP_{rt}$  to construct the total supply of the composite good in each region. First note:

$$VP_{rt} = p_{ot}Y_{ort}^S + p_{et}Y_{ert}^S + p_{st}Y_{srt}^S + p_{at}Y_{2rt}^S + p_{1t}Y_{1rt}^S + OV_{rt}$$

where  $OV_{rt}$  is the value of taxes and counted capital in the region. Assume that  $OV_{rt}$  is a fixed proportion  $\omega$  of  $p_{ot}Y_{ort}^S$  across regions. Then we have:

$$\begin{aligned} \frac{VP_{rt} - (p_{et}Y_{ert}^S + p_{st}Y_{srt}^S + p_{2t}Y_{2rt}^S + p_{1t}Y_{1rt}^S)}{\sum_r (VP_{rt} - (p_{et}Y_{ert}^S + p_{st}Y_{srt}^S + p_{2t}Y_{2rt}^S + p_{1t}Y_{1rt}^S))} &= \frac{(1 + \omega)p_{ot}Y_{ort}^S}{(1 + \omega) \sum_r p_{ot}Y_{ort}^S} \\ &= \frac{p_{ot}Y_{ort}^S}{\sum_r p_{ot}Y_{ort}^S} \end{aligned}$$

Multiplying the result by total supply of the composite good then yields the region-specific supply of the composite good for every region.<sup>49</sup>

## B Model Details

In this section, I discuss details of the model. The first sub-section covers aggregation over individuals, used to produce expressions for aggregate labor supply to each region-sector combination and aggregate product demands. The second sub-section briefly presents the government's aggregate product demands. The third sub-section covers aggregation over parcels of land, used to produce expressions for aggregate product supply and aggregate labor demand in agriculture. The fourth sub-section formally defines equilibrium in the economy. Throughout, for notational simplicity I use  $p_{2t}$  and  $p_{at}$  interchangeably to refer to the price of the other agricultural good.

### B.1 Aggregation Over Individuals

In the main text, I presented the individual maximization problem underlying choices about consumption and labor supply. Here, I present the assumptions needed to go from this to the aggregate labor supplies and product demands used in estimation. While the aggregate product demands do not require it, I must make assumptions about individual heterogeneity to derive the aggregate labor supply expressions from the individual-level optimal choices above. I make the particular distributional assumptions below in order to secure the simplest aggregate expressions possible, given the set-up of the model. After noting these assumptions, I write expressions for the aggregate quantities.

In order to make the aggregation as transparent as possible, I first set out expressions for the optimal hours choice, optimal consumption choices, and associated indirect utility function  $V_{ist}$  for each sector-market combination  $s$ . Given a choice of  $s$ , these are:

$$\begin{aligned} h_{ist} &= (1 - \beta)(T) - \frac{\beta}{w_{st}} M_{it} && \text{for workers} \\ &= 0 && \text{for non-workers} \end{aligned}$$

---

<sup>49</sup>I should note here that the GDP values calculated in this manner will not necessarily correspond with GDP values from the national income accounts. The sum of regional GDP measures differs somewhat from the national totals.

$$\begin{aligned}
y_{ixt} &= \frac{\gamma_{xt}}{p_{xt}(1-\beta)}(w_{st}h_{ist} + M_{it}) \quad \text{for } x = o, a, e, s \\
y_{ikt} &= \frac{\gamma_{kt}}{r_t(1-\beta)}(w_{st}h_{ist} + M_{it})
\end{aligned}$$

$$\begin{aligned}
V_{ist} &= \log[w_{st}^{1-\beta}(T) + w_{st}^{-\beta}M_{it}] + \ln[\beta^\beta] + W_t + \kappa_{st} + \eta_{ist} \quad \text{for workers} \\
&= \log[(T)^\beta(\frac{M_{it}}{1-\beta})^{1-\beta}] + W_t + \kappa_0 + \eta_{i0t} \quad \text{for non-workers}
\end{aligned}$$

where  $W_t$  is a function of the share parameters and national-level prices, as follows:

$$W_t = \sum_b \gamma_{bt} \log\left(\frac{\gamma_{bt}}{p_{bt}}\right) + \gamma_{kt} \log\left(\frac{\gamma_{kt}}{r_t}\right)$$

Next, I note the assumptions used for aggregation. The first set of assumptions involves heterogeneity in non-labor income. I assume non-labor income is given by  $M_{it} = e^{\mu_{it}}$ , where  $\mu_{it}$  represents individual heterogeneity. To provide flexibility in capturing the hours distribution, I assume that the distribution of  $\mu$  is a mixture of two normally distributed random variables  $\mu_1$  and  $\mu_2$ , with means in year  $t$  of  $(\theta_{1t}, \theta_{2t})$  and standard deviations of  $(\sigma_1, \sigma_2)$ . The probability that  $\mu_{it}$  is distributed as  $\mu_m$  is  $\pi_m$  for  $m = 1, 2$ . I refer to this as being of “class  $m$ .” Note that  $\pi_m$  is not time-varying.<sup>50</sup>

It is imperative that total non-labor income per person in the data equal the expectation of individual non-labor income from the model to ensure coherence of the model. This necessitates a method of shifting the  $\theta$  parameters in each year to match the non-labor income for that year. For this purpose, define  $\theta_{it} = \theta_i + \log(\bar{M}_t/\bar{M}_{2005})$ , with  $\bar{M}_t$  and  $\bar{M}_{2005}$  being average non-labor income in period  $t$  and 2005,  $\log(\bar{M}_t/\bar{M}_{2005})$  as a shifting factor for period  $t$  and the constant parameters  $\theta_1$  and  $\theta_2$  reflecting 2005 non-labor income. This ensures that the aggregation is consistent with economy-wide non-labor income.

Regarding heterogeneity in preferences, I assume  $\eta_{ist} \sim$  Type I Extreme Value, and is distributed independently of  $\mu_{it}$ . I normalize the unemployment preference shocks common to all individuals,  $\varphi_{0t}$ , to be zero for all  $t$ . However, I make distributional assumptions on the time-constant portion of unemployment preferences,  $\kappa_0$ . Specifically, I assume that  $\kappa_0$  can actually take one of two values, either  $\kappa_{01} = 0$  or  $\kappa_{02} > 0$ . This discrete random variable – which indicates some preference to be unemployed separate from that captured by the direct

---

<sup>50</sup>This follows the approach of James Heckman & Burton Singer (1984). This method has been widely used in the labor literature, with just one notable example being Zvi Eckstein & Kenneth I. Wolpin (1999).

value of leisure – is distributed independently of all other individual-level random variables, except the probability of being in income class 1 or 2. Let  $d_{1|h} = Pr(\kappa_{0i} = 0|class1)$  and  $d_{1|l} = Pr(\kappa_{0i} = 0|class2)$ , with  $d_{2|l} = 1 - d_{1|l}$ .<sup>51</sup>

