Why Did Sugarcane Growers Suddenly Adopt Existing Technology?

C. Austin Davis*

This version: November 1, 2016
Most recent: `http://goo.gl/ekMvbd`

I investigate the role of regulation and factor prices in the rapid, widespread adoption of mechanical harvesting technology by Brazilian sugarcane growers. I use worker- and establishment-level data to test the effect of regulation using complementary regression discontinuity and difference-in-differences approaches. I find that regulation is, at best, a partial explanation, accounting for no more than one quarter of the dramatic change in harvesting practices. I develop a tipping-point model to show how rising wages may have played an important role even though the change in wages was gradual and the change in harvesting was abrupt; instrumental variables estimates imply that increasing wages alone are sufficient to explain the adoption of green technology.

*Ph.D. Candidate, Department of Economics, University of Michigan. I am grateful for the guidance of my committee: Achyuta Advharyu, Martha Bailey, David Lam, and Paul Rhode. Thanks to Gaurav Khanna, Margaret Lay, and Katherine Lim for their advice and encouragement. I received invaluable feedback from Jacob Bastian, Hoyt Bleakley, Eric Chyn, Morgan Henderson, Prachi Jain, Max Kapustin, Sean McRae, Johannes Norling, Bryan Stuart, seminar participants in the Energy and Environmental Economics Day at the University of Michigan, and seminar participants in the Economic Development Seminar at the University of Michigan. Débora Nogueira and Paula Moura provided indispensable assistance in Brazil. I owe my understanding of sugarcane cultivation and harvesting to Caleb Varner, Marco Ripoli, and Jesse Lopez at John Deere, to Mariana Marquoiri and Jorge Luis Mangolini Neves at the Centro de Technologia Canavieira, and Jose Rodolofo Penatti of AFOCAPI. All errors and omissions are my own.

1 Introduction

In the developing world, a large fraction of the population engages in low-productivity agriculture (Gollin, 2010). Widely available technologies like fertilizer have the potential to improve productivity but are not widely used (Morris et al., 2007). Following this pattern, Brazilian sugarcane growers relied on manual harvesting as late as 2007 even though harvesting machines had been available for many years. But, by 2013, almost all sugarcane was harvested mechanically.[1] This paper studies how this dramatic transformation was achieved.

Recent literature has emphasized several different factors affecting technology adoption in agriculture, from social learning to behavioral biases (Duflo et al., 2011; Conley & Udry, 2010). From interviews with various stakeholders, I identify two other factors of primary importance in Brazilian sugarcane. The first, government regulation, is understudied in this literature and the second, factor prices, is foundational. Manual harvesting was gradually banned by state governments because of pollution associated with the practice. This regulation coincides with the mechanization of harvesting, lending credence to the government's claim that the regulation caused growers to adopt mechanical harvesting. However, I find that the regulation accounts for little of the change in harvesting practices. I find that a strong labor market, where rising real wages made manual harvesting more expensive, is sufficient to explain mechanization.

Both publicly and in interviews, state government advertises regulation as the causal factor driving the adoption of machine harvesting in sugarcane. Between 2002 and 2014, all of the sugarcane-growing states responded to constituent concerns about air pollution by passing gradual bans of the straw burning associated with manual harvesting. These bans, along with enforcement efforts, were covered in national newspapers and they coincided with

[1] Adoption was relatively quick in this case, especially considering that the technology was not new. It took about 22 years for American farmers to completely adopt diesel tractors (White). Mansfield (1961) studies the diffusion of 12 innovations of "outstanding importance" among major firms in several American industries; the average time between initial use and complete adoption is over 18 years. As another point of comparison, den Bulte (2000) finds that widespread consumer adoption of durable goods typically takes 7 to 14 years, depending on characteristics of the good and on the economic and demographic environment at the time of introduction.

the period of rapid mechanization.[2] In various communications, environmental regulators in the largest sugarcane growing state, São Paulo, took credit for mechanization.[3] São Paulo's environmental ministry has a detailed website attesting to the success of the regulation and Brazil's space agency publishes satellite-based monitoring data online.[4]

I test the effects of regulation on harvesting practices using two highly detailed, confidential data sources that provide near universal coverage of the industry. The first data source, known by its Portuguese-language acronym RAIS, captures detailed employment information for all formal-sector workers in Brazil from 1998 to 2014.[5] The second data source is the 2006 Census of Agriculture. I use these data sources to conduct complementary tests of the regulation.

Regression discontinuity estimates from the Census of Agriculture show a small effect of regulation on harvesting practices. I take advantage of an area threshold that exempted small growers to provide complementary evaluations of the regulation, comparing harvesting techniques and input use between unregulated establishments just below the area threshold to regulated establishments just above.[6] If the true effect lies at the extreme of the confidence interval, I find that regulation explains no more than a quarter of the change in harvesting practices. Moreover, regulated farmers show no changes in input use that would be consistent with mechanization.

The regression discontinuity estimates measure the behavior of establishments near the area threshold; I find a similarly small effect of regulation using a difference-in-differences approach which captures the behavior of larger growers. Specifically, I use data from RAIS to

[2] For reporting on passage of the law, see, e.g., Spinelli (2001); Osse (2002); Gazeta Mercantil (2002). For coverage of enforcement efforts, see, e.g. Spinelli (2001); Samora (2006); Credendio (2008); Folha de S. Paulo (2008); Henrique (2008); Coissi (2008); Tomazela (2016); O Globo G1 (2016).

[3] For example, I recently received an email from environmental regulators claiming "As a result of this policy ... 83% of the 2013/2014 crop was harvested without burning." Officials made similar claims during in-person interviews.

[4] See `http://www.ambiente.sp.gov.br/etanolverde/` and `http://www.dsr.inpe.br/laf/canasat/`.

[5] RAIS stands for Relação Anual de Informações Sociais, which roughly translates to Annual Report of Social Information. For historical reasons, most sugarcane workers are formal.

[6] As of 2006, roughly 80 percent of establishments are not regulated as they fall below the 150 hectare threshold. However, regulated establishments control about 80 percent of sugarcane area.

estimate how changes in the stringency of the regulation across states and over time affected changes in county-level labor intensity. Here again, if the true effect lies at the extreme of the confidence interval, I find that regulation can account for at most one quarter of the observed decline in labor intensities.

Besides regulation, what caused the rapid transition of harvesting techniques? My detailed interviews with industry participants suggest that rising wages were an important motivation for mechanized harvesting. Using administrative and survey data, I show that, from 1998 to 2014, increasing labor demand from large sectors like construction helped drive real wages up by almost 50 percent for harvest workers. However, this timing does not obviously support wages as a driver of mechanization. Wages rose continuously from 1998 while widespread mechanization began only in 2007.

I develop a tipping-point model of grower behavior that reconciles these dynamics. For each parcel of land, there is a threshold wage, above which a profit-maximizing grower will harvest the parcel mechanically. This threshold can be different for each parcel, depending on characteristics of the land. If wages are well below the threshold for a majority of parcels, wages may increase steadily without affecting harvesting techniques. Eventually, as wages rise, they will cross the switching thresholds of many parcels, causing widespread mechanization.

To test wages as an explanation for mechanization, I estimate the wage elasticity of labor demand in sugarcane using a set of instruments similar to the instrument developed by Dube & Vargas (2013). These instruments capture shifts in county-level agricultural labor supply. Specifically, for each of the four other crops grown in the sugarcane region, I interact the historical county-level acreage of that crop with a measure of the crop's international price. As a first stage, these instruments predict county-level sugarcane wages. I estimate the wage elasticity of labor demand by regressing the county-level quantity of labor on the predicted wages from the first stage. The estimated elasticity is large, suggesting that the observed increase in wages is sufficient to explain the mechanization of sugarcane harvesting.

As countries become richer, fewer people work in an increasingly productive agricultural sector. This relationship holds across countries and within a country over time; its ubiquity makes it one of the fundamental facts of development (Gollin, 2010). But the nature of this relationship remains uncertain. Do improvements to agricultural productivity stimulate growth in other sectors? Or vice versa? What is the role of policy? Do other factors drive change?

Brazilian sugarcane is fascinating because we observe this development process: a large agricultural sector transitioning from a labor-intensive, low-technology production to capital-intensive, high-technology production. I find that government regulation had an extremely limited role, at least in this context. I provide evidence that growth in other sectors, operating through increased labor demand, motivated sugarcane growers to adopt new technology. These results suggest that, while agriculture is a large part of developing economies, creating labor-market opportunities in other sectors may be an effective way to improve productivity and induce technology adoption in agriculture.

This paper owes much to prior work but offers novel insights by disentangling government and market forces. The research question is similar to various studies of mechanization among farmers in the antebellum United States (David, 1966; Olmstead & Rhode, 1993, 1995). These papers also discuss the importance of factor prices but, beyond differences in time, location, crop and technology, mechanization was not regulated in the historical US. There is also a deep literature on technological adoption in developing country agriculture (Foster & Rosenzweig, 1995; Conley & Udry, 2010; Duflo et al., 2011). However, learning and network effects are primary in these papers and, thanks to the sophistication of Brazilian sugarcane growers and the provision of extension services, these issues are less important here. Finally, prior work has shown that it is difficult to end polluting practices in the developing world; both government regulation and NGO-backed technological fixes have failed (Davis, 2008; Greenstone & Hanna, 2014). At least in this context, reductions to pollution were achieved through development itself, as embodied by higher wages for some of Brazil's poorest workers.

2 Evaluating the Regulation

2.1 Where is Sugarcane Grown?

Although sugarcane is grown in two separate regions of the country, I focus on a group of six contiguous states in South-Central Brazil. These states, which I refer to as the "study region," account for about 80 percent of Brazil's output and experienced the changes in harvesting practices that motivate this paper.[7] By contrast, the smaller, two-state sugarcane growing region in the Northeast has long used a more rudimentary form of mechanized harvesting due to its history and ecology.[8] Figure 1 is a map of Brazil that indicates the study region.

The study region is home to about half of Brazil's 200 million inhabitants and is a major producer of sugarcane, coffee, oranges, soybeans, and maize. From 2006–2010, the study region accounted for 32 percent of world sugarcane production, 21 percent of world coffee production, 23 percent of world orange production, and 4 percent of world maize production.[9] According to household survey data, each crop employs between 0.5 and 2.5 million workers.[10]

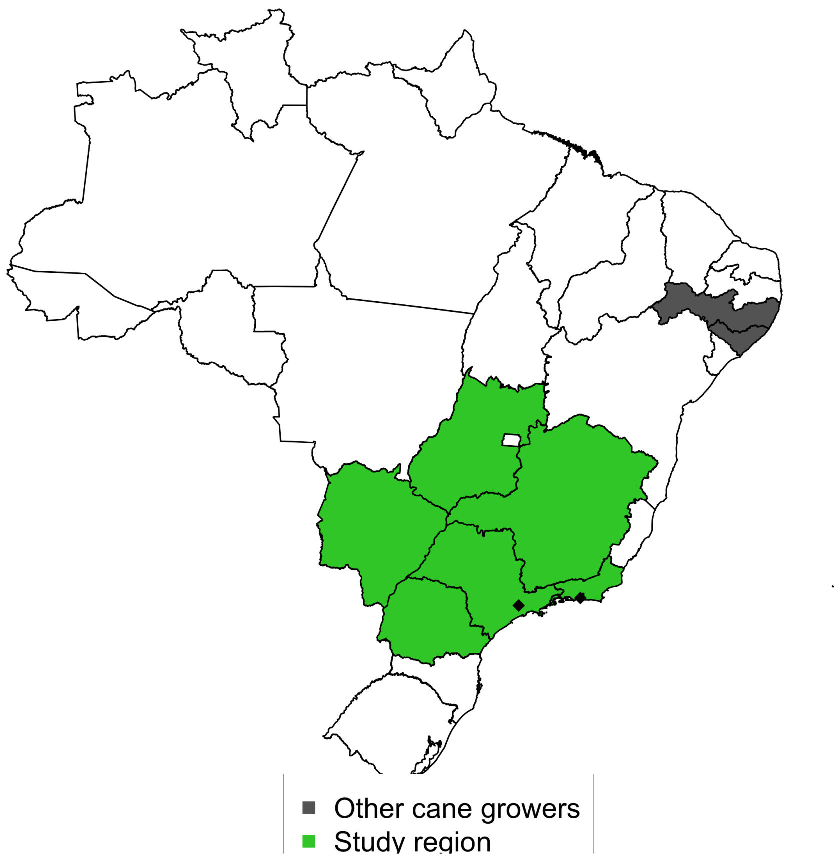
[7] The study region accounted for 80 percent of total sugarcane tonnage between 1990 and 2010, according to the Brazilian Census Bureau's Produção Agrícola Municipal (PAM) data.

[8] The six states in the study region are Goiás, Minas Gerais, Paraná, Mato Grosso do Sul, Rio de Janeiro, São Paulo. The two other sugarcane-producing states are Pernambuco and Alagoas, both located in the northeast region. Pernambuco and Alagoas do not have the optimal soil and climate for growing sugarcane and, according to anecdotal reports, the otherwise unprofitable industry endures thanks to subsidies from the state government.

[9] Brazilian output from PAM. World output from FAO.

[10] Calculated from PNAD.

Figure 1: Major sugarcane producing states and the study region



2.2 Background on Regulation

Sugarcane fields are burned in preparation for manual harvesting but not mechanical harvesting. The resulting pollution motivated state governments to restrict pre-harvest burning, effectively mandating mechanization. By eliminating pests and extra vegetative matter, burning allows harvest workers to move more quickly through the fields, approximately tripling their productivity. Following constituent complaints about salient pollution and health concerns, the state of São Paulo passed a gradual ban of pre-harvest burning in 2002. Under the regulations, property owners are permitted to burn only a fraction of each property.[11] In 2002, owners were permitted to burn 80 percent of each property. The regulation

[11]The regulation applied to each property as listed in the local property registry.

scheduled future reductions to this fraction: 70 percent in 2006, 50 percent in 2011, 20 percent in 2016, and 0 percent in 2021. This regulation became more strict in 2007, advancing all target reductions, with the complete cessation of burning required by 2014.[12] Between 2008 and 2014, the five neighboring states passed similarly structured regulation.

The regulations included meaningful incentives to change grower behavior. Violations of the São Paulo regulation could be punished by large fines. The regulation demands that growers pay a fine for each hectare burned in excess of their allowed fraction which, in 2002, amounted to 13 percent of average per-hectare revenue. The fine was revised upward every year, roughly tracking inflation. Additionally, failure to comply with burning restrictions jeopardized mandatory state environmental licenses. Enforcement strategies vary across the other states. Some impose large fines and threaten jail time while others instead offer incentives for compliance.

Anecdotal evidence suggests that the regulations were salient and enforced. National media publicized the passage of the regulations themselves while also documenting enforcement efforts. The São Paulo government partnered with the Brazilian space agency INPE to monitor harvesting practices via satellite images; this technique has been used to identify and fine violators. Citizens and journalists also reported violations of the law. Small and large growers have received fines throughout the years, according to newspaper reports. In these cases, amounts ranged from $3,000 for a small grower to over $1 million for a large grower.

