Faculdade de Economia,
Administração e Contabilidade
de Ribeirão Preto
Universidade de São Paulo

Texto para Discussão

Série Economia

How green
is sugarcane ethanol?

Marcelo Sant' Anna

Universidade de São Paulo

Faculdade de Economia, Administração e Contabilidade

de Ribeirão Preto

How green is sugarcane ethanol?

Marcelo Sant'Anna[*]

November 8, 2015

JOB MARKET PAPER

Abstract

Biofuels offer one approach for reducing carbon emissions in transportation. However, the agricultural expansion needed to produce biofuels may endanger tropical forests and thus offset the benefits of fossil fuel substitution. Whether this occurs depends on the extent to which increases in biofuels supply arise from gains in yields per acre or expansion in growing areas. I use a dynamic model of land use to disentangle the roles played by acreage expansion and yield increases in the supply of sugarcane ethanol in Brazil. The model is estimated using a panel of 1.8 million fields, which is built using remote sensing (satellite) information of sugarcane activities. My estimates imply that, at the margin, 94% of new ethanol comes from increases in area planted and only 6% from increases in yield. Direct deforestation accounts for 12% of area expansion. Balancing carbon emissions from deforestation and the carbon saved by fossil fuel substitution, I find that it would take about 20 years for the lower emissions from sugarcane ethanol to "pay back" the added emissions from deforestation. As an illustrative policy experiment, I consider the effects of a 5 billion gallon sugarcane ethanol mandate ($\sim$ 3% of US gasoline consumption). Such policy would lead to a 1% price increase and deforestation of about 9,000 sq. km. ($\sim$ ³/₄ the size of Connecticut).

1   Introduction

In the past decade the world has seen an unprecedented debate on climate change, mainly about how to reduce emissions of CO_2 and other greenhouse gases. Reduction of emissions is especially challenging in the transportation sector, which still relies heavily on fossil fuels.[1] Biofuels are an attractive tool for reducing carbon emissions in transportation, as they can be blended with petroleum fuels in unmodified vehicles. This is an advantage of biofuels compared to other lower carbon alternatives for transportation, such as electric vehicles, as it does not require changes in the vehicle stock or the refueling infrastructure. However, the agricultural expansion needed to produce biofuels may endanger tropical forests or other natural habitats and thus offset the alleged environmental benefits of fossil fuels substitution.[2] This deforestation could be reduced if more biofuels are produced by increasing agricultural yields in existing growing areas.

To assess the environmental benefits of biofuels, we need to understand the roles played by acreage expansion and yield increases as we move along the supply curve to meet the increased demand for biofuels feedstock. These issues are particularly important for Brazilian sugarcane ethanol, which accounts for 39% of the world ethanol supply. A large portion of Brazil is covered by tropical ecosystems, such as the Amazon rainforest, with important biodiversity and carbon storage. Figure 1 shows the remaining ecosystems in Brazil alongside existing sugarcane fields in 2013. Those ecosystems could be endangered by the expansion of farmland that would follow an increase in sugarcane ethanol demand.

In this paper, I disentangle the roles played by farmland expansion and agricultural yields in determining the supply of sugarcane ethanol in Brazil. I further estimate the effect of farmland expansion on forests and other untouched ecosystems.

The agronomic technology of sugarcane farming both guides my modeling and provides the means by which yields may respond to prices. Sugarcane is a semi-perennial crop, with declining expected yields over time until fields are replanted and yields restored. Replanting a field or planting sugarcane for the first time requires fixed land preparation investments that only pay off in future seasons, which makes optimal decisions forward-looking. The sugarcane growing technology gives a natural way in which prices can affect average yields through changes in the timing of replanting. By increasing the frequency of replanting, a farmer can boost average yields.

The most common types of biofuels policies, *e.g.*, ethanol mandates, imply permanent shifts in the demand for feedstock used to produce biofuels. Given the fixed costs of land

---

[1]Transportation accounted for 23% of energy-related emissions in 2010 and transport demand is on the rise in the developing world (IPCC, 2014).

[2]The Intergovernmental Panel on Climate Change (IPCC) recognizes a lack of consensus on the role biofuels should play in climate change policies. The IPCC Fifth Assessment Report (IPCC, 2014) recommends support for biofuels be given on a case by case basis.

Figure 1: Sugarcane and remaining forests in Brazil
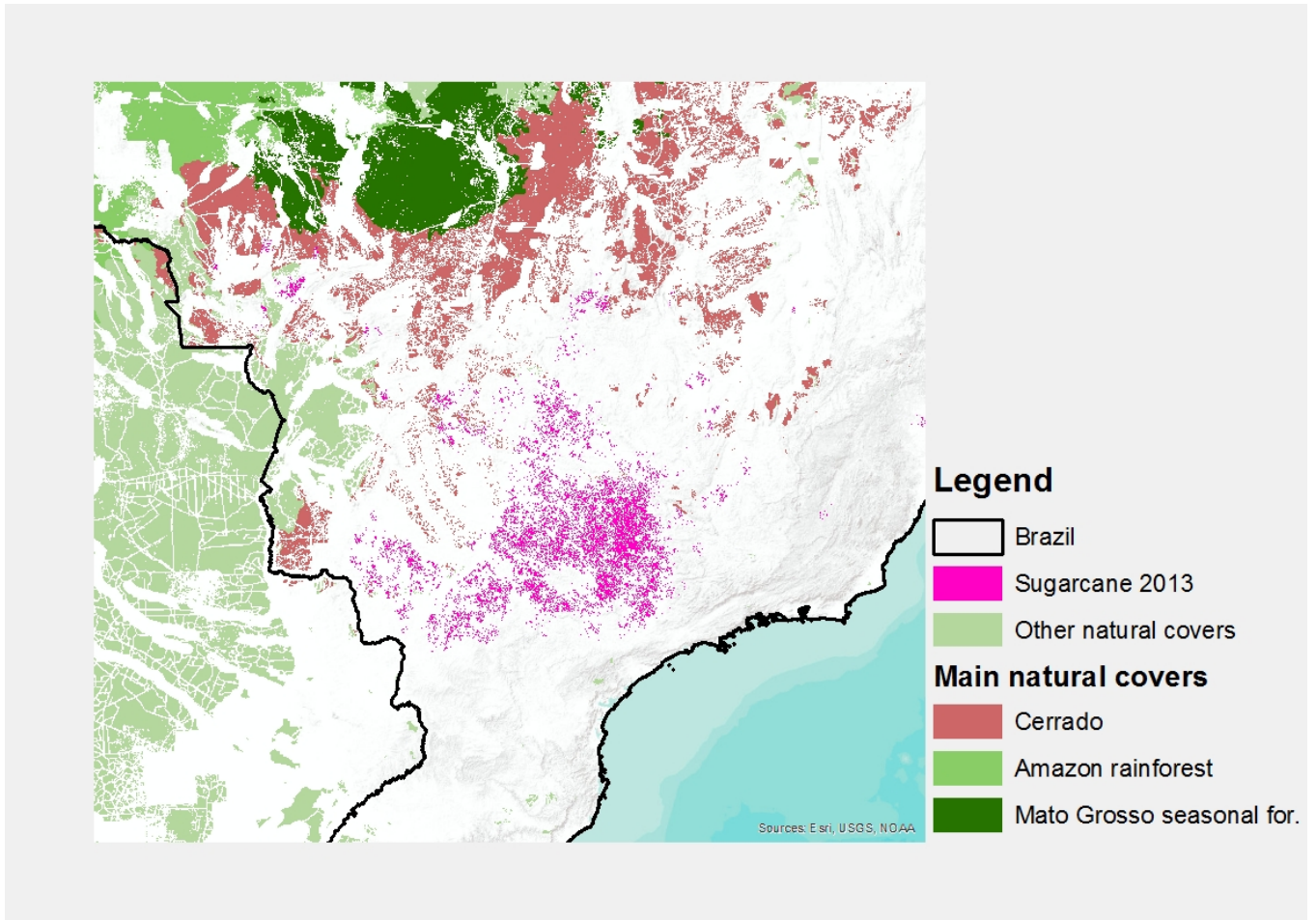
Sources: Information on areas of natural cover: Wild Data Version 2, 2005 (LWP-2): Global Human Footprint Dataset (Geographic). Wildlife Conservation (WCS) and Center for International Earth Science Information Network (CIESIN). Sugarcane areas: CANASAT, Rudorff et al. (2010).

preparation, an increase in price perceived as permanent generates different incentives in terms of sugarcane planting and replanting than transitory price shocks. In this industry, such clean permanent price changes are hard to see in the data. My approach is to use variation in agricultural profitability and replanting patterns to estimate a model that allows me to disentangle the effects of permanent price changes on the intensive (yield) and extensive (acreage) margins.

I incorporate both adoption of sugarcane and replanting in a single dynamic model. In the model, farmers planting sugarcane must decide at every period either to replant their sugarcane fields or not. If they do not replant, expected yields decline next period. If they choose to replant, they pay a fixed replanting cost and sugarcane yields are restored. This feature makes this problem similar to optimal machine replacement (Rust, 1987).[3] Farmers always have the option of switching away from sugarcane production to other uses. Meanwhile, farmers not planting sugarcane must decide at every period whether to plant sugarcane or not. Payoffs from both sugarcane and other uses depend on prices that are allowed to evolve stochastically.

I estimate the model using a unique remote sensing dataset of sugarcane land use in Brazil from the CANASAT project (Rudorff et al., 2010). The project processes satellite images from Landsat to provide detailed maps of sugarcane field activity for the most important sugarcane producing region of Brazil. I complement this land use data with GAEZ/FAO high-resolution potential yield information and other detailed field level land characteristics. Given the sparsity of good transportation network in remote areas of Brazil, transportation costs are an important source of variation in agricultural profitability. Therefore, I model a quality-adjusted transportation network, which is used to measure the transportation cost to destination markets.

I use the model estimates to decompose the sugarcane supply elasticity into farmland expansion (extensive margin) and yield (intensive margin) components. I find a long-run yield elasticity of 0.27, comparable to what is believed an upper bound for most annual crops in developed countries.[4] However, I find a value for the long-run acreage elasticity of 4.38, which is a different order of magnitude compared to the yield elasticity. The combined effects of acreage and yield translates into a high supply elasticity, which suggests small price effects from demand shifts. The high acreage to yield elasticity ratio I find means that, as we move along the supply curve, 94% of new ethanol comes from expansion in farmland and only 6% from yield increase that accrues from a faster replanting rate.

The high long-run acreage elasticity found here contrasts with existing measures of acreage price responses in the literature. Roberts and Schlenker (2013) measure short-run demand

---

[3]There is also an interesting parallel with optimal oil drilling (Kellogg (2014), Anderson et al. (2014)), as oil well productivity also decays exponentially as well pressure recedes.

[4]See Berry (2011), Scott (2013b) and Miao et al. (2015) for discussions on the potential range for annual food crops yield-price elasticities.

and supply elasticities for food crops and find inelastic demand and supply. Scott (2013a) estimates the acreage elasticity using a dynamic model and US land use data. He finds a higher long-run acreage elasticity once forward-looking behavior is taken into account. However, the acreage elasticity found by Scott (2013a) is still an order of magnitude lower than the acreage elasticity I find for sugarcane in Brazil. This difference can be due in part to the very active Brazilian agricultural frontier compared to the already consolidated American farmland. This highlights the danger of extrapolating measures of acreage elasticity from one country to another when evaluating land use changes. Using American numbers for long run acreage responses in Brazil would imply that 45% of new ethanol at the margin would come from yield increases. This would severely understate the environmental costs of biofuels policies as a smaller fraction of new ethanol would be coming from expansion in farmland.

The high-resolution nature of the data together with the estimated model allows me to make predictions about which types of land cover are affected by sugarcane expansion. Out of the total farmland expansion, 12% is predicted to be over forests and other types of natural cover, with the Cerrado ecosystem and the southern fringe of the Amazon rainforest being the most affected. Deforestation can be magnified by indirect land substitution as sugarcane takes over other cropland and pasture. The expansion of sugarcane in areas with previous agricultural use decreases the supply of other agricultural products and is expected to cause further expansion of farmland as the market re-equilibrates at a higher price level.

