Minimizing the costs of biorefinery feedstock by managing perennial crop age: The case of Brazilian sugarcane

Daniel Tregeagle*

David Zilberman

Abstract

We develop and analyze an unexplored mechanism to reduce biorefinery supply chain costs when the feedstock is a perennial crop: adjusting the age-structure, and hence yield, of the perennial feedstock. The non-monotonicity of the age-yield function introduces a non-convexity to the cost-minimization problem. We show that, despite this, the problem has a solution and present analytic and numeric comparative statics, finding that larger refineries are most likely to benefit from optimizing age-structure. The model is calibrated to the sugarcane industry in Brazil. The cost-reductions from optimizing age, compared to the observed regional average age, are less than one percent.

JEL Classification: D24, Q13, Q16

*Email: tregeagle@ncsu.edu. We thank Peter Berck, Larry Karp, Seth Garz, David Romer, Emma Roos-Collins, Leo Simon, Justus Wesseler, Yang Xie, Kelly Zering, three anonymous reviers, and the participants of the 20th International Consortium of Applied Bioeconomy Research Conference for helpful comments and suggestions. Thanks to Leo Simon for sharing code for random parameter draws. Material from this paper formed a chapter of Tregeagle's dissertation.

1 Introduction

Agricultural supply chains are crucial for bringing food from growers to increasingly urban populations. The particular structure of an agricultural supply chain can affect the size and distribution of returns to participants along the chain, as well as the adoption and diffusion of agricultural technologies, and has the potential to transform an economy beyond the agricultural sector (Barrett et al., Forthcoming). While these effects have long been documented by economists and effective frameworks have been developed for analyzing components of the supply chain, there are few studies that provide an "explicit framework for economic principles of supply chain design" (Zilberman, Lu, and Reardon, 2019, p. 289).

Perennial crops provide multiple sources of value including food, such as fruit, nuts, cocoa, and coffee; fuel, including sugarcane ethanol and cellulosic biofuel; agronomic benefits, such as longer growing seasons and more efficient use of water (Glover et al., 2010; Wallace, 2000, p. 1638); and ecological and environmental services, such as carbon sequestration (Kreitzman et al., 2020) and erosion control (Glover et al., 2010; Molnar et al., 2013). Our paper contributes to the literature on supply chain design by developing a novel theoretical framework for analyzing agricultural supply chains for perennial crops, a class of crops with great economic importance worldwide and which have an additional constraint not present in annual crop production: older, higher-yielding plants began as younger, less-productive plants.

For agricultural firms to maximize their profitability, they must optimally design their supply chain. Du et al. (2016) decompose this decision into six components, including questions of how much feedstock and downstream services to produce in-house and how much to outsource, which technologies to use for processing, and what contracts to use for any outsourced production. Answering these questions requires a

two stage model, first determining the minimal supply chain costs for a given output and sourcing arrangement, and second, choosing the profit maximizing level of output, given the minimum cost production function.

In this paper, we focus on the first stage of the supply chain design problem. Minimizing supply chain costs requires minimizing over every relevant dimension. Supply chains with perennial feedstocks have an additional constraint over annual crops: To obtain perennials of a certain age, they must be grown from younger plants. There is a relationship between age and yield. Generally, the yield increases with age, before peaking and declining (Mitra, Ray, and Roy, 1991). The unconstrained grower would wish to have a production system consisting only of plants at the maximum yielding age. However, when incorporating the aging constraint, the yield of such a production system, in which all crops have an identical age, would vary with the relationship between age and yield. For a processor with fixed feedstock processing capacity, holding all else equal, it is better to minimize the deterministic yield variation by keeping a portfolio of multiple of crop ages. Tisdell and De Silva (1986) show that a uniform distribution of ages eliminates any deterministic yield variation due to age and that this distribution can be described by its average age.

But which average age will minimize the supply chain costs? A natural candidate is the yield maximizing distribution identified in Tisdell and De Silva (1986). However, in this paper we show analytically that costs along the supply chain, including replanting and delivery costs, will affect the cost-minimizing age, which is generally older than the yield maximizing age. We ground this model in the context of Brazilian ethanol and sugar production from sugarcane feedstock, a well-developed industry that exemplifies the perennial feedstock supply chain.

1.1 Biofuel supply chains that use perennial feedstocks

Biofuels have great potential to assist efforts in climate change mitigation due to their lower life cycle carbon emissions in comparison to gasoline (Khanna and Crago, 2012). The environmental benefits of biofuel depend on the feedstock used and the production process. In a recent meta-analysis, Hochman and Zilberman (2018) found that corn ethanol reduces greenhouse gas emissions by 11 percent compared to gasoline. In comparison, sugarcane ethanol can reduce greenhouse gas emissions by around seventy five percent (Crago et al., 2010; Manochio et al., 2017) . Cellulosic ethanol can have even greater environmental benefits, with greenhouse gas reductions up to 86 percent (Wang, 2007).

The cost of the supply chain to convert biomass into energy is one of the most important barriers to adoption (De Meyer et al., 2014; Rentizelas, Tolis, and Tatsiopoulos, 2009). Both corn and sugarcane ethanol are 'first-generation' technologies, where the ethanol is fermented directly from sugars in the feedstock. Cellulosic ethanol is a 'second-generation' technology, where non-fermentable complex compounds are first broken down into simpler sugars, which are then fermented. First generation technologies are in widespread use, with the US leading corn ethanol production and Brazil leading sugarcane ethanol production. The adoption of cellulosic ethanol has been hampered by high costs.

The feedstocks for sugarcane and cellulosic biofuels both come from perennial crops, where the feedstock may be harvested for multiple years before being replanted. Brazilian sugarcane is usually grown in a six year cycle (Margarido and Santos, 2012), while miscanthus and switchgrass, two of the most promising cellulosic feedstocks, can be grown for over 10 years before needing to be replanted (Douglas et al., 2009; Heaton, 2010).

Perennial feedstocks tend to come from the local area. Sugarcane is highly per-

ishable after it is harvested and must be processed at the biorefinery (mill) within 24 hours of being harvested. In São Paulo state, the largest sugarcane growing state in Brazil, feedstock travels, on average, around 25 kilometers to the biorefinery (CONAB, 2013b). Cellulosic feedstock, on the other hand, can be dried, so perishability is less of an issue, but it is bulky, making long-distance transportation a substantial cost (Malladi and Sowlati, 2018).

The quantity of feedstock available in the biorefinery's local area will depend on two factors: the planted area and the yield per unit of land. For annual feedstocks, like corn, both the yield and area can be optimized each growing season. Perennial feedstocks, however, have yields that depend on the age of the crop, with maximum yields often being reached some years after the planting season. This imposes an additional constraint on the manager of a biorefinery that uses perennial feedstocks.

As indicated by a series of recent reviews, existing studies of biomass supply chain optimization tend to focus on optimizing the planted area (De Meyer et al., 2014; Malladi and Sowlati, 2018; O'Neill and Maravelias, 2021; Zahraee, Shiwakoti, and Stasinopoulos, 2020). The papers covered by these reviews focus on the area and the location of land to grow the feedstock to supply a local biorefinery, or network of refineries. Mostly they hold yield per unit of land constant, although some allow for heterogeneity between parcels of land and uncertainties in yield. For example, Debnath, Epplin, and Stoecker (2014) solve for the cost-minimizing land area to feed a fixed biorefinery size when yields are subject to stochastic weather shocks. O'Neill and Maravelias (2021) identify three papers where farmer decisions can influence yield through fertilization and/or harvest decisions. A notable exception is Demczuk and Padula (2017), who model the feasibility of Brazilian ethanol production under alternative price scenarios using a simulation model that explicitly accounts for the quantity of sugarcane in each age-category.

1.2 Contributions of the paper

In this paper we present an analytical framework for modeling the supply-chain for a vertically integrated biorefinery using perennial feedstocks, which can be used to minimize the costs of supplying the refinery's feedstock needs. To our knowledge, this is the first study that explicitly includes the optimization of the feedstock crop's age as a control parameter in an analytical model. In the model, both planted area and age-structure can be chosen by the decision maker. Our analysis focuses on a particular subset of possible age-structures known as the 'balanced region'. A balanced region has an equal proportion of land allocated to each age-class. For example, if there were only young and old plants, a balanced region would allocate half the land to young plants and half to old plants. Focusing on the balanced region allows the state of the region to be characterized by the average age, or, equivalently, maximum age. More generally, modeling an n-age-class region and relaxing the 'balanced region' assumption would need n state variables (see Mitra, Ray, and Roy (1991) for a more general discussion of possible age-structures). Furthermore, the balanced region minimizes the year-to-year variation in yields due to age-structure (Tisdell and De Silva, 1986), thereby avoiding the addition of another source of feedstock supply variation. Reducing this variation is an explicit goal of biomass supply chain optimization (Mafakheri and Nasiri, 2014; Sharma et al., 2013; Margarido and Santos, 2012; Debnath, Epplin, and Stoecker, 2015)

Building on the model of processed product costs by Wright and Brown (2007), we develop a model of perennial feedstock production and processing that includes maximum age as a control variable, generating a trade-off between land and yield to feed a processing facility of fixed size. We generate first order conditions for the model for a large class of yield functions and analyze the resulting comparative statics. The lower cost set for this problem is not convex, so we cannot rely on the usual conditions for a solution. We demonstrate that the first order conditions necessarily

have a solution.

We use this model to show that optimal feedstock age is older than the yield-maximizing age (in contrast to Tisdell and De Silva (1986), who suggest using the maximum yield age in a farm-gate only context). We also show that optimal age declines with increases in the size of the processing facility. We present analytical results to explain the trade-offs between age and land. Although our analytical model abstracts away from many of the details of practical supply chains, the majority of existing models of biofuel supply chain optimization are mathematical programming models. While their detail is useful for analyzing policy scenarios, they are less useful for illuminating the general trade-offs along the supply chain.

We calibrate the model to the sugarcane industry in the Brazilian state of São Paulo, a multi-output industry where sugarcane is used to produce ethanol and sugar, and sugarcane fiber byproducts are used to produce electricity. Our simulations show results that are consistent with the model's theoretical predictions, but we find that the cost savings from optimizing age are small relative to the costs of using the observed age in the region. Depending on the size of the mill, cost reductions range from 0.75 percent for small mills to 0.95 percent for large mills.

In what follows, we first develop a theoretical model that incorporates maximum age, planted area, and transportation. Then we identify the conditions for minimizing the cost in this model and show how the optimal planted area and maximum age vary with biorefinery feedstock processing capacity and other parameters. The theoretical results are illustrated with an example from the Brazilian sugarcane industry.

2 An Analytical Model of Perennial Age, Growing Region Area, and Facility Feedstock Processing Capacity

Consider a processing facility of given size that is supplied a feedstock grown by a perennial crop in surrounding fields. Assume that this is a vertically integrated system where a single manager controls the facility and the land and crop management for the feedstock. The manager's problem is to minimize the acquisition cost of feedstock for the facility by choosing how much land to use and how the feedstock is grown on that land. We pose this as a static problem for the manager, which can be interpreted as the long-run, steady-state management strategy for the facility. In this study we neither study the short run dynamics of the manager's problem, nor the choice of facility size in the first place.

Wright and Brown (2007) observed that there are three components to the cost of producing processed product: feedstock cost at the farm gate; feedstock delivery costs; and facility operating costs. The cost minimization problem facing the manager is

$$\min \left[\begin{array}{c} \text{Farm gate} \\ \text{feedstock costs} \end{array} + \begin{array}{c} \text{Feedstock} \\ \text{delivery costs} \end{array} + \begin{array}{c} \text{Processing} \\ \text{costs} \end{array} \right] \quad \text{such that} \quad \begin{array}{c} \text{Feedstock} \\ \text{production} \end{array} = \begin{array}{c} \text{Facility feedstock} \\ \text{processing capacity} \end{array}$$

The feedstock acquisition cost is the sum of only the first two components: farm gate feedstock costs and feedstock delivery costs. We discuss each of these two components in turn to develop a mathematical statement of the manager's objective function.

2.1 Farm gate feedstock costs

The facility requires feedstock for processing. Call the quantity of feedstock arriving at the facility Q. Feedstock production, Q, is the product of growing region area, L, and per-unit land productivity, y, i.e. $Q = y\,L$. The facility feedstock processing capacity is $\bar{Q}$.

2.1.1 Land productivity, y

Since the feedstock is perennial, the productivity of a single plant varies over its lifespan so the total productivity of the region is the weighted sum of the productivities of all the constituent plants.[1] Let $f(a)$—the age-yield function—be the yield per unit land of a-year-old plants.

The age-structure of the region through time can exhibit many different trajectories (see Mitra, Ray, and Roy (1991) for more discussion), but we restrict this analysis to a special type of trajectory: the balanced region. In a balanced region the distribution of plant ages follows a uniform distribution from 0 to the maximum plant age, n (Tisdell and De Silva, 1986). The density of each age is $\frac{1}{n}$. We call this an n-region. We call an *older* region one with a larger n, and a *younger* region one with a smaller n. The average age of an n-region is $\frac{n}{2}$. For example, over the 2009/10 to 2014/15 harvest seasons, São Paulo had an average sugarcane age of 3.76 years. Given the uniform distribution assumption, this average age corresponds to a maximum age of $n = 7.52$.

The balanced region is the supply-variation minimizing steady-state age-structure (Tisdell and De Silva, 1986). There are two reasons to focus on balanced regions. First, balancing a region to minimize supply variation is frequently a direct management objective for perennial crop growers and processing facility operators (Tisdell and De Silva, 1986; Margarido and Santos, 2012; Sharma et al., 2013; Mafakheri and Nasiri, 2014; Debnath, Epplin, and Stoecker, 2015). Second, it allows us to write a simple

model that can focus directly on the trade-offs between age, land, and processing facility capacity, while avoiding the technical details of transition dynamics. Because of this, our analysis must be considered a long-run equilibrium.