Regulators claim these regulations were successful in changing harvesting practices while sugarcane growers use compliance to advertise their environmental stewardship. In verbal and written communications, environmental officials in São Paulo attribute mechanization to the regulation. The state of São Paulo maintains a website about the regulation which makes a number of claims, including: "with [regulation], all the mechanizeable area will be harvested

[12] Technically, the 2007 revision to the law was a voluntary agreement between sugarcane growers and the environmental regulator. The environmental regulator held leverage over sugarcane growers in the form of environmental licensing, other regulations, and lawmakers' threats of stricter legislation. The two parties agreed to more aggressive restrictions on burning while avoiding an uncertain and costly legislative process.

... without burning,"[13] that the regulation avoided millions of tons of pollutants,[14] and that the regulation quadrupled the number of harvesting machines in the state.[15] The Brazilian sugarcane industry's English-language website advertises sustainable practices, writing that "[m]echanization already exceeds 90 percent of the harvest in São Paulo, Brazil's top cane-producing state. It will be the only means of harvesting in São Paulo by 2017, thanks to [regulation]."[16]

Two features of the regulations allow me to evaluate these claims. First, the regulation in São Paulo was much less stringent for growers with less than 150 hectares. I use this variation to evaluate the São Paulo regulation via regression discontinuity. Second, the stringency of regulation varies across states and over time, allowing me to evaluate the regulation in a difference-in-differences framework.

2.3 Regression Discontinuity Evidence

Taking advantage of a size threshold built into the regulation, I estimate the effect of the regulation on harvesting practices via regression discontinuity.

2.3.1 Measuring Size and Harvesting Practices

The 2006 Census of Agriculture is a rich source with which to evaluate the regulation because the data are disaggregated, include a great breadth of information, and offer near-universal coverage. The Census of Agriculture records a variety of information about every agricultural establishment in Brazil, including their location, size, the crops grown, and harvesting technique. Thus, a researcher can identify regulated growers and study a range of relevant outcomes. The Brazilian Census Bureau endeavors to survey every agricultural establishment in the country. Finally, since these data describe each agricultural establishment, the

[13] `http://www.ambiente.sp.gov.br/etanolverde/protocolo-agroambiental/ganhos-ambientais/`

[14] `http://www.ambiente.sp.gov.br/etanolverde/files/2016/06/Etanol-Verde-Relatorio-Safra-15-16.pdf`.

[15] Ibid.

[16] `http://sugarcane.org/sustainability/best-practices`

unit of observation corresponds to the decision-making unit.[17]

The primary outcome in this analysis is an indicator variable for harvesting practices that should respond discontinuously to the regulation. Specifically, the primary outcome assumes a value of one if the establishment used manual harvesting only and zero if the establishment used any mechanical harvesting. Recall that the regulation required every establishment to harvest 20 percent of the land mechanically, so compliant establishments should have a zero. This measure may overstate compliance since it cannot distinguish establishments that mechanize some fraction below the required 20 percent.

While the Census does not record a continuous measure of mechanization, several continuously measured inputs serve as secondary outcomes in Appendix A.3.3. I consider five machine-related outcomes: expenditure on contracting services, fuel expenditure, the number of harvesting machines, machine rental expenditure, the value of all vehicles. I also consider three labor-related outcomes: days paid to temporary workers, the number of temporary workers, and the total number of workers.

The assignment variable is reported establishment area. The Census does not collect administrative data on establishment area, so the Census variable may not be correctly identify treatment status. As such, I test for strategic manipulation of the assignment variable but find no evidence of such manipulation. See Appendix A.3.1 for details.

At the time of the 2006 Census, São Paulo establishments had been regulated for the three previous years, but no other state had yet introduced regulation. Consequently, this analysis focuses exclusively on establishments in the state of São Paulo. Unfortunately, it is not feasible to incorporate additional years since the 2016 Census has not been completed and there are substantive differences between the 1996 and 2006 Censuses.

[17]Given the level of detail and disaggregation, these data are confidential. To access them, I traveled to a Census Bureau facility in Rio de Janeiro. I was permitted to analyze the data only in a secure room and I was only allowed to remove programs and output files, all of which were inspected by Census Bureau employees to ensure the anonymity of respondents.

2.3.2 RD Estimation

The regression discontinuity design compares São Paulo establishments above and below the 150 hectare regulatory threshold. This exercise yields the average treatment effect at the threshold. To the extent that the regulation has different effects for different size establishments, this parameter is most informative about establishments that are close to the 150 hectare threshold.[18]

Using the method described in Calonico et al. (2014), I estimate the treatment effect of the regulation as

$$\begin{aligned}\tau &= \lim_{x \to 150^+} \Big(\mathbb{E}[\, Y_i \mid X_i = x\,] \Big) - \lim_{x \to 150^-} \Big(\mathbb{E}[\, Y_i \mid X_i = x\,] \Big) \\ &= \mu_{Y^+} - \mu_{Y^-}\end{aligned}$$

where Y is an outcome variable, x is establishment area, and the threshold is 150 hectares. Separate local polynomials of degree p are estimated above and below the threshold:

$$\begin{aligned}\hat{\tau}_p(h_n) &= \hat{\mu}_{+,p} - \hat{\mu}_{-,p} \\ \hat{\mu}_{+,p} &= e_0\, \hat{\beta}_{+,p}(h_n) \\ \hat{\mu}_{-,p} &= e_0\, \hat{\beta}_{-,p}(h_n)\end{aligned}$$

with

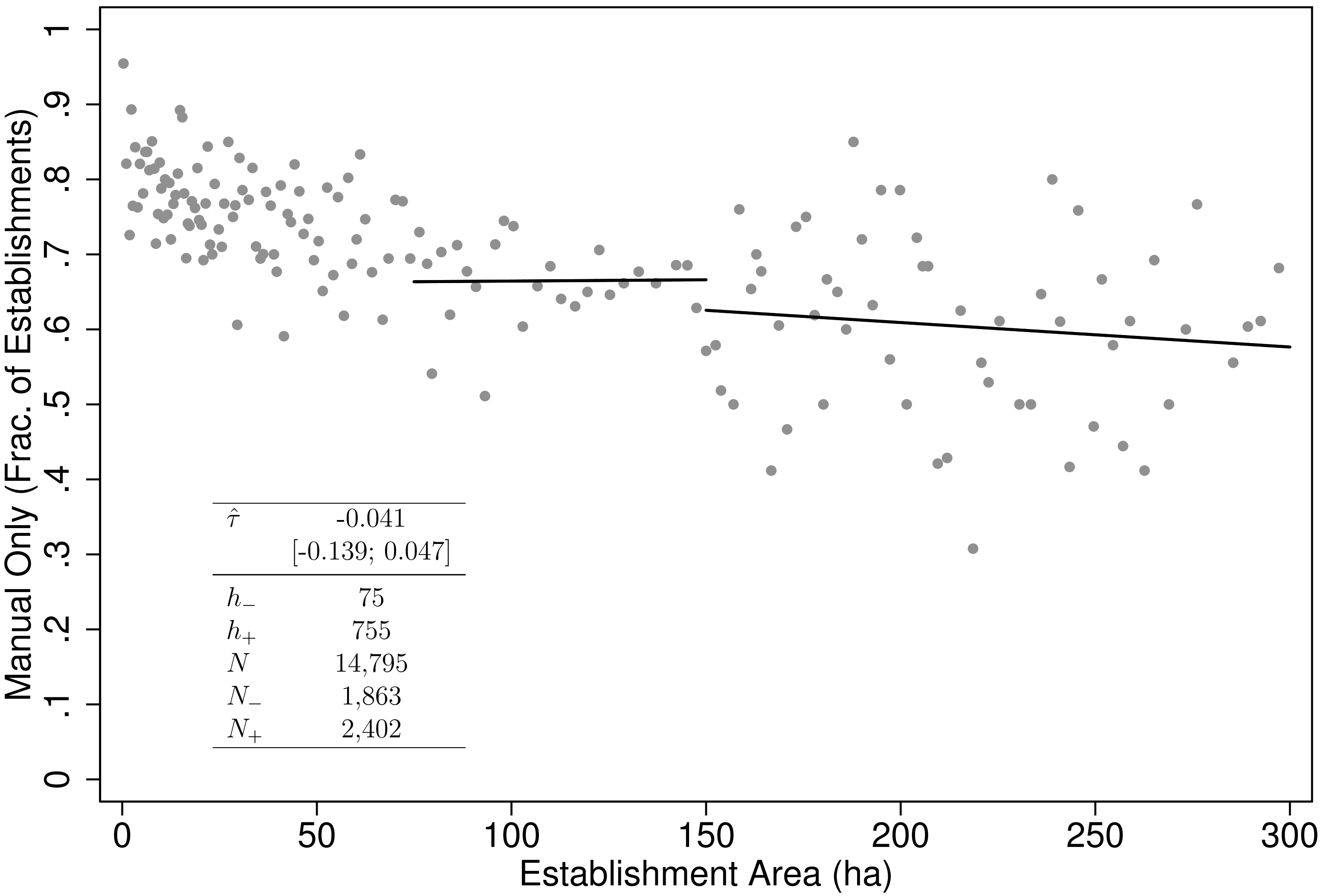
$$\begin{aligned}\hat{\beta}_{+,p}(h) &= \arg\min_{\beta} \sum_{i=1}^{n} \mathbb{1}(X_i \geq 0)\{Y_i - r_p(X_i)'\beta\}^2 K_{h_n} X_i \\ \hat{\beta}_{-,p}(h) &= \arg\min_{\beta} \sum_{i=1}^{n} \mathbb{1}(X_i < 0)\{Y_i - r_p(X_i)'\beta\}^2 K_{h_n} X_i\end{aligned}$$

[18] In Appendix A.4, I include difference-in-diffences estimates which compare the difference between São Paulo establishments above and below the the regulatory threshold, to the same difference among establishments outside of São Paulo. The resulting parameter may be interpreted as an average treatment effect on the treated. The results from this approach are substantively similar to the RD results.

where $r_p(x) = (1, x, ..., x^p)', e_0 = (1, 0, ..., 0)$, e K_{h_n} is a kernel function with a series of bandwidths h_n.

In choosing the parameters of the analysis, I follow the recommendations of Calonico et al. (2014) but varying these choices does not substantively alter the results. Specifically: i) bandwidths for the point estimate are selected using the method developed by Calonico et al. (2014), which minimizes mean squared error, ii) bandwidths for the bias correction are selected using the method developed by Calonico et al. (Forthcoming 2016), which minimizes coverage error, iii) I use the triangular kernel function, iv) I estimate local linear regressions ($p = 1$), and v) variance is estimated using a nearest-neighbor approach clustered by municipality. I use separate bandwidths on either side of the cutoff because the density of X is decreasing in this region; forcing symmetric bandwidths results in many observations below the threshold and few above. Finally, I report a 95% confidence interval instead of an estimated standard error because correct inference requires that the confidence interval be recentered to account for misspecification bias.

Figure 2: Binned Scatter Plot and Local Linear Estimates



Source: 2006 Census of Agriculture.

Data from 2006 Agricultural Census. Local linear estimates are plotted for $x \in [150 - h_-, \min\{300, 150 + h_+\}]$. Bin sizes set to mimic the variance of the underlying data (see Calonico et al. (2015)).

The regression discontinuity estimates in Figure 2 show small, insignificant declines in exclusive manual harvesting. Growers just above the area threshold, i.e. regulated growers, are about 4 percentage points less likely to use manual harvesting only. The magnitude of the estimates suggests that regulation accounts for less than a quarter of the near-complete mechanization observed by 2014. According to the law, no regulated establishment could rely exclusively on manual harvesting. Since more than 60 percent of regulated establishments report manual harvesting only, a regulation that worked as intended would have a much larger treatment effect. More importantly, we observe that virtually all sugarcane harvesting was mechanized by 2014. At the extreme of the confidence interval, the estimates admit the possibility that the regulation caused about 14 percent of growers to shift away from using manual harvesting only. Fourteen percent is small compared to both the design of the regulation and observed changes in behavior.

Even these small estimates may overstate the effect of the regulation since the binary outcome does not capture the extent of mechanization. If the regulation causes some establishments to mechanize, then these establishments may meet or exceed the fraction required by regulation. In this case, the regulation would still account for less than a quarter of the near-complete mechanization observed in later years. But these establishments may instead mechanize a small fraction of their land. I attempt to measure the extent of mechanization by estimating the regression discontinuity using various agricultural inputs as continuous proxy measures of mechanization. These estimates, detailed in Appednix A.3.1, are noisy but show no evidence that regulated establishments used inputs differently than unregulated establishments.

Appendix A.3.1 contains supplementary analysis, including estimates using a range of bandwidths and density tests.

Overall, the evidence shows that regulated establishments were slightly less likely to use manual harvesting alone. However, the regulation is far from sufficient to explain the rapid, widespread change in harvesting techniques.

In interpreting these results, it is important to acknowledge their limitations. These estimates could miss an effect of the regulation under certain circumstances. For example, suppose a small number of large growers switched harvesting techniques because of a targeted government enforcement effort. If the targeted growers are far above the threshold area, they are unlikely to affect the regression discontinuity estimates. However, if these growers control a large fraction of total sugarcane area, they could drive large-scale changes in harvesting practices. This is a real possibility since half of all sugarcane area is controlled by the largest 2 percent of establishments.

In the next section, I address this concern by combining a continuous, size-weighted measure of mechanization with an identification strategy that does not rely on a comparison of establishments near the 150 hectare threshold.

2.4 Difference-in-differences Evidence

2.4.1 Constructing Labor Intensity from Administrative Labor Data and Survey Data on Sugarcane Cultivation

I combine two data sources to construct a continuous, size-weighted measure of mechanization. I measure mechanization as the hours of labor supplied by manual laborers in the sugarcane industry (L) divided by the total area devoted to sugarcane cultivation (T).[19] Labor intensity is calculated for each município in each year.[20] Aggregating to the município level effectively weights mechanization by establishment size; if regulators change the behavior of the largest growers, this change might not be apparent when comparing growers near the 150 hectare threshold, but it should be visible in aggregate labor intensity. The numerator is drawn from confidential, administrative micro data maintained by the Brazilian Ministry of Labor.[21] Known by its Portuguese-language acronym RAIS, the micro data are

[19] I exclude most workers involved in mechanical harvesting by limiting the analysis to workers whose occupation is "manual laborer."

[20] In terms of area, population, and governance, Brazilian municípios are roughly equivalent to US counties.

[21] I combine two variables to calculate L: the length of an employment spell in months and the contracted hours per week. In practice, the vast majority of sugarcane workers report 44 hours per week so the variation

compiled annually and comprise the universe of formal employment.[22] [23] Each record in the dataset corresponds to an employment spell, providing information about the employee and the employer, including the municipality of work, wages, the duration of employment, hours worked, plus detailed industry and job classifications. For each muncípio-year, I measure the quantity of labor L as the sum of hours worked by manual laborers in the sugarcane industry. The denominator T is drawn from the PAM survey conducted by the Brazilian Census Bureau IBGE; IBGE employees contact local producers and other industry participants to determine the land area devoted to sugarcane cultivation in each município-year.