I use available empirical evidence to quantify these indirect effects in deforestation. In order to put this predicted deforestation in perspective, I balance the carbon released by direct and indirect deforestation and the carbon saved by replacing fossil fuels. I find that new ethanol "pays back" in terms of carbon in about 20 years. In contrast, corn ethanol produced in the US is expected to "pay back" in 167 years (Searchinger et al., 2008). Currently there is no consensus about payback times for sugarcane ethanol. Estimates of the sugarcane ethanol carbon payback time vary from 4 to more than 100 years in the scientific literature, depending on the type of land cover affected (Elshout et al. (2015), Gibbs et al. (2008), Fargione et al. (2008)). In this paper, I compute a carbon payback time that brings together the economics of land use and the current scientific knowledge about emissions from land use change.

As an illustrative policy experiment, I discuss the implications of current U.S. Renewable Fuel Standard (RFS) for land use in Brazil. The current standard assigns a total of 5 billion gallons in the Advanced Biofuels (ABF) category that need not be met by cellulosic biofuels. This is equivalent to about only 3% of the U.S. annual gasoline consumption. As of now, Brazilian sugarcane ethanol is the only viable large scale alternative to fill the ABF mandate. I find that a 5 billion gallons shift in the market demand for sugarcane ethanol would imply a modest 1% price increase, but about 2,000 sq. km in direct deforestation. This could be magnified to 9,000 sq. km (~ 3/4 the size of Connecticut) if indirect effects are considered.

This is not the first study to investigate the implications of ethanol policies in the context

of Brazilian sugarcane ethanol (e.g., De Gorter et al. (2013), Elobeid and Tokgoz (2008), Lasco and Khanna (2010) and Nagavarapu (2010)). Other studies in the literature use mainly static general equilibrium to evaluate the effects of policies in the markets for sugar and ethanol using supply elasticities derived from short-run responses to prices. An exception is Nagavarapu (2010), which estimates a static general equilibrium model for land use and labor allocation in the sugarcane industry in Brazil using micro level data on the worker decision, but aggregate data on land use.

The rest of the paper is organized as follows. Section 2 presents a short background of the sugarcane industry in Brazil. Section 3 describes in more detail the land use model. Section 4 presents the data and some descriptive analysis. Section 5 discusses the model estimation. Counterfactuals are discussed in Section 6. Section 7 concludes.

2   Industry background

Sugarcane has long history in Brazil, dating back to colonial times. Once sugar was the most important export commodity in the country, but its importance for the Brazilian economy has faded away over time. In the aftermath of the seventies oil shocks, a government program (PROALCOOL) was created to foster the use of sugarcane ethanol as a replacement for gasoline. Large scale production of sugarcane ethanol has been in place since then. In the past decade, the emergence of flex-fuel vehicles has given a new boost to the sugarcane ethanol industry.

Sugarcane has been historically grown close to the coast in Southeast and Northeast Brazil. Today, more than half of sugarcane in Brazil is produced in the State of São Paulo, where physical conditions are ideal for sugarcane growing. In general, suitable conditions for sugarcane include a warm and rainy growing season and a cooler and drier harvest season. The harvest season in the region studied here goes from April to November, depending on the location and the varieties used. Sugarcane is a semi-perennial crop, which means that after plants are cut, if the roots are untouched, a ratoon or stubble crop will follow. However, yields for the ratoon crop are expected to be lower every time this process is repeated (Crago et al. (2010), Macedo et al. (2008)). For this reason, periodically the field must be replanted so that yields can be restored. Agriculture manuals recommend replanting roughly every 5 years.

After harvest, sugarcane is transported to a nearby mill, which is usually located at close proximity to sugarcane fields. Sugarcane is bulky, so it would be uneconomical to ship cane long distances for milling. Moreover the sugars in the cane deteriorate quickly after it has been cut, so generally mills are not more than 40 km away from source fields. At the mill the sugarcane is crushed and the resulting liquid is either fermented to produce ethanol or processed to produce sugar. Modern mills are also thermal electricity generators.

They produce electricity by burning leftovers from the crushing process. This innovation significantly helped the sugarcane ethanol energy balance (Macedo et al., 2008). Although, there are constant improvements in the milling process, like electricity generation from fibrous materials, the milling technology is well known. There is an active market of equipment and machinery for mills and I know of no technological barriers to entry.

Most modern mills can produce both sugar or ethanol and switch production according to market conditions. This technological feature implies that under perfect competition the price for sugar and ethanol should follow closely together in the medium-run (De Gorter et al., 2013). I consider therefore an unified sugarcane final products market in my analysis, that includes both sugar and ethanol as final outputs. As discussed above, most of the controversy regarding sugarcane ethanol is on aspects related to land use and yields and not on emissions accruing from the industrial processes, as those are well understood.

3   Model

The unit of analysis is a field indexed by i, which is managed by a profit maximizing farmer. In each year, t, farmers must make a decision regarding land use for the next season. This decision is denoted by q_{it}. If farmers are not planting sugarcane, $q_{it} \in \{plant, \ stay\}$, i.e., they can either plant sugarcane or keep their fields in another economic use. For farmers already planting sugarcane, $q_{it} \in \{replant, \ keep, \ out\}$, i.e., they can (i) replant the sugarcane fields, (ii) keep the same plants for next season or (iii) switch land use to another activity.

I denote by $a_{it} \in \{0, 1, \ldots, \bar{a}\}$ the state of fields regarding its sugarcane use. If $a_{it} = 0$, the field is not in sugarcane use, while if $a_{it} \geq 1$, the field is in sugarcane use and a_{it} denotes the sugarcane field age. I denote by $w_{it} \in W$ the exogenous state vector, with information on prices of sugarcane products and alternative crops, land characteristics, transportation costs and distance to existing sugarcane fields. Finally, there is a state vector $\varepsilon_{it} \in \mathbb{R}^5$ which farmers observe but not the econometrician.

The flow payoff is given by:

$$\Pi(a_{it}, w_{it}, q_{it}, \varepsilon_{it}; \theta) = \pi(a_{it}, w_{it}, q_{it}; \theta) + \varepsilon_{it}(q_{it}), \tag{1}$$

where $\pi(a_{it}, w_{it}, q_{it}; \theta)$ is a function that depends only on observed state variables and on a vector of parameters to be estimated, θ. Equation 1 makes it clear that for each choice q_{it}, there is a different associated unobserved state $\varepsilon_{it}(q_{it})$.

I now describe in more detail $\pi(\cdot; \theta)$. For fields not in a sugarcane use, $a_{it} = 0$, the flow

payoff is given by:

$$\pi(0, w_{it}, q_{it}; \theta) = \begin{cases} \delta r_{it}, & \text{if } q_{it} = stay, \\ -\Psi_E(h_i, d_{it}; \theta), & \text{if } q_{it} = plant. \end{cases} \tag{2}$$

The return index r_{it} captures the payoff from non sugarcane uses. I defer the discussion about how the return index r_{it} is constructed to Section 4.4. If farmers choose to stay in other use, $q_{it} = stay$, then $a_{it+1} = a_{it} = 0$. The fixed cost of sugarcane planting $\Psi_E(h_i, d_{it}; \theta)$ depends on the previous land use,[5] h_i, and on the distance between the field and the closest existing sugarcane field at the time d_{it}. As discussed in the previous section, due to cane bulkiness, sugarcane fields are usually not more than 40 km away from a mill. Therefore, d_{it} proxies for mill proximity and $\Psi_E(\cdot; \theta)$ captures the cost of moving the agro-industrial complex (mills and other specific infra-structure) further into the agriculture frontier. The higher cost of planting sugarcane in land farther away from existing sugarcane activities helps to explain the sluggish pattern of sugarcane expansion we see in the data. Note that if farmers decide to plant sugarcane, they reap no sugarcane in the immediate season following their decision. Using the most common plant varieties, after planting (or replanting) sugarcane fields take one and a half years to be ready for harvest. If farmers choose to plant sugarcane, $q_{it} = plant$, then $a_{it+1} = 1$.

The flow payoff for fields in sugarcane use, $a_{it} \geq 1$, is given by:

$$\pi(a_{it}, w_{it}, q_{it}; \theta) = \begin{cases} (p_t^s - tc_i)\kappa\gamma^{a_{it}-1}y_i^s + x_i'\beta, & \text{if } q_{it} = keep, \\ -\Psi_R, & \text{if } q_{it} = replant, \\ \delta r_{it} - \Psi_A, & \text{if } q_{it} = out. \end{cases} \tag{3}$$

If farmers keep the sugarcane fields, their next expected yields are given by the sugarcane potential yields for field i, y_i^s, adjusted by the exponential decay in sugarcane productivity due to field age, $\gamma^{a_{it}-1}$. The parameter κ is a technological conversion factor from sugarcane to final product. Variables p_t^s and tc_i denote, respectively, the final product price and transportation cost to port. The vector x_i keeps track of land specific characteristics, such as climate, elevation, slope and soil type. Those characteristics should affect harvest and upkeep costs (e.g., the amount of fertilizer used), so they are allowed to shift the period return from sugarcane through the term $x_i'\beta$. Naturally, physical land characteristics in x_i should also influence sugarcane yields. However, the effect of those variables in expected yields is assumed

---

[5]I do not observe land use decisions over time for land not in sugarcane. I observe the state of non-sugarcane land use at a single year (2000), Ramankutty et al. (2010a). The curse of dimensionality in dynamic discrete choice models and this limitation in observed other land uses motivates the simple treatment of alternative land uses.

to be fully captured by the potential yield measure y_i^s.[6] Finally, if $q_{it} = keep$, the sugarcane field ages in the next season, i.e., $a_{it+1} = \min\{a_{it}+1, \bar{a}\}$. If farmers decide to replant, they pay a fixed cost Ψ_R. Analogously to sugarcane planting, there are no sugarcane related payoffs in the next season. The replanting decision resets field age: $a_{it+1} = 1$. Switching to other uses gives the farmer the return from those other uses, at a fixed land conversion cost Ψ_A, and sets $a_{it+1} = 0$.

In industry discussions, local weather shocks were usually listed as the most important factor affecting replanting decisions after sugarcane field age. Particularly bad weather may increase the costs of the agricultural operation necessary for field replanting as well as per period field upkeep costs. The effects of these weather shocks are captured by the state variable $\varepsilon_{it}(q_{it})$. Moreover, replanting decisions may not always coincide with the optimal decision from the agronomic point of view. For instance, mills must make sure they have a steady supply of sugarcane, therefore all source fields cannot replant at the same time. In this sense, $\varepsilon_{it}(q_{it})$ captures the additional noise introduced by other operational concerns.

Assumption 1. *The unobserved state variables, $\varepsilon_{it}(q)$, are independently and identically distributed over fields and time.*

Assumption 2. *The evolution of the exogenous state variables w is not affected by farmers decisions and ε, i.e., $F_{w_{it+1}|q_{it},\varepsilon_{it},w_{it}} = F_{w_{it+1}|w_{it}}$.*

Assumption 1 is standard in the dynamic discrete choice literature. Assumption 2 embeds two important underlying features. First, it implies that farmers are price takers, a reasonable assumption for agricultural products markets. Second, it implies that choice specific unobservables ε do not change expectations about the evolution of w.