2.1.2 Allowable age-yield functions

Recall, $f(a)$—the age-yield function—is the yield per unit land of a-year-old plants. We impose the following conditions on the age-yield function to ensure an analytical solution

$$f(a) \text{ is continuous} \tag{1}$$

$$f(0) = 0 \tag{2}$$

$$f(a) \text{ monotonically increases to a maximum, then monotonically decreases} \tag{3}$$

$$\lim_{a \to \infty} a\, f(a) = 0 \tag{4}$$

Assumption (1) aids analysis of the supply-chain optimization problem. Although continuity can only ever be an approximation of an empirical age-yield function, we consider it to be a reasonable assumption, and that it is a worthwhile price to pay to facilitate analysis. Assumption (2) requires that plants are non-yielding when they are planted. This is a reasonable assumption when considering the entire life-cycle of a plant. However, it may be possible for the manager to buy young plants that are yielding when he takes possession of them. We exclude this possibility. Assumption (3) is similar to a standard assumption in the perennial crop theory literature (Mitra, Ray, and Roy, 1991), but it is a little stronger, since it excludes the possibility that plants may have a maturity phase where they produce their maximum yield for several years in a row. Assumption (3), however, allows the age-yield function to become arbitrarily close to this case. Assumption (4) requires the age-yield function approach

zero 'fast enough' as the age of the plant approaches infinity. In particular the assumption requires that the age-yield function approach zero faster than $\frac{1}{x}$. However, this assumption imposes an important modeling restriction. The age-yield function cannot approach a positive constant, which may be an attractive assumption if the plant has a long period of relatively constant yield toward the end of its life.

2.1.3 Feedstock production from an n-region

Assuming a uniform distribution of age, for a region with maximum age n, the yield of feedstock per unit of land is

$$y(n) = \frac{1}{n}\int_0^n f(a)\ da \tag{5}$$

This is referred to as the n-region yield function to distinguish it from, $f(a)$, the age-yield function, or, if unambiguous, simply the yield function. There is a trade-off between marginal and average yield inherent in this function. Since $f(a)$ is non-negative, the integral term is increasing in n. But an increase in n also increases the number of age classes that the yield must be averaged across. Whether y is increasing or decreasing in n depends on the contribution of the marginal plant relative to the average at that n, which is shown by the derivative of y with respect to n.

$$\frac{dy}{dn} \;=\; \frac{1}{n}\left(f(n) - \frac{1}{n}\int_0^n f(a)\ da\right) \;=\; \frac{1}{n}\Big(\underbrace{f(n)}_{\substack{\text{Yield of}\\ \text{additional}\\ n\text{-plant}}} - \underbrace{y(n)}_{\substack{\text{yield of}\\ n\text{-region}}}\Big) \tag{6}$$

The terms in the parentheses are multiplied by $\frac{1}{n}$ because the contribution of any single plant is diluted with an increase in the number of ages in the n-region. Let the maximum age of the uniform distribution that maximizes yield be n_{MSY}. Figure 1 shows an example of the n-region yield and age-yield functions.

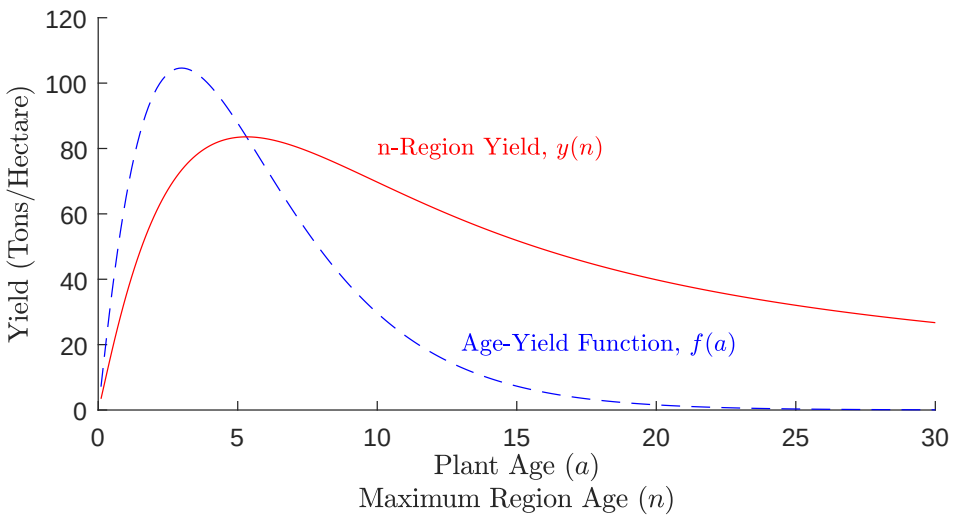

Figure 1: Yield is increasing in n while the yield of the marginal age class, $f(a)$, is more productive than the average of the n-region, $y(n)$.

2.1.4 Farm-gate feedstock costs

Feedstock production costs are separated into age-dependent costs and age-independent costs. Age-independent costs must be incurred per unit land, regardless of the age distribution of the plants on it. Examples include the costs of the manager's time, the rental rate of the land, any irrigation infrastructure etc. Since this cost is *fixed* relative to the age of the plants, it is denoted C_f. Alternatively, other costs depend on the age distribution of land. For example, in an n-region, only $\frac{1}{n}$ of the plants are replanted each year, so the average annual replanting cost is $\frac{1}{n}$ times the cost of replanting an entire unit of land. The age-dependent feedstock cost is denoted C_n.

The classification of a specific production cost into age-dependent or age-independent costs can be unintuitive, especially for activities that are carried out in most but not all years. We illustrate this using a sugarcane enterprise budget prepared for growing a 6-region the South-Central region of Brazil (Teixeira, 2013). In the budget, costs

are divided into five categories, delivery costs, and four that account for farm gate feedstock costs: preparing the soil, planting, harvest, and maintenance of the ratoon.

The total farm gate feedstock costs for a 6 age-class operation is given by

$$\begin{matrix}\text{Total Farm Gate}\\ \text{Feedstock Costs}\end{matrix} = \text{Soil Prep.} + \text{Planting} + 5 \times \text{Harvest} + 4 \times \text{Ratoon}$$

Since the total cost is given for 6 age-classes, the total cost per hectare is

$$\begin{matrix}\text{Total Farm Gate}\\ \text{Feedstock Costs}\\ \text{(Per Hectare)}\end{matrix} = \frac{1}{6} \times \text{Soil Prep.} + \frac{1}{6} \times \text{Planting} + \frac{5}{6} \times \text{Harvest} + \frac{4}{6} \times \text{Ratoon}$$

Assuming that these cost parameters are constant with respect to the number of age-classes we can write the total farm gate feedstock per hectare as a function of the age structure

$$\begin{aligned}\begin{matrix}\text{Farm gate}\\ \text{feedstock costs}\end{matrix}(n) =& \frac{1}{n} \times \text{Soil Prep.} + \frac{1}{n} \times \text{Planting} + \frac{n-1}{n} \times \text{Harvest} + \frac{n-2}{n} \times \text{Ratoon}\\ =& \text{Harvest} + \text{Ratoon} + \frac{1}{n}(\text{Soil Prep.} + \text{Planting} - \text{Harvest} - 2 \times \text{Ratoon})\end{aligned}$$

Hence the age-independent costs, C_f, will be the sum of harvest and ratoon maintenance costs, while the age-dependent costs are the soil preparation and planting costs, less the avoided harvest costs during the planting year and the ratoon maintenance costs for the planting year and the first harvest year.

2.2 Total land, L

The other component determining total feedstock quantity is the area of land controlled by the manager, L. The choice of L determines how many units of land have perennial

feedstock with average yield $y(n)$ on them. This determines total feedstock production, $Q = y(n)\,L$, and total feedstock growing costs, $L\,(C_f + \frac{C_n}{n})$.

2.2.1 Delivery costs

The total area of land also affects the cost of transporting the feedstock from the farm gate to the processing facility (Wright and Brown, 2007). Delivery costs are proportional to the quantity of feedstock multiplied by the average delivery distance.

The average delivery distance is increasing in the area of land around the facility. In the case of a facility surrounded by a circular delivery region with radius r_{max}, following Overend (1982), the distance from the facility to the furthest field is given by

$$L = \pi r_{max}^2 \;\Rightarrow\; r_{max} = \sqrt{\frac{L}{\pi}}$$

The area-weighed average delivery (transportation) distance is $r_{av} = \frac{2}{3} r_{max}$ (Stone, 1991). We express delivery costs as $C_D\, y(n) L^{1.5}$. Bringing another unit of land into the growing region increases both the quantity of feedstock produced and the average distance all feedstock must be transported (delivery costs can equivalently be expressed as $C_D\, Q L^{0.5}$). The increase in feedstock quantity is linear (holding yield constant) and the increase in average delivery distance is proportional to $L^{0.5}$, making the delivery cost function a convex function of growing region area (see appendix B.1 for full derivation).

We make a distinction between the area of land planted (growing region) and the total area of the delivery region. To allow for the possibility that some land in the delivery region is used for other purposes, we allow the planted area to be a linear function of total growing region area, $L = d \times A$ where A is the total delivery region area, and d $(0 < d \leq 1)$ is a density parameter. This facilitates calibrating the model.

2.3 Objective function

Recall the manager's objective is to minimize the cost of feedstock acquisition, given by:

$$\underset{\text{costs}}{\text{Feedstock}} = \underset{\text{feedstock costs}}{\text{Farm gate}} + \underset{\text{delivery costs}}{\text{Feedstock}}$$

Using the notation and formulas developed in the previous section, we can rewrite the feedstock cost function mathematically

$$C(n, L) = \left(C_f + \frac{C_n}{n} \right) L + C_D \, y(n) \, L^{1.5} \tag{7}$$

where n is maximum age, L is area of the growing region, C_f is the age-independent cost per unit of land, C_n is the age-dependent cost per unit of land, and C_D is the delivery cost parameter.

3 Cost Minimization and Comparative Statics

We now return to the manager's optimization problem, minimizing the costs of supplying a processing facility of a given size ($\bar{Q}$):

$$\min_{n,L} \; C(n, L) = (C_f + \frac{C_n}{n})L + C_D \, y(n) \, L^{1.5} \qquad \text{s.t. } y(n)L = \bar{Q} \tag{8}$$

The Lagrangian associated with this cost minimization problem is

$$\mathcal{L}(n, L, \lambda) = (C_f + \frac{C_n}{n})L + C_D \, y(n) L^{1.5} + \lambda(\bar{Q} - y(n)L) \tag{9}$$

3.1 First order conditions

The three first order conditions for the cost minimization problem are

$$\frac{\partial \mathcal{L}}{\partial n} = \frac{-C_n\, L}{n^2} + C_D\, y'(n)\, L^{1.5} - \lambda\, y'(n)\, L = 0 \tag{10}$$

$$\frac{\partial \mathcal{L}}{\partial L} = (C_f + \frac{C_n}{n}) + 1.5\, C_D\, y(n)\, L^{0.5} - \lambda\, y(n) = 0 \tag{11}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \bar{Q} - y(n)\, L = 0 \tag{12}$$

Equations (10) - (12) state that the marginal change in the Lagrangian function with respect to each of the choice variables is necessarily zero at the optimum.

The left hand side of equation (10) shows how the Lagrangian function changes with respect to an increase in the maximum age. There are three components. The first component is the change in age-structure dependent costs (e.g. average replanting costs). This is always negative since the costs are averaged over more ages as n increases. The second component is the change in delivery costs due to the increase in maximum age. This can be either positive or negative depending on the sign of marginal yield, $y'(n)$. If marginal yield is negative, then an increase in maximum age reduces delivery costs since there is less feedstock to deliver. The third term is the penalty for violating the quantity constraint. If $y'(n)$ is non-zero, a change in n changes the quantity of feedstock produced (since we are holding planted area constant). If the constraint was satisfied before the change, then it will now be violated after the change. Generally, λ represents the penalty for violating the constraint by a single unit (at the optimum it represents the change in feedstock costs due to a unit increase in processing facility capacity). So the third term is the product of the per-unit penalty and the change in total feedstock quantity due to the increase in maximum age.

The left hand side of equation (11) shows how the Lagrangian function changes with marginal increase in the planted area. The three components have similar inter-

pretations to the components of equation (10) except that now maximum age is being held constant. The first term is the marginal cost of growing feedstock on an additional unit of land. The second term is the marginal cost of delivery from the additional unit of land. This is an increasing function of total land due to the convexity of delivery costs. Again, the third term is the penalty for violating the capacity constraint, given the penalty per unit, λ.

3.2 Isoquant and isocost curves

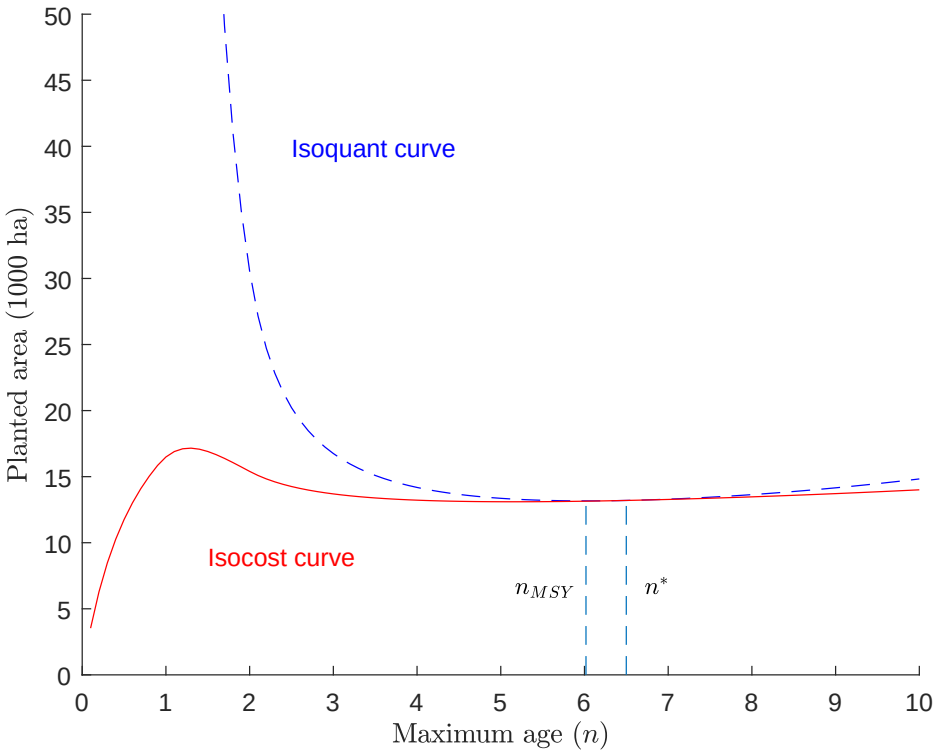

Figure 2: Example isocost and isoquant curves for the cost minimization model. In this instance, the lower contour set for the isocost curve is non-convex.