2.4.2 Identifying Variation from Cross-Sectional and Time-Series Differences in Regulation

Cross-sectional and time-series variation in the stringency of the regulation enable its evaluation via a differences-in-differences approach. The first regulation was introduced in São Paulo in 2002 and became more stringent over time. Other states introduced similarly-structured regulation between 2008 and 2014. The fraction of land that growers were permitted to burn or, equivalently, the fraction of land they were required to mechanize, varies across states and over time. Figure 3 shows the mechanization requirement in each state from 1999 to 2028. A value of zero, shaded red, means there is no mechanization requirement. A value of one, shaded green, means one hundred percent of property area must be mechanized. I use this mechanization requirement (Pct) as the treatment variable in a

is primarily driven by the length of employment.

[22]Owing to the sensitive, identifiable information stored in RAIS, these data are confidential. I obtained permission from the Brazilian Ministry of Labor to store and analyze the data at a secure facility maintained by the University of Michigan.

[23]In interviews, farmers and farm workers indicate that labor in the sugarcane sector is predominantly formal sector and unionized. As a consequence, RAIS captures roughly 60 to 75 percent of the sugarcane and coffee employment recorded in household survey data. Direct comparisons between household survey data, the PNAD, and RAIS are complicated for a several of reasons: i) the unit of analysis in each dataset is different, ii) quantity of labor is measured differently, and iii) many sugarcane workers are seasonal migrants. The figure I report here, 60 to 75 percent, is the national count of employment spells from RAIS divided by national count of individuals from PNAD. I use the national counts because RAIS records the place of work and PNAD records the place of residence. For seasonal migrants, these will not be the same so a single individual might appear in different places in each dataset.

continuous differences-in-differences design. This design assumes that changes in states with no change in stringency provide a counterfactual for states with a change in stringency.

2.4.3 Difference-in-differences Estimation

If regulation caused mechanization, we would expect the labor intensity of sugarcane harvesting to fall in states where the regulation became more stringent as compared to neighboring states without changes in regulation. I estimate the effect of the regulation on labor intensity using a differences-in-differences approach adapted for a continuous treatment variable, namely the stringency of regulation:

$$(L/T)_{j,s,t} = \delta_s + \gamma_t + \omega \mathrm{Pct}_{s,t} + \varepsilon_{j,s,t}, \tag{1}$$

where j indexes municipality, s indexes state, and t indexes year. The outcome L/T is labor intensity, defined as hours of labor contributed by manual workers in the sugarcane industry divided by area harvested. The fixed effects δ and γ capture state-specific and year-specific unobservables. Some specifications include municipality fixed effects instead of state fixed effects. Finally, Pct measures the required fraction of mechanization; these values are shown in Figure 3. The coefficient of interest is ω, which captures how the changing stringency of the regulation affects labor intensity. Some specifications include a lag and lead of Pct to capture anticipatory or delayed responses to the regulation.

The results, presented in Table 1, suggest that regulation can explain no more than a quarter of the observed decline in labor intensity. The reported estimates show how much labor intensity would change, in terms of hours per hectare, moving from unregulated harvesting to a complete ban on burning. The point estimates from columns (1) and (2) suggest that a complete ban on burning would actually increase by about 20 hours per hectare. Assume the true effect of the regulation lies at the lower bound of the 95 percent confidence interval in column 1, i.e. $\omega = 21.43 - 1.96 \times 21.22 = -21.2$. The average value

Figure 3: Required Mechanization as a Fraction of Land Area

	São Paulo	Rio de Janeiro	Goiás	Mato Grosso do Sul	Minas Gerais	Paraná
1999						
2000						
2001						
2002	0.20					
2003	0.20					
2004	0.20					
2005	0.20					
2006	0.30					
2007	0.50					
2008	0.50		0.10		0.80	
2009	0.50		0.10		0.80	
2010	0.70		0.10		0.80	
2011	0.70		0.10	0.17	0.80	
2012	0.70	0.20	0.10	0.34	0.80	
2013	0.70	0.20	0.25	0.50	0.80	
2014	1.00	0.50	0.25	0.67	1.00	
2015	1.00	0.50	0.25	0.84	1.00	0.20
2016	1.00	0.50	0.25	1.00	1.00	0.20
2017	1.00	0.50	0.25	1.00	1.00	0.20
2018	1.00	0.80	0.50	1.00	1.00	0.20
2019	1.00	0.80	0.50	1.00	1.00	0.20
2020	1.00	1.00	0.50	1.00	1.00	0.60
2021	1.00	1.00	0.50	1.00	1.00	0.60
2022	1.00	1.00	0.50	1.00	1.00	0.60
2023	1.00	1.00	0.75	1.00	1.00	0.60
2024	1.00	1.00	0.75	1.00	1.00	0.60
2025	1.00	1.00	0.75	1.00	1.00	1.00
2026	1.00	1.00	0.75	1.00	1.00	1.00
2027	1.00	1.00	0.75	1.00	1.00	1.00
2028	1.00	1.00	1.00	1.00	1.00	1.00

Table 1: The Effect of Regulation on Labor Intensity (DiD)

	(1) L / T	(2) L / T	(3) L / T	(4) L / T
Pct_{t-1}			4.700	2.801
			(17.97)	(18.24)
Pct_{t}	21.43	21.03	40.00	39.99
	(21.22)	(21.68)	(20.89)	(21.06)
Pct_{t+1}			-2.830	1.315
			(20.39)	(20.95)
N	8,263	8,263	6,097	6,097
$\bar{y}$	82.0	82.0	84.7	84.7
Muni FE		Y		Y

Pct is the legal mechanization requirement.
Quantity of labor (L) from RAIS.
Area harvested (T) from PAM.
Outcome Winsorized at the 1^{st} and 99^{th} percentiles.
SEs clustered by municipality.

of Pct, regulation stringency, was 0.573 in 2013. This implies that regulation reduced labor intensity by $0.573 \times 21.2 = 12.2$ hours per hectare in 2013. This amounts to one quarter of the observed reduction; in aggregate, labor intensity declined by 47 hours per hectare between 2007 and 2013.

The estimates are stable with respect to controls, offering some hope that the estimated effect of the regulation is not biased by omitted variables. The estimated effect of the regulation is essentially unaffected by municipality fixed effects.[24] The estimates are also substantively similar when controlling for state-specific time trends (results omitted; available on request). The point estimates are larger in columns 4-6, which include a lag and a lead of regulatory stringency. The larger estimates may result from the smaller sample, since observations from 1999 and 2013 are omitted. In any case, they are not statistically different from the corresponding estimates in columns 1-3.

These results suggest that regulation cannot explain more than a quarter of the observed

[24] Adding controls for município wages for manual laborers in sugarcane also has no affect on $\hat{\omega}$. Results available on request.

decline in labor intensity. The point estimates from columns (1) and (2) suggest that moving from 0 to a 100 percent mechanization requirement would actually increase labor intensity by about 20 hours per hectare.

Compared to the regression discontinuity, the difference-in-differences analysis estimates a conceptually different parameter from a separate data, but both approaches lead to the same conclusion: regulation accounted for a small fraction of the change in harvesting techniques. The difference-in-differences estimates can be thought of as an average treatment effect on the treated, measuring how labor intensity responded to regulation, on average, in regulated counties. That, together with the use of município labor intensity as an outcome, makes this approach sensitive to the behavior of large growers. Still, the effect of the regulation appears limited.

3 The Role of Wages

While regulation may have played some role, the mechanization of sugarcane remains largely unexplained. I turn now to the role of wages. I begin by describing the labor markets that produced a 50 percent increase in real wages for manual laborers in sugarcane between 1999 to 2013. I then present a model that shows how the tipping-point behavior observed in the data can emerge even in a frictionless, full-information environment. Finally, I estimate sugarcane growers' responsiveness to wage changes using an instrumental variables strategy.

3.1 The Market for Unskilled Labor in Brazil, 1999–2013

The late 1980s and early 1990s were a period of political change and economic uncertainty in Brazil. The country emerged from military dictatorship in 1989 and the first democratically elected president was impeached for corruption in 1992. Meanwhile, inflation ranged from 100 percent to over 30,000 percent between 1980 and 1995. As political and economic conditions stabilized in the late 1990s, a period of rapid, sustained growth took hold in the early 2000s.

Sugarcane was part of and subject to a broad-based increase in labor demand; from 2002 to 2013, real wages rose substantially in all occupations and industries while hours worked increased almost everywhere but agriculture. Figure 4 shows that, economy-wide, median hourly wages increased by more than 50 percent in real terms. Hours worked increased by more than 20 percent. Figure 5 shows the growth of real wages and hours worked by industry. Figure 6 gives the same information by occupation. All industries and occupations experienced meaningful wage growth from 2002–2013. The quantity of labor increased in all industries except agriculture and domestic services. The quantity of labor increased in all occupations except agricultural workers. Hours worked in agriculture fell by about 20 percent during this period.[25]

Growers faced steadily increasing wages for sugarcane workers before and during the period of rapid mechanization, while labor supply appears to decline in later years. Even adjusting for inflation, wages for sugarcane workers nearly doubled between 1999 and 2013.[26] Increases in hours worked through 2007 imply increases in labor demand that coincide with a large increase in area harvested. Subsequent decreases in hours, combined with higher wages, suggest contractions in labor supply from 2008 to 2013 (see Figure 7).

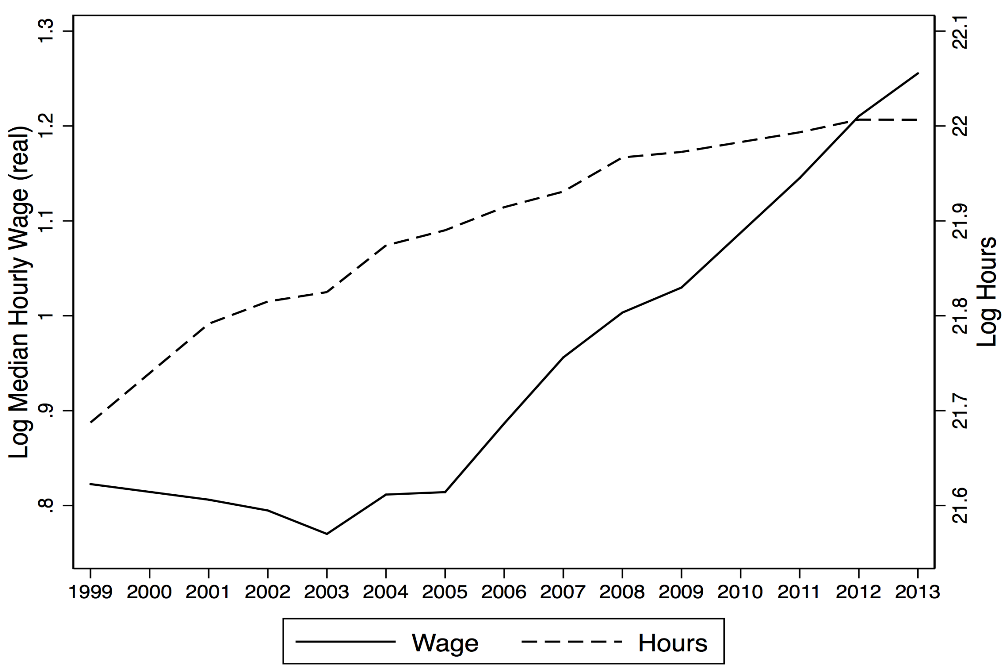
[25]The information in this paragraph, along with Figures 4, 5, and 6, is drawn from a nationally-representative household survey called the Pesquisa Nacional por Amostra de Domicílios (PNAD). Wages and hours worked are from a reference week, typically in early September, that includes the harvest season for many crops, including sugarcane.

[26]Author's calculations using administrative data (RAIS). Results are substantively similar using household survey data (PNAD).

Figure 4: Aggregate Employment and Real Wages from 1999–2013

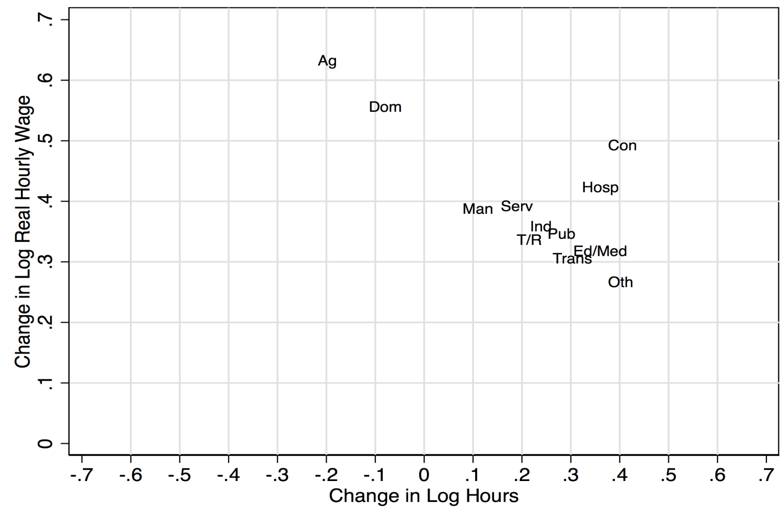
(a) Hours on the horizontal axis to emphasize movements of supply and demand

(b) Years on the horizontal axis to emphasize evolution of each series



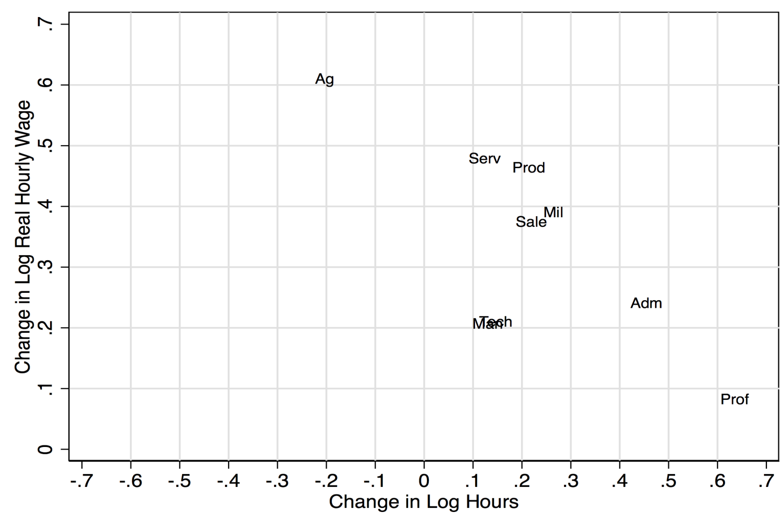
From PNAD micro data; PNAD not conducted in 2010 because that was a census year. Includes all paid workers. Hourly wages measured in 2003 R$.