I assume farmers discount future cash flows using a fixed discount rate $\rho < 1$. Farmers choose q_{it} every period conditional on $(a_{it}, w_{it}, \varepsilon_{it})$ in order to maximize the sum of future discounted flow payoffs:

$$\max_{q_{it}} E\left[\sum_{j=0}^{\infty} \rho^j \Pi(a_{i,t+j}, w_{i,t+j}, q_{i,t+j}, \varepsilon_{i,t+j}; \theta)|a_{it}, w_{it}, \varepsilon_{it}\right].$$

I rewrite below the dynamic optimization problem faced by farmers in the recursive Bellman formulation. In a non sugarcane state, a farmer has two options: leave the land in other use or convert the land to sugarcane. Therefore, the value function at $a_{it} = 0$ is

---

[6]In the model, I assume farmers use a fixed amount of inputs that gives the maximum attainable yield y_i^s. Input choice is an additional margin farmers could use to increase yields. Empirical evidence for high intensity agriculture suggests this margin should be small. For instance, Scott (2013b) finds an upper bound for the yield elasticity of 0.05 from changes in fertilizer use.

$$V_\theta(0, w_{it}, \varepsilon_{it}) = \max \{\pi(0, w_{it}, stay; \theta) + \varepsilon_{it}(stay) + \rho E\left[V_\theta(0, w_{it+1}, \varepsilon_{it+1})|w_{it}\right],$$
$$\pi(0, w_{it}, plant; \theta) + \varepsilon_{it}(plant) + \rho E\left[V_\theta(1, w_{it+1}, \varepsilon_{it+1})|w_{it}\right]\}. \quad (4)$$

At $1 \leq a_{it} \leq \bar{a}$, which are the productive sugarcane field states, the farmer can either keep the current field, replant or switch to other use. In this case, the value function is

$$V_\theta(a_{it}, w_{it}, \varepsilon_{it}) = \max \{\pi(a_{it}, w_{it}, keep; \theta) + \varepsilon_{it}(keep) + \rho E\left[V_\theta(\min\{a_{it} + 1, \bar{a}\}, w_{it+1}, \varepsilon_{it+1})|w_{it}\right],$$
$$\pi(a_{it}, w_{it}, replant; \theta) + \varepsilon_{it}(replant) + \rho E\left[V_\theta(1, w_{it+1}, \varepsilon_{it+1})|w_{it}\right],$$
$$\pi(a_{it}, w_{it}, out; \theta) + \varepsilon_{it}(out) + \rho E\left[V_\theta(0, w_{it+1}, \varepsilon_{it+1})|w_{it}\right]\}. \quad (5)$$

Assumptions 1 and 2 imply the expected continuation value does not depend on the present unobserved state ε_{it}. Moreover, by Assumption 2, current choices do not alter the distribution of w_{it+1} conditional on w_{it}. Let $v_\theta(a_{it}, w_{it}, q_{it})$ be the deterministic component of each choice's value, that is,

$$v_\theta(a_{it}, w_{it}, q_{it}) = \pi(a_{it}, w_{it}, q_{it}; \theta) + \rho E\left[V_\theta(a_{it+1}(a_{it}, q_{it}), w_{it+1}, \varepsilon_{it+1})|w_{it}\right],$$

where $a_{it+1}(a_{it}, q_{it})$ denotes the deterministic age transition given current age and choice. The optimal choice, or policy function, is given by:

$$q_\theta^\star(a_{it}, w_{it}, \varepsilon_{it}) = \arg\max_q v_\theta(a_{it}, w_{it}, q) + \varepsilon_{it}(q).$$

Since ε_{it} is unobserved to the econometrician, given observed state variables and parameters θ, we are not able to precisely determine the optimal choice. We can only recover a conditional choice probability (CCP) given the unobserved state distribution:

$$\Pr(q|a_{it}, w_{it}; \theta) = \int 1\{v_\theta(a_{it}, w_{it}, q) + \varepsilon_{it}(q) \geq v_\theta(a_{it}, w_{it}, q') + \varepsilon_{it}(q') \text{ for all } q'\}dG(\varepsilon_{it}).$$

Assumption 3. $\varepsilon_{it}(q)$ *is independently and identically distributed across alternatives with type 1 extreme value distribution.*

Assumption 3 implies the CCP has the usual logit form:

$$\Pr(q|a_{it}, w_{it}; \theta) = \frac{v_\theta(a_{it}, w_{it}, q)}{\sum_{q'} v_\theta(a_{it}, w_{it}, q')}. \quad (6)$$

The CCP is the basic building block of the likelihood approach I use to estimate the model's vector of parameters θ. Aguirregabiria and Mira (2010) provide a great review of

dynamic discrete choice models and estimation methods available.

4   Data and descriptive analysis

I combine data from several sources to estimate the model. Here I present these data and provide a brief discussion of key stylized facts. I begin by describing the construction of the panel dataset of sugarcane land use. Next, I describe the construction of the distance measure and transportation costs, followed by a brief description of other covariates and its sources. Details are left to Appendix A.

4.1   Tracking sugarcane fields

I use a remote sensing dataset of sugarcane field activity to build the panel of field age $\{a_{it}\}_{i,t}$. The CANASAT project mapped sugarcane fields and replanting decisions in the Center South region of Brazil for all years between 2004 and 2012 (Rudorff et al., 2010). The maps created by the CANASAT project are very detailed and sugarcane fields vary in shape and size.

I built the panel dataset of sugarcane activities by creating a 1 km grid of points covering all the region of interest and tracking land use decisions for each point of the grid over time.[7] The grid extends all geographical micro regions[8] with sugarcane fields in any sample period year. This procedure creates 1,855,224 grid points.

There was a substantial expansion in sugarcane acreage over this sample period. In 2004, less than 1% of the grid points were sugarcane fields. In 2012, this share increased to more than 4%. Almost half of the sugarcane fields in the region are in the State of São Paulo, which has 25% of its territory covered by sugarcane.

Figure 2 shows one specific producing area in detail. The shapes in the figures are sugarcane producing fields. They are colored according to the classification given by the CANASAT project. "Ratoon" refers to sugarcane that has not been replanted in the previous cycle; "replanted" refers to fields that were replanting in the previous cycle; while "replanting" refers to fields being replanted in the current cycle; finally, "expansion" refers to new fields that are for the first time available for harvest.

Land use decisions are tracked for each grid point over time, creating a panel data of sugarcane land use. Figure 3 shows the fraction of fields replanted by field age for various cohorts. The observed mode of replanting is 6 years, which is consistent with industry description of replanting decisions. Replanting at higher ages is also frequent, even though not the

---

[7]There is no standard way in which this grid should be constructed. As discussed in Scott (2013a), there is a trade-off between oversampling and the amount of extra information a thinner grid would provide. I know of no existing result on the optimal way of sampling grid points for this type of problem. I believe that the 1km choice of grid sparsity is a reasonable compromise between these two forces.

[8]Micro regions, defined by the Brazilian Institute of Geography and Statistics, are a disjoint set of municipalities with common social and geographical characteristics.

Figure 2: Sugarcane fields in detail

(a) 2011
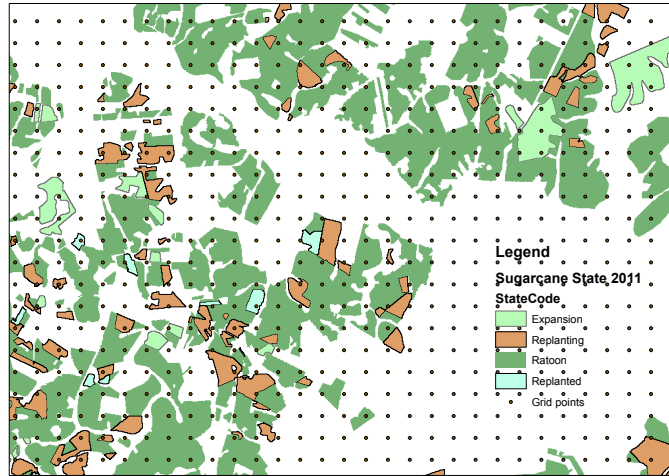
(b) 2012
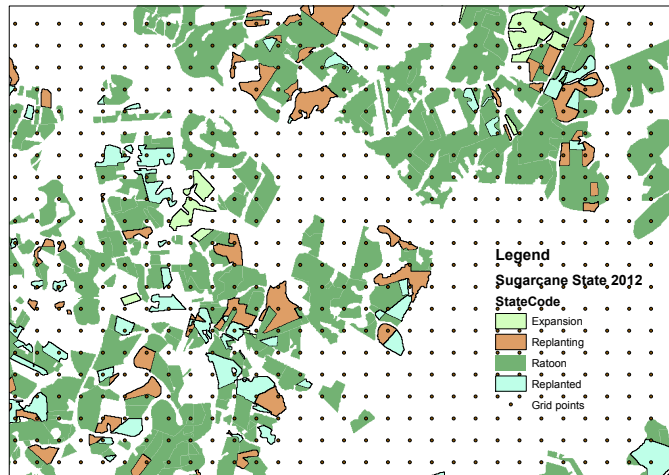
Notes: Maps showing a small fraction of the sample region for two different years (2011 and 2012) for illustrative purposes. The dots in both maps represent the grid points for which sugarcane activities are tracked over time. Colored shapes represent the different classifications given by the CANASAT project.

12

Figure 3: Replanting age by cohort

Note: Fraction of existing sugarcane fields that are reported as replanting in each age for different cohorts by the CANASAT project. The year assigned for each cohort represents the first year the field was available for harvest.

recommended practice by agronomists, who recommend replanting be done not after 5 years. This suggests that achieving higher yields would be possible by increasing the replanting rate.

There is an important caveat about the remote sensing information. Keeping track of the replanting decision requires the observation of 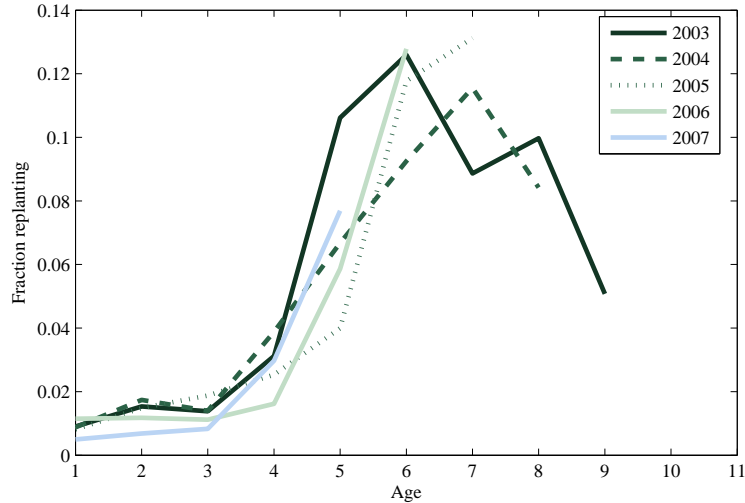more subtle variation in the satellite imagery at specific moments in the year than what is required to simply identify a sugarcane field. Depending on the variety used or on the time of the year replanting is done,[9] the imaging process may fail to identify the replanting activity. So it is expected that some fields are erroneously coded as not replanting in a given year, when in fact they were replanted. I deal with this issue explicitly when estimating the model.

4.2 Distances and transportation cost

I use data on the Brazilian road network from the Ministry of Transportation and the average speed on each road type to adjust for road quality. This allows me to measure road distance between two points taking into account the quality of the road network on the optimal path. More details about the construction of this transportation network are provided in Appendix A.2. Figure 4 shows the map of the available road network in the Center South Region. The network is more dense close to the shore and becomes more sparse as we move further into

---

[9]In general mills have excess capacity in non-harvest months, so it is not uncommon for some fields to be harvested slightly off season.

Figure 4: Road Network
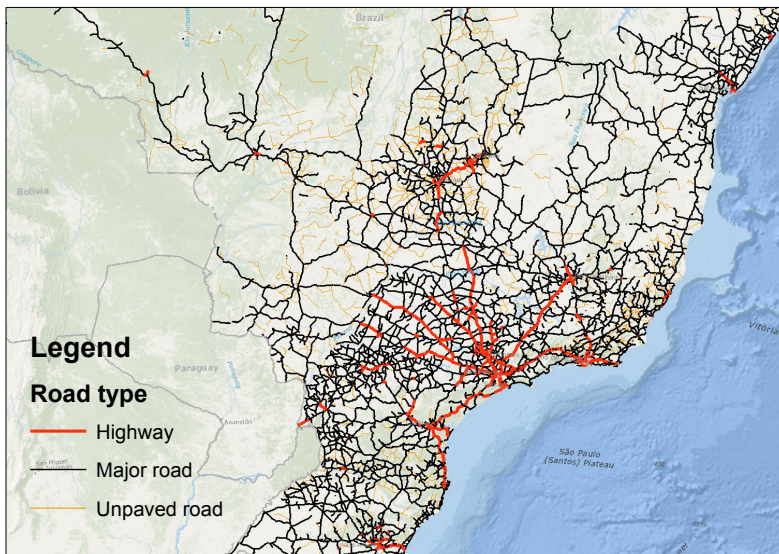
Notes: Map of the road network for South Central Brazil. Own elaboration with road data from the Ministry of Transportation.

the hinterland.

The first use of this distance measure is to compute, for each year, the distance of every grid point to the current closest sugarcane field (variable d_{it}). Figure 5 shows the relation between proximity of existing sugarcane fields and the decision to plant sugarcane for the first time. There is a sharp decline in the probability of sugarcane adoption as we move away from existing fields. This suggests a sluggish pattern of sugarcane expansion, as new sugarcane is usually planted very close to existing fields.

I further use the transportation network to compute transportation costs from every point in the grid to the closest maritime port. I complement the transportation network with actual freight quotes to estimate a simple model of transportation costs. I use freight quotes from SISFRECA (2008), which surveyed transportation quotes for moving sugar from 177 origins to one of three destination ports (Paranaguá, Guarujá and Santos).