Figure 2 shows example isocost and isoquant curves for the constrained minimiza-

tion problem presented in the previous section. The isocost and isoquant functions are presented in (n, L) space, so both are functions of n. The figure was generated in MATLAB using the calibration for sugarcane described in section 4. Although calibrated to the sugarcane industry in São Paulo state, Brazil, this figure displays all the qualitative features of an isocost and isoquant curve of the general constrained minimization problem, as the next sections establish.

Looking at figure 2, two issues arise regarding using the first order conditions to solve the constrained minimization problem. The first is that the domain of the isocost and isoquant function is unbounded to the right. The second is that the lower-cost set of the isocost curve is non-convex. The non-convexity of the lower-cost set in figure 2 provides a counterexample to the statement 'all lower-cost sets from the cost function (equation 8) are convex'. Hence, the lower-cost sets are generally non-convex. Intuitively, we see that the cost function is the sum of a linear term in L and two non-quasiconvex functions in n and L. These two issues mean that we cannot immediately invoke the usual sufficiency conditions for a convex optimization problem, which call for a bounded domain and a convex lower-contour set for the objective function and guarantee that a solution to the first order conditions is also a solution to the original optimization problem.

The first issue can be dispensed with almost immediately by noting that as n approaches infinity, yield approaches zero, requiring the growing region to increase without bound. The marginal cost savings from increasing n and reducing age-independent costs approach zero as n approaches infinity, whereas the marginal cost of expanding the growing region is always positive (and in fact increasing due to the convex delivery costs). Hence it cannot be optimal to let n approach infinity, implying that for each parameter set there must be some upper bound beyond which the optimal n can never be found.

The second issue, a non-convex lower-cost set is dealt with by propositions 1–2.

3.2.1 Ruling out the optimality of $n \leq n_{MSY}$

Proposition 1 *The optimal maximum age, n^* must be strictly greater than the maximum age that maximizes yield, n_{MSY}, i.e. $n^* > n_{MSY}$.*

This proposition shows that production costs can never be minimized while region yield is an increasing function of maximum age due to the presence of age-dependent costs. Intuitively, for each $y(n)$ below the maximum, there is an n that generates an identical $y(n)$ to the left of the maximum and to the right of the maximum. The n to the right of the maximum has a lower average age-dependent cost, since age-dependent costs are dispersed over a larger number of age-classes, and thus is always the lower cost choice. Moreover, the maximum sustainable yield age, n_{MSY}, cannot be cost minimizing, since, at this age, a marginal increase in age has no first-order effect on yield, but will cause a marginal reduction in age-dependent costs. The next section establishes that a solution to the first order conditions exist in the (n_{MSY}, ∞) region. See appendix A for proof.

3.2.2 The existence of a solution to the first-order conditions

Proposition 2 *Given assumptions (1)-(4), a solution, n^*, to the first order conditions exists such that $n^* \in (n_{MSY}, \infty)$.*

The intuition behind this proof is that the derivatives of the isocost and isoquant functions must be equal somewhere on the set (n_{MSY}, ∞). At n_{MSY} the slope of the isocost function is greater than the slope of the isoquant function. Conversely, as n approaches infinity, the slope of the isoquant approaches a positive constant while the slope of the isocost function approaches zero. By continuity and the intermediate

value theorem, there must be some intermediate point where the slopes are equal. See appendix A for proof.

3.3 Comparative Statics

The signs of the comparative statics of the cost-minimization model are presented in table 1. Proofs for these signs are presented in appendix A.1.

As the processing facility size increases, the optimal maximum age decreases, and approaches the maximum sustainable yield age. For $n > n_{MSY}$, a decrease in n increases the yield of the feedstock growing region. Since a fixed quantity of feedstock must be produced, higher yield allows the growing region to be marginally smaller. We have shown that it is always worthwhile to marginally boost yields as the facility size increases. The magnitude of this effect, and whether it is economically important, depends on the particular parameterization of the model.

If the second derivative of the yield function is negative at the optimum, an increase in facility capacity increases growing region area. Intuitively, the increase in feedstock required for an increase in facility capacity must come from some combination of increased land and/or increased yield. An increase in facility capacity always decreases optimal age, and hence increases yield. The question becomes whether the increased yield is sufficient to provide the additional feedstock alone, or must be augmented with additional land. From the proof in the appendix, we see there are 5 terms that affect the sign of $\frac{dL^*}{dQ}$. However, when $y''(n^*) < 0$, a decrease in age has a diminishing increase in yield. In this case, the increase in yield can never be enough to feed the refinery alone, and additional land is always required. By definition, the second derivative of average yield will be negative at the maximum sustainable yield age. Therefore, there is a neighborhood around n_{MSY} in which increases in facility size always increase growing region size.

$\mathbf{x}$	$\frac{\mathbf{dn^*}}{\mathbf{dx}}$	$\frac{\mathbf{dL^*}}{\mathbf{dx}}$
$\bar{Q}$	(< 0)	$y''(n^*) < 0 \Rightarrow (> 0)$
C_f	(< 0)	(< 0)
C_n	(> 0)	(> 0)
C_D	(< 0)	(< 0)

Table 1: Signs of comparative statics of n^* and L^*.

The parameter C_f represents the age-independent cost of feedstock production. An increase in C_f increases the marginal cost of land, so the optimal choice moves away from land and toward yield through a reduction in the maximum age.

The parameter C_n represents the age-dependent costs of feedstock production. This cost includes replanting costs, as well as the cost of maintaining plants in the years after they are planted. The total cost of replanting and maintenance is dependent on the maximum age, since $\frac{1}{n}$ of the region is being replanted and $\frac{n-1}{n}$ is being maintained each year. An increase in the age-dependent cost parameter always increases the attractiveness of an older region, since an older region distributes the age-dependent costs across more ages, reducing the average age-dependent costs. See appendix B for more details.

The parameter C_D represents the costs of delivering feedstock from the field to the processing facility. It affects both costs of yield and of land, since both these variables determine the quantity of feedstock transported. Yield is only indirectly affected through total quantity. Since this analysis is focused on a cost-minimization problem with respect to a fixed facility size, a change in delivery cost only directly affects the marginal cost of land, leaving the quantity of feedstock produced unchanged. Therefore, an increase in the delivery cost increases the marginal cost of land, and like

the effect of C_f, shifts the optimal choice of land and yield towards yield and away from land.

4 Simulations for numerical comparative statics

In the 2011/12 harvest year, indicative of the period of our calibration, Brazil produced 561 million tons of sugarcane, of which 276 million tons was processed into 35.8 million tons of sugar and 285 million tons was processed into 22.8 billion liters of ethanol (CONAB, 2013b). The total gross value of sugarcane production was 39.2 billion Reals, representing 20.1 percent of the total value of production of Brazilian crops (IBGE, 2022). In 2012, just over one million workers were employed by the sector, with 331 710 employed in sugarcane production, 552 874 employed in sugar production, and 207 991 employed in ethanol production, accounting for 3.1 percent of total employment in Brazil and 27.1 percent of agricultural workers (Moraes, Oliveira, and Diaz-Chavez, 2015).

These products were produced from sugarcane grown on 9.6 million hectares of land, representing around 14.1 percent of Brazil's land used for crops (IBGE, 2022). Brazil was the world's largest producer of sugarcane, producing 18.9 percent of the global total, more than the next five largest producing countries combined (FAO, 2022).

More recently, the sector has grown, now producing 657 433 thousand tons of sugarcane in the 2021/22 harvest season. This harvest was processed into 41.5 million tons of sugar and 32.5 billion liters of ethanol (UNICA, 2021). The area used to grow the sugarcane remained almost constant at 10.0 million hectares, but representing a smaller fraction, 11.5 percent, of the total area of crops grown in Brazil. The value of sugarcane produced increased to 75.3 billion Reals (IBGE, 2022).

The state of São Paulo produces more sugarcane than any other Brazilian state,

responsible for 54.5 percent of sugarcane production in 2011/12 (CONAB, 2013b) and 56.6 in 2020/21 (IBGE, 2022).

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 and/or ethanol (James, 2004). Harvesting occurs between April and December, the dry season, and the sucrose content of the cane reaches a maximum in September and October (UNICA, 2012). Mechanized harvesting is replacing manual harvesting, eliminating the need to burn the cane before harvest. A single machine can harvest up to 800 tons of cane in a single day (de Moraes and Zilberman, 2014; Sant'Anna et al., 2016).

After harvest, sugarcane is highly perishable, requiring processing as soon as possible to avoid losing sugar content, know as total recoverable sugar (TRS). Most cane is harvested from fields close to the mill—the average distance in São Paulo in was 25.99km kilometers in 2011/12 (CONAB, 2013b)—and is delivered in less than 24 hours after harvest.

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 prevailing market conditions (Dias et al., 2015).

4.1 Data

The model was calibrated with productivity, cost, and facility capacity data from the São Paulo region of Brazil. Data on facility capacity was obtained from Crago et al. (2010) who obtained facility production costs from 20 mills (mills and facility are used interchangeably in this section), with capacity ranging from 1 to 32 millions tons. These mills represented over 30 percent of installed sugarcane processing capacity in Brazil in 2007-08. While these mills represent only a subset of the 153 mills present

in São Paulo state in 2007-08, the range of 1 to 32 million tons covers 69.9 percent of the number of mills and 91.9 percent of processing capacity the state (CONAB, 2008). This range better represents the most recent CONAB data from 2015-16, when São Paulo had 160 mills. By this time, however, mill capacity had increased, with only 13 percent of mills smaller than 1 million tons, representing only 2.99 percent of processing capacity (CONAB, 2019).

Sugarcane yield data is obtained from Margarido and Santos (2012) who provide an example of planning a sugarcane production system in the Alta Mogiana region of São Paulo. These yield data are similar to those reported by CONAB (2008, 2019). This data is used to fit a piecewise-linear age-yield function, which has four parameters:

$$f(a) = \begin{cases} 0 & \text{if } a \in [0, t_1) \\ \frac{f_{max}}{t_{max} - t_1}(x - t_1) & \text{if } a \in [t_1, t_{max}) \\ \frac{-f_{max}}{t_T - t_{max}}(x - t_T) & \text{if } a \in [t_{max}, t_T) \\ 0 & \text{if } a \in [t_T, \infty) \end{cases} \tag{13}$$

Where a is the age of the plant, t_1 is the age of initial yield, t_{max} is the age of maximum yield, t_T is the age of final yield, and f_{max} is the maximum yield. The fit is shown in figure 6 in appendix B.1.1. Because sugarcane attains its highest yield in the first year of harvest, the piecewise linear function provides an improved fit over other functions used to fit age-yield data that cannot capture this rapid increase, e.g. the Hoerl function (Haworth and Vincent, 1977).

We derived the feedstock cost parameters, C_f and C_n, from Teixeira (2013), and the delivery cost parameter, C_D, from Crago et al. (2010). Teixeira (2013) presents an example enterprise budget for a 6-region sugarcane operation in São Paulo state. The acreage of this representative operation is not stated. Costs are divided into five

categories, delivery costs, and four that account for farm gate feedstock costs: preparing the soil, planting, harvest, and maintenance of the ratoon. The total cost per hectare is

$$\begin{aligned}\text{Total Farm Gate Feedstock Costs (Per Hectare)} =& \frac{1}{6} \times \text{Soil Preparation} + \frac{1}{6} \times \text{Planting} \\ &+ \frac{5}{6} \times \text{Harvest} + \frac{4}{6} \times \text{Ratoon Maintenance}\end{aligned}$$

Assuming that these cost parameters are constant with respect to the number of ages we can write the total farm gate feedstock per hectare as a function of the maximum age:

$$\begin{aligned}\text{Farm gate feedstock costs}(n) =& \frac{1}{n} \times \text{Soil Preparation} + \frac{1}{n} \times \text{Planting} \\ &+ \frac{n-1}{n} \times \text{Harvest} + \frac{n-2}{n} \times \text{Ratoon maintenance}\end{aligned}$$

Substituting Teixeria's numbers (in Reals) from the example budget, the cost function becomes

$$\begin{aligned}\text{Farm gate feedstock costs}(n) =& \frac{656.07}{n} + \frac{4159.83}{n} + \frac{n-1}{n} \times 1273.13 + \frac{n-2}{n} \times 986.54 \\ =& 2259.67 + \frac{1569.69}{n}\end{aligned} \tag{14}$$

Hence for the simulations we use a baseline of $C_f = 2259.67$ and $C_n = 1569.69$.

While Teixeira (2013) does include estimates of delivery costs, he does not include the processing facility size that this example farm is feeding. We therefore turn to Crago et al. (2010) to derive the delivery cost parameter. The total delivery cost from

a growing region is given by

$$\begin{aligned}\overset{\text{Total}}{\underset{\text{Delivery Costs}}{}} =& \text{Average Cost Per Ton Kilometer} \\ & \times \text{Quantity Transported} \\ & \times \text{Average Delivery Distance}\end{aligned}$$

Let δ represent the average delivery cost per ton kilometer (i.e. the average cost to transport one ton of feedstock one kilometer). Crago et al. (2010) report an average transport cost of R$6.7 to transport a ton of feedstock from the farm gate to the mill. The average delivery distance in this study was 22 kilometers so in this case $\delta = 0.3045$. This distance is smaller than the São Paulo state averages reported in 2007-08 (23.87km) and 2015-16 (26.78km) (CONAB, 2008, 2019).