I am now in a position to set out the expression for aggregate labor supply to a particular region/sector and aggregate product demands. The aggregate labor supplies require the introduction of some notation. Use  $s$  to denote the region-sector choice. Note the slight abuse of notation below, in that  $Pr(i \in s|\mu_m)$  actually refers to  $Pr(i \in s|\mu_m, class m)$ . Let  $f(x)$  be the probability density function of a random vector  $x$ . Let  $N_t$  be the total number of individuals in period  $t$ , and  $N_{st}$  be the total number choosing region-sector  $s$  in that period. Then, leaving some of the time subscripts implicit, we have:

$$\begin{aligned}
L_{st}^S &= N_{st}E(h_{ist}|i \in s) \\
&= N_{st}(1 - \beta)(T) - N_{st}\frac{\beta}{w_{st}}E(e^\mu|i \in s) \\
&= N_{st}(1 - \beta)(T) - N_{st}\frac{\beta}{w_{st}}\sum_{m=1}^2\pi_m\int_{-\infty}^{\infty}e^{\mu_m}\frac{Pr(i \in s|\mu_m)}{Pr(i \in s)}f(\mu_m)d\mu_m \\
&= N_t(1 - \beta)(T)Pr(i \in s) - N_t\frac{\beta}{w_{st}}\sum_{m=1}^2\pi_m\int_{-\infty}^{\infty}e^{\mu_m}Pr(i \in s|\mu_m)f(\mu_m)d\mu_m \\
&= N_t\sum_{m=1}^2\pi_m\int_{-\infty}^{\infty}[(1 - \beta)(T) - \frac{\beta}{w_{st}}e^{\mu_m}]Pr(i \in s|\mu_m)f(\mu_m)d\mu_m
\end{aligned}$$

To completely specify the aggregate labor supplies, we need expressions for the last two terms in the integral. Let  $S^1$  be the set of region-sector combinations (excepting unemployment) for which optimal hours of work are positive. Then, leaving time subscripts implicit, the terms in the final expression above are:

$$\begin{aligned}
f(\mu_m) &= \frac{1}{\sigma_m\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\mu_m - \theta_m}{\sigma_m}\right)^2} \\
Pr(i \in s|\mu_m, classm) &= \sum_{j=1}^2 d_{j|classm} \frac{e^{\kappa_s} \beta^\beta [w_s^{1-\beta} T + w_s^{-\beta} e^{\mu_m}]}{e^{\kappa_{0j}} (T)^\beta \left(\frac{e^{\mu_m}}{1-\beta}\right)^{1-\beta} + \sum_{q \in S^1} e^{\kappa_q} \beta^\beta [w_q^{1-\beta} T + w_q^{-\beta} e^{\mu_m}]} \\
&\quad \text{if } s \in S^1 \\
&= 0 \quad \text{if } s \notin S^1
\end{aligned}$$

Therefore, an increase in labor supply to a region-sector  $s$  can come from an increase in the exogenous number of individuals, an increase in the share of individuals in  $s$  conditional on non-labor income, or an increase in hours among workers already in  $s$ . For expositional ease,

<sup>51</sup>These choices regarding the flexible distribution of preferences for unemployment were made after preliminary estimation showed that a simpler model with independence fit the data very poorly.

in the rest of the paper I often use the pair  $(j, r)$  to refer to sector  $j$  in region  $r$ . To reflect this, one can alter the expression above by replacing the subscripts where necessary.

The aggregate product demands are easier to express. These are simply:

$$Y_{xt}^D = \frac{\gamma_x}{p_{xt}(1-\beta)} \left( \sum_{r,j} w_{jrt} L_{jrt}^S + M_t \right) \quad \text{for } x = o, a, e, s$$

$$Y_{kt}^D = \frac{\gamma_k}{r_t(1-\beta)} \left( \sum_{r,j} w_{jrt} L_{jrt}^S + M_t \right)$$

This follows from an application of the same steps as used for labor supply above, though here simplifications stemming from the Cobb-Douglas assumption on preferences produce simpler expressions.

## B.2 Government Product Demand

I assume that the government takes in net tax revenue of  $\tau_t$ , and then spends this intake on the composite good, the other agriculture good, ethanol, and sugar in the same proportions that individuals do. Importantly, the government does not consume capital in the model. Therefore, the consumption shares from above have to be normalized, and the summations below are over all non-capital final goods only:

$$Y_{xt}^G = \frac{\gamma_x}{p_{xt} \sum_j \gamma_j} \tau_t \quad \text{for } x = o, a, e, s$$

$$Y_{kt}^G = 0$$

The scaling of the consumption shares ensures that the government's budget balances.<sup>52</sup>

## B.3 Agricultural Production: Aggregation Over Parcels

In the main text, I presented the parcel-specific profit-maximization problem underlying land allocation decisions. By using assumptions for the distribution of land quality in each region, I now derive expressions for aggregate product supply and aggregate labor demand in each agricultural sector, in each region.

In order to produce the simplest possible expressions for aggregate quantities, I assume that  $u_{jrt}$  for  $j=1,2,3$ , are mutually independent and take the Type I extreme value distribution with parameter  $\gamma$ . That is, the probability density function of  $u_{jrt}$  is given by

---

<sup>52</sup>This is a simplification for the sake of the model, but in reality the government's budget does not actually balance in the national income accounts.

$$\frac{1}{\gamma} e^{-\frac{u_{jrt}}{\gamma}} e^{-e^{-\frac{u_{jrt}}{\gamma}}}. \quad 53$$

Using these assumptions, I integrate over the parcels of land allocated to each use in order to derive the share of land devoted to producing each good ( $S_{jrt}$ ), total effective units of land in each good ( $A_{jrt}$ ), total labor demand in each sector ( $L_{jrt}^D$ ), and total supply of each good ( $Y_{jrt}^S$ ). For convenience, define  $\pi_{jrt}^c$  such that  $\pi_{jrt}^*(u_{jrt}) = \pi_{jrt}^c e^{u_{jrt}}$ . Let  $c_{jrt} = \ln(\pi_{jrt}^c)$  and let  $A_{rt}^*$  be the total amount of land in region  $r$  at time  $t$ . Let  $\Gamma(\cdot)$  denote the gamma function, as opposed to the gamma density.<sup>54</sup>

Beginning with the land shares, and dropping the time subscripts:

$$\begin{aligned} S_j &= Pr(c_j + u_j > c_k + u_k, c_j + u_j > c_p + u_p) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{c_j - c_k + u_j} \int_{-\infty}^{c_j - c_p + u_j} \left(\frac{1}{\gamma}\right)^3 e^{-\frac{1}{\gamma}u_j - \frac{1}{\gamma}u_k - \frac{1}{\gamma}u_p} e^{-e^{-\frac{1}{\gamma}u_j - e^{-\frac{1}{\gamma}u_k - e^{-\frac{1}{\gamma}u_p}}} du_p du_k du_j \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{c_j - c_k + u_j} \left(\frac{1}{\gamma}\right)^2 e^{-e^{-\frac{1}{\gamma}(c_j - c_p + u_j)}} e^{-\frac{1}{\gamma}u_j - \frac{1}{\gamma}u_k} e^{-e^{-\frac{1}{\gamma}u_j - e^{-\frac{1}{\gamma}u_k}}} du_k du_j \\ &= \int_{-\infty}^{\infty} \frac{1}{\gamma} e^{-e^{-\frac{1}{\gamma}(c_j - c_p + u_j)}} e^{-e^{-\frac{1}{\gamma}(c_j - c_k + u_j)}} e^{-\frac{1}{\gamma}u_j} e^{-e^{-\frac{1}{\gamma}u_j}} du_j \\ &= \frac{e^{\frac{c_{jrt}}{\gamma r}}}{\sum_{i=1}^3 e^{\frac{c_{irt}}{\gamma r}}} \end{aligned}$$