The transportation cost model assumes a linear pricing schedule for freight rates. There is a fixed rate FC independent of distance traveled and a per kilometer on highway rate VC. Equation 7 describes the total cost of moving commodities from location i to j.

$$TC_{ij} = FC + VC \times EffectiveDistance_{ij} + \nu_{ij}, \tag{7}$$

where $EffectiveDistance_{ij}$ is on highway equivalent distance between i and j, and ν_{ij} is an error term, assumed exogenous. I compute effective distances for each one of the 177 origin-

Table 1: Transportation cost (2010 US$)

	Estimate	Std. Err.
FC	18.241	1.472
VC	0.039	0.003
N	177	

Notes: Estimated parameters and standard errors from the transportation cost model - Equation (7). Dependent variable is transportation cost between cities from SISFRECA (2008). Regressors are the corresponding effective distances computed using the road network from the Ministry of Transportation and average traveling speeds from SISFRECA (2008), details for the computation of effective distances in Appendix A.2.

destination pairs using the constructed transportation network and estimate fixed and variable costs in equation (7). Results are shown in Table 1. I estimate a US$18.24 fixed cost and US$0.039 per kilometer on highway variable cost of moving one tonne of sugar.

Finally, I compute the effective distances from every point in our grid to the closest maritime port. I consider as available ports all ports, that according to the Ministry of Transportation had reported any trading of sugar or ethanol. I use the estimated model and these effective distances to calculate transportation costs for all grid points (variable tc_i). Figure 5 shows that grid points with a lower transportation cost to ports had a higher conditional probability of planting sugarcane.

4.3 Other field characteristics

I use as sugarcane potential yield, y_i^s, the agro-ecological potential yield from FAO/IIASA (2011). FAO provides high-resolution potential yield for a variety of crops for all regions of the globe. Those potential yield measures take into account a wide range of soil and climate characteristics relevant for agricultural productivity. Figure 5 shows the relation between the measure I use of sugarcane potential yields and sugarcane planting. The conditional probability of sugarcane planting increases sharply for at values of potential yields above 7 ton. DW/ha. The variation in potential yields shifts the agricultural profitability in a similarly to permanent price changes. Therefore, this variation will be valuable to estimate the effects of counterfactual permanent price increases. The noticeable decline in planting at very high yields highlights the importance of accounting for alternative land uses, as potential yields for different crops are correlated.

The dataset used to estimate the model is complemented with extensive information on land characteristics described in Table 2, such as climate variables and elevation. Information on previous land economic use (variable h_i) is from Ramankutty et al. (2010a) and Ramankutty et al. (2010b), which classified all land in the globe into cropland and pasture.

15

Figure 5: Conditional probability of sugarcane expansion
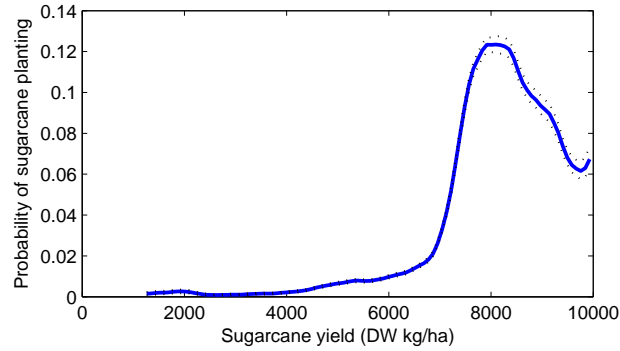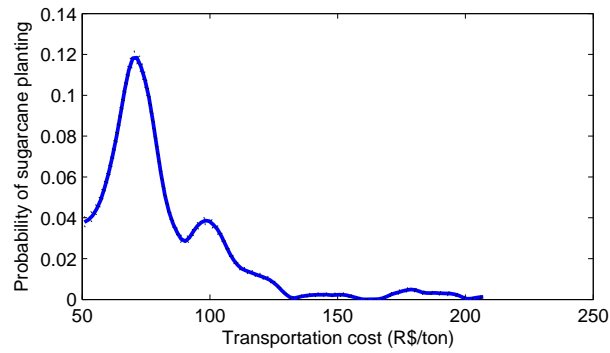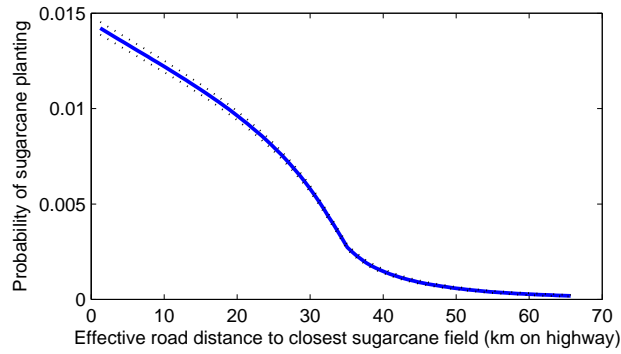
Table 2: Field characteristics

Variable	Mean	Std	Min	Max	p20	p50	p80
Sugarcane pot. Yield (kg DW/ha)	5791	2273	0	12305	3823	5933	7641
Corn pot. Yield (kg DW/ha)	4247	1971	127	11560	2658	3807	5994
Soy pot. Yield (kg DW/ha)	2546	767	0	4443	1943	2645	3241
Transportation cost (US$/ton)	66	26	21	125	42	62	93
Dist. to closest sugar field (km)	217.5	256.9	1.0	1589.2	29.1	133.6	335.9
Share in cropland	0.10	0.16	0.00	1.00	0.00	0.03	0.16
Share in pasture	0.43	0.28	0.00	1.00	0.13	0.44	0.70
Elevation (m)	518	245	0	2260	305	492	732
Prec. growth season (mm)	195.4	44.7	103.8	310.3	153.0	190.8	235.7

Notes: Descriptive statistics of field characteristics. There are 1,855,224 fields (or grid points) in the sample. The characteristics summarized above are available to all fields. Sugarcane, soy and corn potential yields are agroecological potential yields (high inputs) from FAO/IIASA (2011). Transportation cost and distance to closest sugarcane fields are own elaboration (discussed in text and Appendix A.2) using Ministry of Transportation data on roads and survey information from SISFRECA (2008). Cropland and pasture land cover information are respectively from Ramankutty et al. (2010a) and Ramankutty et al. (2010b). Climate and elevation data are from Hijmans et al. (2005).

This is a cross-section data on land use for the year 2000.[10] I note the higher average share of pasture (0.43) in our sample area, which has led many[11] to believe that most of new sugarcane fields would replace old pasture.

4.4 Prices

As argued in Section 2, I consider one single market for sugarcane final products, that includes both sugar and sugarcane ethanol. I use the NYSE price of sugar as the reference final product price. This was was pointed as the relevant reference price for decision making in industry discussions. The converted price series to Brazilian Reais (R$) is shown in Figure 6. The observed up and downward swings in prices provide variation in agricultural profitability that is helpful for the estimation of the model.

Equation 8 defines the return index for land not planting sugarcane. It is a weighted sum of the return of alternative agricultural commodities, with the weights given by the relative importance of those other crops in the field region:

$$r_{it} = \sum_{c \in \{corn, soy\}} p_t^c \alpha_i^c y_i^c, \tag{8}$$

---

[10]Unfortunataly, there is no panel remote sensing information that I know of for other crops or pasture for this region, except for sugarcane.

[11]See for example EPA (2010).

Figure 6: p_t^s (sugar price) and r_{it} (other uses return) series
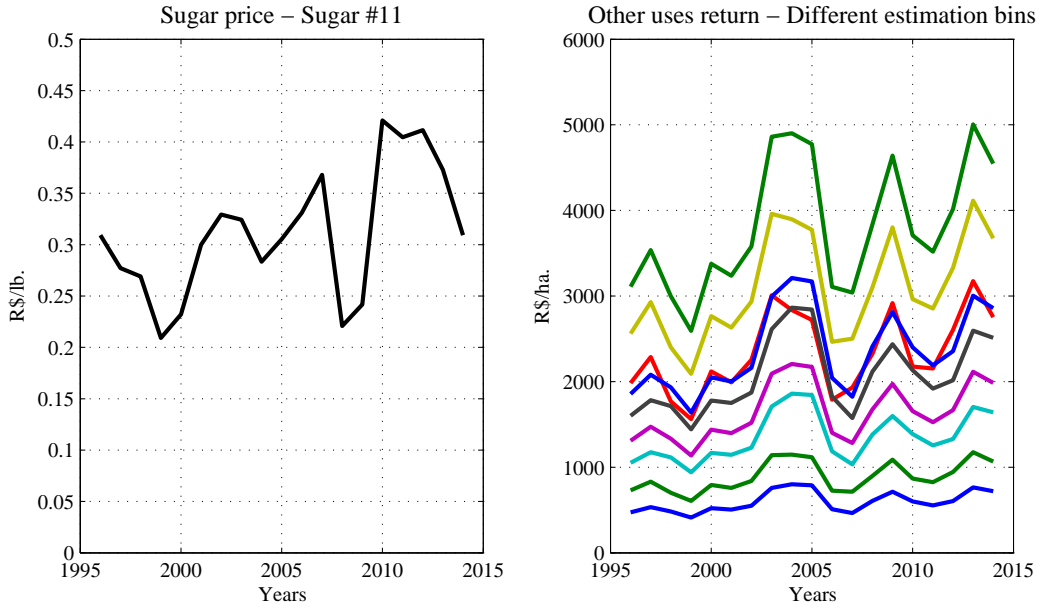
Notes: Prices used are NYSE future contracts (SB1 for sugar, C1 for corn and S1 for soy) all measured in 2010 Reais (R\$/lb), deflated using US and Brazilian CPI. All series from 1995 to 2013.

where p_t^c denotes the price of crop c, α_i^c is a measure of the share[12] of crop c in the region of field i and y_i^c is the FAO/IIASA potential yield of crop c in field i. Corn and soy correspond to more than half the cropland not in sugarcane in this region. Note that this is a continuous state variable since potential yield measures are continuous. For practical estimation purposes, I bin the grid points based on cross section categories for r_{it} based on the expected returns for the alternative crops. Figure 6 plots the index r_{it} over time for the different estimation bins.

5   Estimation

I estimate the parameters in the model by Maximum Likelihood. The goal of the estimation is to recover the vector of parameters θ from observed states $\{w_{it}, a_{it}\}_{it}$ and decisions $\{\tilde{q}_{it}\}_{it}$, where I make the distinction between observed decisions $\tilde{q}_{it}$ and actual decisions q_{it}. This distinction will be more clear shortly. Assumption 2 implies that the evolution of the exogenous state does not depend on current endogenous states or field management decisions.

---

[12]Shares of land used for corn and soy are obtained from the 2006 Agricultural Census and refer to Meso Administrative Regions in the Brazilian Institute of Geography and Statistics (IBGE) classification.

Therefore, we can write the conditional log-likelihood criterion function as

$$\mathcal{L}(\theta; \{\tilde{q}_{it}, w_{it}, a_{it}\}_{it}) = \sum_t \sum_i \log f(w_{it}|w_{it-1}; \theta) + \sum_i \log \left(\Pr(\tilde{q}_i|w_i, a_i; \theta)\right), \qquad (9)$$

where I omit the subscript t to denote the whole vector of decisions and states.

Note that the only exogenous state variables in w_{it} that change over time are (p_t^s, r_{it}). For estimation purposes, I assume (p_t^s, r_{it}) follow an AR(1) process:

$$\begin{bmatrix} p_t^s \\ r_{it} \end{bmatrix} = \begin{bmatrix} k_s \\ k_{r_i} \end{bmatrix} + \begin{bmatrix} \lambda_s & 0 \\ 0 & \lambda_{r_i} \end{bmatrix} \begin{bmatrix} p_{t-1}^s \\ r_{it-1} \end{bmatrix} + \eta_t, \qquad (10)$$

where

$$\eta_t \sim N\left(0, \begin{array}{cc} \sigma_s^2 & 0 \\ 0 & \sigma_{r_i}^2 \end{array}\right).$$

The exogeneity of the w_{it} transition implies that we can estimate k, λ and σ in a separate first step. I then treat those parameters as known when maximizing the second term in the likelihood function (equation 9) with respect to the payoff parameters. This procedure is discussed in more detail now.