The average mill size in Crago et al. (2010) is 4.8 million tons. Given our assumption that the growing region produces the exact quantity required to feed the mill, this implied that the average quantity of feedstock transported was 4.8 million tons. Recall that when calculating the average delivery distance, we must make a distinction between the area of land planted with sugarcane, L, and the area of the growing region, A. Although we are assuming that the growing region is circular, it is not necessarily the case that all the land is planted with sugarcane. In fact, relaxing the link between planted area and growing region area is necessary to correctly calibrate the model to the data in Crago et al. (2010).

Let d be the average density of sugarcane fields in the growing region, and A be the area of the growing region, so $L = d \times A$. The average delivery distance, r_{av}, which is the average radius from the center to points in the circle, is given by the expression

$$r_{av} = \frac{2}{3} r_{max} = \frac{2}{3}\sqrt{\frac{A}{\pi}}$$

Where r_{max} is the radius of the growing region, and π is the mathematical constant. Since the average delivery distance, r_{av}, from Crago et al. (2010) is 22km, the size of the growing region is $A = 342\,119$ ha. The total quantity of feedstock production is the product of the yield, density, and growing region area. Crago et al. (2010) reports an average yield of 75 tons per hectare, so $d = 0.187$. The expression for total delivery costs is

$$\begin{aligned} \underset{\text{Delivery Costs}}{\text{Total}} &= \frac{2\delta}{3}\sqrt{\frac{1}{d \times \pi}} \times y(n)L^{1.5} \\ &= C_D \times y(n)L^{1.5} \end{aligned} \tag{15}$$

For the d and δ derived from Crago et al. (2010), $C_D = 0.2649$.

4.2 Simulation Results

Figure 3 shows a numerical example of the changes in cost minimizing age and planted land area as the facility processing capacity is increased. The shape of the expansion path corresponds to the results of the comparative statics of n^* and L^* as we increase $\bar{Q}$. As facility capacity increases from 1 to 36 million tons, the optimal maximum age decreases from 6.50 to 6.21, corresponding to an increase in the n-region yield from 75.76 tons per hectare to 75.92 tons per hectare. The majority of the additional feedstock is supplied from additions to planted land area.

Over the period 2009-2015, the average age of sugarcane in São Paulo was 3.76 years, which, if the ages are assumed to be distributed uniformly, implies a maximum age of 7.52 (CONAB, 2012, 2013a,b, 2017a,b,c). Denote this maximum São Paulo age as $n_{SP} = 7.52$. Figure 3 shows that as facility size increases, the cost minimizing maximum moves towards the maximum sustainable yield age, and away from the São Paulo maximum age. The difference between n^* and the São Paulo maximum age

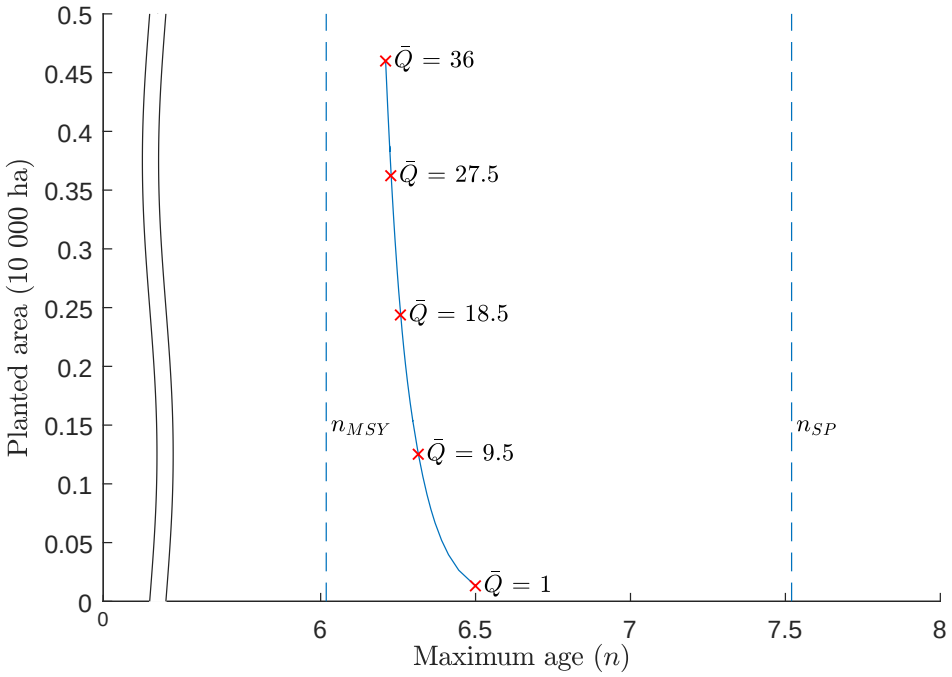

Figure 3: Cost minimizing age and planted land area as processing facility capacity is increased from 1 million tons to 36 million tons. With increased capacity, age approaches $n_{MSY}(= 6.02)$.

implies a difference in average age of between six and eight months. Conditional on the maintained assumptions of the model, what is the difference in feedstock acquisition cost between using an n^*-region and an n_{SP}-region for facility of a given size?

Figure 4 shows the difference in cost between the cost of choosing both n and L to minimize the cost of supplying the processing facility (n^*-region) and the cost of fixing n at n_{SP} and only allowing L to vary (n_{SP}-region). The cost difference ranges from 0.75 percent when supplying a one million ton facility to 0.94 percent for a 36 million ton facility. For the range of facilities presented in the graph, the condition for the comparative static of land with respect to refinery capacity is satisfied and the increased capacity is met with an increase in land and a decrease in optimal age toward

n_{MSY}.

The optimal maximum age is lower for larger facilities. The difference between the optimal maximum age and the fixed São Paulo maximum age, and consequently the difference in costs, increases as the capacity of the facility increases. As the processing facility's capacity increases, the difference in cost increases at a decreasing rate. This increase is approximately log-linear, with a one percent increase in the facility capacity leading to an approximately 0.062 percentage point increase in the cost difference. To put this in perspective, consider a 10 million ton mill. The cost reduction from optimizing age, relative to the São Paulo average age, is 0.872 percent, or 1.137 million Reals. Increasing the capacity of the mill by one percent, to 10.1 million, will increase the cost difference by 0.062 percentage points, or 704 Reals.

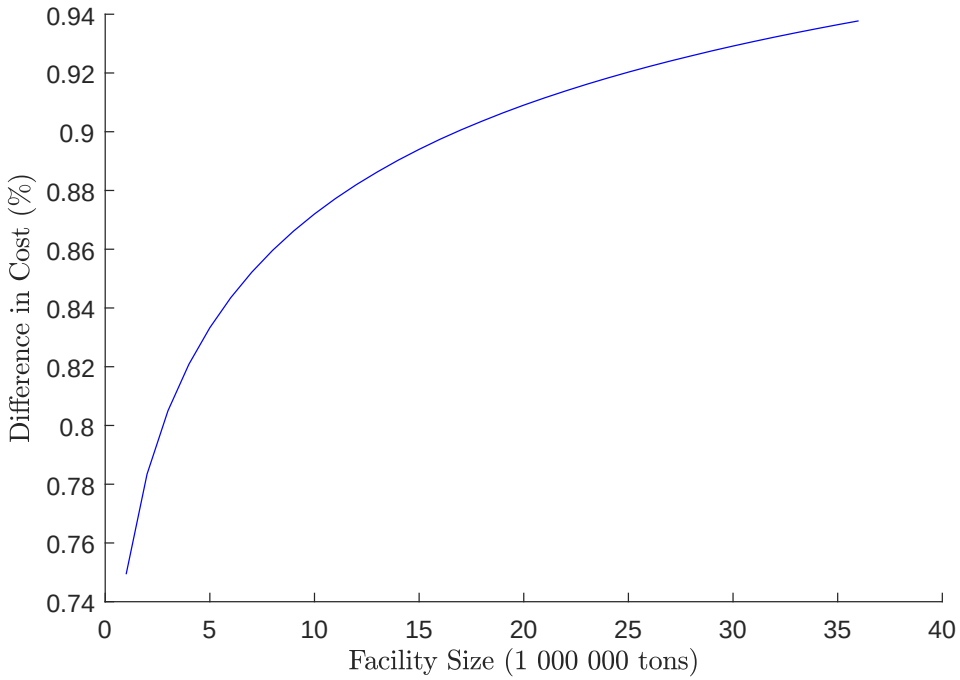

Figure 4: Percentage difference in cost between using the cost-minimizing maximum age (n^*) and implied São Paulo maximum age ($n_{SP} = 7.52$).

4.3 Sensitivity

Table 1 provided qualitative comparative statics of n^* and L^* with respect to facility size and cost parameters. To examine the quantitative magnitude of these derivatives, as well as the effect of changes in the piecewise linear age-yield function parameters, we generated random parameter sets and solved the cost minimization problem for each set. We then used linear regression to estimate the approximate marginal effects of each parameter on total feedstock costs, optimal maximum age, and optimal area.

Simulation parameters were drawn from a uniform distribution generated by scaling draws from a Halton sequence. For cost parameters, we centered the uniform distribution on the calibrated parameter values, with a minimum parameter value of half the calibrated value, and a maximum value of 1.5 times the calibrated value. In the absence of information on plausible ranges for cost and yield parameters, the uniform distribution and bounds were chosen to create sufficient variation to estimate the marginal effects. Mill capacity ranged between 1 and 36 million tons (the range from Crago et al. (2010)).

Each of the four parameters of the piecewise linear age-yield function was allowed to vary independently. In the baseline calibration, the time to initial yield was 1 year after planting and maximum yield was achieved one year after initial yield. The age of final yield was 13 years and the maximum yield was 120 ton/ha. Varying the four parameters independently allows hypothetical age yield relationships to be considered, e.g. what if maximum yield was reached two years after initial yield? The age of maximum yield is defined relative to age of initial yield. If the age of initial yield increases by 1, so will the age of maximum yield. Likewise, the age of final yield is defined relative to the age of maximum yield. The parameter values used in the simulation are reported in table 3 in appendix B.2.

Using MATLAB, we generated 100 000 random parameter sets. For each random

parameter set, the `fmincon` function was used to solve for the cost minimizing values of n and L along with the associated costs for feeding the facility using the cost minimizing age and area. Because of numerical issues or theoretical incompatibility (e.g. $n^* < 0$), 3.9 percent of draws were eliminated.

The simulations allow us to construct numerical comparative statics for all parameters, using linear regression on the simulated data. The results of these regressions are shown in table 2. Each column is a regression of either total cost, optimal age, or optimal land on the parameters and a constant. The natural logarithm of all dependent and independent variables has been taken, so estimated coefficients should be interpreted as elasticities.

The signs of the coefficients are consistent with the theoretical signs presented in table 1, although the coefficients of L^* with respect to age-independent costs and delivery costs are not significantly different from zero. Total costs are convex with increases in facility size due to the corresponding increases in growing region area. Total costs are most sensitive to changes in the delivery cost parameter, with a one percent increase in C_D increasing total costs by 0.75 percent, while total costs are least sensitive to changes in age-dependent costs.

For the age-yield function parameters, a one percent increase maximum yield decreases total costs by 0.63 percent. Increases in time to initial yield and time from initial yield to maximum yield increase total costs as they delay the peak of the n-region yield function. Conversely, increases in the time from maximum yield to final yield decrease total costs, since older plants become relatively more productive.

	$\ln(Cost)$	$\ln(n^*)$	$\ln(L^*)$
$\ln(\bar{Q})$	1.357^{***}	-0.007^{***}	1.000^{***}
	(0.000)	(0.000)	(0.000)
$\ln(C_f)$	0.215^{***}	-0.010^{***}	-0.001
	(0.001)	(0.001)	(0.000)
$\ln(C_n)$	0.019^{***}	0.023^{***}	0.001^{*}
	(0.001)	(0.001)	(0.000)
$\ln(C_D)$	0.745^{***}	-0.015^{***}	-0.001
	(0.001)	(0.001)	(0.000)
$\ln(t_1)$	0.039^{***}	0.115^{***}	0.067^{***}
	(0.000)	(0.000)	(0.000)
$\ln(t_{max})$	0.026^{***}	0.368^{***}	0.054^{***}
	(0.000)	(0.000)	(0.000)
$\ln(t_T)$	-0.119^{***}	0.385^{***}	-0.177^{***}
	(0.001)	(0.001)	(0.001)
$\ln(f_{max})$	-0.626^{***}	-0.007^{***}	-1.000^{***}
	(0.001)	(0.001)	(0.000)
Intercept	Yes	Yes	Yes
Adj. R^2	0.998	0.908	0.998
Num. obs.	96149	96149	96149

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2: Numerical comparative statics. Log-log regression of total cost, optimal age, and optimal land on all parameters.

5 Discussion

We have presented an analytical model of feedstock acquisition cost for a processing facility when the feedstock is derived from a perennial crop, focusing on the trade-off between the cost of frequent replanting, which keeps yields high, and the cost of feedstock delivery from a large growing region. The model yielded non-convex iso-cost curves, arising from the shape of the age-yield function. Despite this, we showed the model has a cost minimizing solution, and provided comparative statics of the solution with respect to changes in processing facility size and cost parameters. Calibrating the model to the sugarcane industry in São Paulo state, Brazil, produced numerical results that conformed to the predictions of the analytical comparative statics.