Total effective land units in use  $j$ ,  $j=1, 2$  can be derived as follows:

$$\begin{aligned} A_j &= S_j A^* E(e^{\lambda_j + u_j} | c_j + u_j > c_k + u_k, c_j + u_j > c_p + u_p) \\ &= S_j A^* \int_{-\infty}^{\infty} e^{\lambda_j + u_j} \frac{\int_{-\infty}^{c_j - c_k + u_j} \int_{-\infty}^{c_j - c_p + u_j} f(u_j, u_k, u_p)}{S_j} du_j \\ &= A^* \int_{-\infty}^{\infty} \int_{-\infty}^{c_j - c_k + u_j} \int_{-\infty}^{c_j - c_p + u_j} \left(\frac{1}{\gamma}\right)^3 e^{\lambda_j + u_j} e^{-\frac{1}{\gamma}u_j - \frac{1}{\gamma}u_k - \frac{1}{\gamma}u_p} e^{-e^{-\frac{1}{\gamma}u_j - e^{-\frac{1}{\gamma}u_k - e^{-\frac{1}{\gamma}u_p}}} d\vec{u} \\ &= \frac{1}{\gamma} A^* e^{\lambda_j} \int_{-\infty}^{\infty} e^{-u_j(\frac{1}{\gamma} - 1)} e^{-e^{-\frac{1}{\gamma}u_j(1 + e^{-\frac{1}{\gamma}(c_j - c_p) + e^{-\frac{1}{\gamma}(c_j - c_k)}})} du_j \\ &= A^* e^{\lambda_j} \int_0^{\infty} v^{-\gamma} e^{-v(1 + e^{-\frac{1}{\gamma}(c_j - c_k) + e^{-\frac{1}{\gamma}(c_j - c_p)}})} dv \\ &= A_{rt}^* e^{\lambda_{jrt}} (S_{jrt})^{1 - \gamma r} \Gamma(1 - \gamma r) \end{aligned}$$

<sup>53</sup>Timmins (2006) assumes use-specific land values are linear in heterogeneous errors that take the extreme value distribution. In preliminary estimation attempts, I allowed  $\gamma$  to vary across region, but the differences among the parameters were very slight. For simplicity, I constrained the parameters to be constant across region.

<sup>54</sup>That is,  $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$ , and the function is defined for  $x \in (0, \infty)$ .

Note the implicit restriction that  $\gamma \in (0, 1)$ . This expression corroborates the intuition that the marginal impact on effective land units of increasing the land share slightly is diminishing in the land share.

Finally, from the expression for  $A_{jrt}$  and the parcel-level labor demands and product supplies, it is easy to see that aggregate labor demand and aggregate product supply in sugarcane for region  $r$  are:

$$\begin{aligned}
L_{1rt}^D &= \left( \frac{p_{1t}\theta_1\alpha_{1rt}^{\frac{1-\theta_1}{1-\rho_{1r}}}}{w_{1rt}r_t^{\frac{\theta_1-\rho_{1r}}{1-\rho_{1r}}}} \right)^{\frac{1}{1-\theta_1}} \left( r_{1t}^{\frac{\rho_{1r}}{1-\rho_{1r}}}\alpha_{1rt}^{\frac{1}{1-\rho_{1r}}} + w_{1rt}^{\frac{\rho_{1r}}{1-\rho_{1r}}}(1-\alpha_{1rt})^{\frac{1}{1-\rho_{1r}}} \right)^{\frac{\theta_1-\rho_{1r}}{\rho_{1r}(1-\theta_1)}} \\
&* A_{rt}^* S_{1rt}^{1-\gamma} \Gamma(1-\gamma) e^{\lambda_{1rt}} \\
Y_{1rt}^S &= \left( \frac{p_{1t}\theta_1}{w_{1rt}r_t} \right)^{\frac{\theta_1}{1-\theta_1}} \left( r_{1t}^{\frac{\rho_{1r}}{1-\rho_{1r}}}\alpha_{1rt}^{\frac{1}{1-\rho_{1r}}} + w_{1rt}^{\frac{\rho_{1r}}{1-\rho_{1r}}}(1-\alpha_{1rt})^{\frac{1}{1-\rho_{1r}}} \right)^{\frac{\theta_1(1-\rho_{1r})}{\rho_{1r}(1-\theta_1)}} \\
&* A_{rt}^* S_{1rt}^{1-\gamma} \Gamma(1-\gamma) e^{\lambda_{1rt}}
\end{aligned}$$

Analogous expressions hold for other agriculture.

For these expressions to be useful in estimation, I need to parameterize  $\lambda_{jrt}$  for  $j = 1, 2, 3$ ,  $\phi_{1rt}$ ,  $\phi_{2rt}$ , and  $\kappa_{rt}$ . Assume that for each  $j$ ,  $\lambda_{jrt}$  takes the following form:

$$\lambda_{jrt} = \lambda_{jr}^0 + \nu_{jrt} \quad \text{for } j = 1, 2, 3$$

where  $\nu_{jrt}$  is a shock to land quality that is common to all parcels of land in the given region. Set  $\phi_{jrt} = \phi_{jr}^0 + \epsilon_{jrt}$  for  $j = 1, 2$ , where  $\epsilon_{1rt}$  and  $\epsilon_{2rt}$  are shocks to production. Finally, let  $\kappa_{rt} = \kappa_r + \nu_{krt}$ , with  $\nu_{krt}$  a shock to the tendency to allocate land to other agriculture.

It is important to note that this framework for agricultural decision-making is quite limiting. For instance, the dynamic nature of production is not considered (see the conclusion in the main text). In another example, the effects of government policy are not considered. For sugarcane, for instance, special allocation of sugar export quota to the Northeast and subsidies to sugarcane producers in the Northeast have persistently led to production in areas that may not otherwise grow sugarcane.<sup>55</sup> In general, time-constant policy variation across regions will be reflected in the estimated production parameters, while time-varying policy changes will appear through the model's error terms. Both outcomes are undesirable. Future work could incorporate policy in a more satisfactory way.

---

<sup>55</sup>While there are some data on region-specific prices for ethanol and sugar (see <http://cepea.esalq.usp.br/>), these data do not use consistent definitions across regions and are only available for the Northeast for 2001 onwards.

## B.4 Equilibrium

Finally, I formally define market equilibrium in the model. An equilibrium involves the following conditions:

*Capital and Labor*

$$Y_{kt}^D + \sum_r (K_{ort}^D + K_{ert}^D + K_{srt}^D + K_{1rt}^D + K_{2rt}^D) = \bar{K}_t$$

$$L_{jrt}^D = L_{jrt}^S \quad \text{for } j = 1, 2, 3 \quad \forall r$$

*Intermediate and Final Goods*

$$\sum_r [Y_{1ert}^D + Y_{1srt}^D] = \sum_r Y_{1rt}^S$$

$$Y_{at}^D + Y_{at}^G = \sum_r Y_{2rt}^S - Y_{at}^X$$

$$Y_{it}^D + Y_{it}^G = \sum_r Y_{irt}^S - Y_{it}^X \quad \text{for } i = o, s, e$$

*Zero-profit Conditions*

$$p_{ot} z_{ort} = \left( \frac{w_{3rt}}{\beta_{ort}} \right)^{\beta_{ort}} \left( \frac{r_t}{1 - \beta_{ort}} \right)^{1 - \beta_{ort}} \quad \forall r$$

$$p_{it} = \frac{p_{1t}}{\alpha_{it}} + \frac{r_t}{\beta_{it}} \quad \text{for } i = e, s$$

*Trade Balance and Budget Constraints*

$$p_{ot} Y_{ot}^X + p_{et} Y_{et}^X + p_{st} Y_{st}^X + p_{2t} Y_{at}^X = 0$$

$$p_{ot} Y_{ot}^D + p_{et} Y_{et}^D + p_{st} Y_{st}^D + p_{2t} Y_{at}^D + r_t Y_{kt}^D = \sum_{r,j} w_{jrt} L_{jrt}^S + M_t$$

$$p_{ot} Y_{ot}^G + p_{et} Y_{et}^G + p_{st} Y_{st}^G + p_{2t} Y_{at}^G = \tau_t$$

where total non-labor income is  $M_t = r_t \bar{K}_t + \sum_r [(1 - \theta_2) p_{2t} Y_{2rt}^S + (1 - \theta_1) p_{1t} Y_{1rt}^S] - \tau_t$ .