Figure 5: Change in Wage and Employment 2002–2013 by Industry



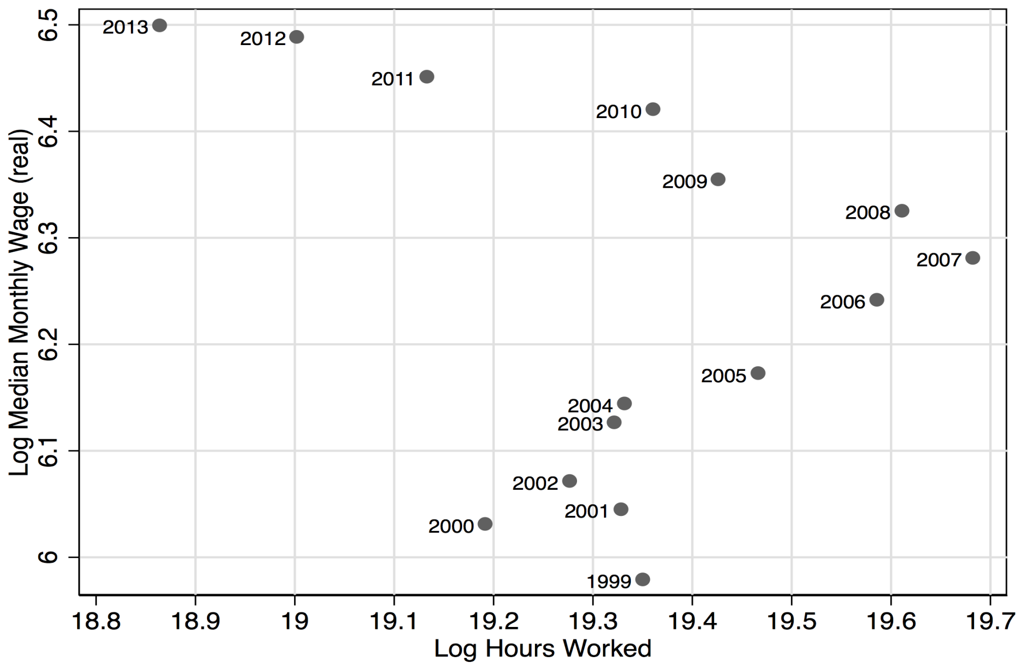
From PNAD micro data. Includes all paid workers. Hourly wages measured in 2003 R$. Industries are Agriculture (Ag), Industry (Ind), Manufacturing (Man), Construction (Con), Trade & repair (T/R), Hospitality (Hosp), Transport, communication, & storage (Trans), Public administration (Pub), Education, health, & social services (Ed/Med), Domestic services (Dom), Other services (Serv), Other (Oth). Changes to PNAD industry codes prohibit easy comparisons to earlier years.

Figure 6: Change in Wage and Employment 2002–2013 by Occupation



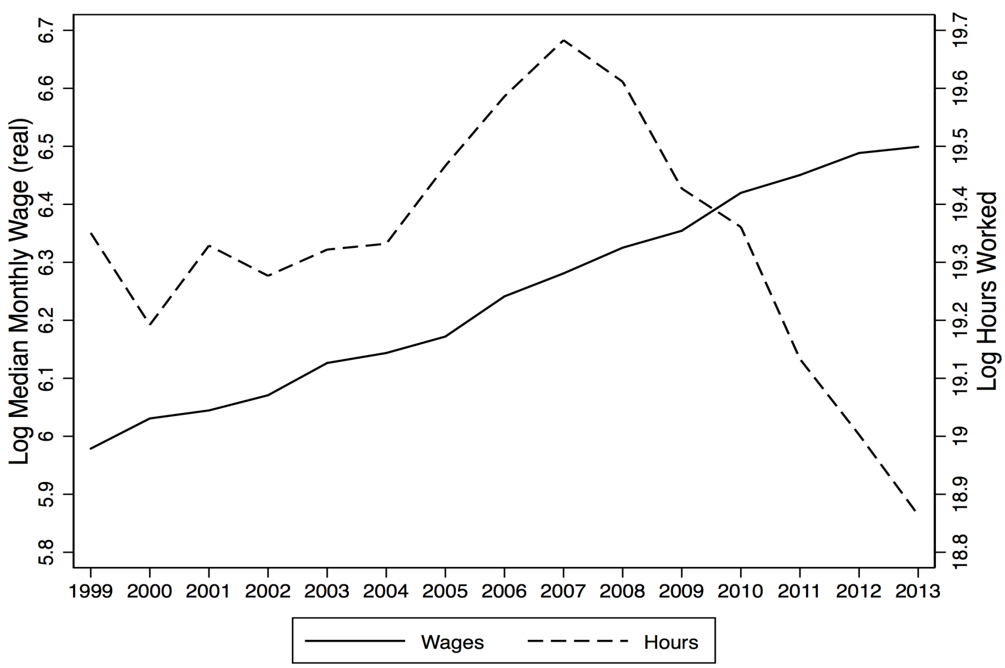
From PNAD micro data. Includes all paid workers. Hourly wages measured in 2003 R$. Occupations are Managers (Man), Professionals in arts and sciences (Prof), Mid-level technicians (Tech), Administrative workers (Adm), Service workers (Serv), Sales and business services (Sale), Agricultural workers (Ag), Production, construction, industrial, repair workers (Prod), Military (Mil). Changes to PNAD occupation codes prohibit easy comparisons to earlier years.

Figure 7: Employment and Wages in Sugarcane from 1999–2013



(a) Hours on the horizontal axis to emphasize movements of supply and demand

(b) Years on the horizontal axis to emphasize evolution of each series



From RAIS administrative data; results from PNAD survey data are substantively similar. Includes workers in the sugarcane-growing states of South-Central Brazil, limited by industry (sugarcane cultivation) and occupation (agricultural worker). Hourly wages measured in 2003 R$.

3.2 Sugarcane Cultivation

To inform a model of technology adoption, I summarize some relevant features of sugarcane cultivation.

Sugarcane growers may harvest manually or mechanically, each with its own inputs. Using the manual technology, unskilled workers cut sugarcane stalks with basic tools like machetes. With the mechanical technology, trained machine operators use sophisticated harvesters to cut and clean the stalks. Because the types of labor and capital are substantively different between the two technologies, I treat them as separate inputs.

The capital and labor shares are different between the two harvesting technologies. The observed capital to labor ratio is higher for the mechanical harvesting technology. In interviews, growers and machine manufacturers report that harvesting the same area requires 70 to 90 percent less labor using the mechanical technology as compared to the manual technology.

The sharing of harvesting machines is extensive so I will assume growers do not face fixed costs of acquiring machines. In practice, there are several ways to share machines, including contract harvest services, machine rental, and land rental. Data indicate that sharing is widespread: according to the 2006 Census of Agriculture, fewer than 10 percent of establishments that use mechanical harvesting actually own a harvesting machine.

The productivity of the manual technology is greatly enhanced by pre-harvest burning. Restrictions on burning effectively decrease total factor productivity of the manual technology. In a burned field, workers can clear 8 to 10 tons of sugarcane per hour. In an unburned field, slowed by heavy dense vegetation, workers clear around 3 tons of sugarcane per hour. While it is possible, pre-harvest burning is almost never combined with mechanical harvesting. Machine harvesters do move slightly faster in a burned field but one important feature of this technology is that, relative to manual harvesting, burning offers a very small increase in productivity.

The productivity of each harvesting technique depends on characteristics of the land. For example, turning large harvesting machines is slow so the size, shape, and layout of a parcel affect the rate of harvesting. Machines must proceed slowly and cautiously along hillsides. Mechanical harvesters cut as little as 450 tons per day on poorly prepared fields and as much as 1,000 tons per day on ideally prepared fields. While the productivity of manual harvesting also depends on parcel characteristics, it is much less sensitive to those characteristics.

3.3 Modeling the Choice of Harvest Technologies

One key fact in the data is that wages rose continuously for many years while mechanical harvesting showed a kind of S-shaped adoption: initially flat and then sharply increasing. S-shaped adoption is a common empirical finding in technology adoption so models of technology adoption inevitably account for this behavior somehow. In one class of models, this behavior originates from information diffusion or learning by doing (see, e.g., Foster & Rosenzweig (1995)). In another class of models, a fixed cost results in a size threshold for adoption and the S-shape comes from the interaction between changing factor prices and the size distribution (see, e.g., David (1969)). Manuelli & Seshadri (2014) develop a model of tractor adoption that, as in this paper, has no fixed costs or information diffusion. They argue that changing factor prices and improvements to the technology itself were the major drivers of adoption.

In the model presented below, I show that S-shaped adoption can be explained by factor prices alone, emerging in a full information environment with no fixed costs and no improvements to technology. S-shaped adoption in this model follows from three assumptions. First, that aggregate production is the sum of two production functions, one for each harvesting technology. Second, that each of the two production functions has constant or increasing returns to scale. These two assumptions imply that, for a given parcel of land, a profit-maximizing grower faces a threshold wage. Above the threshold, the grower will harvest the whole parcel mechanically and below the threshold the grower will harvest manually. The third key assumption is that these thresholds depend on characteristics of the land, like steepness, which have an S-shaped distribution. Thus, steadily rising wages can give rise to abrupt changes in the rate of mechanization.

The productivity of each harvesting techniques depends on characteristics of the land, so the model considers how a grower allocates a homogenous parcel of land between manual and mechanical harvesting techniques. Although analyzing homogenous pieces of land might seem unrealistic, it enhances the model's flexibility. The total area owned by a single grower may be broken into several parcels that are internally homogenous and the grower may decide to harvest each parcel differently.

As described above, growers do not face fixed costs to acquire a harvesting machine. Other fixed costs of mechanization, if any, are incorporated into the productivity of mechanical harvesting. For example, changing the layout of rows can increase the productivity of machine harvesting but, in many cases, it is still possible to harvest mechanically with a suboptimal layout. So, instead of including a fixed cost to change the layout, a parcel with a suboptimal layout will simply have a low productivity of mechanical harvesting.

The model reflects the fact that manual and mechanical harvesting apply fundamentally different types of labor and capital to the same land. The manual harvesting tool is a machete while the mechanical harvesting tool is a sort of combine. Those machines are operated by trained drivers while the manual harvesting is accomplished by unskilled workers. Thus, each technology (manual is denoted p for "person" and mechanical is denoted m for "machine") has separate inputs and factor prices: L_m, K_m, L_p, K_p and w_m, r_m, w_p, r_p. Land, T, is a normalized fixed factor $T_m + T_p = 1$. Production functions are constant elasticity of substitution (CES) with constant returns to scale (CRS):[27]

$$Y_m = A_m\left(\alpha K_m^{-\eta} + \beta L_m^{-\eta} + (1-\alpha-\beta)T_m^{-\eta}\right)^{\frac{-1}{\eta}} \text{ and} \tag{2}$$

$$Y_p = A_p\left(\gamma K_p^{-\zeta} + \delta L_p^{-\zeta} + (1-\gamma-\delta)T_p^{-\zeta}\right)^{\frac{-1}{\zeta}}. \tag{3}$$

Growers solve the following profit maximization problem:

$$\max_{L,K,T} Y_m + Y_p - w_m L_m - r_m K_m - w_p L_p - r_p K_p \tag{4}$$

$$\text{s.t. } T_p + T_m = 1. \tag{5}$$

I interpret the mechanical productivity term A_m as the suitability of a parcel for mechanical harvesting. This parameter will be high for flat parcels with perfectly arranged rows. It will be low for steep parcels or parcels with less-than-ideal layouts. The major source of variation in the manual productivity term A_p will be burning. Manual productivity is high for a burned parcel and low for an unburned parcel. I assume that wages and rental rates are not affected by the decisions made for any individual parcel, i.e. factor prices are exogenous.

In this model, growers will generally allocate an entire parcel to only one harvesting technique. This result follows from the constant returns to scale in each production function combined with linear costs; depending on productivities and factor prices, one harvesting technique will be cheaper than the other.[28] This observation leads to a proposition.

Proposition 1. *There exists a threshold manual wage for each parcel, above which growers mechanize and below which they harvest manually. For some combination of parameters and factor prices, growers are indifferent between techniques, which can be expressed as a*

[27]The CES form offers some generality. Because it is difficult to substitute between capital and labor within each technology, a Leontieff production function may be the most plausible but that is a limiting case of the CES production function.

[28]The same result emerges with increasing returns to scale. With decreasing returns to scale, it is still possible that growers will choose only one technique because land is a fixed factor.

threshold wage:

$$w_p = \phi(A_p, A_m, r_p, r_m, w_m, \alpha, \beta, \eta, \gamma, \delta, \zeta) \tag{6}$$

$$\textit{where } \frac{\partial \phi}{\partial A_p} > 0, \frac{\partial \phi}{\partial A_m} < 0. \tag{7}$$

Naturally, higher wages will encourage mechanization but, for any given parcel, the threshold wage will depend on several parameters. Figure 8a graphically describes a grower's optimal choices. For wages below the threshold, growers mechanize none of the parcel ($T_m = 0$). For wages about the threshold, growers mechanize all of the parcel ($T_m = 1$). Each parcel may have a different threshold wage ϕ which is determined by several parameters. Taking parcel steepness, a determinant of A_m, as an example, flatter parcels will mechanize at lower ages than steeper parcels.

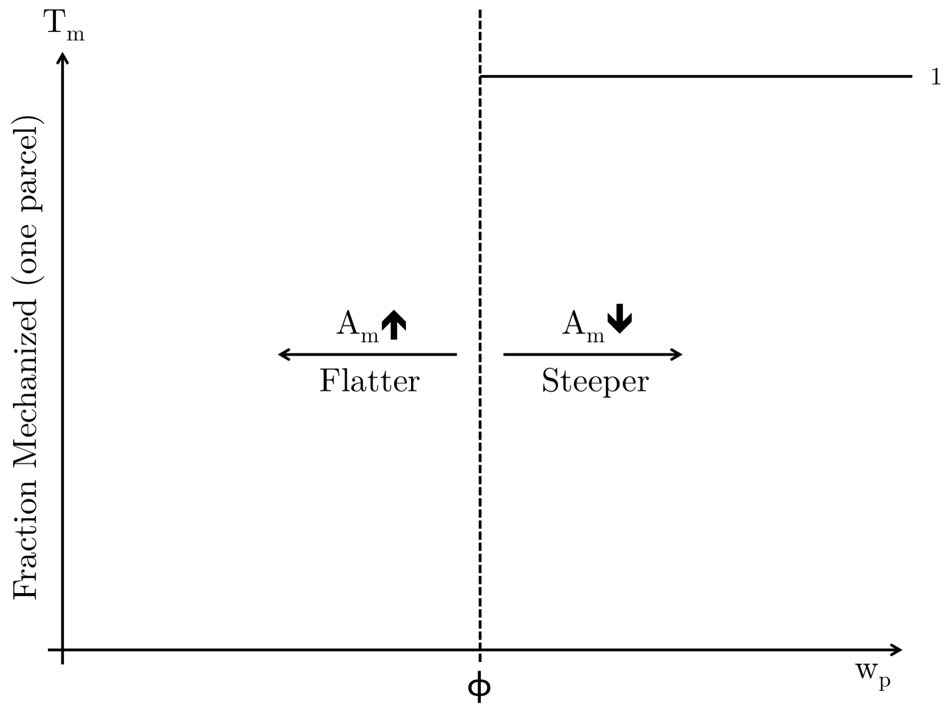
Since each parcel may have a different threshold wage ϕ, the aggregate response to wage changes depends on the distribution of ϕ. Since parcels differ in terms of characteristics like steepness, productivities A_p, A_m differ across parcels. Different productivities imply different thresholds ϕ. Thus, differences across parcels give rise a distribution of threshold wages ϕ.