A surprising result is that the feedstock acquisition cost is not particularly sensitive to the choice of growing region age. Figure 5 shows the cost of feedstock acquisition for an arbitrary maximum age relative to the minimized cost. For $\bar{Q} = 1$, the cost of maximum ages between 4.50 and 9.33 is within five percent of the minimum cost, and, for $\bar{Q} = 32$, the cost of maximum ages between 4.06 and 9.49 is within five percent of the minimum cost. The corresponding average ages range from 2.03 to 4.75. The cost function in this context is an example of a 'flat payoff function', "where even large deviations from optimal decisions make little difference to the payoff" (Pannell, 2006, p. 553), indicating that optimizing the average age of the growing region is unlikely to be a significant factor in feedstock supply chain managers' decisions. Managers may be influenced by other factors including the price paid to sugarcane farmers (relative to other crops), the price of sugar and ethanol, federal and state taxation, credit availability, and strategic factors between mills and sugarcane growers, or between mills themselves (Demczuk and Padula, 2017; Belik et al., 2017; de Moraes and Zilberman, 2014; Sant'Anna et al., 2018).

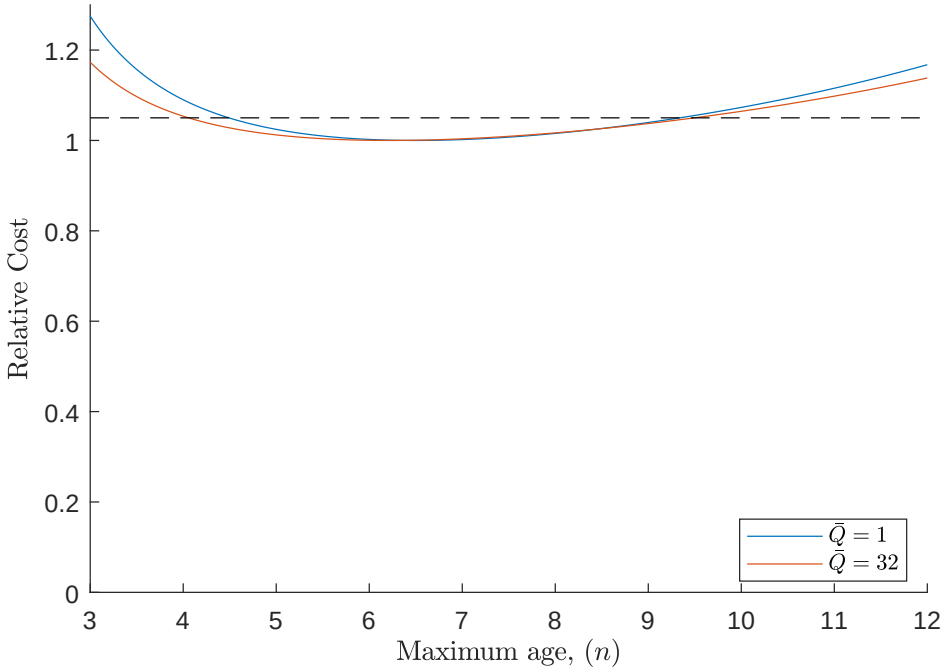

Figure 5: Feedstock acquisition costs are relatively insensitive to age. For $\bar{Q} = 1$, the cost of maximum ages between 4.50 and 9.33 is within five percent of the minimum cost. For $\bar{Q} = 32$, the cost of maximum ages between 4.06 and 9.49 is within five percent of the minimum cost.

Focusing on cost-minimization hides a potential channel for age optimization to affect the management of the growing-processing operation. By keeping the processing facility size fixed, a change in a parameter only causes a substitution between land and age. In a more flexible model that also optimized the facility capacity choice, the same parameter change would have two effects: the substitution effect from the change in the relative cost of land and age; and the size effect, where the change in the parameter alters the optimal size of the facility. These two effects may reinforce or offset each other. For example, the effects may offset when a decrease in delivery costs increases the optimal age through the substitution effect (land is relatively cheaper), and also increases the optimal size of the facility, which would reduce the optimal age, leaving

the net effect ambiguous.

These two effects could be simultaneously analyzed in a net present value maximization model with endogenous processing facility capacity choice, with the cost-minimization model providing a necessary input as feedstock acquisition costs must be minimized for any given facility size. The choice of facility size is one of the two key questions facing an entrepreneur designing a new supply chain, the other being the choice of vertical coordination strategy with input suppliers (Zilberman et al., 2022). The results from this simulation suggest that, for the sugarcane industry in São Paulo, the growing region age is not a substantial consideration in facility size (see figure 4), nor in the choice of vertical coordination strategy.

This model, which assumes vertical integration between the mill and the growing region, focuses on vertical integration, one of the two main vertical coordination strategies present in the Brazilian sugarcane industry, the other being contracting (Sant'Anna et al., 2018). In addition, there is a "well-established sugarcane spot market" in São Paulo (Sant'Anna et al., 2018, p. 808). In São Paulo in 2011/12, mills acquired 56.7 percent of their sugarcane from vertically integrated production and the remaining 43.3 percent from external parties. The nature of sugarcane creates "strong economic incentives for processors to vertically integrate upstream into sugarcane production or at least control some key agricultural decisions and the planning of harvesting and transportation to the mill. This is achieved with a multiple sourcing strategy combining vertical integration with long-term production contracts with growers"(Chaddad, 2016, p. 92-93).

Sant'Anna et al. (2018) argued that a sugarcane mill's choice of vertical coordination strategy is driven by a choice of both operations perspectives, i.e. optimizing sugarcane acquisition costs and TRS content, and strategic perspectives, e.g. securing a stable procurement base, reducing the bargaining power of suppliers, and creating barriers to

entry. These authors measured sugarcane mills' technical efficiency and found that, while vertical integration had a statistically significant impact on efficiency, "[t]he results imply that the decision to backwards vertically integrate is not primarily driven by the desire to increase technical efficiency" (Sant'Anna et al., 2018, p. 805). Given that technical efficiency is not a primary driver of the decision to vertically integrate, and our results showing that average age optimization has a small impact on feedstock acquisition costs anyway, it is unlikely that considerations of average age optimization are an important consideration in the decision to vertically integrate.

More generally, vertical coordination through contracting is likely to lead to a weakly greater average age than under vertical integration. If the the grower were obliged to pay the marginal delivery cost of feedstock, they would face similar incentives to the processing facility and would likely choose the cost-minimizing age, assuming a sufficiently non-'flat' cost function (Pannell, 2006). Conversely, if growers only faced marginal growing costs, but not delivery costs, they would likely choose an average age greater than n^*. On the other hand, sugarcane mills engaged in contracting may indirectly affect growing region age through the length of contract offered. The 'ideal' sugarcane production cycle is six years (Margarido and Santos, 2012; Demczuk and Padula, 2017). Brazilian sugarcane contracts are generally specified as a multiple of this cycle (six years or twelve years) (Sant'Anna et al., 2022), tending to move the maximum age towards six.

Sugarcane mills are generally multi-output, producing ethanol and/or sugar from the sugarcane feedstock. In 2011, 253 of Brazil's mills could produce both ethanol and sugar, 168 produced ethanol as the sole output, and 14 produced sugar only (Walter et al., 2015). The mix of outputs produced by the mill does not affect our modeling scenario, since all mills require a single input, total recoverable sugar (TRS) from sugarcane, as the feedstock for both the ethanol and sugarcane production processes

(van den Wall Bake et al., 2009). In both cases, the mill's interest centers on acquiring sugarcane at the lowest cost (Marques et al., 2015). In a more general net present value maximization model, where the output mix and possibly the mill size were endogenous, the production possibility frontier between outputs, and prices and market structure in the output market would be relevant factors for inclusion, especially since, under the pricing system for independent producers, CONSECANA-SP, the sugarcane price depends on the output mix of the receiving mill (de Moraes and Zilberman, 2014).

The São Paulo calibration focused on minimizing the cost of producing sugarcane and delivering it to the mill. However, mills use the sucrose within the sugarcane, TRS, as the feedstock for producing sugar and/or ethanol (van den Wall Bake et al., 2009). TRS content degrades rapidly in sugarcane after harvest and cane is optimally processed within 48 to 72 hours after harvest to minimize TRS losses (Belik et al., 2017; Sant'Anna et al., 2018). Increasing the size of the growing region can increase the time taken to deliver sugarcane from the field to the mill, and hence increase TRS losses. Our simulation does not incorporate this effect, due to lack of data. However, our focus on yields (tons/ha) and replanting is consistent with the results of structured interviews with Brazilian sugarcane industry experts, who identified the variables most relevant to the production of sugarcane and ethanol as "Farm productivity (yields) (tonnes per hectare); Production costs; Price paid to sugarcane farmers; Replanting the sugarcane fields at the right time in order to maintain high average productivity (yield) levels; Price of regular gasoline to the consumer; [and] Federal and State Taxation" (Demczuk and Padula, 2017, p. 191). In future research TRS losses as a function of delivery distance could be incorporated into this model as an increase in the delivery cost coefficient, C_D, if TRS losses are proportional to delivery distance.

The current model requires that feedstock production exactly match processing facility capacity. This can be thought of as a long-run situation where the facility has

been optimized for the downstream market and the growing region has been optimized for the facility in the absence of strategic reasons to overbuild facilities. However, in reality, managers can run the facility below capacity in the short run, or sell excess feedstock to other processors (assuming delivery costs are not prohibitive). In fact, in São Paulo, the growing region is approximately half the area necessary to fully supply the installed capacity, likely as a deterrent to new entrants (Belik et al., 2017).

The assumption of a steady-state ignores the problem of achieving the optimal age in a least-cost method. Margarido and Santos (2012) present a sugarcane planting sequence for feeding a two million ton facility. This sequence front loads the planting in the first three years of production and plants the equilibrium area each following year, resulting in a stable age nine years after the initial planting. Although the authors note that the initial front loading is necessary because the facility needs to crush a larger amount of sugarcane in the first year, there is no discussion of the optimality of this planting sequence, or comparisons with alternative sequences. Therefore, the cost-minimizing age from this model should be considered a long-run average target, with year-to-year fluctuations around the target. The details of achieving the target should be determined for the particular application.

Although our model was applied to Brazilian sugarcane production, the model's core components (an age-yield function, age-dependent and age-independent costs, and a growing region feeding a central facility) could be adapted to other perennial crop industries, such as biomass ethanol (Yang, Paulson, and Khanna, 2016; McCarty and Sesmero, 2021), cocoa (Mahrizal et al., 2014; Beg et al., 2017), coffee (Das, 2021; de Sousa e Silva et al., 2021) and palm oil (Henson, 2012; Chew et al., 2021).

To implement this model for a different crop, one would need data on the crop's age-yield function, age-dependent and-independent feedstock production costs, and delivery costs. If the shape of the growing region were poorly approximated by a circle,

the delivery cost parameter could be adjusted by using an alternative average distance measure. As this is an analytical model, quantitative results from a calibration should be considered indicative of potential cost reductions from age-structure optimization, rather than a management prescription. In promising contexts, the ideas from this model could be incorporated into a detailed engineering supply chain optimization model.

Table 2 provides an indication of contexts more likely to benefit from a optimization of average age. Benefits from optimization increase in the difference between the region's unoptimized age and the optimal age. For a growing region with a given (unoptimized) age, greater cost reductions would be realized if the region were larger, had higher age-independent costs, lower age-dependent costs, and higher delivery costs. Crops that begin producing quickly after planting, reach maximum yield quickly, and have shorter lifespans are more likely to benefit, as these features would increase the variation in the yield function. Similarly, crops with higher maximum yields are more likely to benefit. The quantitative magnitude of these reductions, however, and whether they rise to the level of economic importance, would need to be evaluated on a case-by-case basis.

6 Conclusion

We have, to our knowledge, presented the first model of optimal perennial crop age when the output is used as a feedstock for a processing facility of a given size. To account for non-convexities in the cost-minimization problem, we proved under certain assumptions on the age-yield function that the first order conditions of the model have a solution. We generated analytical and numeric comparative statics of this solution with respect to facility size and cost parameters. We also provided a simulation calibrated

to the sugarcane industry in São Paulo, Brazil, finding that feedstock acquisition costs are not sensitive to age in that context.

From the results we can conclude that it is indeed *possible* to reduce biorefinery feedstock costs by optimizing the feedstock age, but in the case of the São Paulo sugarcane industry, the benefits are small relative to the continued use of the observed average age in the region. Hence, the importance of age-structure optimization for managers must be considered on a case-by-case basis, depending on whether the parameters of the particular problem generate a large cost saving from optimization relative to the relevant baseline age.

Notes

1. If this model were applied to fruit or nut production, it would be more natural to talk of the age-yield functions of individual trees and the average yields of orchards comprised of many trees.

References

Barrett, C.B., T. Reardon, J. Swinnen, and D. Zilberman. Forthcoming. "Agri-Food Value Chain Revolutions in Low- and Middle-Income Countries." *Journal of Economic Literature*, pp. 1–70.

Beg, M.S., S. Ahmad, K. Jan, and K. Bashir. 2017. "Status, Supply Chain and Processing of Cocoa - A Review." *Trends in Food Science & Technology* 66:108–116.

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.

Chew, C.L., C.Y. Ng, W.O. Hong, T.Y. Wu, Y.Y. Lee, L.E. Low, P.S. Kong, and E.S. Chan. 2021. "Improving Sustainability of Palm Oil Production by Increasing Oil Extraction Rate: A Review." *Food and Bioprocess Technology* 14:573–586.

CONAB. 2008. "Perfil do Setor do Açúcar e do Álcool no Brasil 07/08." Working paper, Companhia Nacional de Abastecimento, Brasília.

—. 2012. "Perfil Do Setor Do Açúcar e Do Álcool No Brasil 09/10." Working paper, Companhia Nacional de Abastecimento, Brasília.

—. 2013a. "Perfil Do Setor Do Açúcar e Do Álcool No Brasil 10/11." Working paper, Companhia Nacional de Abastecimento, Brasília.

—. 2013b. "Perfil Do Setor Do Açúcar e Do Álcool No Brasil 11/12." Working paper, Companhia Nacional de Abastecimento, Brasília.

—. 2017a. “Perfil Do Setor Do Açúcar e Do Álcool No Brasil 12/13.” Working paper, Companhia Nacional de Abastecimento, Brasília.