## C Estimation Details

In the first sub-section, I delineate assumptions about the distribution of production shocks and preference shocks to aggregate quantities in the economy. I also specify the parameters to be estimated. In the second and third sub-sections, I present details on the likelihood functions for the individual-level estimation and the aggregate-level estimation, respectively.

### C.1 Distributional Assumptions and Parameters to be Estimated

I assume that all stochastic errors in the economy are normally distributed, independent of one another. I allow the variance of the shocks to sugarcane land quality and share parameters  $\alpha$  to differ by region. However, the other agricultural shocks share a common variance within a group; that is, all regional shocks to land quality in other agriculture have the same variance, etc.<sup>56</sup> For agriculture,  $\nu_{1rt}$ ,  $\nu_{2rt}$ ,  $\nu_{3rt}$ ,  $\nu_{krt}$ ,  $\epsilon_{1rt}$ , and  $\epsilon_{2rt}$  have mean zero and variances given by  $\sigma_{\nu_{1r}}^2$ ,  $\sigma_{\nu_{2r}}^2$ ,  $\sigma_{\nu_{3r}}^2$ ,  $\sigma_{\nu_{kr}}^2$ ,  $\sigma_{\epsilon_{1r}}^2$ , and  $\sigma_{\epsilon_{2r}}^2$ . For ethanol and sugar production and exports,  $\epsilon_{e1t}$ ,  $\epsilon_{e2t}$ ,  $\epsilon_{st}$ ,  $\epsilon_{xet}$ , and  $\epsilon_{xst}$  have mean zero and variances given by  $\sigma_{\epsilon_{e1}}^2$ ,  $\sigma_{\epsilon_{e2}}^2$ ,  $\sigma_{\epsilon_s}^2$ ,  $\sigma_{\epsilon_{xe}}^2$ , and  $\sigma_{\epsilon_{xs}}^2$ . And for the remaining categories,  $\epsilon_{ort}^i$  has mean zero and variance  $\sigma_{\epsilon_{oir}}^2$  for  $i = p, b$  and  $\delta_{it}$  has mean zero and variance  $\sigma_{\delta_i}^2$  for  $i = e, s, a, k$ . Finally,  $\varphi_{jrt}$  has mean  $z_{jr}$  and variance  $\sigma_{\varphi_{jr}}^2$  for each region-sector combination.

This leaves us with 190 parameters to estimate: (1) Four parameters governing marginal utilities of consumption,  $a_e$ ,  $a_s$ ,  $a_a$ , and  $a_k$ ; (2) 21 land quality parameters  $\lambda_{jr}^0$ , for  $j = 1, 2, 3$  and all regions; (3) Seven land allocation friction parameters,  $\kappa_r$ ; (4) One land quality distribution parameter  $\gamma$ ; (5) Seven  $\phi_{1r}^0$  parameters in sugarcane; (6) Seven  $\phi_{2r}^0$  parameters in other agriculture; (7) Two CES parameters  $\rho_j$  in sugarcane and other agriculture;<sup>57</sup> (8) Two parameters governing decreasing returns to scale in sugarcane and other agriculture  $\theta_j$ ; (9) Seven labor share parameters in composite good production,  $\beta_{or}$ ; (10) Seven TFP parameters in composite good production,  $\phi_{or}$ ; (11) Four parameters governing ethanol and sugar production,  $\alpha_j$  and  $\beta_j$ , for  $j = e, s$ ; (12) Three parameters  $x_e$ ,  $x_s$  and  $\beta_p$  in ethanol and sugar exports; (13) 21 means of the shocks to labor supply,  $z_{jr}$ ;<sup>58</sup> (14) 21 parameters

<sup>56</sup>I typically allow the variance of shocks to vary by region for those shocks that are associated with the most cleanly identified parameters. The shocks to labor shares in composite good production, for instance, are allowed to be region-specific. This is because, as will be clear below, the maximum likelihood estimate of the associated parameter will simply be the mean of a certain transformation of the relevant quantities. The estimate of the variance of the associated shock is simply the average of the squared deviations from the mean.

<sup>57</sup>In preliminary estimation, I allowed these parameters to vary across region. However, given the constraints imposed on these parameters, the maximum value of the likelihood occurred at the upper bound for all regions.

<sup>58</sup>In the complete model, these are not separately identified from the  $\kappa_{jr}$ . These items will play a role in the linearized version of the model, however.

governing the consumer preference for region-sector,  $\kappa_{jr}$ ; (15) Two parameters indicating the propensity to be unemployed,  $d_{1|h}$  and  $d_{1|l}$ ;<sup>59</sup> (16) Five parameters giving the properties of the non-labor income distribution,  $\pi_1$ ,  $\theta_1$ ,  $\theta_2$ ,  $\sigma_1$  and  $\sigma_2$ ; (17) One parameter controlling the preference for leisure,  $\beta$ ; (18) 43 parameters governing the variance of shocks on the production side of the economy; and (19) 25 parameters governing the variance of shocks to preferences.

## C.2 Likelihood of Individual Labor Supply Decisions

For each individual, we potentially observe four choices: whether or not they work, the region of work, the sector of work, and hours. The last three are observed only for those individuals who work. To assure coherence of the individual and aggregate models, the likelihood function conditions on the vector of region-sector wages  $\vec{w}_{jrt}$  and national-level non-labor income  $M_t$ . Let  $d_{it} = 1$  if individual  $i$  works, and equal zero otherwise. The likelihood function is:

$$L = \prod_t \prod_{i=1}^{N_t} [Pr(d_{it} = 0 | \vec{w}_{jrt}, M_t)]^{1-d_{it}} \prod_{r,j} [L_{it}(d_{it} = 1, h_{ijrt}, s_{it} = (r, j) | \vec{w}_{jrt}, M_t)]^{d_{it} 1(s_{it}=(r,j))}$$

where  $h_{ijrt}$  is the hours choice and  $s_{it}$  is the region-sector choice, with  $s_{it}$  taking a value  $(r, j)$ , for  $r = 1, \dots, 7$  and  $j = 1, 2, 3$ . Here,  $1(s_{it} = (r, j))$  is an indicator function taking a value of one if the chosen region-sector is  $(r, j)$ . Let  $S^1$  again indicate the set of region-sector combinations for which optimal hours of work are positive at a given level of non-labor income. Dropping some subscripts for expositional ease, for non-workers the components of the likelihood contribution are:

$$\begin{aligned} Pr(d = 0 | \vec{w}_{jr}, M) &= \sum_z \sum_m d(z|class m) \pi_m \int_{-\infty}^{\infty} \frac{e^{\kappa_{0z}} (T)^\beta \left(\frac{e^{\mu_m}}{1-\beta}\right)^{1-\beta}}{K} f(\mu_m) d\mu_m \\ K &= e^{\kappa_{0z}} (T)^\beta \left(\frac{e^{\mu_m}}{1-\beta}\right)^{1-\beta} + \sum_{p,k \in S^1} e^{\kappa_{pk}} \beta^\beta [w_{pk}^{1-\beta} (T) + w_{pk}^{-\beta} e^{\mu_m}] \end{aligned}$$

For workers we have

$$\begin{aligned} L_{it}(d_{it} = 1, h_{ijrt}, s_{it} = (r, j) | \vec{w}_{jrt}, M_t) &= \sum_m \sum_z \pi_m d(z|class m) \\ &\times P(d_{it} = 1, s_{it} = (r, j) | h_{ijrt}, \vec{w}_{jrt}, M_t, class m, \kappa_{0z}) \\ &\times f(h_{ijrt} | \vec{w}_{jrt}, M_t, class m, \kappa_{0z}) \end{aligned}$$

<sup>59</sup>Note  $\kappa_{02}$  also determines unemployment. However, to ease identification, I simply set  $\kappa_{02}$  to a value of 6 in estimation. This value was chosen based on preliminary estimation.