If the cumulative distribution of ϕ is S-shaped, then steady increases in wages can lead to abrupt changes in the rate of mechanization. In places where the distribution is flat, wage changes will have limited effects on harvesting techniques. This is a situation where, for most farms, manual harvesting is either so cheap or so expensive that small wage changes have no impact on the choice of technique. In places where the distribution is steep, wage changes will have large effects because here many parcels are nearly indifferent between the two techniques. Figure 8b illustrates this point. $G(\phi)$ is a hypothetical CDF of the threshold ϕ across all parcels. A vertical line gives the current manual wage w_p. All parcels with a threshold ϕ below w_p will mechanize. All parcels above will harvest manually. An increase in wage between periods $t = 1$ and $t = 2$ induces very few parcels to mechanize. An identical increase in wage between periods $t = 2$ and $t = 3$ induces many parcels to mechanize as the wage $w_{p,3}$ is now above the threshold for a large fraction of parcels.

This prediction is underpinned by the S shape of the CDF of threshold wages; I find some support for this shape in the data. First, it's worth noting that every CDF has steep and flat portions, unless the underlying variable is uniformly distributed. Beyond that theoretical point, I find S-shaped CDFs in two parcel characteristics that affect the threshold wage through productivity A_m: parcel steepness and parcel size. Recall that steeper plots require machines to move more slowly, lowering A_m and smaller plots also lower A_m be requiring machines to turn more frequently. Figure 9a shows the empirical CDF of parcel grades, Figure 9b shows the empirical CDF of parcel area, and Figure 9c shows the joint density of parcel grade and area.[29] Most sugarcane parcels are observed to have similar steepness and area. If these characteristics are important determinants of the threshold wage, then we would expect most sugarcane parcels to have similar threshold wages. Future drafts will present direct evidence regarding the relationship between grade, area, and threshold wages.

Thus, profit maximizing growers, even in the absence of information frictions or fixed costs, may respond to continuously rising wages by sharply increasing the rate of mechanization.

[29]Note that the unit of observation is a parcel and not an establishment as in the Census of Agriculture. A parcel is defined as a unbroken area of sugarcane cultivation that is harvested using the same technique. One establishment may be composed of many parcels.



(a) Land allocation under the model

(b) Harvesting techniques and the distribution of ϕ

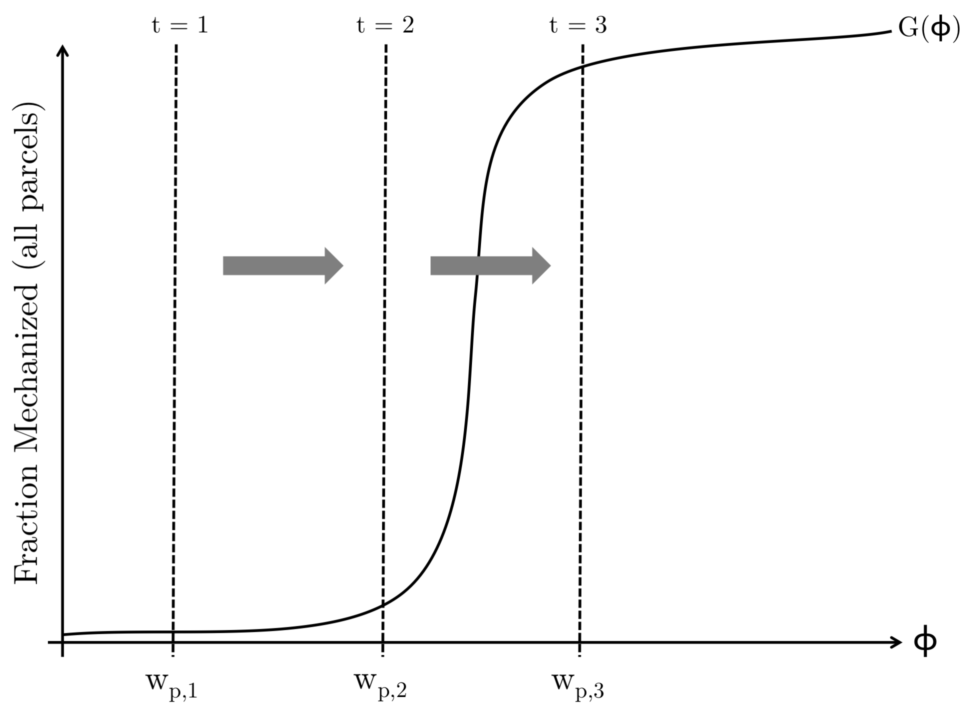
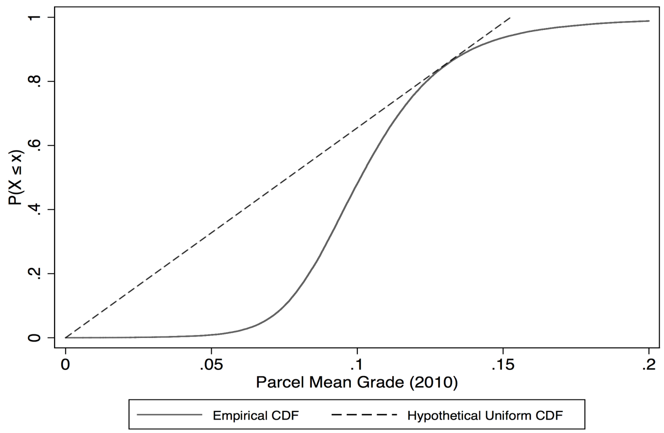
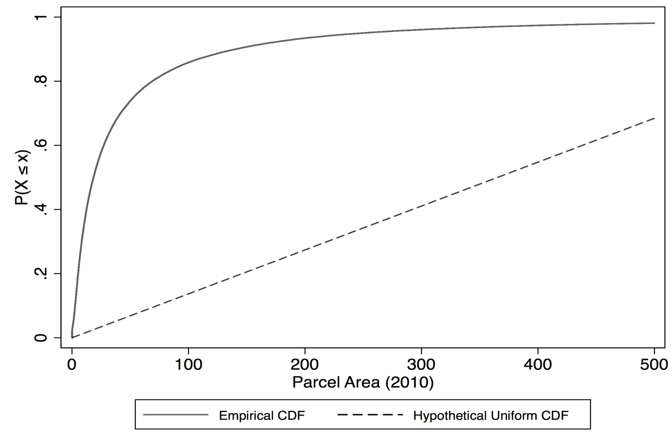


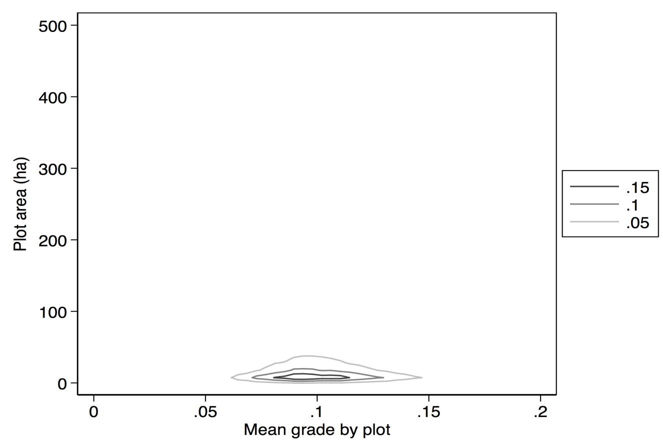
Figure 9: Empirical Support for S-Shaped Adoption



(a) Parcel-level CDF of Steepness (2010)



(b) Parcel-level CDF of Area (2010)



(c) Joint Density of Steepness and Area (2010)

3.4 Estimating The Effect of Wages on Mechanization

This section will estimate the change in labor intensity that would be expected solely due to the observed change in wages. Measuring the effect of wages requires exogenous variation to solve a basic problem of reverse causality, namely that mechanization itself could affect wages by shifting labor demand. This section details a method for estimating the elasticity of labor demand in sugarcane using instrumental variables based on fluctuations in labor demand from other agricultural sectors.

The study region is a major producer of sugarcane, coffee, oranges, soybeans, and maize. From 2006–2010, the study region accounted for 32 percent of world sugarcane production, 21 percent of world coffee production, 23 percent of world orange production, and 4 percent of world maize production.[30] According to household survey data, each crop employs between 0.5 and 2.5 million workers.[31]

I construct instruments for sugarcane wages that are conceptually similar to the instrument used in Dube & Vargas (2013); exogenous fluctuations in the markets for other crops generate variation in the wage for sugarcane workers. The intuition is straightforward. There are four other crops that i) are grown in the same region as sugarcane, ii) have a large land area devoted to their cultivation, iii) employ at least as many workers as sugarcane, and iv) employ similar types of workers as sugarcane. They are maize, soybeans, coffee, and oranges. Shocks to the production of these crops will likely shift the supply of labor facing sugarcane producers. For instance, unfavorable rains in Indonesia lead to lower coffee output, raising the international price. Coffee farmers hire more labor, increasing their output with extra pruning and tending. Because the newly-hired coffee workers might have harvested sugarcane, low coffee output in Indonesia generates a contraction in the labor supply faced by sugarcane farmers.

Taking coffee as an example, the instruments are constructed as the interaction between

[30] Brazilian output from PAM. World output from FAO.
[31] Calculated from PNAD.

two components: coffee output from the top three other producers interacted with the historical area cultivated of coffee. The first component of each instrument is the output from the top 3 producers of coffee, excluding Brazil.[32] This provides time-varying shifts in the labor supply faced by sugarcane farmers. The magnitude of that shift will depend on the importance of coffee to the local economy, offering a source of cross-sectional variation. Therefore, the second component of each instrument is the historical area cultivated of coffee, calculated as the 1994–1998 average.

I include a set of lags for each crop to account for the possibility that agricultural labor demand does not respond to contemporaneous price shocks. The speed of the response might vary by crops so I use several different lag structures. For the temporary crops maize and soy, which can be harvested the same season they are planted, I use contemporaneous and previous year data. For the permanent crops coffee and oranges, which take 3 to 5 years to bear fruit, I use contemporaneous data and four lags. The vector of excluded instruments Z is given by

$$\begin{aligned} Z_{j,t} \equiv \{&\mathrm{Int}_j^{maize} \times Y_t^{maize}, \mathrm{Int}_j^{maize} \times Y_{t-1}^{maize}, \\ &\mathrm{Int}_j^{soy} \times Y_t^{soy}, \mathrm{Int}_j^{soy} \times Y_{t-1}^{soy}, \\ &\mathrm{Int}_j^{cof} \times Y_t^{cof}, \mathrm{Int}_j^{cof} \times Y_{t-1}^{cof}, ..., \mathrm{Int}_j^{cof} \times Y_{t-4}^{cof}, \\ &\mathrm{Int}_j^{orng} \times Y_t^{orng}, \mathrm{Int}_j^{orng} \times Y_{t-1}^{orng}, ..., \mathrm{Int}_j^{orng} \times Y_{t-4}^{orng}\}, \end{aligned} \tag{8}$$

where Y represents logged output from the top 3 producers of crop c excluding Brazil. Intensity of cultivation is defined as the historical average area of crop c harvested in municipality j.

[32] The international price might influence agricultural labor supply more directly but this region of Brazil grows 20 to 30 percent of the world's coffee, soybeans, oranges, and sugarcane. Therefore, Brazilian output can have meaningful impacts on the international price of those crops. Imagine there's a frost in Brazil that damages both coffee and sugarcane plants. As a consequence, international coffee prices rise and sugarcane farmers hire less labor. In this case, the frost is directly affecting both the international price and the outcome. International prices cannot satisfy the exclusion restriction required of a valid instrument. I use output from other large producers to address this concern. However, this strategy may not totally eliminate the confounding effects of weather if those large producers change their output meaningfully in response to events in Brazil.

3.4.1 Measuring Wages and Quantities Labor, Constructing the Instrument

As in Section Section 2.4.1, I turn to confidential administrative data (RAIS) for annual, município-level wages and quantities of labor. Recall that these data include the universe of formal-sector workers and most sugarcane workers are in the formal sector. Using detailed occupation and industry codes, I measure the wages and employment of only the relevant workers: manual laborers in the sugarcane industry. I draw other-country crop output from the UN Food and Agriculture Oragnization's FAOSTAT database. The PAM survey described in Section 2.4.1 contains annual, município-level area harvested for all crops; I draw contemporaneous sugarcane area and historical area for other crops are from PAM.

3.4.2 Estimating the Elasticity of Labor Demand via IV

The first-stage regression takes the form:

$$\log w^S_{j,t} = \gamma_0 + \gamma_1 X_{j,t} + \Gamma Z_{j,t} + u, \tag{9}$$

where the superscript S indicates sugarcane, j indexes municipality, and t indexes year. I measure wages w^S as the median real wage for manual laborers employed in the sugarcane industry. This study is primarily concerned with the rapid decline in labor intensity associated with mechanization. Sugarcane area harvested increase meaningfully between 1999 and 2013. To control for the associated increase in labor demand, X includes the natural logarithm of sugarcane area harvested.

The second stage regression is given by:

$$\log L^S_{j,t} = \beta_0 + \beta_1 X_{j,t} + \beta_2 \widehat{\log w^S_{j,t}} + \varepsilon \tag{10}$$

The outcome variable L^S is the hours worked by manual laborers employed by sugarcane growers. The object of interest is β_2 which, when multiplied by the observe wage change, will provide an estimate of the change in labor predicted by the change in wages.