—. 2017b. “Perfil Do Setor Do Açúcar e Do Álcool No Brasil 13/14.” Working paper, Companhia Nacional de Abastecimento, Brasília.

—. 2017c. “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.

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.

Das, S. 2021. “Post-Harvest Processing of Coffee: An Overview.” *Coffee Science* 16:1–7.

De Meyer, A., D. Cattrysse, J. Rasinmäki, and J. Van Orshoven. 2014. “Methods to optimise the design and management of biomass-for-bioenergy supply chains: A review.” *Renewable and Sustainable Energy Reviews* 31:657–670.

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.

de Sousa e Silva, J., A.P. Moreli, S.M.L. Donzeles, S.F. Soares, and D.G. Vitor. 2021. “Harvesting, Drying and Storage of Coffee.” In L. Louzada Pereira and T. Rizzo Moreira, eds. *Quality Determinants In Coffee Production*. Cham: Springer International Publishing, Food Engineering Series, pp. 1–64.

Debnath, D., F.M. Epplin, and A.L. Stoecker. 2014. “Managing Spatial and Temporal Switchgrass Biomass Yield Variability.” *Bioenergy Research* 7:946–957.

—. 2015. "Switchgrass procurement strategies for managing yield variability: Estimating the cost-efficient D (downtime cost) L (land to lease) frontier." *Biomass and Bioenergy* 77:110–122.

Demczuk, A., and A.D. Padula. 2017. "Using System Dynamics Modeling to Evaluate the Feasibility of Ethanol Supply Chain in Brazil: The Role of Sugarcane Yield, Gasoline Prices and Sales Tax Rates." *Biomass and Bioenergy* 97:186–211.

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.

Douglas, J., J. Lemunyon, R. Wynia, and P. Salon. 2009. "Planting and Managing Switchgrass as a Biomass Energy Crop." Working paper No. 3, United States Department of Agriculture, Natural Resources Conservation Service, Plant Materials Program, Sep.

Du, X., L. Lu, T. Reardon, and D. Zilberman. 2016. "Economics of Agricultural Supply Chain Design: A Portfolio Selection Approach." *American Journal of Agricultural Economics* 98:1377–1388.

FAO. 2022. "FAOSTAT."

Glover, J.D., J.P. Reganold, L.W. Bell, J. Borevitz, E.C. Brummer, E.S. Buckler, C.M. Cox, T.S. Cox, T.E. Crews, S.W. Culman, L.R. DeHaan, D. Eriksson, B.S. Gill, J. Holland, F. Hu, B.S. Hulke, A.M.H. Ibrahim, W. Jackson, S.S. Jones, S.C. Murray, A.H. Paterson, E. Ploschuk, E.J. Sacks, S. Snapp, D. Tao, D.L. Van Tassel, L.J. Wade, D.L. Wyse, and Y. Xu. 2010. "Increased Food and Ecosystem Security via Perennial Grains." *Science* 328:1638–1639.

Haworth, J.M., and P.J. Vincent. 1977. "Medium Term Forecasting of Orchard Fruit Production in the EEC: Methods and Analyses." Working paper, Eurostat, Brussels.

Heaton, E. 2010. "Giant Miscanthus for Biomass Production." Working paper, Iowa State University Extension.

Henson, I.E. 2012. "Ripening, Harvesting, and Transport of Oil Palm Bunches." In *Palm Oil*. Elsevier, pp. 137–162.

Hochman, G., and D. Zilberman. 2018. "Corn Ethanol and U.S. Biofuel Policy 10 Years Later: A Quantitative Assessment." *Americal Journal of Agricultural Economics* 100:570–584.

IBGE. 2022. "Produção Agrícola Municipal." Dataset available at: https://sidra.ibge.gov.br/tabela/5457#resultado.

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

Khanna, M., and C.L. Crago. 2012. "Measuring Indirect Land Use Change with Biofuels: Implications for Policy." *Annual Review of Resource Economics* 4:161–184.

Kreitzman, M., E. Toensmeier, K.M.A. Chan, S. Smukler, and N. Ramankutty. 2020. "Perennial Staple Crops: Yields, Distribution, and Nutrition in the Global Food System." *Frontiers in Sustainable Food Systems* 4.

Mafakheri, F., and F. Nasiri. 2014. "Modeling of biomass-to-energy supply chain operations: Applications, challenges and research directions." *Energy Policy* 67:116–126.

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.

Malladi, K.T., and T. Sowlati. 2018. "Biomass Logistics: A Review of Important Features, Optimization Modeling and the New Trends." *Renewable and Sustainable Energy Reviews* 94:587–599.

Manochio, C., B. Andrade, R. Rodriguez, and B. Moraes. 2017. "Ethanol from Biomass: A Comparative Overview." *Renewable and Sustainable Energy Reviews* 80:743–755.

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.

Marques, T.A., L.C.G. Neves, E.M. Rampazo, E.L. Deltrejo Junior, F.C. Souza, and P.A.A. Marques. 2015. "TRS Value of Sugarcane According to Bioenergy and Sugar Levels." *Acta Scientiarum. Agronomy* 37:347.

McCarty, T., and J. Sesmero. 2021. "Contracting for Perennial Energy Crops and the Cost-Effectiveness of the Biomass Crop Assistance Program." *Energy Policy* 149:112018.

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.

Molnar, T., P. Kahn, T. Ford, C. Funk, and C. Funk. 2013. "Tree Crops, a Permanent Agriculture: Concepts from the Past for a Sustainable Future." *Resources* 2:457–488.

Moraes, M.A.F.D., F.C.R. Oliveira, and R.A. Diaz-Chavez. 2015. "Socio-Economic Impacts of Brazilian Sugarcane Industry." *Environmental Development* 16:31–43.

O'Neill, E.G., and C.T. Maravelias. 2021. "Towards Integrated Landscape Design and Biofuel Supply Chain Optimization." *Current Opinion in Chemical Engineering* 31:100666.

Overend, R.P. 1982. "The average haul distance and transportation work factors for biomass delivered to a central plant." *Biomass* 2:75–79.

Pannell, D.J. 2006. "Flat Earth Economics: The Far-Reaching Consequences of Flat Payoff Functions in Economic Decision Making." *Review of Agricultural Economics* 28:553–566.

Rentizelas, A.A., A.J. Tolis, and I.P. Tatsiopoulos. 2009. "Logistics issues of biomass: The storage problem and the multi-biomass supply chain." *Renewable and Sustainable Energy Reviews* 13:887–894.

Sant'Anna, A.C., J.S. Bergtold, A. Shanoyan, M.M. Caldas, and G. Granco. 2022. "Biofuel Feedstock Contract Attributes, Substitutability and Tradeoffs in Sugarcane Production for Ethanol in the Brazilian Cerrado: A Stated Choice Approach." *Renewable Energy* 185:665–679.

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.

Sharma, B., R.G. Ingalls, C.L. Jones, and A. Khanchi. 2013. "Biomass supply chain design and analysis: Basis, overview, modeling, challenges, and future." *Renewable and Sustainable Energy Reviews* 24:608–627.

Stone, R.E. 1991. "Technical Note—Some Average Distance Results." *Transportation Science* 25:83–90.

Teixeira, F.L.D.S. 2013. "Custo Médio Operacional da Lavoura de Cana-de-açúcar em Reais."

Tisdell, C.A., and N.T.M.H. De Silva. 1986. "Supply-Maximising and Variation-Minimizing Replacement Cycles of Perennial Crops and Similar Assets: Theory Illustrated by Coconut Cultivation." *Agricultural Economics* 37:243–251.

UNICA. 2012. "Final Report of the 2011/2012 Harvest Season—South-Central Region." Working paper, The Brazilian Sugarcane Industry Association (UNICA).

UNICA. 2021. "Harvest 2021/22 - Harvest Update of the South-Central Region."

van den Wall Bake, J.D., M. Junginger, A. Faaij, T. Poot, and A. Walter. 2009. "Explaining the Experience Curve: Cost Reductions of Brazilian Ethanol from Sugarcane." *Biomass and Bioenergy* 33:644–658.

Wallace, J. 2000. "Increasing agricultural water use efficiency to meet future food production." *Agriculture, Ecosystems & Environment* 82:105–119.

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. 2015. "Brazilian Sugarcane Ethanol: Developments so Far and Challenges for the Future." In P. D. Lund, J. Byrne, G. Berndes, and I. A. Vasalos, eds. *Advances in Bioenergy*. Oxford, UK: John Wiley & Sons, Ltd, pp. 373–394.

Wang, R. 2007. "The Optimal Consumption and the Quitting of Harmful Addictive Goods." *The B.E. Journal of Economic Analysis & Policy* 7.

Wright, M., and R.C. Brown. 2007. "Establishing the optimal sizes of different kinds of biorefineries." *Biofuels, Bioproducts and Biorefining* 1:191–200.

Yang, X., N.D. Paulson, and M. Khanna. 2016. "Optimal Mix of Vertical Integration and Contracting for Energy Crops: Effect of Risk Preferences and Land Quality." *Applied Economic Perspectives and Policy* 38:632–654.

Zahraee, S.M., N. Shiwakoti, and P. Stasinopoulos. 2020. "Biomass Supply Chain Environmental and Socio-Economic Analysis: 40-Years Comprehensive Review of Methods, Decision Issues, Sustainability Challenges, and the Way Forward." *Biomass and Bioenergy* 142:105777.

Zilberman, D., L. Lu, and T. Reardon. 2019. "Innovation-Induced Food Supply Chain Design." *Food Policy* 83:289–297.

Zilberman, D., T. Reardon, J. Silver, L. Lu, and A. Heiman. 2022. "From the Laboratory to the Consumer: Innovation, Supply Chain, and Adoption with Applications to Natural Resources." *Proceedings of the National Academy of Sciences* 119:e2115880119.

A Proofs of Propositions and Lemmas

Lemma 1 *The optimal maximum age must be greater than or equal to the maximum sustainable yield age, i.e* $n^* \geq n_{MSY}$.

Proof of Lemma 1.

From our assumptions on the age-yield function in section 2.1.2, for all yields less than the maximum yield, there are two ages that generate that yield. That is, for all $\bar{y} \in (0, y(n_{MSY}))$, there exist $n_{\bar{y}}^- < n_{MSY} < n_{\bar{y}}^+$, such that $y(n_{\bar{y}}^-) = y(n_{\bar{y}}^+) = \bar{y}$.

On the graph of the isoquant, these two n values generate the same area, $\bar{L}$, since $L = \frac{\bar{Q}}{y(n)}$ so $\frac{\bar{Q}}{y(n_{\bar{y}}^-)} = \frac{\bar{Q}}{y(n_{\bar{y}}^+)} = \bar{L}$.

Now compare the costs of these two n values.

$$\begin{aligned} C(n_{\bar{y}}^-, \bar{L}) - C(n_{\bar{y}}^+, \bar{L}) &= \left(C_f + \frac{C_n}{n_{\bar{y}}^-}\right)\bar{L} + C_D\, y(n_{\bar{y}}^-)\bar{L}^{1.5} - \left(C_f + \frac{C_n}{n_{\bar{y}}^+}\right)\bar{L} - C_D\, y(n_{\bar{y}}^+)\bar{L}^{1.5} \\ &= \left(C_f + \frac{C_n}{n_{\bar{y}}^-}\right)\bar{L} + C_D\, \bar{y}\bar{L}^{1.5} - \left(C_f + \frac{C_n}{n_{\bar{y}}^+}\right)\bar{L} - C_D\, \bar{y}\bar{L}^{1.5} \\ &= C_n\bar{L}\left(\frac{1}{n_{\bar{y}}^-} - \frac{1}{n_{\bar{y}}^+}\right) (> 0) \end{aligned}$$

Hence for any level of yield, the cost minimizing maximum age is greater than or equal to the maximum sustainable yield age, i.e. $n^* \geq n_{MSY}$. ■

Lemma 2 *The minimum of the isoquant is located at* n_{MSY}.

Proof of Lemma 2.

The isoquant is defined by $y(n)L = \bar{Q}$. This can be rewritten so that L is a function of n, i.e. for a particular level of feedstock production $L = \frac{\bar{Q}}{y(n)}$. The minimum of this function (i.e. the least quantity of land necessary to produce the desired quantity)

occurs when the derivative of this function is set to zero.

$$\left.\frac{dL}{dn}\right|_{\text{isoquant}} = \frac{-\bar{Q}y'(n)}{[y(n)]^2} = 0 \Leftrightarrow y'(n) = 0$$

From the conditions imposed on the age-yield function in section 2.1.2 there is a unique maximum of the yield function located at n_{MSY}. Hence the unique minimum of the isoquant function occurs at n_{MSY}. ■

Lemma 3 *The minimum of the isocost curve is located at $n < n_{MSY}$.*

Proof of Lemma 3.

The isocost curve is defined by a level set of the cost function: $C(n, L) = \bar{C}$. We wish to locate the set of local extrema of the isocost curve, where L is expressed as a function of n. This set is a subset of the critical points of $\frac{dL}{dn}$.

Totally differentiate the cost function:

$$\left[\left(C_f + \frac{C_n}{n}\right) + 1.5\, C_D\, y(n)\, L^{0.5}\right] dL + \left[\frac{-C_n\, L}{n^2} + C_D\, y'(n)\, L^{1.5}\right] dn = 0$$

Thus

$$\left.\frac{dL}{dn}\right|_{\text{isocost}} = \frac{C_n\, L/n^2 - C_D\, y'(n)\, L^{1.5}}{\left(C_f + \frac{C_n}{n}\right) + 1.5\, C_D\, y(n)\, L^{0.5}} = 0 \Leftrightarrow C_n\, L/n^2 = C_D\, y'(n)\, L^{1.5}$$

since all the terms in the denominator are non-negative. The only term in this last equality that can change sign is $y'(n)$. All other terms are constrained to be non-negative. Hence the equality cannot be satisfied if $y'(n) < 0$, which occurs when $n > n_{MSY}$. Also, if $n = n_{MSY}$ it must be that $L = 0$ for the equality to be satisfied. If $L = 0$ we have $C(n_{MSY}, 0) = 0$, so for any positive level of cost $(n_{MSY}, 0)$ is not an element of the graph of the isocost function, and n_{MSY} cannot be a critical point.