Note that after conditioning on hours, the income class is irrelevant in the probability of region-sector. After conditioning on income class, the value of  $\kappa_{0z}$  is irrelevant in the density. Dropping some subscripts, for workers the components above are then:

$$\begin{aligned}
P(d = 1, s = (r, j) | h, \kappa_{0z}, \vec{w}_{jr}, M) &= \frac{e^{\kappa_{jr}} \beta^{\beta-1} [w_{jr}^{1-\beta} (T-h)]}{D(z)} \\
D(z) &= e^{\kappa_{0z}} (T)^\beta \left[ \frac{w_{jr}}{\beta(1-\beta)} [(1-\beta)(T) - h] \right]^{1-\beta} \\
&+ \sum_{p,k \in S^1} e^{\kappa_{pk}} \beta^\beta \left[ w_{pk}^{-\beta} (T)(w_{pk} - w_{jr}) + \frac{w_{jr} w_{pk}^{-\beta}}{\beta} (T-h) \right] \\
f(h | class m, M) &= \frac{1}{\sigma_m \sqrt{2\pi}} \frac{\exp\left(-\frac{1}{2\sigma_m^2} \left[ \ln\left(\frac{w_{jr}}{\beta} [(1-\beta)(T) - h]\right) - \theta_m \right]^2\right)}{(1-\beta)(T) - h}
\end{aligned}$$

Conditioning on  $M$  in the last expression is important because  $\theta_m$  for each year depends on non-labor income in the year, as discussed above.

### C.3 Likelihood Function for Aggregate Quantities

As noted in the text, the joint density of  $\vec{V}_t$  is:

$$f(\vec{V}_{1t}, \vec{V}_{2t}) = g(\vec{\psi}_{1t}(\vec{V}_{1t}, \vec{V}_{2t}))g(\vec{\psi}_{2t}(\vec{V}_{1t}, \vec{V}_{2t})) \det \begin{pmatrix} \frac{\partial \psi_1}{\partial V_1} & \frac{\partial \psi_1}{\partial V_2} \\ \frac{\partial \psi_2}{\partial V_1} & \frac{\partial \psi_2}{\partial V_2} \end{pmatrix}$$

since  $\vec{\psi}_1$  and  $\vec{\psi}_2$  are independent. In fact, the errors within the vectors  $\vec{\psi}_1$  and  $\vec{\psi}_2$  are independent as well, which breaks apart this density even further.

The next step in constructing the likelihood function involves finding expressions for the error vectors  $\vec{\psi}_1$  and  $\vec{\psi}_2$  in terms of the endogenous variables. These expressions can then be substituted into the densities above in the change of variables formula. There are convenient expressions for  $\vec{\psi}_1$ :

$$\begin{aligned}
\epsilon_{1rt} &= -\rho_{1r} \ln\left(\frac{w_{1rt}}{r_t}\right) + (1 - \rho_{1r}) \ln\left(\frac{\theta_1 p_{1t} Y_{1rt} - w_{1rt} L_{1rt}}{w_{1rt} L_{1rt}}\right) - \phi_{1r} \\
\nu_{1rt} &= \ln\left(\frac{Y_{1rt}}{A_{rt}^* (S_{1rt})^{1-\gamma} \Gamma(1-\gamma)}\right) - \frac{\theta_1}{1-\theta_1} \ln\left(\frac{p_{1t} \theta_1}{w_{1rt} r_t}\right) \\
&- \frac{\theta_1 (1-\rho_{1r})}{\rho_{1r} (1-\theta_1)} \ln\left(r_{1t}^{\frac{\rho_{1r}}{1-\rho_{1r}}} \alpha_{1rt}^{\frac{1}{1-\rho_{1r}}} + w_{1rt}^{\frac{\rho_{1r}}{1-\rho_{1r}}} (1-\alpha_{1rt})^{\frac{1}{1-\rho_{1r}}}\right) - \lambda_{1r}
\end{aligned}$$

$$\begin{aligned}\epsilon_{2rt} &= -\rho_{2r} \ln\left(\frac{w_{2rt}}{r_t}\right) + (1 - \rho_{2r}) \ln\left(\frac{\theta_2 p_{2t} Y_{2rt} - w_{2rt} L_{2rt}}{w_{2rt} L_{2rt}}\right) - \phi_{2r} \\ \nu_{2rt} &= \ln\left(\frac{Y_{2rt}}{A_{rt}^* (S_{2rt})^{1-\gamma} \Gamma(1-\gamma)}\right) - \frac{\theta_2}{1-\theta_2} \ln\left(\frac{p_{2t} \theta_2}{w_{2rt} r_t}\right) \\ &\quad - \frac{\theta_2(1-\rho_{2r})}{\rho_{2r}(1-\theta_2)} \ln\left(r_{2t}^{\frac{\rho_{2r}}{1-\rho_{2r}}} \alpha_{2rt}^{\frac{1}{1-\rho_{2r}}} + w_{2rt}^{\frac{\rho_{2r}}{1-\rho_{2r}}} (1-\alpha_{2rt})^{\frac{1}{1-\rho_{2r}}}\right) - \lambda_{2r}\end{aligned}$$

$$\nu_{3rt} = \gamma \ln(1 - S_{1rt} - S_{2rt}) - \ln(S_{1rt}) + \ln\left(\frac{p_{1t} Y_{1rt} (1 - \theta_1)}{A_{rt}^* \Gamma(1 - \gamma)}\right) - \lambda_{3r}$$

$$\nu_{krt} = \ln\left(\frac{S_{2rt}}{S_{1rt}}\right) + \ln\left(\frac{p_{1t} Y_{1rt} (1 - \theta_1)}{p_{2t} Y_{2rt} (1 - \theta_2)}\right) - \kappa_r$$

$$\epsilon_{st} = \ln\left(\frac{p_{1t} \beta_s}{p_{st} \beta_s - r_t}\right) - \ln(\alpha_s)$$

$$\epsilon_{e1t} = \ln\left(\frac{p_{1t} \beta_s Y_{et}}{\beta_s (p_{1t} \sum_r Y_{1rt} - p_{st} Y_{st}) + r_t Y_{st}}\right) - \ln(\alpha_e)$$

$$\epsilon_{e2t} = \ln\left(\frac{r_t \beta_s Y_{et}}{\beta_s (p_{et} Y_{et} + p_{st} Y_{st} - p_{1t} \sum_r Y_{1rt}) - r_t Y_{st}}\right) - \ln(\beta_e)$$

$$\epsilon_{ort}^b = \ln(p_{ot} Y_{ort} - w_{3rt} L_{3rt}) - \ln(w_{3rt} L_{3rt}) - \beta_{or}$$

$$\begin{aligned}\epsilon_{ort}^p &= \log\left(\frac{r_t}{p_{ot}}\right) + \log\left[\frac{p_{ot} Y_{ort}}{p_{ot} Y_{ort} - w_{3rt} L_{3rt}}\right] \\ &\quad + \frac{w_{3rt} L_{3rt}}{p_{ot} Y_{ort}} \left[\log\left(\frac{p_{ot} Y_{ort} - w_{3rt} L_{3rt}}{r_t L_{3rt}}\right)\right] - \phi_{or}\end{aligned}$$

$$\delta_{at} = \ln\left(\frac{p_{2t} (\sum_r Y_{2rt} - Y_{2t}^X)}{p_{ot} \sum_r Y_{ort} + p_{2t} Y_{2t}^X + p_{et} Y_{et}^X + p_{st} Y_{st}^X}\right) - a_a$$

