Predicting Perennial Crop Yields Using the Replant Rate: The Case of Sugarcane in Brazil

8/21/2023

Abstract

This paper presents a novel and parsimonious method of predicting the dynamic yield impacts of a change in a perennial crop's replant rate using only data on the crop's age-yield function. We test the econometric specification implied by this model on Brazilian sugarcane data and find that it explains approximately one third of yield variation during the study period of 2005 to 2013, lending support to the hypothesis that reductions in the renewal rate after the financial crisis in 2008–9 and subsequent compensatory replanting contributed to this yield decline. The framework presented here is flexible and can be applied to any other perennial crop, so long as data on the age-yield function is available.

Keywords: Brazil, Biofuel, Sugarcane, Perennial Crop, Yield Prediction.

1 Introduction

Sugarcane production in Brazil, being the key feedstock of Brazil's ethanol industry, expanded rapidly in the 2000's, leading Brazil to become a major producer and exporter of ethanol. This was an explicit policy goal of Brazil's government at the time, with President Luiz Inacio Lula da Silva declaring that Brazil wanted to become the "Saudi Arabia of Biofuel" (Globo, 2007). By 2010, ethanol production accounted for around 2.3 percent of Brazil's GDP (Valdes, 2011). However, Brazilian sugarcane production growth slowed at the end of 2000's, stagnating from the period between 2010 and 2014, even reversing in 2011. Alongside the reduction in production growth, investment in sugarcane processing capacity declined. In the mid-2000's when the industry was booming, a net of 27 new sugarcane-processing mills opened in 2008, but from 2011 to 2014 a net of ten mills closed each year (UNICA, 2014).

Decomposing production changes into area and yield effects, we find that the production change from 2010 onwards was yield driven. What happened to yields? Why did Brazil's average sugarcane yield drop by approximately 10 tons/ha in 2011? Industry observers provide several explanations, including that growers and mills struggled to obtain credit in the wake of the 2008 global financial crisis, that the region faced bad weather, and that the average age of production increased (Leahy, 2012; Crooks and Meyer, 2011; da Silva, 2016; Ewing, 2013b,c,a, 2014; Moreira, 2015; Walter et al., 2016). For example, The Economist (2012, para. 2) stated that "Poor weather, and cash-strapped growers delaying their replanting after the 2008 credit crunch, have recently squeezed production."

In this study, we focus on the role of replanting in Brazilian sugarcane yields over the decade from 2005 to 2013. To do so, we develop a theoretical model of perennial crop yields as a function of their age-distribution. Replanting decisions affect the age-distribution and thus the trajectory of yields. The model allows predictions of the future trajectory of yield in response to a change in the replant rate. The model is applicable to a wide variety of perennial crops, such as coffee, cocoa, tree nuts, and tree fruit, allowing for an arbitrary number of age-classes, and an arbitrary yield in each age-class.

Using this model, we develop an econometric specification to quantify the effect of replant rate changes on yields, leveraging yield data from the Brazilian Institute of Geography and Statistics, and data on area replanted from the CANASAT project, a remote sensing effort led by the Brazilian National Institute for Space Research from 2005 to 2013. The econometric results are consistent with the theoretical model, and explain approximately one third of the yield variation over this period.

Existing sugarcane yield prediction models do not emphasize the dynamic impacts of replanting on forecasting yield (Alvarez et al., 1982; Pagani et al., 2017; Ferraciolli, Bocca, and Rodrigues, 2019). While these studies highlight sugarcane age as an important predictive factor for yields, they confine their interest to predicting yields for the upcoming season. Crucially important for planting decisions, this time horizon can obscure the impacts of replanting decisions on yields over intermediate time horizons (2–5 years), which is more relavant for investment decisions by firms and policy design by lawmakers and regulators.

This paper also contributes to the literature on perennial supply response by focusing on the effect of the replant rate, rather than the area replanted, as is more common (e.g. French and Bressler, 1962; French and Matthews, 1971; French, King, and Minami, 1985; Akiyama and Trivedi, 1987; Knapp and Konyar, 1991). Moreover, recent research has suggested that replanting strategies based on a percentage of total acreage (”phased replanting”) can provide perennial crop growers a conceptually simple strategy to generate higher and less volatile income streams (Mahrizal et al., 2014).

This article is organized as follows. Section 2 provides background and context for the Brazilian sugarcane ethanol industry. Section 3 decomposes sugarcane production changes into area and yield effects, identifying yield as the primary determinant of production since 2010. Section 4 develops and analyzes a theoretical model of perennial crop yields as a function of the replant rate. Sections 5 and 6 present the application of this model to Brazil. Sections 7 and 8 discuss the results and conclude.

2 Brazilian Sugarcane Industry

In the 2014-15 harvest year, Brazil produced 532 million tons of sugarcane, processed into 35.5 million tons of sugar (of which 24.2 million tons were exported) and 28.4 billion liters of ethanol (of which 1.4 billion liters were exported) (UNICA, 2015). This harvest was grown on 10.9 million hectares of land, a small fraction of Brazil’s 330 million hectares of arable land, but a more sizable fraction of its 60 million hectares of cultivated land. Brazil is by far the world’s largest producer of sugarcane, producing a greater mass of sugarcane in 2015 than the next 6 largest producing countries combined.

The sugarcane sector plays a substantial role in Brazil's economy. In 2015, the sugarcane sector's revenue was greater than US$70 billion, which is around 3.5 percent of Brazil's GDP, while exports of processed sugar and ethanol were valued at US$10.2 billion. Just over 1 million workers were directly employed by the sugarcane sector, which is just under 1 percent of Brazil's labor force (UNICA, 2015).

Brazilian sugarcane is processed into either sugar or ethanol. For a liquid fuel, Brazilian sugarcane ethanol has particularly low carbon emissions, with Crago et al. (2010) estimating that, on an energy equivalent basis, it reduces carbon emissions by 74 percent relative to gasoline, and its life-cycle emissions are about half that of corn ethanol.

In 2015, 91 percent of the area planted with sugarcane in Brazil was in the south-central region, and 9 percent was in the north-east.[1] Although the north-east is the oldest growing region in Brazil, with cultivation dating back to the 1500s, the growth of the industry in modern times has been centered in the South-Central growing region (Sant'Anna et al., 2016).

In 2015, 98 percent of the sugarcane grown in the South-Central region comes from 6 states: Goías, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Paraná and São Paulo. São Paulo is by far the largest producer, responsible for 60 percent of sug-

---

[1] `http://www.unicadata.com.br` Accessed: 31 Dec, 2016

arcane production. The next largest producing state, Minas Gerias, accounts for 11 percent of production.[2]

Sugarcane is a perennial grass, usually grown in rotations of 4-8 years, that is harvested and sent to local mills for processing into sugar or ethanol (James, 2004). Harvesting takes place between April and December, the dry season, and the sucrose content of the cane reaches a maximum in August and September. Mechanized harvesting is replacing manual harvesting, eliminating the need to burn the cane. A single machine can harvest up to 800 tons of cane in a single day (de Moraes and Zilberman, 2014).

After it is cut, sugarcane is highly perishable, needing to be processed in a mill as fast as possible to avoid losing sugar content. Most cane is collected from fields close to the mill—in 2014, the average distance from sugarcane fields to a mill in São Paulo was 26.3km (CONAB, 2017)—and sugar losses are minimized if the cane is processed within 48-72 hours after being cut (Belik et al., 2017; Sant'Anna et al., 2018).

At the mill, the sugarcane stalks are crushed. The resulting fiber, along with some cane straw, is burned to produce electricity, while the juice is purified and processed into sugar and/or ethanol, depending on the configuration of the mill and the market conditions at the time (Dias et al., 2015; Sant'Anna et al., 2016). In 2015, there were 369 sugarcane mills operating nationwide with 81 percent of these located in the South-Central region (UNICA, 2016; CONAB, 2019). Across Brazil, 70.3 percent of

---

[2] `http://www.unicadata.com.br` Accessed: 31 Dec, 2016

mills were capable of producing both sugar and ethanol, while 26.4 percent specialized in ethanol only, and the remaining 3.3 percent produced only sugar (CONAB, 2019). On average, approximately half of the total recoverable sugar (TRS) available for processing is converted to sugar and half to ethanol, with small (2-3 percent) fluctuations around this mean (Sant'Anna et al., 2016).

3 Decomposing Sugarcane Production into Area and Yield Effects

[Figure 1 about here.]

Around 2010, Brazil's sugarcane production ended a decade long period of steady growth. This period was followed by a decline in production of around 10 percent in 2011, relative to 2010, and a reduction in the growth trend. What accounts for this change in trend? In particular, how much of this change can be attributed to changes in planted area, and how much to changes in yield? Visual inspection of the area and yield panels in figure 1a suggests that area growth has been the main driver of overall production growth, but that yield deviations bear more responsibility for the production pattern after 2010.

Production changes are the sum of area and yield changes (Babcock, 2015). Figure 1b shows the decomposition of changes in production into area, yield and mixed effects using the discrete time decomposition method of Alauddin and Tisdell (1986). We exclude the result for 2005 from the graph since there was practically no change in production between 2004-2005 (nearly two orders of magnitude smaller than the next smallest production change), where an increase in area was almost exactly offset by a decline in yield. Such a small production change led to a small denominator when the decomposition shares were normalized and an outlier when placed on the graph.