Hence for any positive level of cost, any extrema of the isocost function must occur when $n < n_{MSY}$. ■

Lemma 4 *The isocost curve has a positive slope for all* $n \geq n_{MSY}$.

Proof of Lemma 4.

This follows immediately from the proof of lemma 3 since the expression for the slope of the isoquant curve is strictly positive for all $n > n_{MSY}$. ■

Proof of proposition 1.

The optimal n *must be strictly greater than* n_{MSY}*, i.e.* $n^* > n_{MSY}$.

The isocost curve has a positive slope for all $n \geq n_{MSY}$ (lemma 4). The isoquant curve has a zero slope at n_{MSY} (lemma 2). Hence the isocost and isoquant curves cannot be tangential at n_{MSY}, so $n^* \neq n_{MSY}$. Combining this with lemma 1 gives us the result.

■

Lemma 5 *Assumptions (1)-(4) imply that* $0 < \lim_{n\to\infty} \int_0^n f(a)\, da < \infty$

Proof of Lemma 5.

We can split $\lim_{n\to\infty} \int_0^n f(a)\, da$ in two by partitioning its domain:

$$\begin{aligned}\lim_{n\to\infty} \int_0^n f(a)\, da &= \lim_{n\to\infty} \int_0^k f(a)\, da + \lim_{n\to\infty} \int_k^n f(a)\, da \\ &= \int_0^k f(a)\, da + \lim_{n\to\infty} \int_k^n f(a)\, da\end{aligned}$$

Now consider $\int_0^k f(a)\, da$. The age-yield function is bounded below by 0 by construction ($f(a)$ represents a physical quantity). Assumptions (1)-(3) imply that $f(a)$ is bounded above. Hence $f(a)$ is bounded on the domain $[0, k]$ for all $k \in \mathbb{R}_{>0}$. Thus $0 \leq \int_0^k f(a)\, da < \infty$ since this is the integral of a bounded positive function on a finite domain.

We must consider two possibilities when analyzing $\lim_{n\to\infty} \int_k^n f(a)\ da$: either $f(a) > 0$ for all $a \in \mathbb{R}_{\geq 0}$, or there exists some $\hat{k} \in \mathbb{R}_{\geq 0}$ such that for all $a > \hat{k}$ $f(a) = 0$. In the first case, we must establish that $\lim_{n\to\infty} f(a)$ approaches zero fast enough that $\lim_{n\to\infty} \int_k^n f(a)\ da$ is not infinite. Assumption (4) implies that there exist $k \in \mathbb{R}_{\geq 0}$ and $p > 1$ such that for all $a > k$ $f(a) < \frac{1}{a^p}$ (if such k and p did not exist, $\lim_{n\to\infty} a\, f(a)$ would either be strictly positive, or infinite). Thus

$$\lim_{n\to\infty} \int_k^n f(a)\ da < \lim_{n\to\infty} \int_k^n \frac{1}{a^p}\ da < \infty$$

since integrals of the form $\int_k^\infty \frac{1}{x^p}\ dx$ are convergent if and only if $p > 1$. In the second case, $\lim_{n\to\infty} \int_{\hat{k}}^n f(a)\ da = 0$, and $\lim_{n\to\infty} \int_0^n f(a)\ da = \int_0^k f(a)\ da$. Thus $0 \leq \lim_{n\to\infty} \int_k^n f(a)\ da < \infty$

Assumption 3 implies that $f(a)$ is strictly positive on some subset of $\mathbb{R}_{\geq 0}$ with non-empty interior. Hence $\lim_{n\to\infty} \int_0^n f(a)\ da > 0$

Therefore $0 < \lim_{n\to\infty} \int_0^n f(a)\ da < \infty$ ■

Proof of proposition 2.

Given assumptions (1)-(4), a solution, n^*, to the cost minimization problem exists such that $n^* \in (n_{MSY}, \infty)$.

Sketch of the proof: We have already demonstrated that n^* must be greater than n_{MSY}. At n_{MSY} the slope of the isocost curve is strictly positive and the slope of the isoquant curve is zero. We show that as n approaches infinity, the slope of the isocost curve approaches zero, while the slope of the isoquant curve approaches a positive value. By continuity the slope functions must cross at least once, and hence there must exist at least one point where the isocost and isoquant curves are tangent to each other.

We begin by showing that the slope of the isocost curve approaches zero as n approaches infinity. The slope of the isocost function when L is written as a function of n (as derived in lemma 2)

$$\left.\frac{dL}{dn}\right|_{\text{isocost}} = \frac{C_n\, L(n)/n^2 - C_D\, y'(n)\, L(n)^{1.5}}{\left(C_f + \frac{C_n}{n}\right) + 1.5\, C_D\, y(n)\, L(n)^{0.5}}$$

To take the limit of this expression as n approaches infinity, we need to know how $L(n)$ on the isocost function behaves as n approaches infinity. The isocost function is defined as

$$C(n, L) = (C_f + \frac{C_n}{n})L + C_D\, y(n)\, L^{1.5} = \bar{C}$$

This implicitly defines L as a function of n.

$$C(n) = (C_f + \frac{C_n}{n})L(n) + C_D\, y(n)\, L(n)^{1.5} = \bar{C}$$

Now we take the limit of this expression as $n \to \infty$ and solve for the unknown value L_∞.

$$\begin{aligned}
&\lim_{n\to\infty} (C_f + \frac{C_n}{n})L(n) + C_D\, y(n)\, L(n)^{1.5} = \bar{C} \\
&\quad \Rightarrow C_f\, L_\infty = \bar{C} \\
&\quad \Rightarrow L_\infty = \frac{\bar{C}}{C_f} \qquad \text{A constant}
\end{aligned}$$

Returning to the derivative of the isocost function

$$\lim_{n\to\infty} \left.\frac{dL}{dn}\right|_{\text{isocost}} = \lim_{n\to\infty} \frac{C_n\, L(n)/n^2 - C_D\, y'(n)\, L^{1.5}}{\left(C_f + \frac{C_n}{n}\right) + 1.5\, C_D\, y(n)\, L(n)^{0.5}}$$

$$\begin{aligned} &= \frac{0-0}{C_f + 0 + 0} && \text{Since } y(n) \text{ and } y'(n) \text{ both approach 0,} \\ &= 0 && \text{and } L(n) \text{ approaches a constant as } n \to \infty \end{aligned}$$

Now we show that under a certain condition the slope of the isoquant function approaches a positive constant as $n \to \infty$. The isoquant function is given by $y(n)\, L = \bar{Q}$ and can be rewritten as

$$\begin{aligned} L &= \frac{\bar{Q}}{\frac{1}{n}\int_0^n f(a)\, da} \\ &= \frac{\bar{Q}\, n}{\int_0^n f(a)\, da} \end{aligned}$$

The slope of the isoquant function is given by

$$\left.\frac{dL}{dn}\right|_{\text{isoquant}} = \frac{\bar{Q}\, \left(\int_0^n f(a)\, da - n\, f(n)\right)}{\left[\int_0^n f(a)\, da\right]^2}$$

The limit of the slope as n approaches infinity is

$$\lim_{n\to\infty} \left.\frac{dL}{dn}\right|_{\text{isoquant}} = \lim_{n\to\infty} \frac{\bar{Q}\,\left(\int_0^n f(a)\,da - n\,f(n)\right)}{\left[\int_0^n f(a)\,da\right]^2}$$

$$= \bar{Q}\,\frac{\lim_{n\to\infty}\left(\int_0^n f(a)\,da - n\,f(n)\right)}{\lim_{n\to\infty}\left[\int_0^n f(a)\,da\right]^2} \qquad \text{since } \lim_{n\to\infty}\int_0^n f(a)\,da > 0 \text{ (Lemma 5)}$$

$$= \bar{Q}\,\frac{\overbrace{\lim_{n\to\infty}\int_0^n f(a)\,da - \lim_{n\to\infty} n\,f(n)}^{+}}{\underbrace{\lim_{n\to\infty}\left[\int_0^n f(a)\,da\right]^2}_{+}}$$

since $\lim_{n\to\infty}\int_0^n f(a)\,da > n\,f(n) = 0$ and $0 < \lim_{n\to\infty}\int_0^n f(a)\,da < \infty$ (Lemma 5)

$$\implies 0 < \lim_{n\to\infty} \left.\frac{dL}{dn}\right|_{\text{isoquant}} < \infty$$

Now define a function that returns the difference in the slopes of the isocost and isoquant functions, $h(n) = \left.\frac{dL}{dn}\right|_{\text{isocost}} - \left.\frac{dL}{dn}\right|_{\text{isoquant}}$. Since both constituent functions are continuous on the the interval (n_{MSY}, ∞), $h(n)$ is also continuous on this interval. At the maximum yield age $h(n_{MSY}) > 0$ (from lemmas 2 and 4) and, as we have just shown, when n approaches infinity the limit of $h(n)$ is strictly less than zero. Hence by the intermediate value theorem, there must exist some $n^* \in (n_{MSY}, \infty)$ such that $h(n) = 0$, and the isocost and isoquant curves are tangent to one another. ■

A.1 Proofs of comparative static results

Proof of $\frac{dn^*}{d\bar{Q}} < 0$.

As processing facility size increases, the optimal age decreases, i.e. $\frac{dn^*}{d\bar{Q}} < 0$.

Totally differentiating $g(n, \bar{Q})$ (The derivative of the cost function when the constraint is used to eliminate L — derived in the proof of proposition 2) gives us an expression for the desired comparative static

$$\frac{dn^*}{d\bar{Q}} = \frac{-g_{\bar{Q}}}{g_n}$$

At an optimum the second order condition for a minimum must hold, so g_n must be positive. Hence

$$\text{sign}\left(\frac{dn^*}{d\bar{Q}}\right) = -\,\text{sign}(g_{\bar{Q}})$$

Differentiating $g(n, \bar{Q})$ with respect to $\bar{Q}$, and evaluating at the optimum yields

$$g_{\bar{Q}} = -0.25 \underbrace{[y(n^*)]^{-1.5}}_{+} \underbrace{y'(n^*)}_{-} \underbrace{\bar{Q}^{-0.5}}_{+} \qquad (> 0)$$

Hence

$$\frac{dn^*}{d\bar{Q}} < 0$$

■

Proof of $y''(n*) < 0 \;\Rightarrow\; \frac{dL^*}{d\bar{Q}} > 0$.

The change in optimal growing region size with respect to a change in processing facility capacity is generally ambiguous, but if $y''(n^*) < 0$, then increased processing facility capacity leads to increase growing region size, i.e. $\frac{dL^*}{d\bar{Q}} > 0$.

To analyze this comparative static of the constrained cost minimization problem using the substitution method we need to define the inverse yield function, $g(y) = n$ ($g^{-1}(n) = y(n)$). Since the yield function is not surjective, we can only define and analyze the inverse yield on a subset of the domain. Fortunately, as shown by proposition 1, the optimal n is found in the subset $n > n_{MSY}$. On this subset the yield function is bijective, and we are guaranteed the existence of $g(y)$.

Using the constraint on processing facility capacity ($y(n)L = \bar{Q} \Rightarrow y(n) = \frac{\bar{Q}}{L}$ and $n = g(\frac{\bar{Q}}{L})$) we can rewrite the cost function as a function of growing region only.

$$C(n(L), L) = \left(C_f + \frac{C_n}{g\left(\frac{\bar{Q}}{L}\right)} \right) L + C_D \, \bar{Q} L^{0.5}$$

The first order condition with respect to a minimum is

$$\frac{dC}{dL} = C_f + \frac{C_n}{g\left(\frac{\bar{Q}}{L}\right)} + \frac{\bar{Q} \, C_n \, g'\left(\frac{\bar{Q}}{L}\right)}{L \left[g\left(\frac{\bar{Q}}{L}\right) \right]^2} + (0.5) C_D \, \bar{Q} \, L^{-0.5} = 0$$

Cross multiply by $L \left[g\left(\frac{\bar{Q}}{L}\right) \right]^2$

$$h(L) = C_f \, L \left[g\left(\frac{\bar{Q}}{L}\right) \right]^2 + C_n \, L \, g\left(\frac{\bar{Q}}{L}\right) + C_n \, \bar{Q} \, g'\left(\frac{\bar{Q}}{L}\right) + (0.5) C_D \, \bar{Q} \, \left[g\left(\frac{\bar{Q}}{L}\right) \right]^2 L^{0.5} = 0$$

Totally differentiating $h(n, \bar{Q})$ gives us an expression for the desired comparative static

$$\frac{dL^*}{d\bar{Q}} = \frac{-h_{\bar{Q}}}{h_L}$$

At an optimum the second order condition for a minimum must hold, so g_n must be

positive. Hence

$$\text{sign}\left(\frac{dL^*}{d\bar{Q}}\right) = -\,\text{sign}(h_{\bar{Q}})$$

$$h_{\bar{Q}} = (0.5)C_D\left[g\left(\frac{\bar{Q}}{L}\right)\right]^2 L^{0.5} \qquad (>) \qquad (16)$$

$$+\,2(0.5)C_D\,\bar{Q}\,g\left(\frac{\bar{Q}}{L}\right)g'\left(\frac{\bar{Q}}{L}\right)L^{-0.5} \qquad (<) \qquad (17)$$

$$+\,2C_f\,g\left(\frac{\bar{Q}}{L}\right)g'\left(\frac{\bar{Q}}{L}\right) \qquad (<) \qquad (18)$$

$$+\,2C_n g'\left(\frac{\bar{Q}}{L}\right) \qquad (<) \qquad (19)$$

$$+\,\frac{C_n\,\bar{Q}g''\left(\frac{\bar{Q}}{L}\right)}{L} \qquad \text{(Ambiguous)} \qquad (20)$$

If $g''\left(\frac{\bar{Q}}{L}\right) < 0$ at L^*, the term 20 in $h_{\bar{Q}}$ is negative.