$$\delta_{et} = \ln\left(\frac{p_{et} (Y_{et} - Y_{et}^X)}{p_{ot} \sum_r Y_{ort} + p_{2t} Y_{2t}^X + p_{et} Y_{et}^X + p_{st} Y_{st}^X}\right) - a_e$$

$$\delta_{st} = \ln\left(\frac{p_{st} (Y_{st} - Y_{st}^X)}{p_{ot} \sum_r Y_{ort} + p_{2t} Y_{2t}^X + p_{et} Y_{et}^X + p_{st} Y_{st}^X}\right) - a_s$$

$$\begin{aligned}\delta_{kt} &= \ln\left(\frac{K_t^{DL}}{p_{ot} \sum_r Y_{ort} + p_{2t} Y_{2t}^X + p_{et} Y_{et}^X + p_{st} Y_{st}^X}\right) \\ &\quad - \ln\left(\frac{p_{ot} \sum_r Y_{ort} + p_{2t} \sum_r Y_{2rt} + p_{et} Y_{et} + p_{st} Y_{st} - \tau_t}{p_{ot} \sum_r Y_{ort} + p_{2t} \sum_r Y_{2rt} + p_{et} Y_{et} + p_{st} Y_{st}}\right) - a_k\end{aligned}$$

$$\begin{aligned}\epsilon_{xet} &= Y_{et}^X - x_e - \beta_D (t_e p_{et} + d_e) (I^*) - \beta_p p_{pt} \\ &\quad + \beta_S (t_e p_{et} + d_e) (I^*)\end{aligned}$$

$$\epsilon_{xst} = Y_{st}^X - x_s - \alpha_D p_{st} + \alpha_S p_{st}$$

where

$$\begin{aligned}
 r_t &= \frac{1}{\bar{K}_t} \left[ M_t + \tau_t - \sum ((1 - \theta_1)p_{1t}Y_{1rt} + (1 - \theta_2)p_{2t}Y_{2rt}) \right] \\
 p_{st}Y_{st} &= \sum_j \sum_r w_{jrt}L_{jrt} + M_t + \tau_t - p_{et}Y_{et} - \sum_r p_{ot}Y_{ort} - \sum_r p_{2t}Y_{2rt} - K_t^{DL}
 \end{aligned}$$

Using the ratio of production to labor in each agricultural sector, one can obtain expressions for  $\alpha_{1rt}$  and  $\alpha_{2rt}$  in terms of the endogenous variables. These expressions can be substituted in above to ensure that only exogenous and endogenous variables appear on the right-hand side of all equations.

## D Additional Tables and Figures

Figure 6: Sugarcane Cultivation Area, Eight Quantiles in 2007 (Source: IBGE)

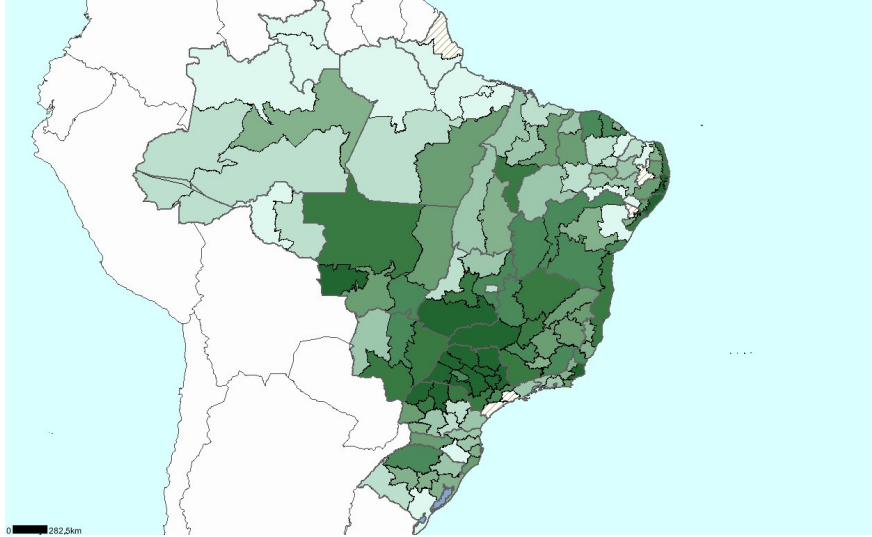


Figure 7: Sugarcane Cultivation Area, Eight Equal-Sized Intervals in 2007 (Source: IBGE)



Note: In both figures, darker areas indicate higher values.