Looking at the results of the decomposition, there are three distinct periods. First, from 1990–91 to 2003-04, the effects of area and yield are relatively equal, with neither effect dominating the production trajectory. In the second period, from 2004–05 to 2010–11, there is a decoupling between area and yield changes. During this period, production growth is driven almost entirely by area growth, and the contribution of yield to growth is small, or slightly negative. Also during this period, area driven production growth increases from 2003–04 to 2009–10 after which the effect size declines. The third period, from 2011–12 to 2014–15, is a period of highly variable yield effects. During this period the magnitude of the yield effects dominate the area effects, with unusually large negative yield contributions in 2011-12 and 2014-15. Area driven growth is positive during this period, but mostly continues the decreasing trend started in 2009–10. Throughout the entire time horizon, the mixed effect plays an insubstantial role in explaining changes in production.

The next section develops a model of yield changes as a function of the replant rate to explain how the changes in yield seen post-2010 could be explained by changes in the replant rate.

4 The Yield Trajectory after a Change in the Replant Rate

Perennial crops, such as sugarcane, can be grown and harvested for multiple years before they need to be replanted. Over their lifespan, the yield of the crop changes with time, following the *age-yield function.* Following Mitra, Ray, and Roy (1991) the age-yield function can be decomposed into three phases: the establishment phase (increasing yield), the peak phase (constant, maximal yield), and the declining phase (decreasing yield). The particular age-yield function will vary depending on the crop, the growing location, the farm management practices, pest pressure, temperature, and water availability, among other factors.

To illustrate the idea of an age-yield function, figure 2 shows an example for the Alta Mogiana region[3] of São Paulo state, Brazil (Margarido and Santos, 2012). The establishment phase occurs in the year of planting (year 0). The peak occurs in the first year after planting and lasts for only one year. The declining phase begins in the second year after planting and continues until the 6th year. Since Brazilian sugarcane tends to be renewed by or before its 6th year, we are not aware of data on the age-yield relationship for Brazilian sugarcane for higher years.

[Figure 2 about here.]

Margarido and Santos (2012) identify the key features of sugarcane yield dynamics,

---

[3] The Alta Mogiana region is in the north-east of São Paulo state. It is located within the Ribeirão Preto mesoregion, which is included in analysis below.

mainly that the yield trajectory will be non-monotonic in response to a change in the renewal rate:

> It is important to point out that after large decreases in planting or in renovation, there is a significant increase in total production in the next year, but drastic reduction in the second year, because of two factors: i) part of the first cut cane (1/7), which is used for seedlings, is not used for sowing, and therefore, it is added to the next growing season; ii) because of renovation itself, which if it is not carried out, increases the cutting area in the following year. (Margarido and Santos, 2012, p. 12)

Although they identify this key feature, they leave several questions unanswered. What happens in subsequent years? What will be the new equilibrium level of production? An econometric model of the effect of renewal rates on sugarcane yield requires answers to these questions—both to correctly specify the model and also to provide testable hypotheses. The remainder of this section develops a general model of yield trajectories as a function of changes in the renewal rate. This model uses an exogenous renewal rate—it is not determined by an optimization model. The model is applicable for any perennial crop, and is applied to a representation of Brazilian sugarcane to generate testable hypotheses for this specific case.

Before considering the dynamics of the yield of a perennial crop we must first consider the dynamics of its age-structure. Age-structure is the division of the plants in a growing region into different age-classes. We study the simplest model of age-structure dynamics, where there is a fixed plot of land (size normalized to 1) divided into sub-plots of different ages.

Let x_{st} be the area of land allocated to age-class s in year t. Under the natural dynamics of this system (i.e. without human intervention) the canes will enter the next oldest age-class next year,[4] that is, $x_{s,t} = x_{s-1,t-1}$. Following Mitra, Ray, and Roy (1991) and Salo and Tahvonen (2004) we assume the existence of some oldest age-class, S, creating $S+1$ age-classes in total (freshly planted cane is denoted by x_0). This makes the analysis tractable by imposing a finite number of age-class variables. It is also reasonable—the oldest age-class could simply be a zero yield class for plants that are dead or non-yielding from old age.

On top of this baseline aging process, consider the possibility of replanting, meaning replacing an old plant with a fresh seed, seedling, or cutting. When replanting s-year-old plants, r_{st}, land is moved from age-class s to age-class 0. When considering a single replanting decision, this implies two linked dynamic equations: $x_{0,t} = r_{st}$ and $x_{st} = x_{s,t-1} - r_{st}$. Not all land allocated to a single age-class needs to be replanted at once, and replanting happens at the start of a period, with yield being realized at the end of that period. The replant choice variable is constrained to be between 0 and x_{st}.

Combining the natural and artificial dynamics of the system and summing up across all age-classes yields the following system of dynamic equations, which is illustrated

---

[4] We assume away any loss between years, e.g. due to weather damage, or pest damage etc. The model could be extended by including a loss parameter, $\alpha < 1$ between transitions, i.e. $x_{s,t} = \alpha x_{s-1,t-1}$

in figure 3:

$$x_{0,t} = \sum_{s=0}^{S} r_{st}$$

$$x_{s,t} = x_{s-1,t-1} - r_{st} \text{ for } 1 \leq s < S$$

$$x_{S,t} = x_{S-1,t-1} + x_{S,t-1} - r_{St}$$

[Figure 3 about here.]

Let x_{st} be an *active age-class* if $x_{st} > 0$. In principle, land from any age-class could be replanted, implying that without further restrictions there could be an age-structure with active and inactive age-classes interleaved among each other. Mitra, Ray, and Roy (1991, section 3) demonstrate that a profit maximizing orchard manager will replant old age-classes in preference to young age-classes if the crop follows a single-peaked age-yield relation—like the one shown in figure 2.

The active-age contiguity result of Mitra, Ray, and Roy (1991) allows the dynamics of a stationary system to be studied in terms of the renewal rate. In a stationary system the state of the system remains unchanged from period to period. Let $\boldsymbol{x}_t$ be the vector of land allocations across all age-classes in period t, so, in a stationary system, $\boldsymbol{x}_t = \boldsymbol{x}_{t+1}$. To achieve this state, a constant fraction of the land must be renewed each year.

Proposition 1 *In a stationary system a constant fraction of the land must be renewed each year.*

Proof. Since $\boldsymbol{x}_t = \boldsymbol{x}_{t+1}$, it follows that $x_{1t} = x_{1,t+1}$. Using the equation of motion for land in the first age-class to write this in terms of replanting decisions, $\sum_{s=1}^{S} r_{s,t-1} = \sum_{s=1}^{S} r_{st}\ \forall t$. Hence the aggregate quantity of land replanted in each period must be constant in a stationary system. ■

In a stationary system there will be equal quantities of land allocated to all but the oldest active age-class, i.e. $\boldsymbol{x}_t = \boldsymbol{x}_{t+1}$ and $x_{0t} = x_{1t} = \ldots = x_{s-1,t} \geq x_{st}$.

Let R be the replant rate, that is the fraction of the land that is renewed at the start of the year. For each replant rate $R \in [0, 1]$ there exists a corresponding stationary system, denoted $\mathbf{x}(R)$, defined as:

$$\mathbf{x}(R_t) = \begin{cases} \text{for } \left\lceil \frac{1}{R} \right\rceil < S & \begin{cases} x_{st} = R & \text{for } s < \left\lceil \frac{1}{R} \right\rceil \\ x_{st} = 1 - R\left(\left\lceil \frac{1}{R} \right\rceil\right) & \text{for } s = \left\lceil \frac{1}{R} \right\rceil \\ x_{st} = 0 & \text{otherwise} \end{cases} \\ \\ \text{for } \left\lceil \frac{1}{R} \right\rceil \geq S & \begin{cases} x_{st} = R & \text{for } s < S \\ x_{st} = 1 - R\,S & \text{for } s = S \end{cases} \end{cases} \tag{1}$$

where $\lceil . \rceil$ is the ceiling function. This characterization assumes a constant, unit area of land.

Proposition 2 *If the replant rate is held constant at $\bar{R}$, then an arbitrary plantation will reach the stationary state described by equation (1) in at most* $\min(\left\lceil \frac{1}{R} \right\rceil, S)$

periods.