The remote sensing exercise may fail to capture replanting if it happens in an unusual period of the year or depending on the sugarcane varieties used. This creates classification error, as some fields will be classified as not replanting, $\tilde{q} = $ keep, when indeed replanting happened. If a field is not coded as replanting, I cannot be sure this actually reflects classification error or just a long sugarcane cycle. If left untreated, this issue could bias upwards our estimated cost of field replanting. I treat classification error as a field specific unobserved state $u_i \in \{1, 2\}$. If $u_i = 1$, there is no classification error on field i remote sensing observations. In this case, $\tilde{q}_{it} = q_{it}$, *i.e.*, observed and actual decisions are the same. If $u_i = 2$, replanting decisions on field i are not observed, that is, $\tilde{q}_{it} = $ keep even when $q_{it} = $ replant. Note there is only classification error for replanting, the decision to plant sugarcane is not subject to observational problems. I assume $\Pr(u_i = 1) = \mu$ for all i. We can write the conditional probability of observed choices as:

$$\Pr\left(\tilde{q}_i|w_i, a_i; \theta\right) = \mu \prod_t \Pr\left(\tilde{q}_{it}|w_{it}, a_{it}, u_i = 1; \theta\right) + (1 - \mu) \prod_t \Pr\left(\tilde{q}_{it}|w_{it}, a_{it}, u_i = 2; \theta\right). \quad (11)$$

The choice probability $\Pr\left(\cdot|w_{it}, a_{it}, u_i = 1; \theta\right)$ is exactly the model's CCP (equation 6) for all $a_{it} \in \{0, 1, \ldots, \bar{a}\}$, since in this case there is no classification error. However, $\Pr\left(\cdot|w_{it}, a_{it}, u_i = 2; \theta\right)$ is only equal to the model CCP when $a_{it} = 0$, that is, for fields not planting sugarcane. For $a_{it} > 0$ and $u_i = 2$, only $\tilde{q}_{it} = $ keep is coded, so $\Pr\left(\text{keep}|w_{it}, a_{it}, u_i = 2; \theta\right) = 1$.

The model's CCP depends in principal on the full solution of the dynamic discrete choice

19

problem since $v_\theta(a_{it}, w_{it}, q)$ depends on the continuation value $E\left[V_\theta(a_{it+1}(a_{it}, q_{it}), w_{it+1}, \varepsilon_{it+1})|w_{it}\right]$. Solving the dynamic discrete choice problem by value function contraction at every different likelihood evaluation is computationally demanding. I use the Nested Pseudo Likelihood (NPL) method proposed by Aguirregabiria and Mira (2002) to circumvent this problem. Instead of solving the dynamic problem at every different likelihood evaluation, this method uses a single contraction in the space of CCPs. I embed the NPL algorithm with an Expectation Maximization (EM) step to account for classification error following Arcidiacono and Miller (2011).

Standard errors for payoff parameters are computed by bootstrap. The bootstrap procedure follows the estimation steps. For each bootstrap repetition, I re-estimate transportation costs to ports using a bootstrap sample of freight quotes. Additionally, I re-estimate the transition process (equation 10) using a parametric bootstrap sample of prices. I then re-estimate the full model with a block bootstrap sample of observations. I discuss estimation details in Appendix B.

5.1   Estimation results

Table 3 shows the estimates for the processes in equation (10). In the first column we present the bin number for each category of r_{it}. The second and third column show the values of the weighted yields for corn and soy that define each bin. The other columns show estimates and standard errors for the auto-regressive processes. The estimated transition for p_{ts} is shown at the bottom of the table.

There is no sign of violation in the stationarity assumption in any of the processes, even though standard errors are relatively high. There is a trade-off here between using a longer price series in dollars and a shorter series measured in the Brazilian currency, Real (R\$), which was only adopted in 1994. I opt for the second, since this seems to be the appropriate reference price, especially in terms of volatility, for decision makers in this market. Moreover, it is not clear that using very old price information adds much to the analysis given changes in market dynamics in recent decades.[13] The cost of relying on relatively recent price information is a short time series. Imprecision in the estimation of those processes will be taken into account when we compute standard errors for the dynamic model estimates.

In the model section, the functional form for the cost of sugarcane planting, $\Psi_E(h_i, d_{it}; \theta)$, was left unspecified. In estimation, I use the following empirical specification for this cost:

$$\Psi_E(h_i, d_{it}; \theta) = \sum_{l=\{crop, pasture, other\}} \psi_h^l 1\{h_i = l\} + \psi_d^1 d_{it} 1\{d_{it} \le 40\} + \psi_d^2 1\{d_{it} > 40\}. \quad (12)$$

---

[13]For instance, it was common in the past to observe price spikes whenever there were stock-outs (Deaton and Laroque, 1992). Those type of stock-outs were not seen in recent decades.

Table 3: Exogenous state variables transition (std. err.)

(a) r_{it} transition

Est. Bin	k_{r_i}		λ_{r_i}		$\sigma^2_{r_i}$	
1	0.241	(0.144)	0.604	(0.228)	0.010	(0.004)
2	0.396	(0.233)	0.561	(0.251)	0.022	(0.009)
3	1.208	(0.686)	0.486	(0.285)	0.160	(0.065)
4	0.524	(0.307)	0.621	(0.213)	0.051	(0.024)
5	0.665	(0.396)	0.603	(0.228)	0.073	(0.034)
6	1.444	(0.841)	0.537	(0.262)	0.264	(0.111)
7	0.793	(0.462)	0.624	(0.209)	0.122	(0.057)
8	0.929	(0.550)	0.614	(0.220)	0.153	(0.071)
9	1.681	(0.990)	0.564	(0.250)	0.392	(0.170)

(b) p^s_t transition

k_s		λ_s		σ^2_s	
0.161	(0.087)	0.482	(0.278)	0.003	(0.001)

Notes: Estimation of the transition process in equation (10). Standard errors (asymptotic) in parenthesis. Prices used are NYSE future contracts (SB1 for sugar, C1 for corn and S1 for soy) all measured in 2010 Reais (R\$/lb), deflated using US and Brazilian CPI. All series from 1995 to 2013.

Table 4 shows the results from the NPL estimation of the payoff functions (equations 2 and 3) and fixed cost parameters. I find a steep rate of yield decay from sugarcane field age of 0.79, which represents an expected yield half life of approximately 3 seasons. Current available information about yield decay for sugarcane is restricted to surveys for specific regions and years. For instance, Crago et al. (2010) finds a less steep yearly decay rate of 0.86 using survey information restricted to the state of São Paulo in 2007. Their assessment, however, only considers the first 5 harvests. Taking into account that replanting takes one season, a decay rate of 0.79 implies that the replanting age that maximizes average expected yields is 3 years. This is in line with recommended agronomic practices, that suggest 3 years as a minimum age for field replanting. However, given the positive fixed cost of replanting, we rarely see such short sugarcane cycles, as observed in Figure 3.

I estimate a lower conversion cost for land previously being used for crops in comparison to land used for pasture or in other use. Consistent with descriptive evidence discussed before, I find a high penalty in the cost of planting sugarcane from the distance to existing sugarcane fields. This penalty implies that the increase in fixed cost associated with moving away from existing sugarcane fields by only 10 km is approximately 40% of the revenue associated with the first sugarcane cut of an average sugarcane field.

Table 4: Payoff and fixed cost estimates

Parameter	Estimate	Std. Err.	Parameter	Estimate	Std. Err.
γ	0.791	0.013	ψ_h^{other}	3.401	0.909
δ	0.052	0.021	$\psi_h^{pasture}$	3.301	0.913
κ	0.692	0.053	$\psi_h^{cropland}$	3.142	0.897
β			ψ_d^1	0.062	0.003
Elevation	-0.267	0.019	ψ_d^2	3.581	0.104
Elevation sq	0.033	0.003	Ψ_A	6.262	1.006
Prec. Growth	-0.178	0.050	Ψ_R	5.067	0.125
Prec. Growth sq.	0.022	0.006	μ	0.7528	0.000

Notes: Estimates and standard errors for payoff and fixed cost parameters in equations (2) and (3). Estimation by ML using the NPL algorithm. Standard errors computed by bootstrap.

6 Counterfactuals

6.1 Counterfactual elasticities

I use the estimated model parameters to compute a supply elasticity from permanent price changes, which is then decomposed in yield and acreage components. There is a challenge in evaluating responses to permanent price changes in this setting. The importance of proximity to other fields for sugarcane adoption creates path dependence in the expansion pattern of new sugarcane fields. Therefore, initial conditions matter to determine the system evolution. Motivated by this, I compute price elasticities in the following way. Starting from the current state of all fields in the sample at the last year available (2012), I simulate a small 1% price increase in the average of the process governing the evolution of sugar price. I compare the evolution of this system under the new price regime to the baseline case of no price increase to calculate acreage and yield effects attributable to the price increase. The acreage effect refers to the additional number of grid points in sugarcane in the counterfactual simulation in comparison to the baseline case. The yield effect refers to the change in average output per field that follows the price increase.

Long-run supply side elasticities are reported in Table 5. I find an acreage elasticity ξ^L of 4.3, which is 16 times higher than the estimated yield elasticity ξ^Y, for which I find a value of 0.27. This acreage to yield elasticity ratio means that, at the margin, only 6% of the increase in long-run supply comes from higher yields, the other 94% comes from acreage expansion. The combined effect of acreage and yield elasticities translates into a supply elasticity of 4.6.

The small contribution of yields at the margin does not mean the yield effect generated by a more intensive pattern of replanting studied here is small. In fact, the yield elasticity ξ^Y is actually in the ball park of estimates for different crops.[14] The novelty here is actually

---

[14]There is little consensus however on the magnitude of yield elasticities. Scott (2013b) uses indirect evidence

Table 5: Long-run elasticities

Supply	ξ^S	4.6158
Acreage	ξ^L	4.3372
Yield	ξ^Y	0.2721

Notes: Long-run supply elasticities computed by simulating (2000 simulations) the evolution of the fields in our dataset after a 1% price increase.

the very high acreage elasticity. In comparison, Miao et al. (2015) find a corn yield elasticity of 0.23 and an acreage elasticity of 0.45 using US data. In the context of global agriculture supply, Roberts and Schlenker (2013) find a much lower acreage elasticity of 0.1. Using a forward looking model of land use in the US, Scott (2013a) finds a higher acreage elasticity of 0.3, still much lower than the value found here.

Three effects combine to determine the high acreage elasticity for sugarcane in Brazil I document here in comparison to other studies of land use. First, I use a dynamic model to derive the long-run elasticity, while some of the previous studies have only focused on short run acreage responses (Roberts and Schlenker (2013), Miao et al. (2015)). In this sense my results go in the same direction as Scott (2013a), who finds higher acreage price elasticity once forward looking behavior is taken into account. Second, as in Miao et al. (2015), this paper focuses on a single crop, so there is the possibility of cross crop substitution. This is in contrast to Roberts and Schlenker (2013) and Scott (2013a), which aggregate agricultural markets and consider responses to an aggregate price index. Third, Brazil has a very active agricultural frontier, with large extents of undeveloped land, still shy from realizing its agricultural potential.

In this sense, my results highlight the pitfalls of extrapolating cross-country measures of acreage elasticity to study land use. Using the acreage elasticity estimated by Scott (2013a) for the US in our analysis would imply that 45% new ethanol at the margin would come from the intensive margin. This would imply less expansion in farmland as we move along the supply curve and thus a downward bias in expected deforestation.

Figure 7 shows the evolution of output, acreage and yield elasticities in the first years after the permanent price shock. The permanent price change encourages expansion of sugarcane to new areas. The expansion of sugarcane in comparison to the baseline translates into acreage elasticities reported in Figure 7. Those new sugarcane fields have higher expected yields than the average pool of sugarcane fields, so yield elasticity jumps in the first periods. As those new areas age, the yield elasticity converges to its long-run value, which reflects the more intensive pattern of replanting that follows the permanent price increase. The output elasticity

from fertilizer use and finds that yield elasticities for corn are unlikely to be larger than 0.04. Miao et al. (2015) presents a good summary of other results in the literature, which range from not statistically significant to 0.61, depending on the crop studied and methodological approach.