Figure 8: Prices in the Sugarcane Industry, 1981-2005

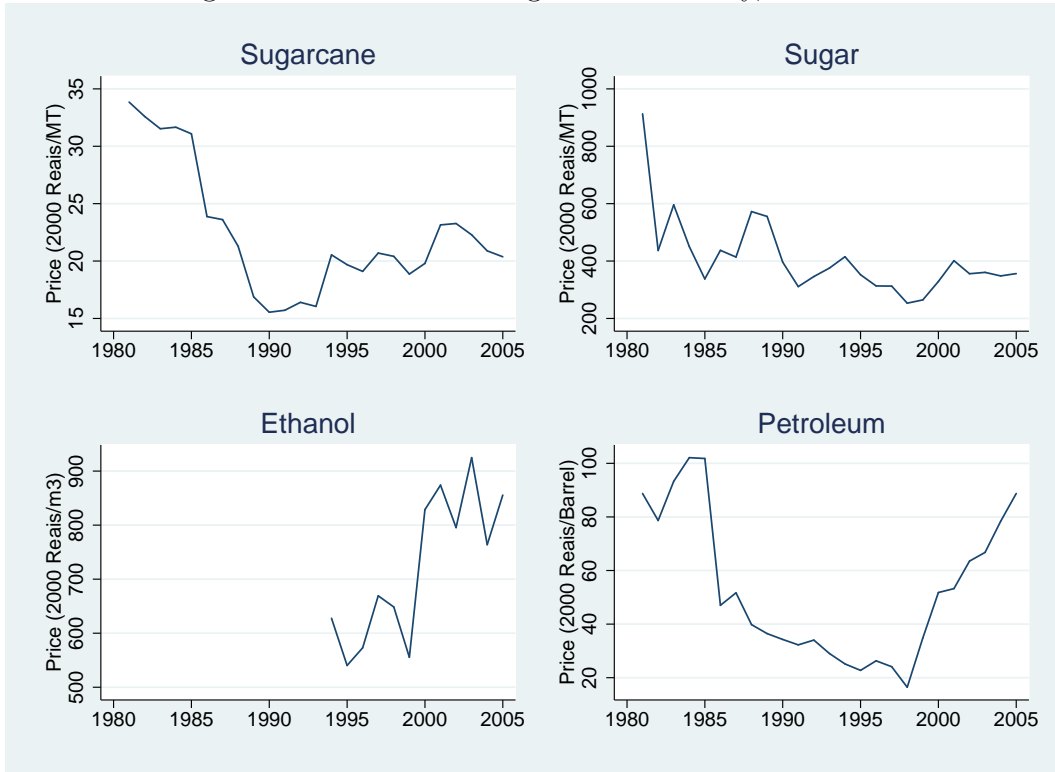


Figure 9: Prices, Relative to 1995

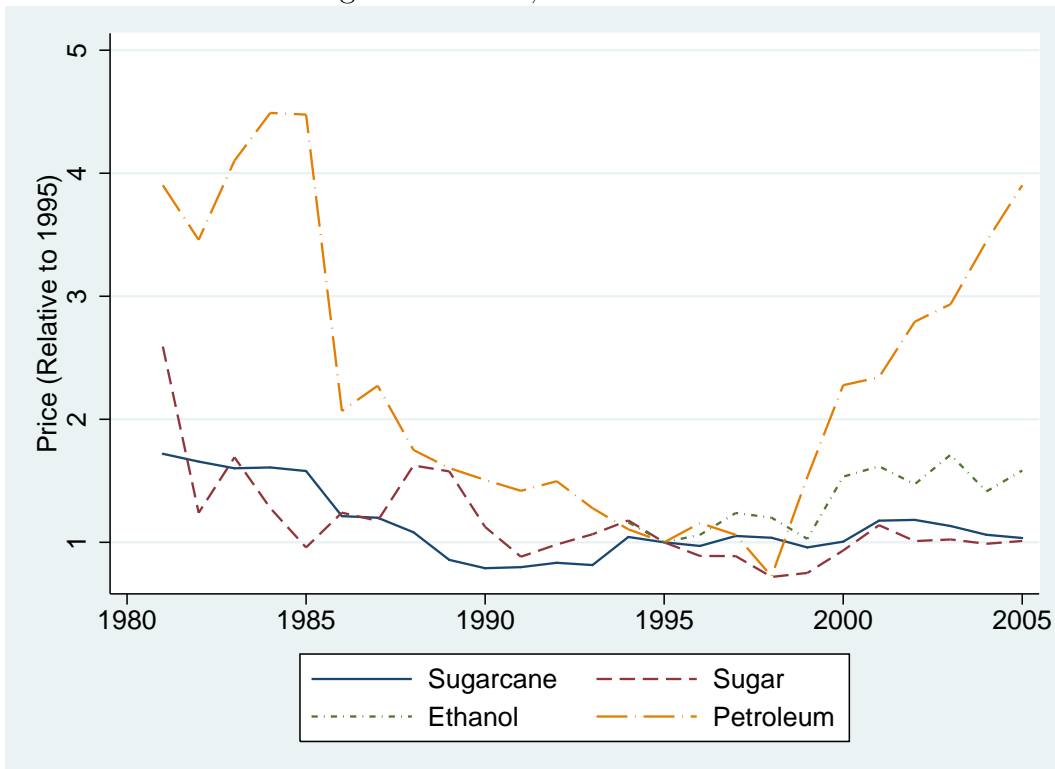


Figure 10: Agricultural Prices in 1981-2005, Relative to 1981

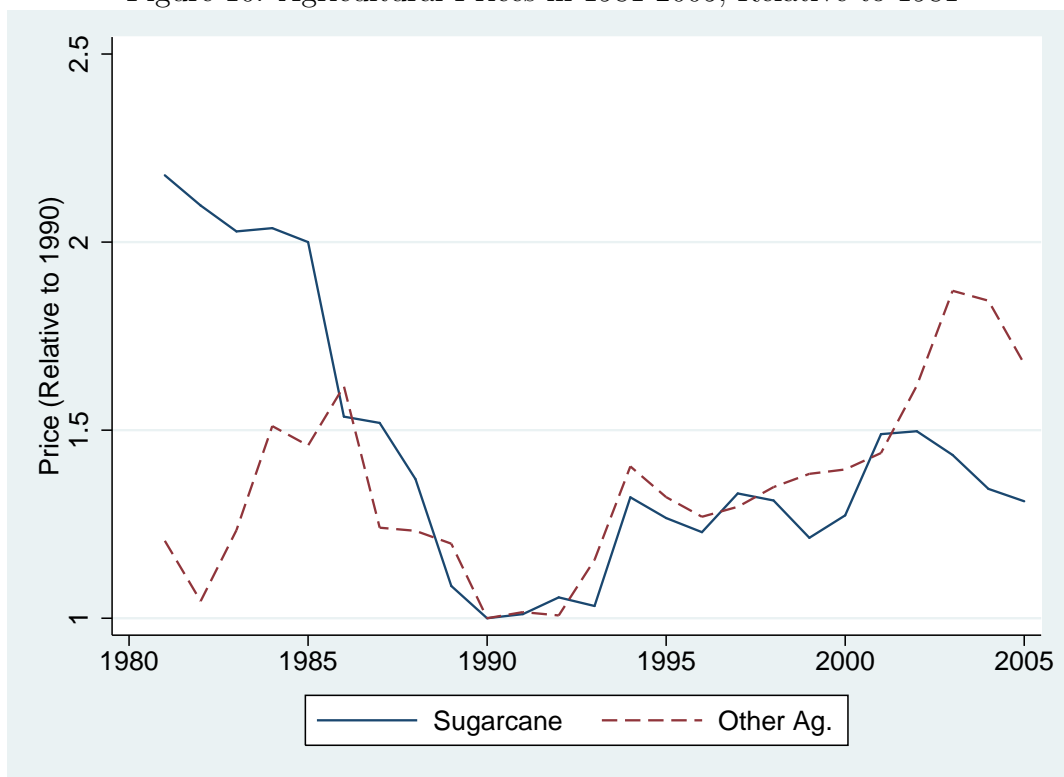


Figure 11: Median Hourly Wages, 1981-2005

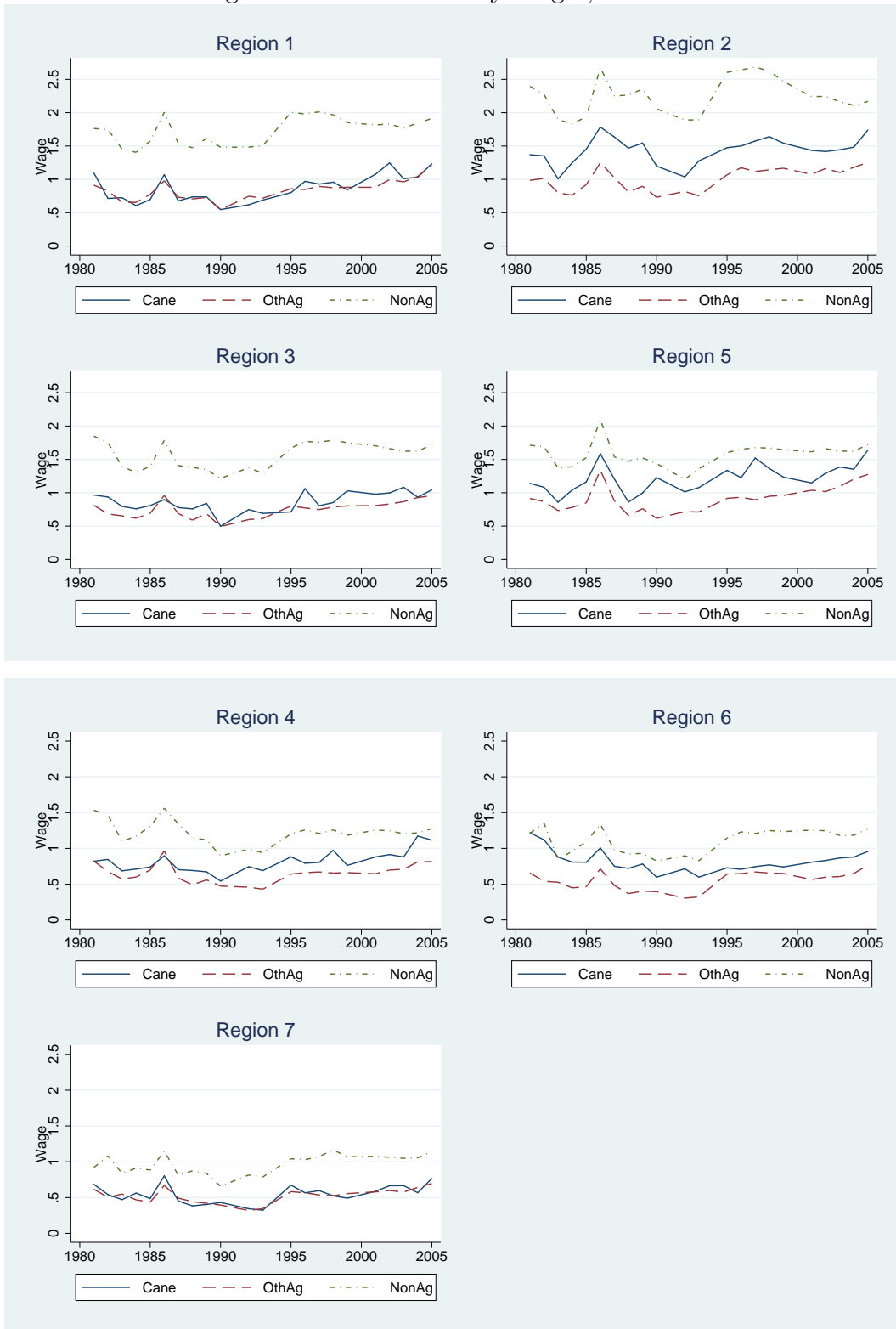


Table 13: Raw Data

| Category       | Available in Data   |
|----------------|---|
| Production     | $Y_{1rt}^S, Y_{2rt}^S, Y_{ert}^S, Y_{srt}^S$                                |
| Prices         | $p_{1t}, p_{et}, p_{st}, p_{2t}, p_{ot}, w_{jrt}$                           |
| Labor Hours    | $L_{1rt}, L_{2rt}, L_{3rt}$   |
| GDP Components | $K_t^{D*}, Y_{et}^{X*}, Y_{st}^{X*}, Y_{2t}^{X*}, G_t^*, VP_{rt}, GDP_{rt}$ |
| Land Shares    | $S_{1rt}, S_{2rt}, S_{3rt}$   |

Table 14: Use of Raw Data to Construct Quantities

| Can Use  | To Construct                   |
|--|--------------------------------|
| $GDP_{rt}$   | $GDP_t, GDP_t^N, GDP_t^*$      |
| $Y_{et}^{X*}, Y_{st}^{X*}, Y_{2t}^{X*}$ , zero North net ex. for each good                         | $Y_{et}^X, Y_{st}^X, Y_{2t}^X$ |
| $G_t^*$ , government budget balance  | $\tau_t^*$                     |
| $\tau_t^*, GDP_t, GDP_t^N, GDP_t^*$  | $\tau_t^N, \tau_t$             |
| $Y_{ert}^S, Y_{et}^X, Y_{srt}^S, Y_{st}^X, Y_{2rt}^S, Y_{2t}^X$                                    | $Y_{et}^D, Y_{st}^D, Y_{2t}^D$ |
| $r_t K_t^{D*}, GDP_t^*, \tau_t^*, GDP_t, \tau_t$   | $r_t K_t^D$                    |
| $GDP_t, Y_{et}^D, Y_{st}^D, Y_{2t}^D, r_t K_t^D, p_{1t}, p_{et}, p_{st}, p_{2t}, p_{ot}$           | $Y_{ot}^D$                     |
| $Y_{et}^X, Y_{st}^X, Y_{2t}^X, p_{1t}, p_{et}, p_{st}, p_{2t}, p_{ot}$ , trade balance             | $Y_{ot}^X$                     |
| $Y_{ot}^X, Y_{ot}^D$   | $Y_{ot}^S$                     |
| $L_{1rt}, L_{2rt}, L_{3rt}, w_{1rt}, w_{2rt}, w_{3rt}$   | $WB_t$                         |
| $GDP_t, WB_t, \tau_t$  | $M_t$                          |
| $VP_{rt}, p_{et} Y_{ert}^S, p_{st} Y_{srt}^S, p_{2t} Y_{2rt}^S, p_{1t} Y_{1rt}^S, p_{ot} Y_{ot}^S$ | $p_{ot} Y_{ort}^S$             |

Table 15: Wages in 2005 and Simulated Environments (2000 Reais)

|              | Sugarcane |       |       | Other Ag. |      |       | Non-Agriculture |       |      |       |       |       |
|--------------|-----------|-------|-------|-----------|------|-------|-----------------|-------|------|-------|-------|-------|
|              | 2005      | Sim 1 | Sim 2 | Sim 3     | 2005 | Sim 1 | Sim 2           | Sim 3 | 2005 | Sim 1 | Sim 2 | Sim 3 |