Proof. Start with a system in an arbitrary state. Let the replant rate be set to $\bar{R}$ at the start of period $t = 0$. Thus $x_{0,0} = \bar{R}$. In each subsequent period x_{0t} will be set to $\bar{R}$. Hence after $\min(\left\lceil \frac{1}{\bar{R}} \right\rceil, S)$ periods the fraction of land in each of the age-classes 0 to $\min(\left\lceil \frac{1}{\bar{R}} \right\rceil, S) - 1$ will be equal to $\bar{R}$, and, assuming a constant quantity of land, age-class $\min(\left\lceil \frac{1}{\bar{R}} \right\rceil, S)$ must contain $1 - \bar{R}\,(\min(\left\lceil \frac{1}{\bar{R}} \right\rceil, S))$ units of land. This corresponds to the stationary-state in equation 1. ■

Given this dynamic yield system, what happens to the stationary-state yield after a one-off, persistent change to the replant rate? As equation (1) shows, when the replant rate is changed it is possible that the number of active age-classes also changes. If the replant rate increases sufficiently, the older active age-classes will become inactive, and, conversely, if the replant rate decreases sufficiently, previously inactive age-classes will activate. For the analysis below, we only consider small, i.e. marginal, changes to the replant rate. In the case of a marginal increase[5,6] in the replant rate it is not possible for the number of active age-classes to decrease, since for any $R \in [0, 1]$ there exists an $\varepsilon > 0$ such that $\left\lceil \frac{1}{R+\varepsilon} \right\rceil = \left\lceil \frac{1}{R} \right\rceil$.

Proposition 3 *Equilibrium yield increases after an increase in the renewal rate if and only if*

$$\frac{f_0 + f_1 + \ldots + f_{s-1}}{s} - f_s > 0$$

---

[5] It is possible for a marginal decrease in the replant rate to increase the number of active age-classes, but only if $\left\lceil \frac{1}{R} \right\rceil = \frac{1}{R}$. The set of such R has Lebesgue measure zero, and can thus be neglected for all practical purposes.

[6] See appendix A for the case with non-marginal increases in the replant rate large enough to decrease the number of active age-classes.

Proof. Equation (1) implies that all but the oldest active age-classes will have R units of land allocated to them, and the oldest age-class will contain $1-R\,(\min(\lceil \frac{1}{R} \rceil, S))$ units of land, which allows the yield equation to be rewritten as a function of the renewal rate:

$$\begin{aligned} yield &= f_0x_0 + f_1x_1 + \ldots + f_sx_s \quad \text{Where } s \text{ is the oldest active age-class} \\ yield &= f_0R + f_1R + \ldots + f_{s-1}R + f_s(1 - s\,R) \end{aligned}$$

where f_i is the productivity of age-class i. The set $\{f_0, \ldots, f_S\}$ is the age-yield function.

The derivative with respect to R represents the change in stationary-state yield with respect to a change in the renewal rate.

$$\frac{d\ yield}{dR} = f_0 + f_1 + \ldots + f_{s-1} - f_s\,s$$

This expression is positive if and only if $\frac{f_0+f_1+\ldots+f_{s-1}}{s} - f_s > 0$. ■

That is, an increase in the replant rate increases stationary-state yield if and only if the average productivity of all but the oldest age-class is greater than the productivity of the oldest age-class, or, equivalently, if having more land allocated to the oldest age-class reduces the average yield.

It is not enough to know the change in stationary-state yield from a marginal change in the replant rate, since to specify an econometric model one needs to know the trajectory followed by yield to the new stationary-state. Proposition 2 says that the new stationary-state will be reached in at most s periods. Hence, for each of those periods ($0 \leq t \ \leq s$) does yield, y_t, increase or decrease relative to the yield before the change, y_{-1}?

Proposition 4 *The change in yield t years after an increase in the replant rate, relative to the yield prevailing before the change, y_{-1}, is given by*

$$\frac{d(\Delta yield_{t,-1})}{dR} = \sum_{i=0}^{t}(f_i - f_s)$$

Proof. At the beginning of period $t = 0$, let the replant rate change from R to R' and let $\Delta R = R' - R$. The yield $t - 1$ years after the renewal rate change is given by:

$$yield_{t-1} = f_0 R' + \ldots + f_{t-1} R' + f_t R + \ldots + f_{s-1} R + f_s(1 - R\,(s-t) - R'\,t)$$

Similarly, after t years, the yield will be given by:

$$yield_t = f_0 R' + \ldots + f_{t-1} R' + f_t R' + \ldots + f_{s-1} R + f_s(1 - R\,(s-(t+1)) - R'\,(t+1))$$

The change in yield from $t - 1$ to t ($yield_t - yield_{t-1} = \Delta yield_t$) is given by:

$$\Delta yield_{t,t-1} = f_t R' - f_t R + f_s(1 - R\,(s-(t+1)) - R'\,(t+1)) - f_s(1 - R\,(s-t) - R'(t))$$

Simplifying and collecting like terms gives:

$$\begin{aligned}\Delta yield_{t,t-1} =& f_t\left(R' - R\right) - f_s\left(R' - R\right)\\ =& \Delta R(f_t - f_s)\end{aligned}$$

Hence, $\frac{d\,yield_{t,t-1}}{dR} = \lim_{\Delta R \to 0} \frac{\Delta yield_{t,t-1}}{\Delta R} = \frac{\Delta R(f_t - f_s)}{\Delta R} = (f_t - f_s)$.

The net change t years after a change in the replant rate is the sum of these year-to-year marginal changes

$$\frac{d(\Delta yield_{t,-1})}{dR} = \sum_{i=0}^{t} \frac{d(\Delta yield_{i,i-1})}{dR} = \sum_{i=0}^{t}(f_i - f_s)$$

■

With the formulae developed in proposition 4, we can use the Margarido and Santos (2012) yield function to generate qualitative and quantitative predictions about the effect of a marginal change in the replant rate on Brazilian sugarcane yields.

In the Brazilian case, $f_0 = 0$ and $f_1 > f_2 > \ldots > f_s > f_0$. Thus

$$\frac{d(\Delta yield_{0,-1})}{dR} = f_0 - f_s < 0$$

and

$$\frac{d(\Delta yield_{t,t-1})}{dR} = f_t - f_s > 0, \quad \forall t \text{ such that } 0 < t < s$$

Figure 4a presents these year-on-year changes using the Margarido and Santos (2012) age-yield function, showing the qualitative shape predicted above, with the first year-on-year change being negative, and the remainder being positive, each positive change being smaller than the last. Figure 4b shows the *net* change in yield t years after a change in the replant rate, relative to the yield before the change. For Brazilian sugarcane, the change trajectory is a concave, monotonically increasing function of time since the change, with the same-year effect negative, the one-year effect slightly negative, and the subsequent effects positive until the new stationary-state is reached 5 years after the change, stabilizing the yield at its new level. The shape of the age-yield relationship determines the shape of this curve—the roughly zero net effect in the year following the replant rate increase is an artifact of the yield in the oldest age-class being roughly halfway between the yield of the first two age-classes.

[Figure 4 about here.]

5 Empirical Methodology

Transitioning from the theoretical model to an empirical model needs a change in perspective. The theoretical model explores the *future* impacts of a change to the *current* replant rate, while an empirical model is restricted to using data from the past. The question for the empirical model is ”in which previous year(s) could a change in the replant rate have affected the current yield?”, thereby changing the focus to explaining current yield as a function of previous changes, or lags, of the replant rate.

The relationship between the change in the replant rate and its effect on current and future yields is given by proposition 4. For the econometric equation we examine the effect of current and past changes in the replant rate on the current yield. The regression equation is

$$y_{it} = \sum_{l=0}^{L} \beta_l ReplantRate_{i,t-l} + \boldsymbol{\alpha}\mathbf{X}_{it} + v_i + u_{it} \tag{2}$$

This equation implies that the yield in region i in period t is a function of L lags of the replant rate, including the contemporaneous replant rate, a vector of region and time specific co-variates, a region-specific fixed effect, reflecting unobservable, unchanging differences in yield across regions, and an idiosyncratic shock. Total area is included as a control in the specifications below since the theoretical model includes the assumption that total area was unchanging over time.

As proposition 2 shows, a sugarcane plantation managed in the manner of the theoretical model in section 4 will take $\min(\lceil \frac{1}{R} \rceil, S)$ years to reestablish a stationary state after a shock to the replant rate. For the study region and period in Brazil, the replant rate varies between 5.7 and 12 percent (see figure 5b), implying that the time to equilibrium, and hence the number of lags of replant rate that affect current yield, may be anywhere between 9 and 18 years if S is not binding. However, since we have observed no data suggesting that Brazilian sugarcane is cultivated beyond the 6^{th} year, we assume that $S = 6$ is binding.

Under the maintained hypothesis that the theoretical model is correct, the sign predictions from figure 4b will hold for the econometric equation. The lag of the replant rate from t years ago should have the same impact on current yields as the impact of a change in the replant rate now on yields t years in the future. For the Brazilian application, the coefficient on the contemporaneous replant rate should have a negative effect on current yield, the coefficient on the first lag of replant rate should have a negative coefficient close to zero, and the coefficients on the remaining lags should be positive and increasing to a magnitude similar to the absolute value of the coefficient on the contemporaneous replant rate.

Replanting rates may exhibit serial correlation. The serial correlation may be positive, where a low rate last year may be followed by a low rate this year due to a persistent shock, e.g. credit constraints spanning multiple years. Alternatively, the serial correlation may be negative, where a low replant rate last year leads to a high rate this year to compensate for the previous low rate. However, this is not necessarily an issue for this regression; an issue arises if the *idiosyncratic errors*, u_{it}, display autocorrelation.