Figure 7: Elasticities on the path to the long-run
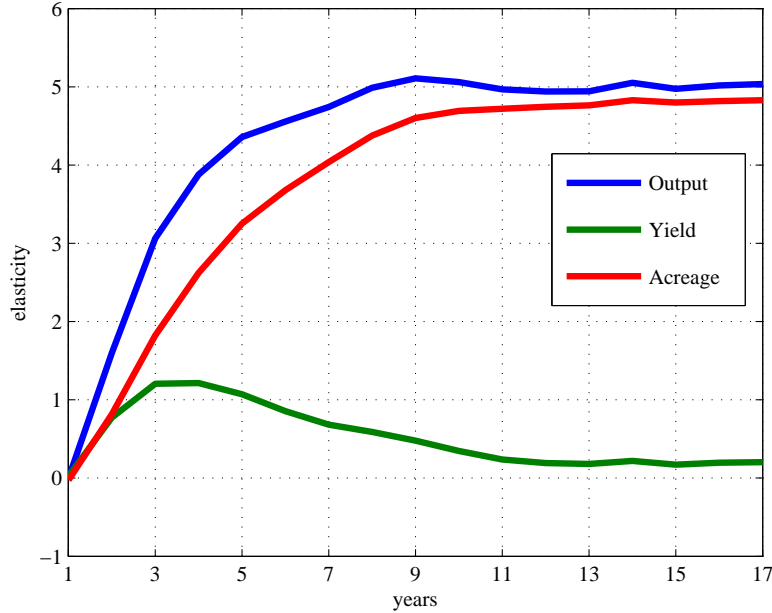
Notes: Output, yield and acreage elasticities path to long-run. Elasticities computed by simulating (2000 simulations) the evolution of the fields in our dataset after a 1% price increase.

combines the effects of yield and acreage elasticities. This explains why the output elasticity is steeper than acreage elasticity in the first periods and the faster speed of convergence to the long-run, in comparison to acreage. Although, the focus of this paper is not the short-run dynamics of this market, it is still interesting to note how the yield and acreage effects studied here combine in the short-run to generate a faster response of output in comparison to acreage.

6.2 Deforestation and carbon "payback" times

The acreage and yield elasticities reported in Table 5 suggest that almost all new sugarcane produced following demand shifts in ethanol would come from new growing areas (extensive margin) and not from more intensive replanting cycles (intensive margin). This is a reason for environmental concern, since land use change accounts for a significant part of world greenhouse gas emissions.[15]

In fact, most of the controversy regarding the use of ethanol biofuels comes from the trade-off between the one-shot emission from land conversion and the carbon emissions avoided over

---

[15]According to IPCC (2014), deforestation accounted for 12% of global anthropogenic CO_2 emissions.

time by replacing fossil fuels by a renewable source. A standard measure used to describe this trade-off is the carbon payback time, *i.e.*, the time it takes for the benefits from replacing fossil fuels to compensate the land use change emissions.[16] For the case of sugarcane ethanol there is little consensus for carbon payback times. This could vary from 5 to more than 100 years, depending on assumptions regarding the type of land cover substituted by sugarcane (Searchinger et al. (2008), Elshout et al. (2015), Gibbs et al. (2008), Fargione et al. (2008)). If the land converted to sugarcane fields comes from areas of natural cover with high carbon storage, *e.g.*, tropical forests, the carbon payback time is going to be high. In turn, if land with a lower carbon retention is converted, *e.g.*, cropland and pasture, the carbon payback time will be smaller.

I use the estimated model to predict the direct effect of sugarcane expansion in different natural ecosystems from permanent price shifts. Table 6 shows a decomposition of acreage changes on different areas of natural cover and on existing cropland and pasture. The first column reports acreage elasticities of sugarcane for the different land covers. The second column reports the share of sugarcane predicted by the model in the long-run. The second to last column shows the share of each type of land converted to sugarcane at the margin in the long-run.

The last column in Table 6 reports expected carbon emissions from each natural cover for an 1 km^2 sugarcane expansion at the margin.[17] Even though the Cerrado region is the one with the highest predicted decrease in natural cover, emissions of the same magnitude are predicted to come from land conversion of the two ecosystems connected to the Amazon rainforest, the Madeira-Tapajós and Mato Grosso seasonal forests. This reflects the higher carbon density in the Madeira-Tapajós and Mato Grosso forests compared to the Cerrado.

A few caveats are in order. First, the model does not distinguish between areas of natural cover and cleared areas not in cropland or pasture. The predicted expansion reported ignores specific fixed costs of land clearing and environmental regulation that could restrict deforestation. In this sense, my measure of deforestation and carbon emissions are worse case scenarios. However, this is not the case for pasture and cropland conversion to sugarcane, as I allow the fixed cost of sugarcane planting to vary depending on these two types of land use. Moreover, my analysis focuses only on the Center-South region of Brazil, which includes only the Southern fringe of the Amazon rainforest, which is likely to be the most affected by the expansion of farmland. For a specific treatment of the demand for deforestation in the entire Brazilian Amazon rainforest, see Souza-Rodrigues (2015).

The carbon emission reported in Table 6 represents only direct deforestation by sugarcane expansion, which could be aggravated by indirect land use substitution. Note that, out of

---

[16]It is important to note that the carbon payback time is not a sufficient measure for a normative analysis. However, several studies of biofuels have used this measure, allowing for comparison across studies.

[17]The assessment of carbon emissions comes from IPPC guidelines for the evaluation of emissions from land use change (IPCC, 2006); I leave the details of this computation to Appendix C.

Table 6: Acreage elasticity decomposition

	Acreage Elasticity	Baseline Share	Total Area km^2	Share of Expansion	Δ Carbon (ton CO_2)
Wild area classes					
Cerrado	5.14	0.074	159,927	4.7%	2,204
Chiquitano dry forest	5.62	0.073	23,571	0.8%	353
Madeira-Tapajós forest	4.25	0.132	81,611	3.5%	2,499
Matogrosso forest	5.52	0.171	45,449	3.3%	2,345
Atlantic forest	2.63	0.393	867	0.1%	49
Total wild			311,425	12.3%	7,451
Cropland vs. Pasture					
Pasture	4.10	0.187	799,467	47.4%	
Cropland	4.71	0.217	188,335	14.8%	

Notes: *Baseline Share* refers to the share of sugarcane in each land class predicted in the long-run. *Share of Expansion* refers to the share of marginal sugarcane acreage increments that come from different types of land cover. $\Delta Carbon$ gives predicted emissions (in tonnes of CO_2) from land use change in different wild covers from a 1 sq. km increase in sugarcane acreage. See Appendix C for details on the computation of carbon emissions.

sugarcane expansion at the margin, 62% is expected to come from areas already in some economic use, either in cropland or pasture. The substitution of these areas for sugarcane will arguably decrease the supply of agricultural products and may cause additional deforestation as prices of those products increase. The amount of additional land use change from this indirect channel will crucially depend on the relative importance of demand and supply elasticities for the agricultural commodities displaced. Estimates in the literature for those elasticities (Roberts and Schlenker, 2013) suggest that 2/3 of productive areas displaced by sugarcane could move further into the agriculture frontier, causing additional deforestation.[18]

In order to put those values of carbon emissions from land use change in perspective, one should consider the carbon that could be saved over time by replacing fossil fuels by sugarcane ethanol. Not taking into account emissions from land use change, sugarcane ethanol is supposed to avoid 84% of carbon emissions from gasoline, or $1,979 \, kg \, CO_2$ equivalent per m^3 (Macedo et al., 2008). This measure already includes the use of fossil fuels and energy in the several stages of the production of sugarcane ethanol, including farm operations and the manufacturing of ethanol from sugarcane.

I balance the carbon that is expected to be released by direct and indirect deforestation with the carbon saved by fossil fuel substitution to compute long-run carbon payback times

---

[18]There is little evidence however on the direct relationship between the expansion of sugarcane and deforestation in the Amazon region through indirect land use change (de Sá et al., 2013). Nevertheless, this indirect land use change could happen in other parts of the world if the agricultural commodities affected are traded internationally.

Table 7: Emissions and payback time (1,000L new ethanol)

ILUC scenario	0%	66%	100%
CO2 emissions (tonnes)	9.6	41.5	57.9
Payback time (years)	4.8	20.9	29.2

Notes: CO_2 equivalent emissions from deforestation and carbon payback times computed under different assumptions about indirect use substitution. All values computed based on the marginal effects of deforestation reported in Table 6. The *x% ILUC* scenarios consider the direct emissions plus the emissions corresponding to x% of the displaced area in pasture and cropland on areas of natural cover.

for sugarcane ethanol. Table 7 shows payback times for different assumptions on the magnitude of the indirect land use change (ILUC). I find 4.8 years payback for sugarcane ethanol, considering only direct deforestation effects. Assuming that 2/3 of the expansion over other cropland and pasture converts into further deforestation, I find a higher payback time of 20.9 years. This time could be increased to almost 30 years in the extreme case in which all cropland and pasture converted to sugarcane ends up expanding into forests. This extreme case would only be realized if the ratio between supply and demand elasticities for the commodities displaced by sugarcane is infinite. This extreme case provides a reasonable upper bound for the sugarcane ethanol carbon payback time.

The sensitivity of the carbon payback times to indirect land use change computed here merits extra attention and should be a topic of future research. The deforestation from indirect land use change might be pivotal in a welfare assessment of sugarcane ethanol policies. However, the carbon payback times in any scenario are dwarfed compared to US corn ethanol, which pays back in 167 years (Searchinger et al., 2008). This difference is primarily due to a lower carbon efficiency in corn ethanol production in comparision to sugarcane.

6.3 Ethanol mandates

In this section I use the estimated model to discuss the effects of biofuels policies. There is a range of ethanol policies in the world today. Here I focus on the most common ones: those that shift the demand for ethanol by establishing ethanol blending standards in transportation fuel. Brazil, like many countries, establishes a fixed proportion of ethanol blend in gasoline. Although some States in the U.S. have similar rules, the Federal policy establishes only an aggregate volume of ethanol to be blended to gasoline every year. For simplicity, I study here the effects of shifts in the demand curve for ethanol that would be implied by those mandates.

There are some important caveats in the analysis that follows. First, the results so far concern only the supply side of sugarcane ethanol. Evaluating the effects of demand shifts in this market will require some knowledge of the demand side. I do not estimate a demand elasticity in this paper; instead, I rely on existing results from the literature and assess the robustness

Table 8: U.S. 2007 Renewable Fuel Standard (Billions of gallons)

	Total volume	Advanced biofuels	Cellulosic biofuels
2014	18.15	3.75	1.75
2018	26	11	7
2022	36	21	16

of my findings. Second, I estimate the long-run price increase that follows the demand shift using a static equilibrium model. This is in contrast to my measure of supply elasticity that was derived using a dynamic model. I believe that a static equilibrium framework provides a parsimonious environment to study the long-run effects of ethanol mandates.

As an illustrative policy experiment, I consider the effect of the 2007 EISA,[19] which establishes increasing mandates for ethanol in a nested system. EISA sets an increasing total mandate for renewable fuels from 18 billion gallons in 2014 to 36 billion gallons in 2022 but is subject to EPA rulemakings, which can allow for lower standards each year.[20] Table 8 describes the mandated volumes of biofuels for selected years. The first column shows the total volume of biofuels to be blended. The second column defines the amount that must be met with advanced biofuels and the third column, the amount of the advanced biofuels mandate that must be met by cellulosic biofuels.

In order to meet the mandate, biofuels must achieve at least a 20% reduction in life cycle greenhouse[21] emissions in comparison with gasoline and diesel. Under the non-advanced biofuel category falls almost all corn based ethanol produced in the US. Advanced biofuels must achieve at least a 50% reduction in greenhouse gas emissions in comparison with fossil fuels. Finally, cellulosic biofuels refer to biofuels derived from any cellulose that achieve a 60% gain in terms of greenhouse gas emissions. There was not much interest from the private sector in cellulosic ethanol mainly because the necessary conversion technology is still too costly for large scale application (Bracmort, 2012). As consequence, EPA has been continuously using its statutory authority to the waive cellulosic biofuels mandate on the basis of "insufficient supply."

Brazilian sugarcane ethanol is classified by the EPA as an advanced biofuel. In fact, it

---

[19]Energy Independence and Security Act.

[20]The EPA has the stutory authority to waive the biofuels mandate in the cases of "insufficient supply" or "economic harm." For example, in 2014 the EPA set the final volume of ethanol in the standard to 15.9 billion gallons, below the volume predicted in EISA. It is likely that EPA will continue to set lower stadard volumes of ethanol in the following years.

[21]According to EISA 2007 sec. 201, "'lifecycle greenhouse gas emissions' means the aggregate quantity of greenhouse gas emissions (including direct emissions and significant indirect emissions such as significant emissions from land use changes), as determined by the Administrator, related to the full fuel lifecycle, including all stages of fuel and feedstock production and distribution, from feedstock generation or extraction through the distribution and delivery and use of the finished fuel to the ultimate consumer, where the mass values for all greenhouse gases are adjusted to account for their relative global warming potential."