Table 2 below presents the estimation results from the second stage. The estimates in the first row show the increase in log hours associated with a one log point increase in area harvested. The second row displays estimates of the wage elasticity of labor demand. Columns (1)-(5) estimate equation (10) using different sets of instruments. Column (1) uses the maize instruments only. Columns (2)-(4) use soybeans only, coffee only, and oranges only. Column (5) uses the instruments from all four crops. Column (6) uses the instruments from all four crops but adds a squared wage term to capture the curvature in labor demand predicted by the model.[33]

Table 2: IV Estimates of Wage Elasticity of Labor Demand

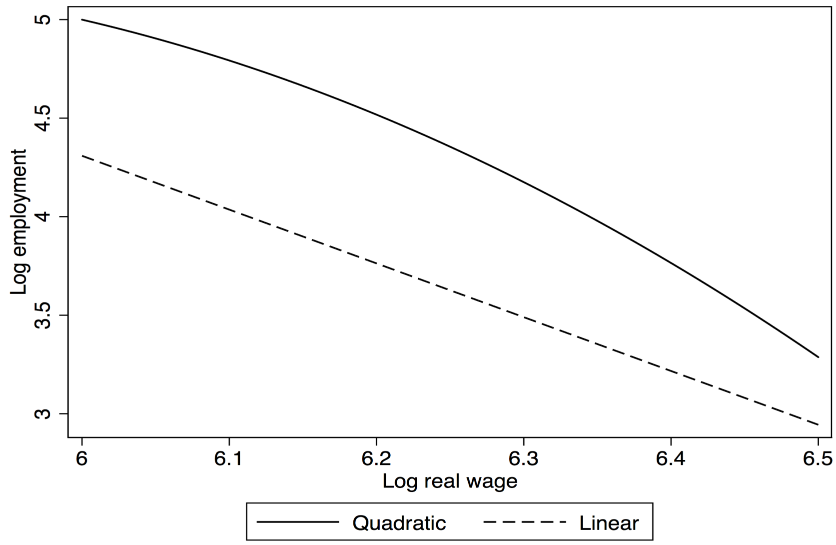
	(1) Maize	(2) Soybeans	(3) Coffee	(4) Oranges	(5) All	(6) All
$\log(A^S)$	0.852***	0.598***	0.813***	0.868***	0.840***	0.794***
	(0.0653)	(0.0900)	(0.0438)	(0.0560)	(0.0405)	(0.0446)
$\log(w^S)$	-2.938**	1.423	-2.204***	-3.201***	-2.729***	38.86*
	(0.992)	(1.477)	(0.551)	(0.726)	(0.420)	(17.01)
$[\log(w^S)]^2$						-3.383*
						(1.387)
N	8,233	8,233	8,179	8,179	8,179	8,179
F stat	19.9	13.9	15.1	5.1	9.5	

SEs clustered by county.

The estimates of the wage elasticity of labor demand are statistically significant with a consistent and credible magnitude. Columns (1)-(4) use only the instruments associated with one crop. Each of the four estimates relies on a different source of variation; the cross-sectional intensities and price series for each crop are are different. In three out of the four crops, the estimated elasticities are negative, statistically significant, and similar in magnitude. The estimate from soybeans is imprecise; although the point estimate is

[33]The sample includes the years 1999-2013. About 400 municipalities appear in the early years, increasing to almost 800 by the end of the sample. This increase likely corresponds to the approximate doubling of sugarcane cultivation during that time but, as the labor data come from an administrative dataset, it could also be the result of increased reporting.

Figure 10: Quadratic and Linear IV Results



positive 1.4, the lower bound of the 95 percent confidence interval is negative 1.5, roughly. For oranges, the excluded instruments are not as strong as one would like, with a first-stage F-statistic of 5.1. Column (5) uses the instruments for all crops, again finding a strongly significant elasticity near -3.

There is some evidence of the curvature in labor demand that is predicted by the model. Column (6) accounts for this curvature by adding the square of log wages as second endogenous variable. As shown in Figure 10, the models in columns (5) and (6) make roughly similar predictions for the quantity of labor over the observed range of wages.[34] However, the model from column (6) predicts a smaller elasticity at lower wages and a larger elasticity at higher wages. This pattern is consistent with both the model and the aggregate pattern of mechanization.

3.4.3 Expected Change in Labor Intensity from Observed Change in Wages

We can predict the change in labor intensity attributable to the wage changes alone by multiplying the observed wage change by the estimated elasticity. This calculation assumes

[34]The log of real median wages for sugarcane workers was near 6 in 1999 and close to 6.5 by 2013.

higher wages result from a decrease in labor supply; area harvested is meant to control for any shifts in labor demand. Differencing the second stage, equation (10) yields:

$$\widehat{\Delta \log L^s} = \hat{\beta}_1 \times \Delta \log A^s + \hat{\beta}_2 \times \Delta \log w^S \tag{11}$$

Now, I substitute the estimates for $\hat{\beta}_1, \hat{\beta}_2$ from Column (5) above, the change in the log of aggregate area cultivated for $\Delta \log A^s$, and the change in the log of aggregate median real wages for $\Delta \log w^s$:

$$= 0.84 \times 0.7 + -2.7 \times 0.5 = -0.76 \text{ with } 95\% \text{ CI } [-1.13; -0.39] \tag{12}$$

Given an observed change in the log of aggregate quantity of labor of -0.49, the wage changes are sufficient to predict the decline in labor intensity.

4 Conclusions

The course of development involves a shift from low productivity agriculture that employs many to high productivity agriculture that employs few. Brazilian sugarcane offers a window into this process with the recent adoption of mechanized harvesting.

As lawmakers contemplated regulation, economists predicted that mechanization would dramatically depress employment (Osse, 2002). Instead, the adoption of labor-saving technology coincided with a period of tightness in the labor market, characterized by large wage increases and increases in aggregate employment. By making manual harvesting more expensive, these wage changes potentially contributed to mechanization but mechanization began after years of rising wages.

Wages are not the only candidate explanation; beginning in 2002, state governments passed regulation that prioritized reductions to pollution and its associated health benefits. At the predicted cost of hundreds of thousands of jobs, sugarcane growers were obligated to reduce and eventually halt the pre-harvest burning that facilitates manual harvesting. Increasingly stringent regulation coincides with the adoption of machine harvesting.

However, the empirical analysis developed in this paper argues that development itself, in the form of rising wages for some of the poorest workers in Brazil, pulled labor out of agriculture. A range of econometric evaluations find limited evidence that the regulation contributed to mechanization. By contrast, my estimate of the elasticity of labor demand suggests that the observed change in wages is sufficient to explain mechanization.

These empirical results are supported by a theoretical framework that reconciles the trends in wages and mechanization. The model has two key insights. First, that there exists a threshold wage for each parcel of land; for wages below that threshold it is cheaper to harvest manually but for wages above that threshold it is cheaper to harvest mechanically. Therefore, if wages do not rise above the threshold, they can increase without causing a change in harvesting techniques. The second insight is that each parcel of land will have a different threshold based on the characteristics of that parcel. Because each parcel has a different threshold wage, mechanized and manual harvesting will coexist at certain wages.

This paper argues that mechanization was caused by increasing wages without thoroughly investigating why wages were increasing. That wages were increasing in spite of widespread mechanization is surprising and it suggests large increases in labor demand, especially for low-skilled workers. Future research will explore changes in Brazilian labor demand, identifying the sectors responsible and exploring the consequences for the wage distribution. A companion paper studies the health benefits that might be expected from the reduction in burning.

Another way to view these results is as a success of sustainable development. It is often assumed that more economic output means more pollution but, in this context, major development markers, like wages for the poor and agricultural productivity, improved alongside environmental outcomes. Sugarcane farmers once burned an area the size of New Jersey every year. Mechanization has significantly curtailed the air pollution associated with sugarcane harvesting even as agricultural workers earn substantially more. Development may well have led to a better environment in this case.

References

Athey, S., & Imbens, G. W. (2006). Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica*, *74*(2), 431–497.
URL http://dx.doi.org/10.1111/j.1468-0262.2006.00668.x

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (Forthcoming 2016). On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *Journal of the American Statistical Association*.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, *82*(6), 2295–2326.
URL http://dx.doi.org/10.3982/ECTA11757

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2015). Optimal Data-Driven Regression Discontinuity Plots. *Journal of the American Statistical Association*, *110*(512), 1753–1769.
URL http://dx.doi.org/10.1080/01621459.2015.1017578

Cattaneo, M. D., Jansson, M., & Ma, X. (2016). Simple Local Regression Distribution Estimators with an Application to Manipulation Testing. Working paper.
URL http://www-personal.umich.edu/~cattaneo/papers/Cattaneo-Jansson-Ma_2016_LocPolDensity.pdf

Coissi, J. (2008). Usinas lideram ranking da poluição ambiental na região. *Folha de S. Paulo*.

Conley, T. G., & Udry, C. R. (2010). Learning about a New Technology: Pineapple in Ghana. *The American Economic Review*, *100*(1), pp. 35–69.
URL http://www.jstor.org/stable/27804921

Credendio, J. E. (2008). Usinas lideram ranking de multas por poluição em SP. *Folha de S. Paulo*.

David, P. A. (1966). The Mechanization of Reaping in the Ante-bellum Midwest. *Industrialization in Two Systems*, *30*.

David, P. A. (1969). *A Contribution to the Theory of Diffusion*. Memorandum: Center for Research in Economic Growth. Research Center in Economic Growth, Stanford University.
URL https://books.google.com/books?id=OC5JGwAACAAJ

Davis, L. W. (2008). The Effect of Driving Restrictions on Air Quality in Mexico City. *Journal of Political Economy*, *116*(1), pp. 38–81.
URL http://www.jstor.org/stable/10.1086/529398

den Bulte, C. V. (2000). New Product Diffusion Acceleration: Measurement and Analysis. *Marketing Science*, *19*(4), 366–380.
URL http://dx.doi.org/10.1287/mksc.19.4.366.11795

Dube, O., & Vargas, J. F. (2013). Commodity Price Shocks and Civil Conflict: Evidence from Colombia. *The Review of Economic Studies*, *80*(4), 1384–1421.
URL http://restud.oxfordjournals.org/content/80/4/1384.abstract

Duflo, E., Kremer, M., & Robinson, J. (2011). Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya. *American Economic Review*, *101*(6), 2350–90.
URL http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.6.2350

Folha de S. Paulo (2008). Usina diz que recorreu de multa ambiental.

Foster, A. D., & Rosenzweig, M. R. (1995). Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture. *Journal of Political Economy*, (pp. 1176–1209).

Gazeta Mercantil (2002). O fim da queima da palha da cana.

Gollin, D. (2010). Chapter 73 Agricultural Productivity and Economic Growth. vol. 4 of *Handbook of Agricultural Economics*, (pp. 3825 – 3866). Elsevier.
URL http://www.sciencedirect.com/science/article/pii/S1574007209040730

Greenstone, M., & Hanna, R. (2014). Environmental Regulations, Air and Water Pollution, and Infant Mortality in India. *American Economic Review*, *104*(10), 3038–72.
URL http://www.aeaweb.org/articles.php?doi=10.1257/aer.104.10.3038

Henrique, B. (2008). Cetesb multa produtores de cana na região de Ribeirão. *O Estado de S. Paulo*.

Mansfield, E. (1961). Technical change and the rate of imitation. *Econometrica*, *29*(4), 741–766.
URL http://www.jstor.org/stable/1911817

Manuelli, R. E., & Seshadri, A. (2014). Frictionless Technology Diffusion: The Case of Tractors. *American Economic Review*, *104*(4), 1368–91.
URL http://www.aeaweb.org/articles?id=10.1257/aer.104.4.1368

McCrary, J. (2008). Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test . *Journal of Econometrics*, *142*(2), 698 – 714.
URL http://www.sciencedirect.com/science/article/pii/S0304407607001133

Morris, M., Kelly, V. A., Kopicki, R. J., & Byerlee, D. (2007). *Fertilizer Use in African Agriculture*. The World Bank.
URL http://elibrary.worldbank.org/doi/abs/10.1596/978-0-8213-6880-0

O Globo G1 (2016). Polícia Ambiental usa imagens de satélite para identificar queimadas.

Olmstead, A. L., & Rhode, P. W. (1993). Induced Innovation in American Agiculture: A Reconsideration. *Journal of Political Economy*, (pp. 100–118).

Olmstead, A. L., & Rhode, P. W. (1995). Beyond the threshold: An analysis of the characteristics and behavior of early reaper adopters. *The Journal of Economic History*, *55*, 27–57.

Osse, J. S. (2002). Colheita mecânica de cana-de-açúcar fecha 300 mil vagas. *Folha de S. Paulo*.

Samora, R. (2006). Usinas são multadas em R$ 1,6 mi por queima irregular de cana. *Reuters*.

Spinelli, E. (2001). Usinas mantêm queimadas na região. *Folha de S. Paulo*.

Tomazela, J. M. (2016). Polícia Ambiental adota critério que anistia queima de cana em SP. *O Estado de S. Paulo*.

White, W. J. (????). Economic History of Tractors in the United States. `https://eh.net/encyclopedia/economic-history-of-tractors-in-the-united-states/`. Accessed: 2016-06-16.

A Appendix

A.1 Labor Demand

It is possible to derive an analytical expression for labor demand by assuming a form of the production function and assuming a joint distribution of threshold wage and parcel area.

I begin by assuming that the production functions are Leontieff. From section 3.3, farmers choose to harvest manually when the manual wage w_p is below the threshold wage ψ. For parcels which are harvested manually, i.e. for which $\psi \geq w_p$, a Leontieff production function has two implications i) farmers will devote the entire parcel area to manual harvesting ($T_p = T$), ii) L_p will always be employed in a fixed proportion to land area $L_p = \lambda_p T_p = \lambda_p T$. Thus, for any given parcel, labor demand is given by:

$$L_p = \begin{cases} 0 & \text{if } \psi \leq w_p \\ \lambda_p T & \text{if } \psi > w_p \end{cases} \tag{13}$$

Moving from the parcel-level labor demand to aggregate labor demand requires an assumption about the joint distribution of ψ and land area T. For analytical convenience, I assume that ψ, T have a joint normal distribution:

$$\begin{pmatrix} \psi \\ T \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_\psi \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_\psi^2 & \rho\sigma_\psi\sigma_T \\ \rho\sigma_\psi\sigma_T & \sigma_T^2 \end{pmatrix}\right] \tag{14}$$

Aggregate labor demand will be given by (with i indexing N total parcels):

$$\mathbb{E}\Big[\sum_i L_p\Big] = \sum_i \mathbb{E}\big[L_p\big] = N\,\mathrm{P}(\psi > w_p)\,\lambda_p\,\mathbb{E}\big[T \mid \psi > w_p\big] \tag{15}$$

$$= N\left(1 - \Phi\left(\frac{w_p - \mu_\psi}{\sigma_\psi}\right)\right)\lambda_p\,\mathbb{E}\big[T \mid \psi > w_p\big] \tag{16}$$

where $\Phi()$ is the standard normal CDF. To derive an expression for the last expectation, begin with the conditional expectation of a jointly distributed normal variable

$$\mathbb{E}\big[T \mid \psi = w_p\big] = \mu_T + \rho\frac{\sigma_T}{\sigma_\psi}\Big(\mathbb{E}\big[\psi \mid \psi = w_p\big] - \mu_\psi\Big) \tag{17}$$