To test for the presence of serial correlation in the idiosyncratic errors we perform the Wooldridge test of serial correlation for panel data models, as implemented for the STATA software package by Drukker (2003). The STATA implementation of the Wooldridge test by Drukker (2003) reports the p-value from a test of whether the coefficient from a regression of the residual on its lag is equal to -0.5, with the null hypothesis being that the coefficient is equal to -0.5. The p-values for the alternative lag specifications, estimated with the Brazilian data, are presented in table 1.

If the assumptions underlying the panel data model hold, particularly that serial correlation is either not-present, or adequately controlled by the use of clustered standard errors, the β coefficients can be interpreted as follows: β_l represents the marginal effect of a one unit increase in the replant rate l years ago on yields in period t, all else being equal.

To measure the effect of changes in the replant rate on yields this study uses a dataset of sugarcane planted area, replanted area, and yields from the 2005-06 to the 2013-14 growing year in 30 mesoregions[7] of the South-Central sugarcane growing region of Brazil, comprised of the states: Espírito Santo, Goías, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Paraná, Rio de Janerio, Rio Grande do Sul, Santa Catarina, and São Paulo. There are 74 mesoregions in the south-central region. The final dataset used mesoregions from the states Goias (GO), Mato Grosso (MT), Mato Grosso do Sul (MS), Minas Gerais (MG), Parana (PR), and São Paulo (SP), which accounted for over 99 percent of sugarcane production in the South-Central region of Brazil in the 2014-2015 growing year.

Data for quantity of sugarcane produced, yield, and planted area were downloaded from the IBGE website on 4 Jan, 2017. The IBGE data included the planted area (hectares), production (tons), and average yield (kilograms/ha, which was converted to tons/ha), for each mesoregion in the South-Central region, by year. These data are

---

[7] Mesoregions are a statistical (but not administrative) subdivision of Brazilian states. Created by the Brazilian Institute of Geography and Statistics (IBGE – *Instituto Brasilleiro de Geografia e Estatística*), the mesoregions attempt to subdivide the states into regions with similar "social processes", conditioned by their "natural setting" and the degree of "communication and place network". There are 136 mesoregions in Brazil.

collected by IBGE in the Produção Agrícola Municipal (PAM) survey. This survey is conducted annually and collects agricultural production data at the municipality level. This data is estimated by an IBGE agent in each municipality through consultation with agricultural technicians, large producers and their own knowledge of the industry (Instituto Brasileiro de Geografia e Estatística, 2018). Through centralizing data collection in a single respondent per municipality, there is a greater potential for biased reporting, compared to agricultural surveys in which many, randomly sampled producers in a region complete their surveys (e.g. the Crops/Stocks survey from the USDA's National Agricultural Statistics Service). This potential for bias is unlikely to effect the econometric analysis in this paper for two reasons. First, on average across Brazil, there are 40.9 municipalities per mesoregion, so biases in any individual municipality are likely to be canceled out through aggregation. Second, the empirical specifications used below include mesoregion fixed effects. Any bias still present at the mesoregion level that is constant over time will be absorbed by the fixed effects. However, any mesoregion-level biases that are changing over time and systematically correlate with the replant rate still have the potential to bias the coefficient estimates.

The data on area replanted was obtained from the CANASAT project,[8] run by the Brazilian National Institute for Space Research (INPE – *Instituto Nacional de Pesquisas Espaciais*). The CANASAT project uses satellite data to classify sugarcane growing regions into one of four classes: ratoon, canes that are growing from established rootstock; expansion, area freshly converted from non-sugarcane use;

---

[8] `http://www.dsr.inpe.br/laf/canasat/en/tables.html` Accessed: 25 August, 2014

under-renovation, canes that have been replanted, but not yet harvested; and renovated, the first harvest of freshly replanted canes. The CANASAT project collected and released data for the 2003-04 to 2013-14 harvest years.

The datasets were merged in STATA, dropping any mesoregions with a zero total-cultivated area each year, resulting in a balanced panel of 270 observations across 30 mesoregions and 9 years, from harvest year 2005-06 to harvest year 2013-14. Harvest year 2004-05 was dropped because the area replanted was not reported in all states except São Paulo, since that was the year monitoring began for those states. In 2013-14 the total production from these 30 mesoregions was 668 million tons. Total production in Brazil that year was 768 million tons. These mesoregions represent 87 percent of Brazil's total sugarcane production in 2013-14.

6 Results

Figure 5a shows the area-weighted average yield and figure 5b shows the area-weighted percent replanted across the 30 mesoregions in the sample for the years 2005 to 2013. Area-weights were used to ensure that average yield correctly measures the total production divided by the total area. For each year in the sample, a weight was assigned to each mesoregion, representing the proportion of total cultivated area that mesoregion provided over the entire sample, that is:

$$\text{weight}_i = \frac{\sum_t \text{area}_{it}}{\sum_t \sum_i \text{area}_{it}}$$

[Figure 5 about here.]

[Table 1 about here.]

[Figure 6 about here.]

Table 1 shows six models, varying the number of lags, L, from 0 to 5. Figure 6a provides a graphical representation of the replant rate coefficient estimates presented in table 1. In the models with 0 to 4 lags, the Wooldridge autocorrelation test results imply that the null hypothesis of no autocorrelation can be rejected at the 1 percent significance level, indicating the presence of autocorrelation in these models. To control autocorrelation in all 6 models, clustered standard errors were calcualted, with the mesoregion used as the unit of clustering. Clustering the standard errors allows for an arbitrary correlation structure within the cluster, accommodating the autocorrelation detected by the Wooldridge test. Clustering at the mesoregional level still maintains the assumption of independence of errors between the mesoregions. However, since some of the farmers' replant rate choices are likely to be influenced by state- and national-level factors (e.g. the credit crisis), this independence assumption is unlikely to hold in practice. Hence, the standard error estimates are a lower bound—the actual error is likely to be larger.

The coefficient on the contemporaneous replant rate is negative in all the models, and significant at the 1 percent level in all but the 5 lag model. In most cases it is around -0.5, implying that a 1 percentage point increase in the replant rate decreases yields by approximately 0.5 tons/ha in the same year. In almost all the models the coefficients on the lags are positive, the exceptions being the coefficient on the first lag in the 1, 2, and 5 lag models. In each of these cases the coefficient is not significantly different from zero. The coefficient on the first lag is a tight zero in each of the 5 models that include it. The coefficient on the second lag is positive in 3 of the 4 models that include it. Coefficients on lags further from $t = 0$ have larger standard errors, although the results suggest that the coefficient magnitudes are either constant or returning to zero after the peak at 2 lags.

In each model the coefficient on area planted was negative, but also statistically indistinguishable from zero at a 5 percent significance level. The R^2 values for the models ranged from 0.18 to 0.44, naturally increasing as more lags were added. The higher lag models (models 2–5) explained around one third of the variation in sugarcane yield during the sample, implying that there are other, omitted factors, such as bad weather as suggested in the introduction, that play an important role in explaining sugarcane yields in the South-Central region of Brazil. The R^2 results reported from the regressions are the *within* R^2 values.

7 Discussion

Figure 6b shows the theoretical prediction from figure 4b and the estimated coefficients for each of the six regression models tested. There is a striking consistency between the regression coefficient estimates from the six models and the theoretical predictions. Generally, the theoretical prediction is within the 95 percent confidence interval for most of the coefficients from most of the models. For the first three coefficient estimates (no lag, 1st lag, and 2nd lag) the theoretical prediction is within the 95 percent confidence interval of all but one of the the coefficient estimates, the exception being the no-lag coefficient from the 5-lag model, which is lying closer to zero than the theoretical prediction would place it. For the first three coefficients, their point estimates are generally higher than the theoretical prediction, although the prediction lies with in the 95 percent confidence interval. For the second three coefficients, the results are weaker, with the theoretical predictions falling outside, or close to the edge of, the 95 percent confidence intervals of the estimated coefficients. In each case the theoretical prediction is higher than the point estimate for each of the coefficients. A possible reason for the greater discrepancy between the predictions and the estimates for the higher lags is the smaller sample sizes that each of these models used. Adding an additional lagged variable reduces the sample size by 30 observations. So the no-lag model has 270 observations, while the higher lag models have only 180–120 observations to work with, reducing the precision of the estimates.

[Figure 7 about here.]

Figure 7 compares the actual average yield across the 30 mesoregions of the sample against the predicted yield from the 4-lag variant of the model. This graph shows that changes in the replant rate explain a substantial share of the yield variation, but clearly other factors are also important for explaining yields. This is reflected by the R^2 value of 0.35 for the 4-lag model. Figure 7 was generated using the Margarido and Santos (2012) age-yield relationship shown in figure 2.

The 4-lag model was chosen for the prediction because it is the only one of the higher lag models (3–5 lags) that is consistent with the theoretical predictions for each of its coefficients. The 3-lag model's coefficient on the 3-lag variable is significantly different from the theoretical prediction, while the 5-lag model's coefficient on the no-lag coefficient is significantly different from the theoretical prediction.

However, this preferred specification has limitations. In particular, the average age-yield relationship across the region may be different. Also, the model allows the age-yield relationship to vary across the regions only by a mesoregion-specific scalar, i.e. all mesoregions have an age-yield relationship with the same relative differences between the age-classes, but the level of all the age-classes is shifted up or down by a common factor. If the age-yield relationships across the mesoregions have different relative differences between the age-classes, the regression equation only captures the average of these individual age-yield relationships. This makes predictions from the regression model valid for the sample region as a whole, but less so for specific, individual mesoregions.