Table 9: Effects of sugarcane ethanol mandates

5 billion gallons mandate

Mandate/World output:	6.45%		
Mandate/Brazil output:	9.71%		
Demand elasticity	**-0.05**	**-0.2**	**-0.9**
Δ % price	1.38%	1.33%	1.16%
Δ % yield	0.37%	0.36%	0.31%
Δ Total Acreage (km^2)	17,830	17,277	15,093
$-\Delta$ Natural (km^2)	2,202	2,133	1,864

Notes: Aggregate effects of a 5 billion gallons sugarcane ethanol mandate. Mandate share of world and Brazil output were calculated using FAOSTAT 2013 information of sugarcane total output and the model baseline predicted change in Brazilian supply. We use 86.3 L/tonne of cane (Macedo et al., 2008).

is the sole large scale source of Advanced Biofuels in the EPA classification. I consider a counterfactual shift in the world demand of sugarcane ethanol of 5 billion gallons, which is the volume in the advanced biofuel category that needn't be met by cellulosic biofuels $(21 - 16 = 5$, in 2022, Table 8).

Table 9 shows aggregate effects on prices, acreage expansion and yields in Brazil of a 5 billion gallons mandate of sugarcane ethanol. I use a baseline demand elasticity of -0.2 from Elobeid and Tokgoz (2008).[22] Policy effects are not sensitive to demand elasticities in the inelastic range. The low price responses and comparatively high acreage responses are driven primarily by the high acreage to yields elasticity ratio I find. This analysis suggests that a 5 billion gallon mandate, which is equivalent to about 3% of US gasoline consumption, could imply about 2,000 sq. kilometers in direct deforestation. Following the previous analysis of indirect deforestation, this could be magnified to 9,000 sq. kilometers if indirect effects are considered.

7 Conclusion

This paper studies the economics of land use for the sugarcane ethanol production to quantify the environmental effects of biofuel policies. The expansion in ethanol production may endanger tropical forests, which could offset the carbon savings accrued over time by the replacement of fossil fuels. I use a dynamic model of land use that encompasses both adoption of sugarcane and replanting decisions, which is crucial to disentangle acreage and yield responses to policy changes.

I find a high acreage-price elasticity, which implies much higher acreage to yield elasticity ratios than found in previous studies. The results suggest that ad-hoc assumptions on the

---

[22]Roberts and Schlenker (2013) find a comparable values of demand elasticities for agricultural products.

pattern of land expansion can severely bias the evaluation of the merits of sugarcane ethanol as a greener replacement for fossil fuels. The most detrimental environmental effects from biofuels demand shifts come from land use change associated with an expansion of biofuels crops. If the acreage to yield elasticity ratio is low, there is scope for yields to absorb the increase in demand. However, if the acreage elasticity is higher than the yield elasticity, the increase in ethanol supply comes primarily from new producing areas.

In the case of Brazilian sugarcane ethanol, I find that the extensive margin (acreage) dominates the intensive margin (yield). This results in large acreage expansion following an increase in feedstock demand and a comparatively small increase in yields. I use the high-resolution nature of the dataset and the estimated model to predict the direct effects on natural land cover and associated carbon emissions. I discuss how indirect deforestation caused by crop and pasture substitution could aggravate land use change emissions. I find lower carbon payback times for sugarcane ethanol compared to US corn ethanol, but these are somewhat sensitive to the importance of indirect land use change, which points in the direction of important future research.

References

Victor Aguirregabiria and Pedro Mira. Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models. *Econometrica*, 70(4):1519–1543, 2002.

Victor Aguirregabiria and Pedro Mira. Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, 156(1):38–67, 2010.

Treb Allen and Costas Arkolakis. Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, 129(3):1085–1139, 2014.

Soren T Anderson, Ryan Kellogg, and Stephen W Salant. Hotelling under pressure. *NBER Working Paper*, (20280), 2014.

Peter Arcidiacono and Robert A Miller. Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, 79(6):1823–1867, 2011.

Steven Berry. Biofuels policy and the empirical inputs to GTAP models. *Discussion Paper*, 2011.

Kelsi Bracmort. Meeting the renewable fuel standard (RFS) mandate for cellulosic biofuels: Questions and answers. *CRS Report for Congress*, (R41106), 2012.

Christine L Crago, Madhu Khanna, Jason Barton, Eduardo Giuliani, and Weber Amaral. Competitiveness of Brazilian sugarcane ethanol compared to US corn ethanol. *Energy Policy*, 38(11):7404–7415, 2010.

Harry De Gorter, Dusan Drabik, Erika Kliauga, and Govinda R Timilsina. An economic model of Brazil's ethanol-sugar markets and impacts of fuel policies. *World Bank Policy Research Working Paper*, (6524), 2013.

Saraly Andrade de Sá, Charles Palmer, and Salvatore Di Falco. Dynamics of indirect land-use change: empirical evidence from Brazil. *Journal of Environmental Economics and Management*, 65(3):377–393, 2013.

Angus Deaton and Guy Laroque. On the behavior of commodity prices. *The Review of Economic Studies*, 59(1):1–23, 1992.

Amani Elobeid and Simla Tokgoz. Removing distortions in the US ethanol market: What does it imply for the United States and Brazil? *American Journal of Agricultural Economics*, 90(4):918–932, 2008.

PMF Elshout, R van Zelm, J Balkovic, M Obersteiner, E Schmid, R Skalsky, M van der Velde, and MAJ Huijbregts. Greenhouse-gas payback times for crop-based biofuels. *Nature Climate Change*, 2015.

EPA. *Renewable Fuel Standard Program (RFS2) Regulatory Impact Analysis*. 2010.

FAO/IIASA. *Global Agro-ecological Zones (GAEZ v3.0)*. FAO and IIASA, Rome, Italy and Laxenburg, Austria, 2011.

Joseph Fargione, Jason Hill, David Tilman, Stephen Polasky, and Peter Hawthorne. Land clearing and the biofuel carbon debt. *Science*, 319(5867):1235–1238, 2008.

Holly K Gibbs, Matt Johnston, Jonathan A Foley, Tracey Holloway, Chad Monfreda, Navin Ramankutty, and David Zaks. Carbon payback times for crop-based biofuel expansion in the tropics: The effects of changing yield and technology. *Environmental Research Letters*, 3(3):034001, 2008.

R.J. Hijmans, S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25:1965–1978, 2005.

V Joseph Hotz and Robert A Miller. Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3):497–529, 1993.

IPCC. *2006 IPCC Guidelines for National Greenhouse Gas Inventories*, volume 4. IPCC, Hayama, Japan, 2006.

IPCC. *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, UK, 2014.

Ryan Kellogg. The effect of uncertainty on investment: Evidence from Texas oil drilling. *The American Economic Review*, 104(6):1698–1734, 2014.

Christine Lasco and Madhu Khanna. US-Brazil trade in biofuels: Determinants, constraints, and implications for trade policy. In *Handbook of Bioenergy Economics and Policy*, pages 251–266. Springer, 2010.

Isaias C Macedo, Joaquim EA Seabra, and João EAR Silva. Green house gases emissions in the production and use of ethanol from sugarcane in Brazil: The 2005/2006 averages and a prediction for 2020. *Biomass and bioenergy*, 32(7):582–595, 2008.

Ruiqing Miao, Madhu Khanna, and Haixiao Huang. Responsiveness of crop yield and acreage to prices and climate. *American Journal of Agricultural Economics*, 2015.

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

N. Ramankutty, A.T. Evan, C. Monfreda, and J.A. Foley. *Global Agricultural Lands: Croplands, 2000.* Data distributed by the Socioeconomic Data and Applications Center (SEDAC): http://sedac.ciesin.columbia.edu/es/aglands.html. (Accessed 09/19/2014)., 2010a.

N. Ramankutty, A.T. Evan, C. Monfreda, and J.A. Foley. *Global Agricultural Lands: Pasture, 2000.* Data distributed by the Socioeconomic Data and Applications Center (SEDAC): http://sedac.ciesin.columbia.edu/es/aglands.html. (Accessed 09/19/2014)., 2010b.

Michael J. Roberts and Wolfram Schlenker. Identifying supply and demand elasticities of agricultural commodities: Implications for the US ethanol mandate. *American Economic Review*, 103(6):2265–2295, 2013.

B. F. T. Rudorff, D. A. Aguiar, W. F. Silva, L. M. Sugawara, M. Adami, and M. A. Moreira. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using landsat data. *Remote sensing*, 2(4):1057–1076, 2010.

John Rust. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55(5):999–1033, 1987.

Paul T. Scott. Dynamic discrete choice estimation of agricultural land use. *Working Paper*, 2013a.

Paul T. Scott. Indirect estimation of yield-price elasticities. *Working Paper*, 2013b.

Timothy Searchinger, Ralph Heimlich, Richard A Houghton, Fengxia Dong, Amani Elobeid, Jacinto Fabiosa, Simla Tokgoz, Dermot Hayes, and Tun-Hsiang Yu. Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science*, 319(5867):1238–1240, 2008.

SISFRECA. *SISFRECA 2008 Yearbook.* ESALQ-LOG, Piracicaba, Brazil, 2008.

Eduardo A. Souza-Rodrigues. Deforestation in the Amazon: A unified framework for estimation and policy analysis. *Working Paper*, 2015.

A. Valente, V. Tani, A. Coelho, M. Pimenta, P. Tome, R. Reibnitz, and R. Queiroz. Determinacao das velocidades médias de operacao para o ano de 2006. Technical report, 2008.

Table 10: Sugarcane fields in Center South Brazil

State	MG	SP	PR	MS	MT	GO	Total
2004	699	9936	311	155	231	699	12031
2005	1618	15749	1298	598	726	1299	21288
2006	2928	24498	2480	1180	1096	2200	34382
2007	4469	33088	3657	2218	1538	3625	48595
2008	5563	37018	4162	3480	1857	4877	56957
2009	6240	39520	4414	4318	2001	5673	62166
2010	6931	41619	4572	5104	2081	6390	66697
2011	7819	44457	4884	5930	2234	7542	72866
2012	8790	48466	5297	7011	2516	8945	81025
# Grid points	398489	200963	106363	311772	511391	326246	1855224

Notes: Grid points with sugarcane fields by State and year. State abbreviation: *MG*, Minas Gerais; *SP*, São Paulo; *PR*, Paraná; *MS*, Mato Grosso do Sul; *MT*, Mato Grosso; *GO*, Goiás.

Appendix

## A Data

A.1 Panel of sugarcane land use

I create a $1km$ grid of points encompassing all micro-regions in Center South Brazil that had reported sugarcane areas in the remote sensing dataset. I exclude urban areas as well as water bodies, such as lakes and large rivers. Moreover, I exclude grid point in the Pantanal eco-region. There is a strict ban on sugarcane growing in this area. In fact, I found no reports of sugarcane growing in this region. The resulting grid extends through 6 States in Brazil. Table 10 documents the number of grid points planting sugarcane in each State and year.

The remote sensing maps only inform new fields as well as replanting decisions for existing fields. Therefore I can only track the age of a given sugarcane field if it is a new field in my sample or if it replants. As it can be noted in Table 10 most sugarcane fields by 2012 were planted for the first time during the sample period.

A.2 Distances and transportation cost

Using existing road information from the Ministry of Transportation, we build a road network that covers the entire region of interest. We classify each road type in terms of average traveling speed of trucks relative to highways (the best road type available). This gives a measure of relative cost of traveling over each point in the map. We consider 1 km on

Figure 8: Predicted transportation cost and effective distance
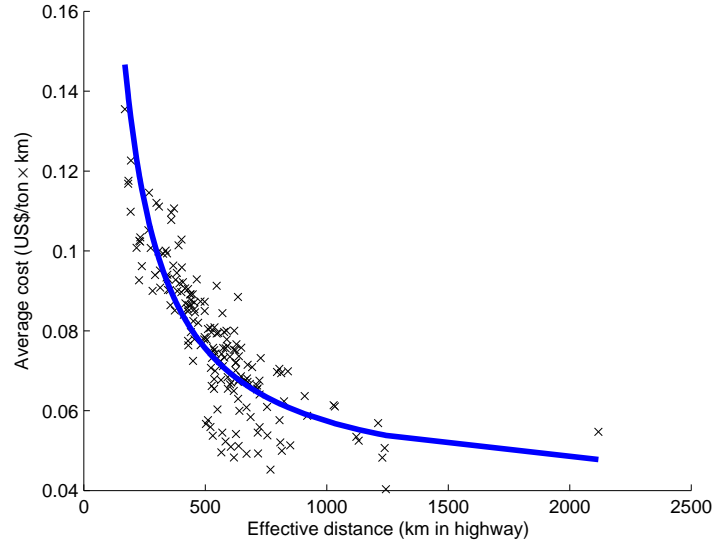
highway to be equivalent to 1.25 km on a major road, 1.66 km on unpaved road and 2.5 km everywhere else. Those values are informed by Valente et al. (2008), which surveys traveling speed of different types of vehicles. As an example, suppose the best route between i and j is 1 km on a dirt road, then we consider that effective distance between i and j to be 1.66 km.