| Parana       | 1.23      | 3.56  | 1.32  | 4.41      | 1.21 | 2.37  | 1.85            | 1.43  | 1.91 | 1.34  | 1.29  | 1.46  |
| Sao Paulo    | 1.74      | 4.34  | 4.05  | 2.82      | 1.25 | 0.63  | 1.55            | 1.08  | 2.17 | 1.47  | 1.41  | 1.62  |
| Minas Gerais | 1.04      | 1.96  | 3.18  | 3.86      | 0.96 | 1.09  | 1.67            | 3.56  | 1.72 | 1.21  | 1.17  | 1.32  |
| Bahia        | 1.12      | 0.85  | 0.89  | 5.37      | 0.82 | 1.42  | 0.82            | 0.87  | 1.28 | 0.85  | 0.82  | 0.94  |
| Mato Grosso  | 1.64      | 2.27  | 5.21  | 2.78      | 1.28 | 1.17  | 3.48            | 0.56  | 1.72 | 1.19  | 1.14  | 1.30  |
| Pernambuco   | 0.96      | 1.84  | 2.39  | 2.89      | 0.77 | 2.20  | 0.76            | 4.35  | 1.28 | 1.01  | 0.99  | 1.07  |
| Maranhao     | 0.77      | 4.15  | 4.67  | 6.00      | 0.70 | 2.68  | 0.88            | 0.67  | 1.15 | 0.93  | 0.91  | 0.98  |

Table 16: Aggregate Hours in 2005 and Simulated Environments (Annual)

|              | Sugarcane |        |        | Other Ag. |         |         | Non-Agriculture |          |          |           |           |          |
|--------------|-----------|--------|--------|-----------|---------|---------|-----------------|----------|----------|-----------|-----------|----------|
|              | 2005      | Sim 1  | Sim 2  | Sim 3     | 2005    | Sim 1   | Sim 2           | Sim 3    | 2005     | Sim 1     | Sim 2     | Sim 3    |
| Parana       | 69.24     | 18.51  | 169.74 | 26.82     | 5175.14 | 991.92  | 3223.21         | 6090.88  | 20916.77 | 7320.23   | 43817.83  | 42873.41 |
| Sao Paulo    | 373.94    | 138.45 | 157.72 | 819.59    | 1765.36 | 8921.33 | 1789.94         | 4366.31  | 34870.87 | 69270.78  | 16498.45  | 27751.12 |
| Minas Gerais | 128.48    | 82.93  | 27.53  | 42.34     | 4319.71 | 2858.26 | 1913.91         | 370.32   | 28689.62 | 113.17    | 108169.18 | 2795.40  |
| Bahia        | 63.80     | 500.70 | 425.68 | 4.71      | 3777.05 | 732.56  | 6620.60         | 5647.27  | 7888.91  | 1106.41   | 1432.41   | 20.35    |
| Mato Grosso  | 59.88     | 81.87  | 12.97  | 117.77    | 2365.25 | 2530.36 | 444.36          | 28025.30 | 10268.13 | 2342.74   | 17.84     | 25429.32 |
| Pernambuco   | 535.24    | 208.28 | 92.49  | 122.39    | 2952.58 | 109.50  | 5552.31         | 30.99    | 9516.87  | 52605.26  | 5764.35   | 81999.10 |
| Maranhao     | 68.27     | 0.97   | 0.69   | 0.80      | 3995.02 | 97.92   | 3577.84         | 8035.80  | 8652.22  | 136942.85 | 35682.63  | 226.22   |