The theoretical analysis treats yield as a function of replant rate, all else being equal. This assumption may not hold for the econometric analysis. The econometric analysis studies the effect of a replant rate change, holding constant the total cultivated area, other lags of replant rate (for those included in the model), and mesoregion specific fixed effects, such as soil quality. However there are other variables that may affect the yield that were not controlled. Some of the omitted variables include weather, input use, harvesting method, sugarcane variety, and pest damage. If any of these variables are unchanging over time, they will be captured by the mesoregion fixed effects. The components that are changing over time may bias the coefficient estimates, if they are systematically correlated with the replant rate.

The empirical analysis is conducted at the mesoregion level, so the resulting yields are averaged across fields with different characteristics, such as soil type, sugarcane variety, harvesting method, within that mesoregion. In this case, the age-yield function from the theoretical model should be considered an average age-yield function, representing the average yield for each age class across all the fields in the region. However, the theoretical predictions assume the age-yield function is constant over time. Factors affecting yield that are constant over the study period, such as soil quality, will be controlled by the mesoregion fixed effects (Schlenker and Roberts, 2009; Cooper, Tran, and Wallander, 2017). Time varying factors, such as the proportion of fields harvested mechanically, remain uncontrolled in our baseline specification.

As a robustness check, we re-estimated the model using mesoregion-specific time trends, which capture broad changes in average yields over the study period (see appendix B). Productivity growth in sugarcane has been approximately linear over time (Chaddad, 2016). These time trends will capture slower changes to the industry, such as increases in the mechanization rate or the adoption of new sugarcane varieties, but will not capture year-to-year shocks such as whether. These shocks remain in the idiosyncratic error term.

Our preferred specification, the four-lag model, is not completely robust to the inclusion of the mesoregion-specific time trends. In particular, the coefficient on contemporaneous replanting, β_0, is no longer significantly different from zero, and is significantly different from the theoretical prediction. The predictions remain within the 95-percent confidence intervals of the other coefficients.

The vertical integration between the sugarcane fields and the mills will not affect the results here. The management of sugarcane fields in Brazil is usually undertaken by one of two entities. Either the fields are controlled by the mill in a vertically integrated operation, or they are operated by independent producers who sell to mills through contracts or a spot market (Chaddad, 2016; Sant'Anna et al., 2018). While the ownership structure might affect the decision when to replant fields (Tregeagle and Zilberman, 2023), it is unlikely to affect the results of this analysis, since the analysis takes the replanting decisions as given, then explores the impacts of these decisions on future yields. The motivation for the decision, once taken, does not affect the yield dynamics explained by changes in the replant rate.

In the preferred specification, approximately one third of the yield variation is explained by the econometric model with replant rate lags and area. The model under-predicts the yield peak in 2009 and over-predicts the yield trough in 2011 and 2012. This is consistent with the view that the yield decline in 2011 and 2012 was a 'perfect storm' of factors, including lack of investment in replanting, adverse weather conditions, and changing international market conditions (Walter et al., 2016). A key insight from the analysis in this paper is the lag between changing replanting rates and the resultant impacts on average yield. In figure 5, replant rates increase in 2011 and 2012 from the minimum in 2010. Yields, however, continued to decline in 2011 and 2012. This framework predicts that yields will subsequently increase, which was the case in 2013 and 2014 (UNICA, 2014).

8 Conclusion

This paper presented a novel and parsimonious method of predicting the dynamic impacts of the change in the replant rate of a perennial crop using only data on the crop's age-yield function. We tested the econometric specification implied by this model on Brazilian sugarcane data and found that it explains approximately one third of the yield variation during the study period from 2005 to 2013, lending support to the hypothesis that reductions in the renewal rate after the financial crisis in 2008–9 and subsequent compensatory replanting contributed to the yield decline.

Counterintuitively, the model predicts that an increase in the replanting rate will decrease yields in the short-term, as more land is allocated to sugarcane that takes time to provide its initial yield. Thus, the efficacy of policies to increase replanting should only be evaluated after several years, so that the initial yield decline has passed.

The framework introduced in this study highlights the dynamic impacts of replanting decisions. It is not intended as a comprehensive prediction tool, since many important variables identified by earlier studies are not included. It does, however, serve to illustrate how changes in replanting decisions can have counterintuitive impacts on yield in the short-term. The framework offered in the paper could be used to improve qualitative intuition and quantitative forecasts for sugarcane yields over a medium-term time horizon. Moreover, the framework presented here is flexible and can be applied to any other perennial crop, so long as data on the age-yield function is available.

References

Akiyama, T., and P.K. Trivedi. 1987. “Vintage Production Approach to Perennial Crop Supply: An Appliation to Tea in Major Producing Countries.” *Journal of Econometrics* 36:133–161.

Alauddin, M., and C. Tisdell. 1986. “Decomposition Methods, Agricultural Productivity Growth and Technological Change: A Critique Supported by Bangladeshi Data.” *Oxford Bulletin of Economics and Statistics* 48:353–372.

Alvarez, J., D.R. Crane, T.H. Spreen, and G. Kidder. 1982. “A Yield Prediction Model for Florida Sugarcane.” *Agricultural Systems* 9:161–179.

Babcock, B.A. 2015. “Extensive and Intensive Agricultural Supply Response.” *Annual Review of Resource Economics* 7:333–348.

Belik, W., B.B. Perosa, S.R.F. Figueira, and A. Koga-Vicente. 2017. “Milling Capacity and Supply Competition in Sugar-Ethanol Industry in São Paulo, Brazil.” *Geografia* 42:39–56.

Chaddad, F. 2016. “Agriculture in Southeastern Brazil.” In *The Economics and Organization of Brazilian Agriculture*. Elsevier, pp. 73–110.

CONAB. 2017. “Perfil Do Setor Do Açúcar e Do Álcool No Brasil 14/15.” Working paper, Companhia Nacional de Abastecimento, Brasília.

—. 2019. “Perfil Do Setor Do Açúcar e Do Álcool No Brasil 15/16.” Working paper, Companhia Nacional de Abastecimento, Brasília.

Cooper, J., A.N. Tran, and S. Wallander. 2017. "Testing for Specification Bias with a Flexible Fourier Transform Model for Crop Yields." *American Journal of Agricultural Economics* 99:800–817.

Crago, C.L., M. Khanna, J. Barton, E. Giuliani, and W. Amaral. 2010. "Competitiveness of Brazilian sugarcane ethanol compared to US corn ethanol." *Energy Policy* 38:7404–7415.

Crooks, E., and G. Meyer. 2011. "Brazilian imports of US ethanol soar." *Financial Times*, may, pp. Online.

da Silva, H.J.T. 2016. "Ethanol Production Costs in Brazil Fall Not Enough to Generate Profits." *World Ethanol & Biofuels Report* 14:216–218.

de Moraes, M.A.F.D., and D. Zilberman. 2014. *Production of Ethanol from Sugarcane in Brazil From State Intervention to a Free Market*. Springer International Publishing.

Dias, M.O.d.S., R. Maciel Filho, P.E. Mantelatto, O. Cavalett, C.E.V. Rossell, A. Bonomi, and M.R.L.V. Leal. 2015. "Sugarcane Processing for Ethanol and Sugar in Brazil." *Environmental Development* 15:35–51.

Drukker, D.M. 2003. "Testing for serial correlation in linear panel-data models." *Stata Journal* 3:168–177.

Ewing, R. 2013a. "Big crop, policy revive Brazil's sugar-ethanol mills." *Reuters*, Feb, pp. Online.

—. 2013b. “Brazil ethanol returns to US as biofuel rules pave way.” *Reuters*, Sep, pp. Online.

—. 2014. “Brazil’s big cane mills cut costs, capex against losses.” *Reuters*, Feb, pp. Online.

—. 2013c. “Brazil’s rapid recovery in sugar output to slow.” *Reuters*, Oct, pp. Online.

Ferraciolli, M.A., F.F. Bocca, and L.H.A. Rodrigues. 2019. “Neglecting Spatial Autocorrelation Causes Underestimation of the Error of Sugarcane Yield Models.” *Computers and Electronics in Agriculture* 161:233–240.

French, B.C., and R.G. Bressler. 1962. “The Lemon Cycle.” *Journal of Farm Economics* 44:1021–1036.

French, B.C., G.A. King, and D.D. Minami. 1985. “Planting and Removal Relationships for Perennial Crops: An Application to Cling Peaches.” *American Journal of Agricultural Economics* 67:215–223.

French, B.C., and J.L. Matthews. 1971. “A Supply Response Model for Perennial Crops.” *American Journal of Agricultural Economics* 53:478–490.

Globo. 2007. “Brasil quer ser ’Arábia Saudita’ do biocombustível, diz Lula.” *Globo*, Mar, pp. Online.

Instituto Brasileiro de Geografia e Estatística, ed. 2018. *Pesquisas agropecuárias*, 3rd ed. No. 6 in Séries Relatórios metodológicos, Rio de Janeiro: IBGE, Instituto Brasileiro de Geografia e Estatística.