From this measure of relative traveling cost, we search for the cheapest path given the different costs of traversing each parcel. This is implemented using ArcGIS cost distance tools. For an excellent description of this type of least cost path algorithm, see Allen and Arkolakis (2014).

Since we use the estimated transportation cost model (equation 7) to extrapolate freight rates for all points in the grid, it is important that the transportation cost model fits the data well. In fact, the transportation model fit is reasonably good. Figure 8 examines the fit of the transportation cost model, where I plot predicted transportation cost by the model in the solid line and actual quotes.

The calculated map of transportation costs is shown in Figure 9. We can see a big variation in projected transportation costs to the closest port ranging from US$18 per tonne close to destination ports to more than US$100 per tonne on the agriculture frontier. I extract the

Figure 9: Sugar transportation cost to nearest port (tc_i)
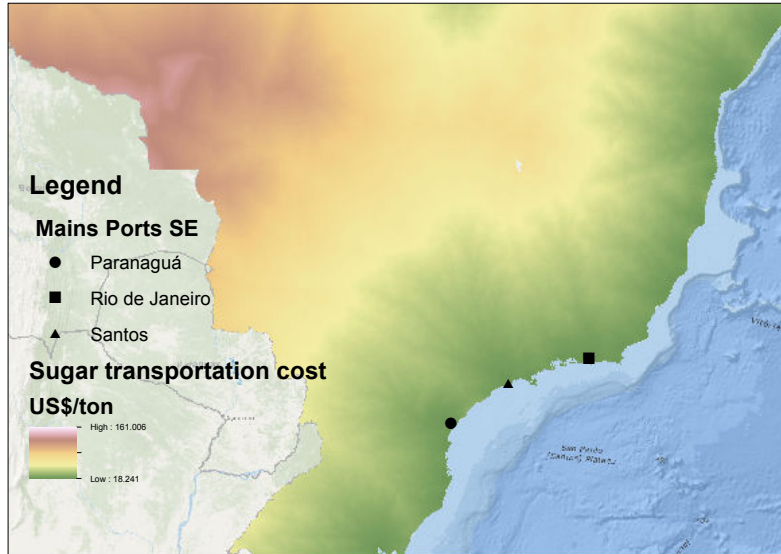
Notes: Sugar transportation cost (US$/ton) and main destination ports in South Central Brazil. Variable tc_i in equation 3. Transportation cost is calculated for each grid point using the estimated transportation cost model; equation (7) and Table 1.

value of transportation cost from this map for every point in the grid.

A.3 Other land characteristics

Potential yields (y_i^c). I use agro-ecological potential yields from FAO/IIASA (2011), v3.0. GAEZ provides potential yield infomation for different levels of input use and irrigation system. For all crops, c, I use values for rain-fed and high-input use. Figure 10 depicts the sugarcane potential yield information. A high yield region is noticeable where most current sugarcane fields are located in the center of the map. There is also a high yield area to the north of the map that coincides with the Southern fringe of the Amazon rainforest.

Previous economic use (h_i). I use information on previous economic land uses from Ramankutty et al. (2010a) and Ramankutty et al. (2010b). This dataset informs for every point in the globe a predicted probability that the land is covered by pasture and the probability the land is covered by cropland. Those probabilities are informed at a 10 km resolution. In model estimation, I draw for every grid point a land cover (cropland, pasture or other) using the probabilities informed by the dataset.

Climate and elevation (x_i). Climate and elevation are extracted for all grid points from Hijmans et al. (2005). This is a high-resolution climate dataset with information on monthly temperature and rainfall. This climate information is too rich to be used fully in estimation.

Figure 10: Sugarcane potential yield (y_{is})
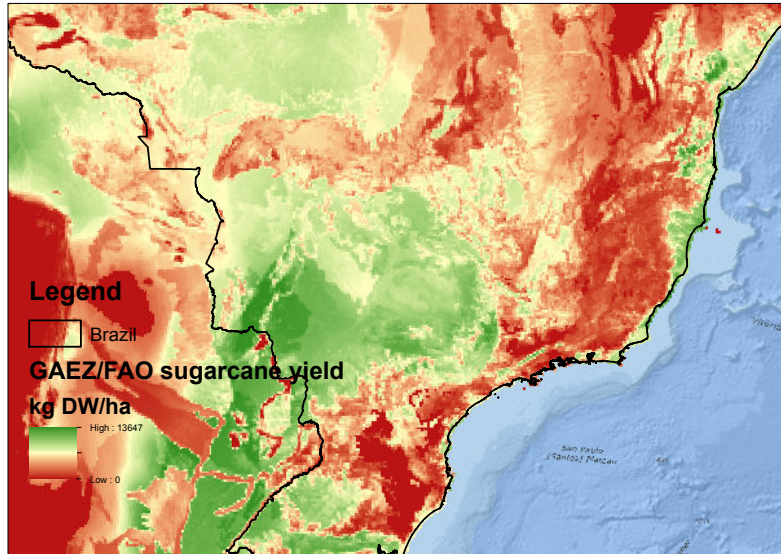
Notes: Sugarcane agro-ecological potential yield (high inputs). Source: FAO/IIASA (2011).

I use averages over the growth (Nov-Apr) and harvest seasons (May-Oct) of those climate variables climate variables in estimation.

B   Estimation

Structural parameters in the payoff functions are estimated using the Nested Pseudo Likelihood (NPL) algorithm proposed by Aguirregabiria and Mira (2002). The algorithm is modified to incorporate misreports in replanting in the spirit of Arcidiacono and Miller (2011).

The processes that govern the evolution of the sugar price and outside options return are estimated offline. The NPL algorithm builds on the method proposed by Hotz and Miller (1993), which suggest a clever use of Conditional Choice Probabilities (CCP) to reduce the computational burden imposed by "full solution" alternatives, *e.g.*, the Nested Fixed Point algorithm (Rust (1987)).

Aguirregabiria and Mira (2002) show that in many dynamic discrete choice problems there is a contraction mapping in the space of conditional choice probabilities:[23]

$$\mathbf{p} = \Upsilon_\theta(\mathbf{p}),$$

where the contraction mapping is naturally dependent on the vector of parameters describing

---

[23]This is analogous to the contraction mapping in the space of value functions.

the optimization problem θ.

The algorithm I use is a modification of NPL proposed by Arcidiacono and Miller (2011) that is used here to control for misreports in replanting. Given technical difficulties in the classification of replanting, for some fields that replant will not have this activity coded in our panel dataset. We consider two types of regime. A field can either not be subject to misreport in replanting ($u_i = 1$), in which case whenever replanting happens, this activity will be observed in the dataset, or it can be subject to misreport in replanting ($u_i = 2$), in which case I will not see replanting in case it happens. There is no indication of classification error in the adoption of sugarcane. Let μ be the probability there is no classification error. Naturally, $1 - \mu$ denote the probability in the population that a field is subject to classification error. μ is treated as unknowns and estimated by Maximum Likelihood with the other parameters.

I now describe the estimation algorithm. Let $\tilde{q}_i$ denote the observed decisions for field i. Start the algorithm at step $K = 0$ with an initial try of the CCP estimate $\mathbf{p}_0$ and for $\Pr(u_i = \bar{u}) = \mu_0^{\bar{u}}$, for $\bar{u} \in \{1, 2\}$. For every step $K \geq 1$,

1. Update each field regime ex-post probability using Bayes rule:

$$\Pr\left(u_i = \bar{u}|\tilde{q}_i, \omega_i, a_i\right) = \frac{\mu_{K-1}^{\bar{u}}\mathbf{p}_{K-1}(\tilde{q}_i|\omega_i, a_i, \bar{u})}{\sum_{u=\{1,2\}} \mu_{K-1}^u \mathbf{p}_{K-1}(\tilde{q}_i|\omega_i, a_i, u)}, \text{ for } \bar{u} = 1, 2.$$

2. Compute estimates of payoff parameters

$$\hat{\theta}_K = \arg\max_\theta \sum_{i=1}^n \sum_{\bar{u}=\{1,2\}} \Pr\left(u_i = \bar{u}|\tilde{q}_i, \omega_i, a_i\right) \ln \Upsilon_\theta(\mathbf{p}_{K-1})(\tilde{q}_i|\omega_i, a_i, \bar{u}). \quad (13)$$

3. Use estimates $\hat{\theta}_K$ to update the CCPs:

$$\mathbf{p}_K = \Upsilon_{\hat{\theta}_K}(\mathbf{p}_{K-1}).$$

4. Update μ^1 and μ^2:

$$\mu_K^u = \frac{1}{N}\sum_{i=1}^N \Pr\left(u_i = u|\tilde{q}_i, \omega_i, a_i\right), \text{ for } u = 1, 2.$$

5. Iterate until convergence in $\mathbf{p}$, θ and μ is reached.

Note that if replanting was observed for a given field i, then $\mathbf{p}_K(\tilde{q}_i|\omega_i, a_i, 2) = 0$, and so $\Pr\left(u_i = 1|\tilde{q}_i, \omega_i, a_i\right) = 1$. Moreover, fields in the replanting classification error regime have, naturally, zero probability of having replanting coded, hence $\mathbf{p}_K(\tilde{q}_i|\omega_i, a_i, 2) = 1$. Existing sugarcane fields in the beginning of the sample for which replanting was never observed in the 10 years of data available are treated as being in the classification error state, $u_i = 2$. This

Table 11: Estimation settings

ρ	0.9			
$\bar{a}$	9			
Discrete state space:				
Transp. cost (R\$/lb)	0.027	0.043	0.058	0.084
Dist. to closest sugar field (km)	1.5	9.0	27.5	>40
y^s (kg DW/ha)	2119	5089	6601.5	9784.5
Altitude (m)	168.5	414	586.5	1484
Prec. Growth Season (mm)	131.3	174.8	207.0	266.7
p^s (R\$/lb)	0.188	0.270	0.353	0.435

Notes: Estimation settings for NPL estimation. ρ is the discount factor and $\bar{a}$ is the cap in sugarcane field age. Discretized values for each continuous state variables are shown. With the exception of the distance to sugarcane fields, all classifications are based in equally sparse quantiles, so bins for each discrete category have the same number of observations.

avoids an initial condition problem for those fields, as we can only observe a field age a_i if i is a new field or if it replants. Although there are reports of sugarcane fields going more than 10 years without renovation, this is an abnormal situation in this industry.

Intuitively, the algorithm re-weights observations in the likelihood (13), taking into account the presence of classification error. For instance, a field with an observed old age that has never replanted will have a lower weight in estimation than a younger field that has never replanted, conditional on other observables being the same. This is because the former has a higher ex-post probability of being type $u = 2$ (classification error) than the latter.

Continuous state variables are discretized for ML estimation purposes. Table 11 shows the estimation settings, assumed values for the discount factor ρ and the maximum sugarcane field age $\bar{a}$. We also report the discretized values for the continuous state variables. With the exception of the distance to sugarcane fields, all classifications are based in equally sparse quantiles, so bins for each discrete category have the same number of observations.

C   Carbon emissions from deforestation

We compute the decrease in carbon stocks from cleared forest land using the guidelines set in IPCC (2006). The suggested formula for the initial change in carbon stocks converted to another land is

$$\Delta C = -\sum_f B_f \cdot (1 + R_f) \cdot \Delta A_f \cdot CF_f, \tag{14}$$

where B_f stands for the above-ground biomass stock (tonnes d.m./ha) of type f natural cover, R_f is the ratio between above and below ground biomass (tonne root/tonne shoot), ΔA_f is the change in land cover and CF_f is the carbon fraction of dry matter.

Table 12: Parameters for carbon stock assessment

f	B_f	R_f	CF_f
Tropical rain forests	300	0.37	0.47
Tropical moist deciduous forest	220	0.24	0.47

Note: Values of above-ground biomass stock B_f (tonnes d.m./ha), ratio of above-below ground biomass R_f and carbon fraction of dry matter CF_f used in the assessment of carbon emissions from deforestation.

Table 12 provides the values used in equation (14) to assess the effects of deforestation in carbon emissions.