Then, by the Law of Iterated Expectations,

$$\mathbb{E}\big[T \mid \psi > w_p\big] = \mu_T + \rho\frac{\sigma_T}{\sigma_\psi}\left(\mathbb{E}\big[\psi \mid \psi > w_p\big] - \mu_\psi\right) \tag{18}$$

$$= \mu_T + \rho\frac{\sigma_T}{\sigma_\psi}\left(\mu_\psi + \sigma_\psi\frac{\phi\left(\frac{w_p-\mu_\psi}{\sigma_\psi}\right)}{1-\Phi\left(\frac{w_p-\mu_\psi}{\sigma_\psi}\right)} - \mu_\psi\right) \tag{19}$$

$$= \mu_T + \rho\sigma_T\frac{\phi\left(\frac{w_p-\mu_\psi}{\sigma_\psi}\right)}{1-\Phi\left(\frac{w_p-\mu_\psi}{\sigma_\psi}\right)} \tag{20}$$

where $\phi()$ is the standard normal PDF. Returning to equation 16,

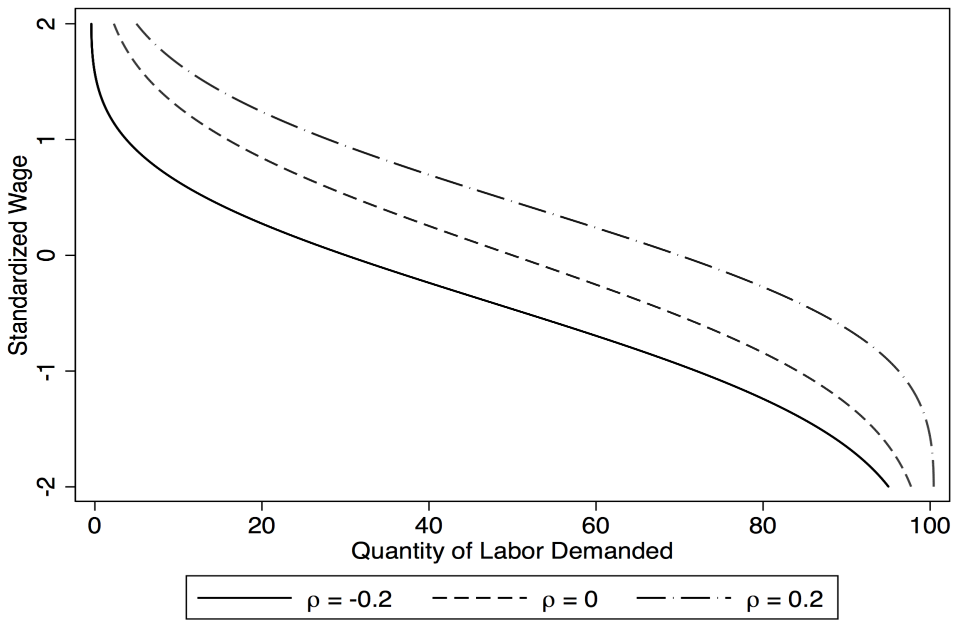
$$\mathbb{E}\Big[\sum_i L_p\Big] = \sum_i \mathbb{E}\big[L_p\big] = N\left(1-\Phi\left(\frac{w_p-\mu_\psi}{\sigma_\psi}\right)\right)\lambda_p\left(\mu_T + \rho\sigma_T\frac{\phi\left(\frac{w_p-\mu_\psi}{\sigma_\psi}\right)}{1-\Phi\left(\frac{w_p-\mu_\psi}{\sigma_\psi}\right)}\right) \tag{21}$$

$$= \underbrace{N\lambda_p\mu_T}_{\text{scale}}\left[\underbrace{\left(1-\Phi\left(\frac{w_p-\mu_\psi}{\sigma_\psi}\right)\right)}_{\text{Crossing thresholds}} + \underbrace{\rho\,\frac{\sigma_T}{\mu_T}\,\phi\Big(\frac{w_p-\mu_\psi}{\sigma_\psi}\Big)}_{\text{threshold area corr.}}\right] \tag{22}$$

We can interpret this expression as the combination of three effects. The first term, labeled "scale," scales the quantity of labor based on the number of parcels N and the labor required to harvest a parcel of average area, $\lambda_p\mu_T$. The second term, labeled "crossing thresholds," measures the fraction of parcels that are harvesting manually, i.e. those with thresholds ψ above the observed wage w_p. The final term, labeled "threshold area corr.," adjusts quantity of labor based on the correlation between parcel threshold ψ and parcel area T. Because parcels only harvest manually if the threshold ψ is above the observed wage w_p, labor is only demanded by parcels in a truncated portion of the threshold distribution. If parcel thresholds ψ are correlated with parcel area T, knowing that a parcel threshold ψ is above the observed wage w_p also reveals something about the area of parcels that are demanding labor. For example, if large parcels have lower thresholds, i.e. T and ψ are negatively correlated, we must adjust the quantity of labor to reflect the fact that manually harvested parcels will be the smallest parcels.

This expression provides general intuition about the competing forces that determine aggregate demand for sugarcane harvest labor. However, deriving this expression required two non-trivial assumptions. Leontieff production functions are reasonable in this context as inputs are not readily substitutable. The joint normal distribution, however, is less defensible. The wage threshold ψ is unobservable and the distribution of parcel areas is not normal. The resulting labor demand has an undesirable feature: it can have a positive slope if parcel area T has a high coefficient of variation and / or the correlation ρ is large in absolute value. Figure 11 plots Equation 22 for several values of ρ with a coefficient of variation equal to 2.5, which lies at the low end of the observed value across the years 2006–2010.

Figure 11: Labor Demand with Different Correlations ρ



A.2 Regression Discontinuity Evidence: The Steepness Threshold

Below, I describe a supplemental regression discontinuity analysis that takes advantage of another threshold in São Paulo's regulation. In addition to the 150 hectare area threshold, growers were exempted from strict regulation if their plots had a steepness of at least 12 percent.

Exploiting the steepness threshold built into the regulation, I evaluate the regulation with a regression discontinuity design with burning as the outcome. From satellite data analyzed by the Brazilian space agency INPE, I directly observe the location and harvesting method for all sugarcane cultivation in the state of São Paulo from 2006 to 2012. I calculate the mean slope of each plot using a high-resolution digital elevation model produced by NASA.[35]

[35] In principle, the area of 150 hectares offers another threshold which might be tested using a RD analysis. Unfortunately, the parcels identified in the satellite data do not necessarily correspond to legally defined property boundaries. Comparing the distribution of areas in the satellite data to the distribution of farm areas from the 2006 Agricultural Census, parcels in the satellite data are generally smaller than the farm area. Since the regulation was meant to apply to each farm, using satellite-measured area would systematically understate the running variable, undermining any resulting RD estimates. In future drafts, I will exploit

As in Section 2.3.2, I use the method described in Calonico et al. (2014). The outcome Y takes a value of one if the field is unburned and zero otherwise and the running variable parcel steepness.[36]

We would expect to observe a discontinuity in harvesting practices at the grade and size thresholds if i) farmers at each threshold would choose manual harvests in the absence of the regulation and ii) the regulation was effective. We would expect these discontinuities to appear in 2002 and to widen in 2006 and 2007, corresponding to changes in the law.

The regulation does not appear to affect harvesting practices near the threshold; there is no statistically significant difference in burning propensity across the area or steepness thresholds. Figure 12 show the binned averages, a fitted global polynomial control function, and the threshold for 2007.[37] This figure is largely representative of the other years; somewhere between 40 and 60 percent of fields are burned and there is no jump in burning at the thresholds. Figure 13 plot the estimated average treatment effect at the threshold for each year, along with 95 percent confidence intervals. The estimates represents a percentage point change in the likelihood of mechanization associated with being subject to the regulation, i.e. below the steepness threshold. If the regulation caused mechanization below the threshold, we would expect statistically significant positive results. Instead, we find negative point estimates, none of which is significantly different from zero.

The estimates are somewhat imprecise but they rule out effects of a size we would expect if the regulation were binding and well-enforced. The RD estimates the change in the likelihood that a parcel will be mechanized moving from the unregulated to the regulated side of the threshold. Because the area of parcels will be equal, in expectation, on either side of the threshold, the estimate also corresponds to the fraction of area mechanized because of the regulation.[38] Between 2007 and 2009, the agreement required the mechanization of 50 percent of farm area for farms below 12 percent grade and no mechanization of farms above 12 percent grade. Thus, if the regulation were binding and well-enforced near the threshold, we would expect a RD estimate of 0.5 in those years. The upper limit of the confidence interval is only about one quarter that size. Similar logic and conclusions apply

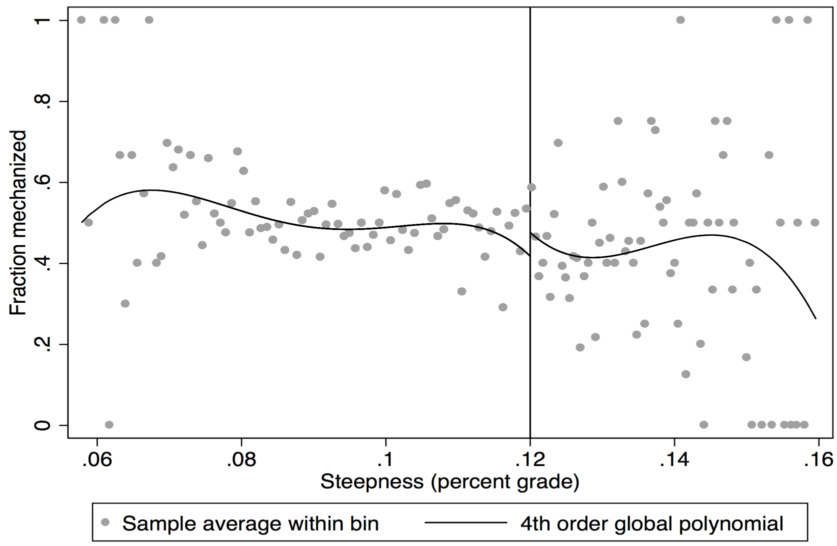
this 150 hectare area threshold using confidential micro data from the 2006 Agricultural Census. This issue will also affect the measurement of steepness. However, with steepness, the splitting of farms may just add noise to the measured running variable, rather than systematically mismeasuring it.

[36] I calculate parcel steepness using a digital elevation map from NASA's Shuttle Radar Topography Mission. These data record elevation on a 30 meter by 30 meter grid. I calculate percent grade at each grid point and take the mean of all grid points within each parcel. Results do not change substantively if I use the 75^{th} percentile instead of the mean.

[37] The global polynomial is for illustrative purposes only. The RD estimates are the difference between intercepts of a local polynomial.

[38] Running the RD analysis with area as the outcome variable reveals no statistically significant differences in area across the threshold.

Figure 12: Burning Propensity and Fitted Polynomial (2007)



Excludes plots below the 150 ha threshold for area; Triangular kernel; BW and bin sizes estimated as in Calonico, Cattaneo, and Titiunik.

to the estimates from 2006 and 2010.

There are a several potential explanations for these results. One is that the regulation had no effect, at least not near the thresholds. Another is that I mismeasure the running variable. For this analysis, each "parcel" is a polygon outlined from a satellite image that may or may not correspond to the unit at which the policy was applied. As outlined above, this limitation of the data introduces noise into the running variable. This noise will decrease the precision of the estimates and attenuate the point estimates so, if there is a small effect of the regulation, the analysis might fail to detect it. A final explanation is that, instead of changing harvesting practices, farmers on the regulated side threshold switched crops or lay fallow. I will consider changes in land use as in outcome in future drafts.

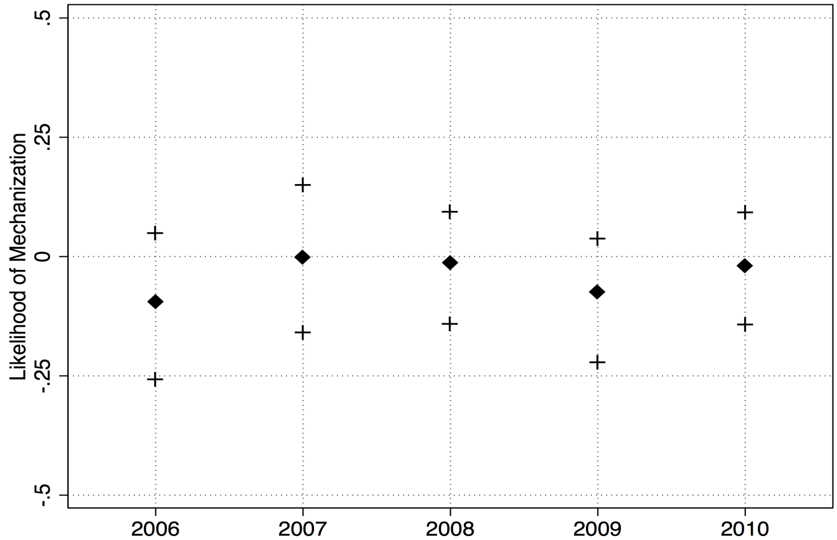
A.3 Supplemental Analysis for Regression Discontinuity Estimates Using Size Threshold

A.3.1 Bandwidths and Running Variable Density

This section will address two issues relevant to the validity of regression discontinuity estimates: i) continuity of the running variable at the threshold and ii) bandwidth selection.

The density of establishment area is discontinuous at the regulatory threshold; rather than manipulation to avoid regulation, this discontinuity appears to the result of heaping

Figure 13: Estimated Average Treatment Effect at the Threshold by Year



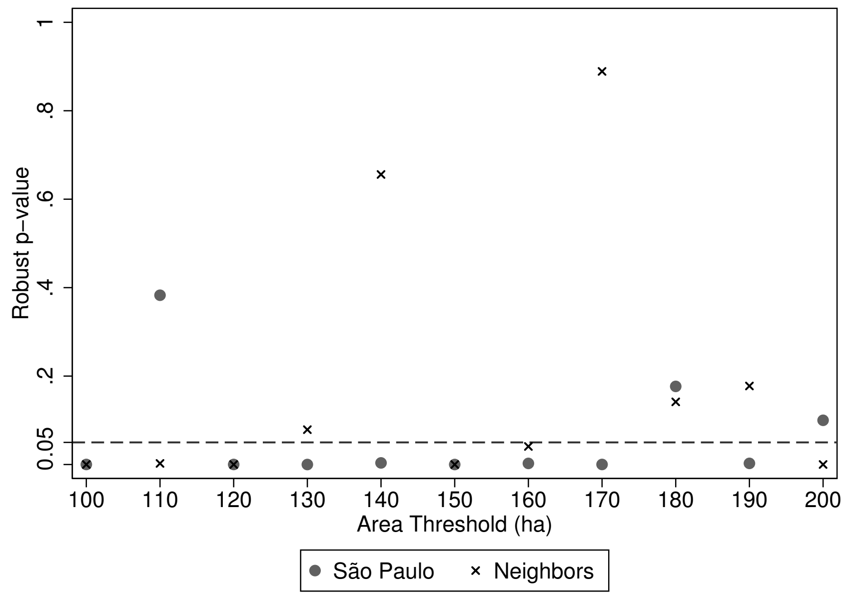
Excludes plots below the 150 ha threshold for area; Triangular kernel; BW and bin sizes estimated as in Calonico, Cattaneo, and Titiunik.

at multiples of 10. I test for continuity of the density using the procedure described by Cattaneo et al. (2016) but results are substantively similar using McCrary (2008). Figure 14 shows the p-values for tests of continuity conducted at every multiple of 10 between 100 and 200 hectares, both inside and outside of São Paulo. At the 95 percent level, the test rejects continuity for 8 of 11 area thresholds inside São Paulo. Outside São Paulo, among neighboring states without a burning regulation at the time, the test rejects continuity for 7 of 11 area thresholds. The histogram shown in Figure 15 makes clear that observations are concentrated near multiples of 5. Moreover, the histogram does not suggest a large mass of establishments strategically underreporting their area to avoid mechanization.