James, G., ed. 2004. *Sugarcane*. Blackwell Science.

Knapp, K.C., and K. Konyar. 1991. "Perennial Crop Supply Response: A Kalman Filter Approach." *American Journal of Agricultural Economics* 73:841–849.

Leahy, J. 2012. "Brazil sugar production costs rise." *Financial Times*, May, pp. Online.

Mahrizal, L.L. Nalley, B.L. Dixon, and J.S. Popp. 2014. "An Optimal Phased Replanting Approach for Cocoa Trees with Application to Ghana." *Agricultural Economics* 45:291–302.

Margarido, F.B., and F. Santos. 2012. "Agricultural Planning." In F. Santos, A. Borém, and C. Caldas, eds. *Sugarcane Bioenergy, Sugar and Ethanol – Technology and Prospects*. Ministry of Agriculture, Livestock and Food Supply, chap. 1, pp. 7–21.

Mitra, T., D. Ray, and R. Roy. 1991. "The economics of orchards: an exercise in point-input, flow-output capital theory." *Journal of Economic Theory* 53:12–50.

Moreira, M. 2015. "Stop and Go in Brazilian Ethanol Sector." Presented at Berkeley Bioeconomy Conference, 2015.

Pagani, V., T. Stella, T. Guarneri, G. Finotto, M. Van Den Berg, F.R. Marin, M. Acutis, and R. Confalonieri. 2017. "Forecasting Sugarcane Yields Using Agro-Climatic Indicators and Canegro Model: A Case Study in the Main Production Region in Brazil." *Agricultural Systems* 154:45–52.

Salo, S., and O. Tahvonen. 2004. “Renewable Resources with Endogeneous Age Classes and Allocation of Land.” *American Journal of Agricultural Economics* 86:513–530.

Sant’Anna, A.C., J.S. Bergtold, A. Shanoyan, G. Granco, and M.M. Caldas. 2018. “Examining the Relationship between Vertical Coordination Strategies and Technical Efficiency: Evidence from the Brazilian Ethanol Industry.” *Agribusiness* 34:793–812.

Sant’Anna, A.C., A. Shanoyan, J.S. Bergtold, M.M. Caldas, and G. Granco. 2016. “Ethanol and Sugarcane Expansion in Brazil: What Is Fueling the Ethanol Industry?” *International Food and Agribusiness Management Review* 19:163–182.

Schlenker, W., and M.J. Roberts. 2009. “Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change.” *Proceedings of the National Academy of Sciences* 106:15594–15598.

The Economist. 2012. “Brazilian Brew.” *The Economist*, Jan, pp. Online.

Tregeagle, D., and D. Zilberman. 2023. “Minimizing the Costs of Biorefinery Feedstock by Managing Perennial Crop Age: The Case of Brazilian Sugarcane.” *Journal of Agricultural and Applied Economics*, Jun., pp. 1–23.

UNICA. 2014. “Final report of 2013 / 2014 harvest season South-Central region.” Working paper, UNICA, São Paulo.

UNICA. 2014. “Final Report of the 2013/2014 Harvest Season—South-Central Region.” Working paper, The Brazilian Sugarcane Industry Association (UNICA).

UNICA. 2015. “Sugarcane: One Plant, Many Solutions.” Working paper, UNICA, São Paulo.

—. 2016. “Virtual Mill.” Working paper, UNICA, São Paulo.

Valdes, C. 2011. “Brazil’s Ethanol Industry: Looking Forward.” Working paper, USDA-ERS.

Walter, A., M.V. Galdos, F.V. Scarpare, M.R.L. Verde Leal, J.E. Abel Seabra, M.P. da Cunha, M.C. Araujo Picoli, and C.O.F. de Oliveira. 2016. “Brazilian Sugarcane Ethanol: Developments so Far and Challenges for the Future.” In *Advances in Bioenergy*. John Wiley & Sons, Ltd, chap. 24, pp. 373–394.

Part I

Appendices

A Calculating the Yield Change following a Discrete Change in the Replant Rate that Changes the Number of Active Age-Classes

A.1 The effect of a discrete increase in the renewal rate that reduces the number of active age-classes by one

Unlike the marginal change case, a discrete change in the replant rate from R to R' can change the number of active age-classes. Here we show the effect of an *increase* in the replant rate on the yield transition trajectory.

Say that at time t there are $s+1$ active age-classes (where $s+1 = \lceil \frac{1}{R} \rceil$), and that at time $t+1$ the number of active age-classes declines to s. What is the change in yield?

The yield at time t is

$$yield_t = f_0 R' + \ldots + f_t R' + f_{t+1} R + \ldots + f_{s-1} R + f_s(1 - R(s-t) - R'(t))$$

and the yield at time $t+1$ is

$$yield_{t+1} = f_0R' + \ldots + f_tR' + f_{t+1}R' + \ldots + f_{s-1}(1 - R((s-1)-(t+1)) - R'(t+1))$$

Notice how the oldest active age-class at $t+1$ is now $s-1$, and that in the $s-1$ land allocation equation the R term is multiplied by $(s-1)-(t+1)$. This is because there are now $s-1$ other active age-classes.

The change in yield between t and $t+1$ is given by the difference between these two expressions

$$\Delta yield_t = f_{t+1}R' - f_{t+1}R + f_{s-1}(1 - R((s-1)-(t+1)) - R'(t+1)) - f_{s-1}R - f_s(1 - R(s-t) - R'(t))$$

which, after simplifying, becomes

$$\Delta yield_t = (f_{t+1} - f_{s-1})\Delta R + (f_{s-1} - f_s)(1 - R(s-t) - R'(t))$$

The first term in this expression is the 'within age-class yield effect' and the second term is the 'between age-class yield effect' which exists due to the change in the number of active age-classes. Notice that the 'within yield effect' is not exactly the same as the case when there was no change in the number of age-classes. The yield of the $t+1^{\text{th}}$ age-class is now being compared to the $s-1^{\text{th}}$ age-class, not the s^{th}.

A.2 The effect of a discrete increase in the replant rate that reduces the number of active age-classes by n

The change in the replant rate must be big enough to change the number of active age-classes by n *in one time step*, otherwise the formula in section A.1 is sufficient with a redefinition of s each time step.

Say that at time t there are $s+1$ active age-classes, and that at time $t+1$ the number of active age-classes declines to $s+1-n$. What is the change in yield?

The yield at time t is

$$yield_t = f_0R' + \ldots + f_tR' + f_{t+1}R + \ldots + f_{s-1}R + f_s(1 - R(s-t) - R'(t))$$

and the yield at time $t+1$ is

$$yield_{t+1} = f_0R' + \ldots + f_tR' + f_{t+1}R' + \ldots + f_{s-n}(1 - R((s-n) - (t+1)) - R'(t+1))$$

The change in yield between t and $t+1$ is given by the difference between these two expressions

$$\begin{aligned}\Delta yield_t =& f_{t+1}R' - f_{t+1}R \\ &+ f_{s-n}(1 - R((s-n) - (t+1)) - R'(t+1)) - f_{s-n}R - f_{s-n+1}R - \ldots \\ &- f_{s-1}R - f_s(1 - R(s-t) - R'(t))\end{aligned}$$

which, after simplifying, becomes

$$\Delta yield_t = (f_{t+1} - f_{s-n})\Delta R + (f_{s-n}(n-1) - \sum_{i=1}^{n-1} f_{s-n+i})R + (f_{s-n} - f_s)(1 - R(s-t) - R'(t))$$

B Robustness check using mesoregion-specific time trends

The robustness of the results was examined by restimating equation 2 after adding a mesoregion-specific time trend. The modified regression equation is given by

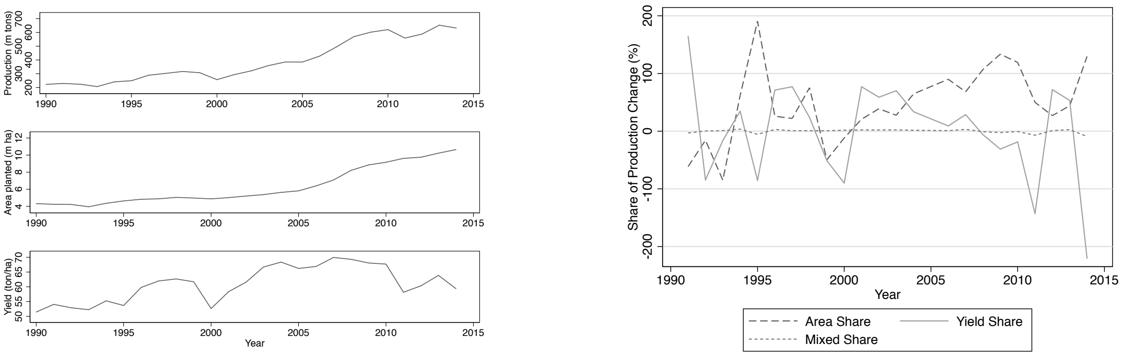
$$y_{it} = \sum_{l=0}^{L} \beta_l ReplantRate_{i,t-l} + \boldsymbol{\alpha}\mathbf{X}_{it} + \beta_i\ t + v_i + u_{it} \tag{3}$$

The results from estimating the modified regression equation are given in table 2 and figure 8.