Aside: Rewriting this condition in terms of $y(n^*)$

This condition on the second derivative of the inverse yield function is not particularly intuitive. We can rewrite this condition in terms of $y(n^*)$ which makes it much easier to understand. To do this we must rewrite this condition on the second derivative of an inverse function in terms of the original function. The relationship between the second derivative of a function and its inverse is

$$(f^{-1})''(f(x)) = \frac{-f''(x)}{[f'(x)]^3}$$

For the inverse yield function this becomes

$$g''\left(\frac{\bar{Q}}{L}\right) = g''(y(n^*)) = \frac{-y''(n^*)}{[y'(n^*)]^3}$$

So

$$g''\left(\frac{\bar{Q}}{L}\right) < 0 \;\Leftrightarrow\; \frac{-y''(n^*)}{[y'(n^*)]^3} < 0$$

Since $y'(n^*) < 0$ and the cubing operation preserves sign, this inequality is satisfied if and only if $y''(n^*) < 0$.

Returning to the proof

Given that $y''(n^*) < 0$, we now show that term (16) plus term (17) is negative.

$$(16) + (17) = (0.5)C_D \left[g\left(\frac{\bar{Q}}{L}\right)\right]^2 L^{0.5} + 2(0.5)C_D\,\bar{Q}\,g\left(\frac{\bar{Q}}{L}\right)g'\left(\frac{\bar{Q}}{L}\right)L^{-0.5}$$

Extract common factors

$$(16) + (17) = \underbrace{(0.5)C_D\,g\left(\frac{\bar{Q}}{L}\right)L^{0.5}}_{>0}\left[g\left(\frac{\bar{Q}}{L}\right) + 2\bar{Q}\,g'\left(\frac{\bar{Q}}{L}\right)L^{-1}\right]$$

Therefore

$$\text{sign}\left((16) + (17)\right) = \text{sign}\left(g\left(\frac{\bar{Q}}{L}\right) + 2\bar{Q}\,g'\left(\frac{\bar{Q}}{L}\right)L^{-1}\right)$$

Substitute the definition of $\bar{Q} = y(n)\,L$

$$\begin{aligned}
&g\left(\frac{y(n)\,L}{L}\right) + 2y(n)\,L\,g'\left(\frac{y(n)\,L}{L}\right)L^{-1} \\
=&g(y(n)) + 2y(n)\,g'(y(n)) \\
=&n + \frac{2y(n)}{y'(n)} \qquad\qquad \text{since } g(.) \text{ is inverse of } y(.)
\end{aligned}$$

Recall $y'(n) = \frac{f(n)-y(n)}{n}$, so

$$\begin{aligned} n + \frac{2y(n)}{y'(n)} =& n + \frac{2n\,y(n)}{f(n)-y(n)} \\ =& n\left(1 + \frac{2y(n)}{f(n)-y(n)}\right) \\ =& n\left(1 - \frac{2y(n)}{y(n)-f(n)}\right) \end{aligned}$$

For $n > n_{MSY}$, $y(n) > f(n) \geq 0$, hence $\frac{2y(n)}{y(n)-f(n)} > 1$, so $(16) - (17) < 0$, $h_{\bar{Q}} < 0$, and $\frac{dL^*}{d\bar{Q}} > 0$. ■

Proof of remaining comparative statics.

See table on page 20

As explained in the proofs the previous two comparative statics, the sign of the comparative static of n^* and L^* with respect to any exogenous variable x can be found by analyzing the sign of the relevant derivative of the first order condition, i.e.

$$\text{sign}\left(\frac{dn^*}{dx}\right) = -\,\text{sign}(g_x) \quad \text{and} \quad \text{sign}\left(\frac{dL^*}{dx}\right) = -\,\text{sign}(h_x)$$

We now present and sign the expressions of g_x for the parameters of interest.

$$g_{C_f} = -\frac{\bar{Q}\,\overbrace{y'(n^*)}^{-}}{y(n^*)^2} \qquad (>0)$$

$$g_{C_n^*} = \underbrace{\frac{-\bar{Q}}{n^*\,y(n^*)}}_{-}\underbrace{\left[\frac{1}{n^*} + \frac{y'(n^*)}{y(n^*)}\right]}_{+} \quad (<0) \qquad \text{Since } \varepsilon_{y(n^*)>-1} \Rightarrow \tfrac{1}{n^*} + \tfrac{y'(n^*)}{y(n^*)} > 0 \quad \text{(Prop 2)}$$

$$g_{C_D} = \underbrace{(-0.5)}_{-} \bar{Q}^{1.5}\, y(n^*)^{-1.5}\, \underbrace{y'(n^*)}_{-} \qquad (>0)$$

We now present and sign the expressions of h_x for the parameters of interest.

$$h_{C_f} = L(n^*)^2 \qquad (>0)$$

$$h_{C_n} = Ln^* + \frac{\bar{Q}}{y'(n^*)} \quad (<0) \qquad \text{Since } \varepsilon_{y(n^*)>-1} \Rightarrow Ln^* + \frac{\bar{Q}}{y'(n^*)} < 0 \quad \text{(Prop 2)}$$

$$h_{C_D} = (0.5)\,(n^*)^2\,\bar{Q}\,L^{0.5} \qquad (>0)$$

■

B Calibration of the Cost-Minimization Problem

A piecewise linear spline function is used to estimate the age-yield function. With this functional form, the cost minimization problem has 9 parameters for which values must be found.

B.1 Brazilian Sugarcane

B.1.1 Piece-wise Linear Age-Yield Function: t_1, t_{max}, t_T, f_{max}

The piece-wise linear age-yield function has four parameters. Parameter t_1 designates the age at which the yield first becomes positive, t_{max} is the age at which maximum yield is achieved, and t_T is the age at which yield returns to zero. During the increasing phase, between t_1 and t_{max} the function is a positive affine and during the decreasing

phase, between t_{max} and t_T, the function is negative affine.

To estimate the parameters, we fit the linear-piecewise function to age-yield data obtained from Margarido and Santos (2012). Figure 6 shows the original data and the fitted age-yield function.

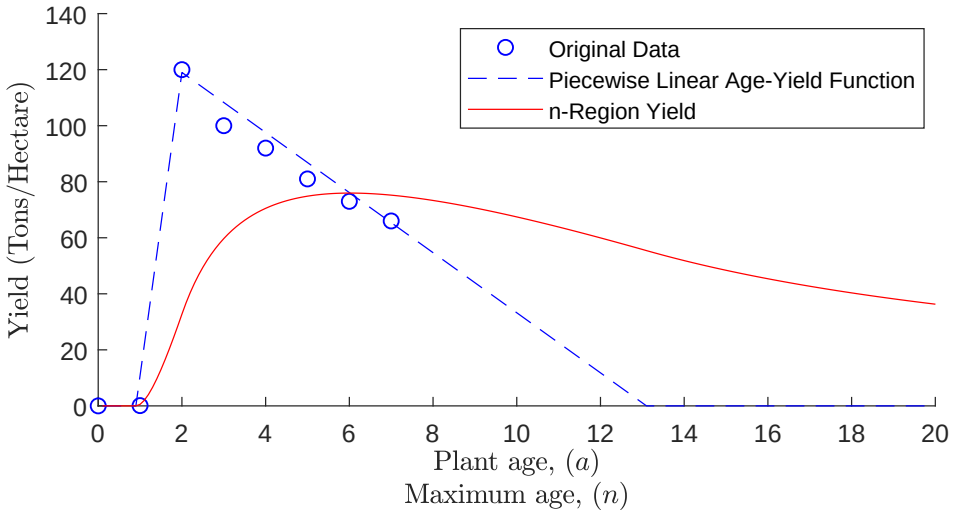

Figure 6: Fitting the piecewise-linear age-yield function to the Brazilian age-yield data from Margarido and Santos (2012).

The parameter values obtained are $t_1 = 1$, $t_{max} = 2$, $t_T = 13$, and $f_{max} = 120$.

B.1.2 Farm-gate Cost parameters: C_f, C_n

We derived the feedstock cost parameters, C_f and C_n, from Teixeira (2013), and the delivery cost parameter, C_D, from Crago et al. (2010).

Teixeira (2013) presents an example operating budget for a 5-cut (6-age-class) sugarcane operation in São Paulo state, where they assume that 80 percent of the cane is harvested burned, and 20 percent is harvested raw. Costs are divided into five categories, delivery costs, and four that account for farm gate feedstock costs: preparing the

soil, planting, harvest, and maintenance of the ratoon. The total farm gate feedstock costs for a 6 hectare operation is given by

$$\begin{aligned}\text{Total Farm Gate Feedstock Costs} =& \text{Soil Preparation} + \text{Planting} \\ &+ 5 \times \text{Harvest} + 4 \times \text{Ratoon maintenance}\end{aligned}$$

Since the total cost is given for 6 hectares, the total cost per hectare is

$$\begin{aligned}\text{Total Farm Gate Feedstock Costs (Per Hectare)} =& \frac{1}{6} \times \text{Soil Preparation} + \frac{1}{6} \times \text{Planting} \\ &+ \frac{5}{6} \times \text{Harvest} + \frac{4}{6} \times \text{Ratoon Maintenance}\end{aligned}$$

Assuming that these cost parameters are constant with respect to the number of age-classes we can write the total farm gate feedstock per hectare as a function of the age structure

$$\begin{aligned}\text{Farm gate feedstock costs}(n) =& \frac{1}{n} \times \text{Soil Preparation} + \frac{1}{n} \times \text{Planting} \\ &+ \frac{n-1}{n} \times \text{Harvest} + \frac{n-2}{n} \times \text{Ratoon maintenance}\end{aligned}$$

Substituting Teixeria's numbers (in Reals) from the example budget, the cost function becomes

$$\text{Farm gate feedstock costs}(n) = \frac{656.07}{n} + \frac{4159.83}{n} + \frac{n-1}{n} \times 1273.13 + \frac{n-2}{n} \times 986.54$$

Which on rearranging becomes

$$\text{Farm gate feedstock costs}(n) = 2259.67 + \frac{1569.69}{n}$$

Hence for the simulations we use a baseline of $C_f = 2259.67$ and $C_n = 1569.69$.

B.1.3 Delivery Cost parameter: C_D

While Teixeira (2013) does include estimates of delivery costs, he does not include the processing facility size that this example farm is feeding. We therefore turn to Crago et al. (2010) to derive the delivery cost parameter.

The total delivery cost from a growing region is given by

$$\begin{aligned}\overset{\text{Total}}{\underset{\text{Delivery Costs}}{}} =& \text{Average Cost Per Ton Kilometer} \\ &\times \text{Quantity Transported} \\ &\times \text{Average Delivery Distance}\end{aligned}$$

Let δ represent the average delivery cost per ton kilometer (i.e. the average cost to transport one ton of feedstock one kilometer). Crago et al. (2010) report an average transport cost of R\$6.7 to transport a ton of feedstock from the farm gate to the mill. The average delivery distance in this study was 22 kilometers so in this case $\delta = 0.3045$.

The average mill size in Crago et al. (2010) is 4.8 million tons. Given our assumption that the growing region produces the exact quantity required to feed the mill, this implied that the average quantity of feedstock transported was 4.8 million tons.

When calculating the average delivery distance, we must make a distinction between the area of land planted with sugarcane, L, and the area of the growing region, A. Although we are assuming that the growing region is circular, it is not necessarily the case that all the land is planted with sugarcane. In fact, relaxing the link between planted area and growing region area is necessary to correctly calibrate the model to the data in Crago et al. (2010).

Let d be the average density of sugarcane fields in the growing region, and A be

the area of the growing region. Hence

$$\mathrm{L} = d \times A$$

The average delivery distance is given by the expression

$$r_{av} = \frac{2}{3} r_{max} = \frac{2}{3}\sqrt{\frac{A}{\pi}}$$

Since the average delivery distance, r_{av}, from Crago et al. (2010) is 22km, the size of the growing region is $A = 342\,119$ ha.

We calculate the density parameter from

$$\text{Total Quantity} = \text{Yield} \times \text{Density} \times \text{Growing Region Area}$$

Crago et al. (2010) reports an average yield of 75 tons per hectare. So we calculate the density as

$$4800000 = 75 \times d \times 342119 \Rightarrow d = 0.187$$

Hence the expression for the total delivery cost becomes

$$\begin{aligned}
\underset{\text{Delivery Costs}}{\text{Total}} &= \delta \times Q \times r_{av} \\
&= \delta \times Q \times \frac{2}{3}\sqrt{\frac{A}{\pi}} \\
&= \delta \times Q \times \frac{2}{3}\sqrt{\frac{L}{d \times \pi}} \\
&= \frac{2\delta}{3}\sqrt{\frac{1}{d \times \pi}} \times Q \times \sqrt{L} \\
&= \frac{2\delta}{3}\sqrt{\frac{1}{d \times \pi}} \times y(n) L\sqrt{L} \\
&= \frac{2\delta}{3}\sqrt{\frac{1}{d \times \pi}} \times y(n) L^{1.5} \\
&= C_D \times y(n) L^{1.5}
\end{aligned}$$

For the d and δ derived from Crago et al. (2010), $C_D = 0.2649$.

B.2 Calibrated parameters and ranges used in simulations

	Parameter	Min Value	Calibration	Max Value
Yield	t_1	0	1	2
	t_{max}	$t_1 + 1$	2	$t_1 + 5$
	t_T	$t_{max} + 7$	13	$t_{max} + 13$
	f_{max}	60	120	180
Cost	C_f	1129.84	2259.67	3389.51
	C_n	784.85	1569.69	2354.54
	C_D	0.13	0.26	0.40
Capacity	$\bar{Q}$	1 000 000	19 000 000	36 000 000

Table 3: Support for random parameters used in cost minimization. The parameters are drawn from a uniform distribution centered on the Brazilian calibration