[Table 2 about here.]

[Figure 8 about here.]

List of Figures



(a) Sugarcane production, planted area, and yield series for Brazil from 1990 to 2014.

(b) Decomposition of yearly changes in sugarcane production into area driven changes, yield driven changes, and mixed effect changes

Figure 1: Changes in sugarcane production after 2010 were driven by yield changes.

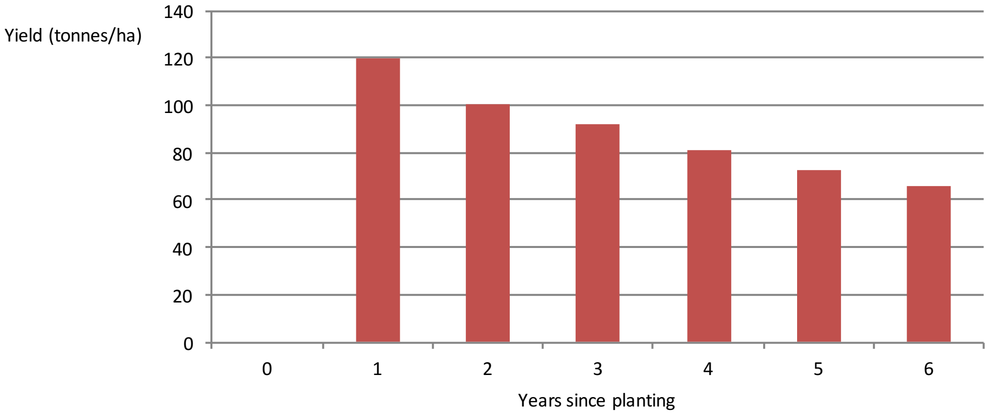


Figure 2: Age yield relationship taken from Margarido and Santos (2012). Freshly planted canes provide no yield (year 0).

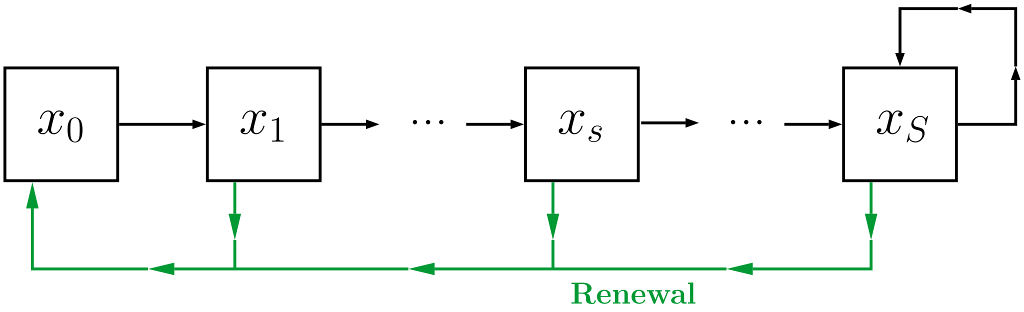
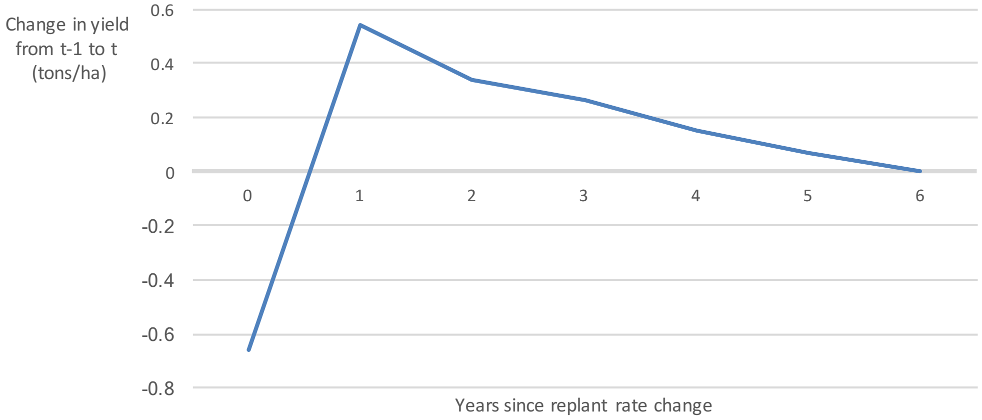
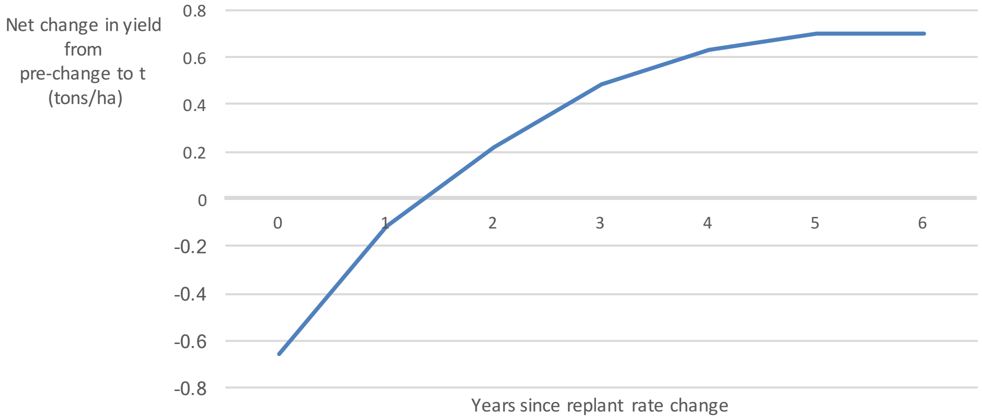


Figure 3: Diagrammatic representation of the dynamics of the area of land in each age-class.

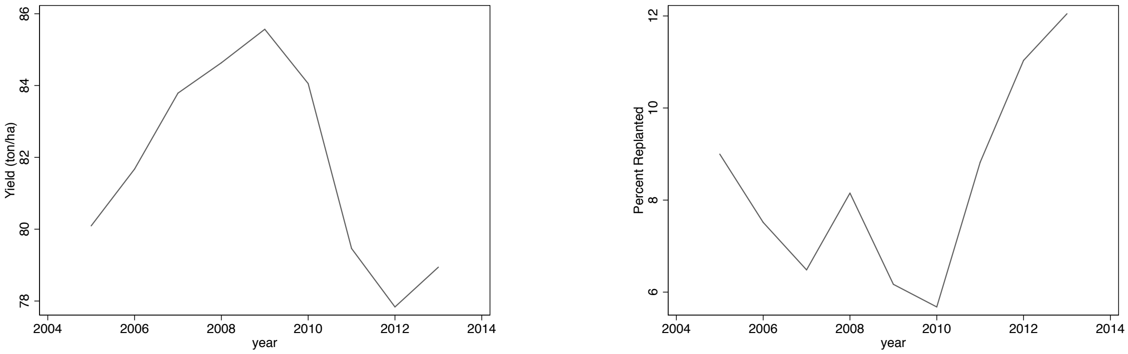


(a) Marginal *year-to-year* changes in sugarcane yield t years since a change in the replant rate.



(b) Marginal *net* changes in sugarcane yield t years since a change in the replant rate.

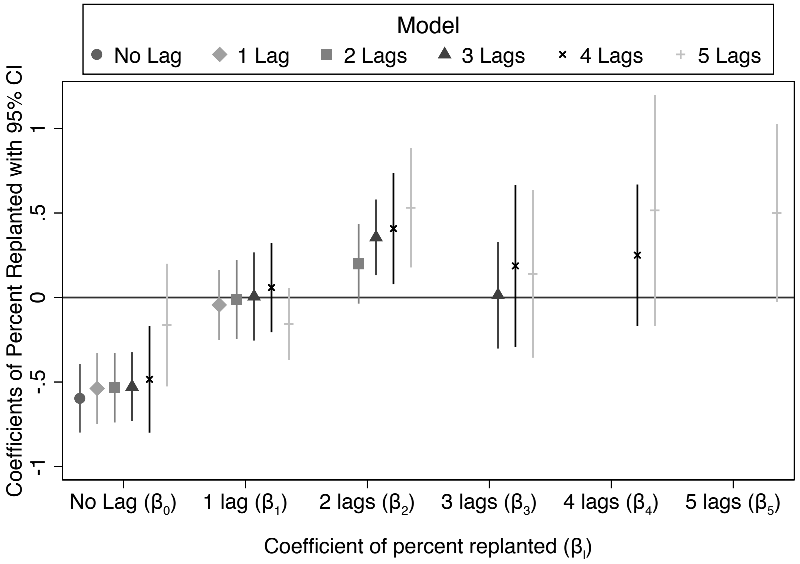
Figure 4: The change in yield t years after a 1 percentage point increase in the replant rate. Graph generated using the São Paulo age-yield relationship from figure 2.



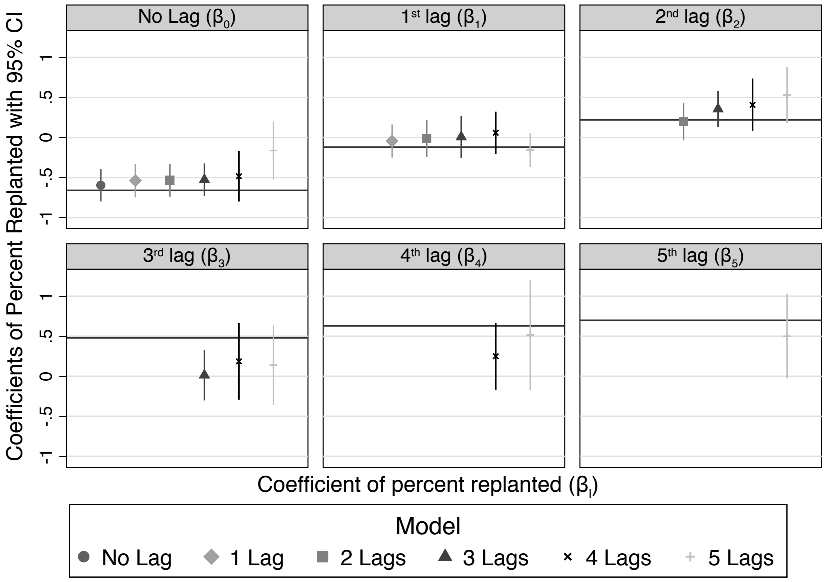
(a) Average yield

(b) Average replant rate

Figure 5: Average yield and replant rate across the 30 mesoregions in the sample.



(a) Coefficient estimates and 95 percent confidence intervals from all 6 models



(b) Coefficient estimates compared to the theoretical predictions in figure 4b

Figure 6: Two views of the coefficients of the 6 models. Figure 6a shows the coefficient estimates relative to zero. Figure 6b shows the coefficient estimates relative to the predictions from figure 4b.

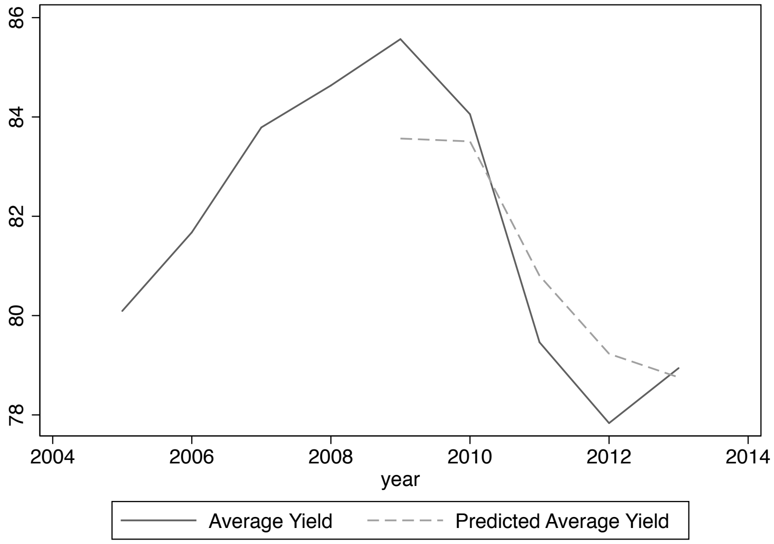
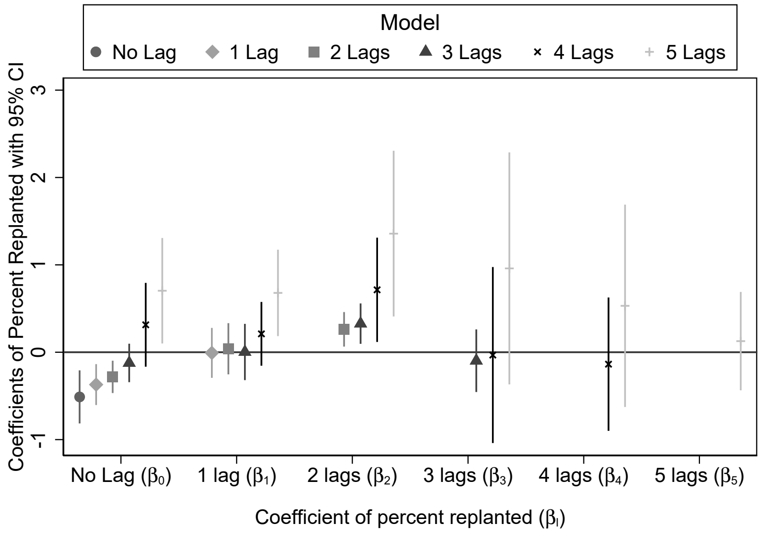
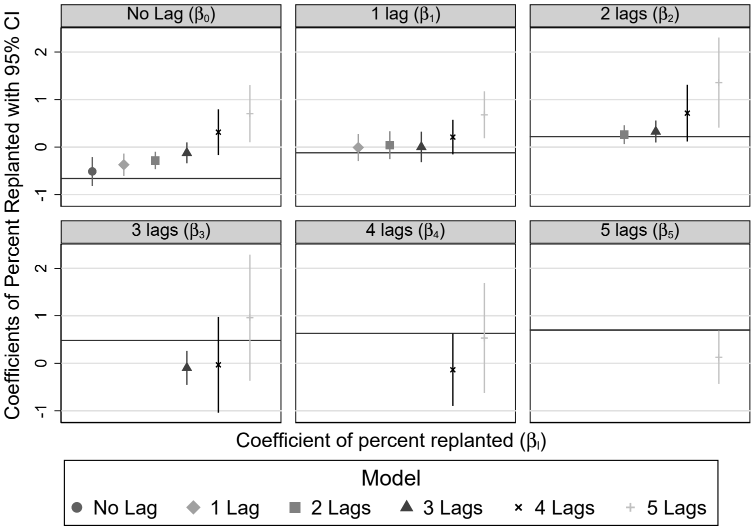


Figure 7: Actual and predicted yields for the 30 South-Central mesoregions in the sample using the 4-lag variant of the regression model.



(a) Coefficient estimates and 95 percent confidence intervals from all 6 models when mesoregion-specific time trends are included.



(b) Coefficient estimates compared to the theoretical predictions in figure 4b when mesoregion-specific time trends are included.

Figure 8: Two views of the coefficients of the 6 models when mesoregion specific time trends are included. Figure 6a shows the coefficient estimates relative to zero. Figure 6b shows the coefficient estimates relative to the predictions from figure **??**.

List of Tables

	No Lag	1 Lag	2 Lags	3 Lags	4 Lags	5 Lags
% Replanted	-0.5968***	-0.5384***	-0.5335***	-0.5280***	-0.4841***	-0.1630
% Replanted - Lagged one year		-0.0441	-0.0110	0.0061	0.0586	-0.1575
% Replanted - Lagged two years			0.1995*	0.3557***	0.4076**	0.5309***
% Replanted - Lagged three years				0.0140	0.1872	0.1402
% Replanted - Lagged four years					0.2506	0.5154
% Replanted - Lagged five years						0.4998*
Area Planted (1000 ha)	-0.0046	-0.0095*	-0.0169*	-0.0279	-0.0264	-0.0228
Mesoregion Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dep Variable	82	82	82	82	81	80
R-squared	0.176	0.219	0.304	0.346	0.346	0.438
Autocorrelation†	0.000	0.000	0.000	0.000	0.001	0.288
N	270	240	210	180	150	120

Standard errors clustered at the mesoregion level

$\dagger$ p-values of Wooldrige serial correlation test where H_0: No serial correlation (see Drukker (2003))

Mesoregions weighted by their average share of cultivated area

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

Table 1: Results from estimating equation (2) with the Brazilian dataset using 0–5 lags of replant rate

	No Lag	1 Lag	2 Lags	3 Lags	4 Lags	5 Lags
% Replanted	-0.5116***	-0.3704***	-0.2822***	-0.1226	0.3140	0.7038**
% Replanted - Lagged one year		-0.0076	0.0395	0.0027	0.2104	0.6791***
% Replanted - Lagged two years			0.2619**	0.3273***	0.7144**	1.3575***
% Replanted - Lagged three years				-0.0971	-0.0322	0.9596
% Replanted - Lagged four years					-0.1368	0.5314
% Replanted - Lagged five years						0.1271
Area Planted (1000 ha)	0.0289***	0.0369***	0.0399**	0.0670**	-0.0146	-0.0090
Mesoregion Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Mesoregion-Specific Time Trends	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dep Variable	81.78	82.00	82.04	81.75	81.17	80.07
R-squared	0.448	0.588	0.675	0.702	0.700	0.802
Autocorrelation†	0.000	0.000	0.000	0.000	0.001	0.288
N	270	240	210	180	150	120

Standard errors clustered at the mesoregion level

† p-values of Wooldrige serial correlation test where H_0: No serial correlation (see Drukker (2003))

Mesoregions weighted by their average share of cultivated area

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

Table 2: Results from estimating equation (2) with the Brazilian dataset using 0–5 lags of replant rate and including mesoregion specific time trends