Heaping is unlikely to introduce a material bias in the RD estimates. Imagine that the regulator and the grower know the establishment's true area but, responding to the Census, the grower rounds area to the nearest multiple of 5. The RD estimate assumes that establishments who report 150 regulated when, in this scenario, some are not. In principle, the RD could underestimate the effect of regulation by counting some unregulated establishments as regulated. However, this bias is probably small because, in spite of the statistically significant discontinuity in the density, less than 40 observations report exactly 150 hectares, only a fraction of these would be misclassified, the estimation samples include over 4,000 observations, and the outcome is binary.

Heaping may explain the discontinuities in the density of the running variable but it does not rule out manipulation; if manipulation is present, it suggests that the already small RD

Figure 14: p-values for Continuity of the Density of the Running Variable



Data from 2006 Census of Agriculture.

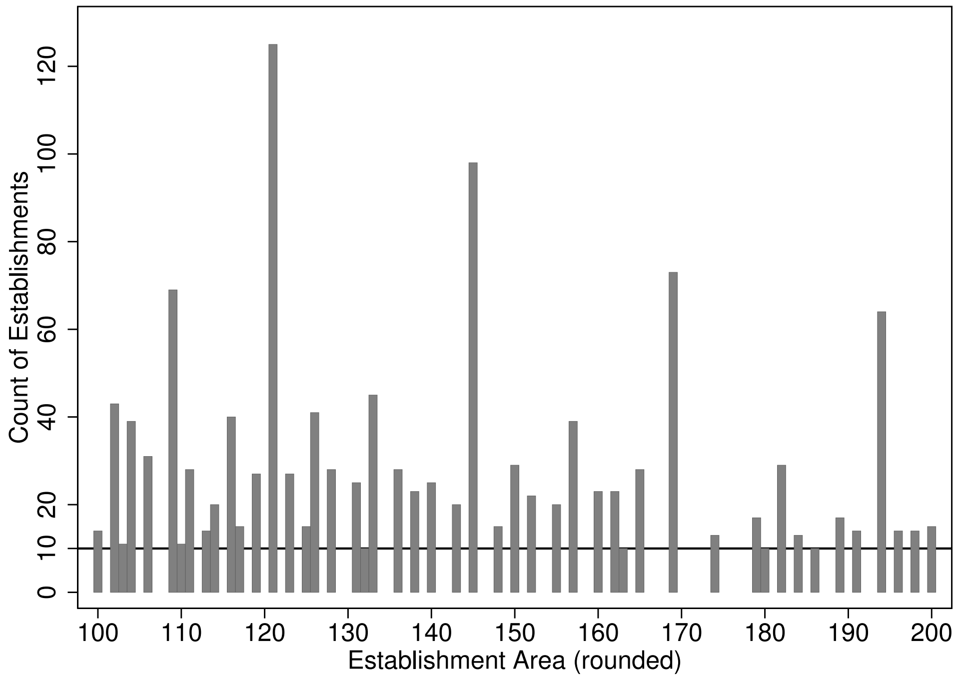
estimates may be an upper bound on the effect of regulation. Perhaps some of the observations reporting 145 hectares truly have 150. Intuition suggests that such manipulation would lead to an overestimate of the effect of regulation. The growers most likely to strategically underreport their area are the growers with the highest costs of complying with the regulation and, consequently, the least likely to comply. Since manipulation would remove from the "treatment" group the establishments that are least likely to comply, manipulation would introduce an upward bias into the RD estimate.

Bandwidth selection has important impacts on point estimation and inference. Generally speaking, a larger bandwidth increases the precision of a point estimate at the cost of increasing bias. Recent work on RD estimation by Cattaneo and coauthors argues that estimating this misspecification bias is critical for correct inference. Confidence intervals must be recentered, to adjust for the bias, and scaled, to account for the variability introduced by estimating the bias. Estimating the bias also necessitates selecting a bandwidth. Since each bandwidth may be different on either side of the threshold, point estimation and inference can use up to 4 bandwidths.

Because bandwidth selection is important, I report estimates based on systematic, objective procedures that produce bandwidths with desirable properties. To estimate the RD parameter τ, I use the bandwidth selection procedure from Calonico et al. (2014) which balances bias and variance to minimize MSE. To generate confidence intervals, I use the bandwidth selection procedure from Calonico et al. (Forthcoming 2016), which minimizes coverage error. For point estimating and inference, I choose separate bandwidths on either side of the threshold. I do so because the density of the running variable is decreasing rapidly around the cutoff, resulting in a large number of observations below the threshold and few above. Symmetric bandwidths would give a precise but biased estimate below the threshold and an imprecise but relatively unbiased estimate above the threshold. Using separate bandwidths ensures that estimates above and below the threshold optimally balance bias and variance.

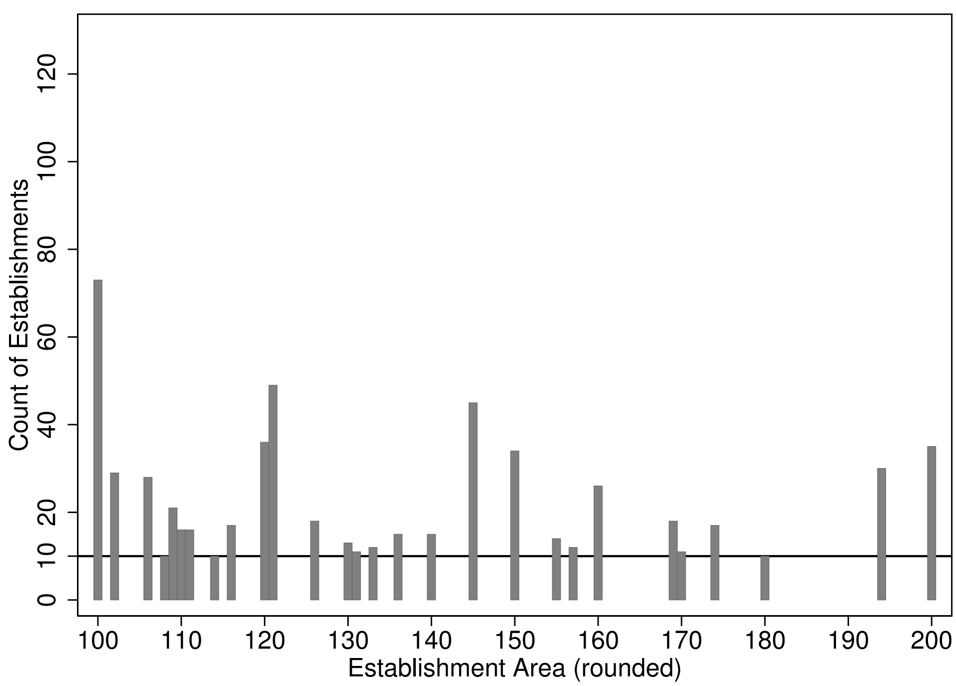
The regulation is unable to explain mechanization under conservative assumptions and a range of bandwidths. Figure 16 shows the point estimates with confidence intervals for a range of symmetric bandwidths, including manually selected and MSE-optimal bandwidths. I show symmetric bandwidths here for ease of exposition and because they yield larger point estimates and wider confidence intervals. Since the evidence overall argues against a large effect for the regulation, I take these estimates as conservative. No point estimate is larger than 10 percentage points in absolute value and the point estimates tend to zero as the bandwidth increases. None of the 95 percent confidence intervals exclude zero. The confidence intervals do exclude a decrease of more than 30 percentage points in manual harvesting, an increase of more than 30 percentage points in manual and mechanical, and an increase of more than 20 percentage points in mechanical harvesting. Recall that, by 2014, virtually all harvesting was done mechanically. To account for this change, the regulation would have to reduce manual harvesting by roughly 65 percentage points, with corresponding changes in the other outcomes

Figure 15: Histogram of Establishment Area (rounded to nearest hectare)



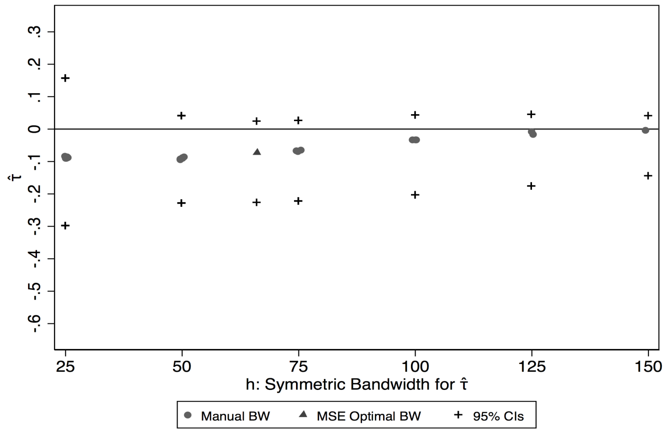
(a) São Paulo

(b) Neighbors

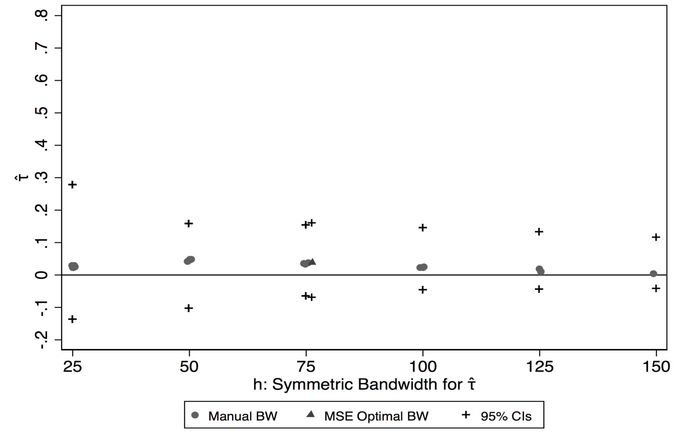


Data from 2006 Census of Agriculture. Disclosure rules prevent me from displaying bins with less than 10 observations. This level is marked on the graph. The true value for missing bins may be anywhere below the line.

Figure 16: Point Estimates and CIs at Various Symmetric Bandwidths

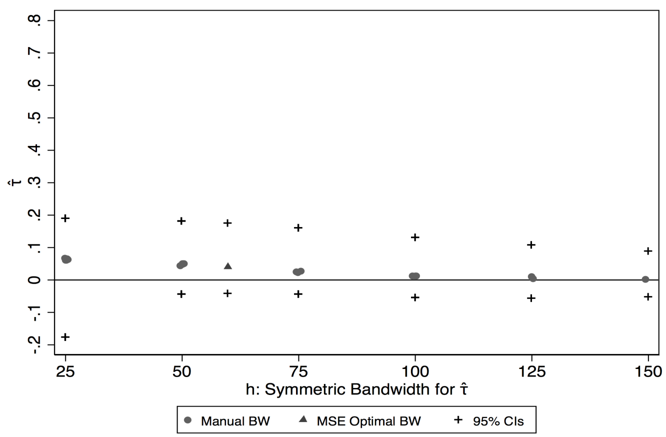


(a) Manual only



(b) Manual and mechanical

(c) Mechanical only



Data from 2006 Agricultural Census. Point estimate from local polynomial of degree 1. Bias correction from local polynomial of degree 2. Variance estimated using nearest neighbor method clustered by municipality. There are multiple point estimates at each manually selected bandwidth h; these correspond to different bandwidths b used to generate robust bias-corrected confidence intervals. The bias-correction bandwidth b must be at least as large as the point-estimate bandwidth h. For each symmetric h, I estimate confidence intervals using $h \leq b \in \{25, 50, 75, 100, 125, 150\}$ and display the widest.

Table 3: The Effect of Regulation on Harvesting Techniques (RD)

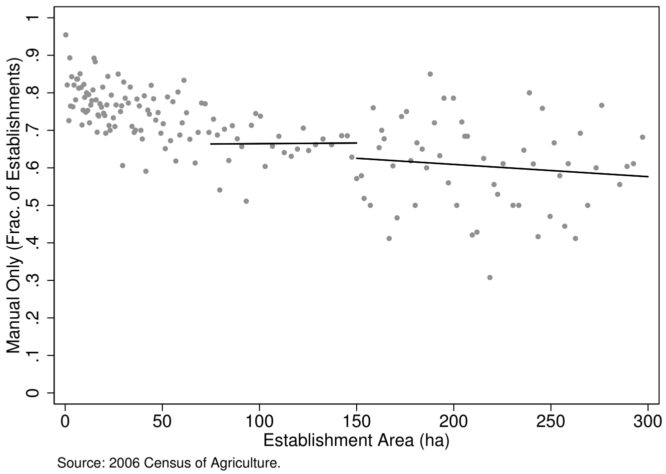
	$\mathbb{1}$(Manual)	$\mathbb{1}$(Both)	$\mathbb{1}$(Mechanical)
$\hat{\tau}$	-0.041	0.013	0.025
	[-0.139; 0.047]	[-0.068; 0.099]	[-0.035; 0.093]
h_-	75	87	78
h_+	755	697	1,270
N	14,795	14,795	14,795
N_-	1,863	2,375	2,070
N_+	2,402	2,348	2,594

Point estimate from local polynomial of degree 1. Bias correction from local polynomial of degree 2. Variance estimated using nearest neighbor method clustered by municipality. Reported bandwidths h selected to generate the MSE optimal point estimates $\hat{\tau}$. The CER optimal CIs use somewhat smaller bandwidths. Assignment variable is Establishment Area.

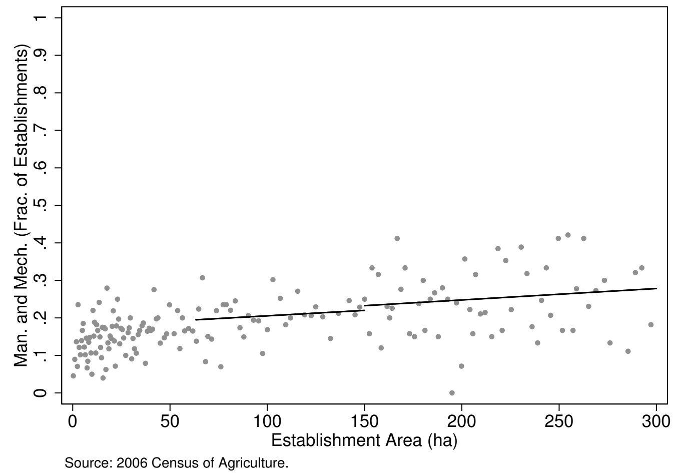
A.3.2 The Full Set of Binary Outcomes

The 2006 Census of Agriculture asks respondents to report whether they use manual harvesting only, mechanical harvesting only, or both. This is the only direct question on harvesting practices. I convert this question into three indicator variables: $\mathbb{1}$(Manual), $\mathbb{1}$(Mechanical), and $\mathbb{1}$(Both). For brevity, Section 2.3.2 presents the results of estimation for only the first of those three indicator variables. In Table 3 and Figure 17, I provide the results for all three. Estimation procedures are the same as described in Section 2.3.2.

Figure 17: Binned Scatter Plots and Local Linear Estimates

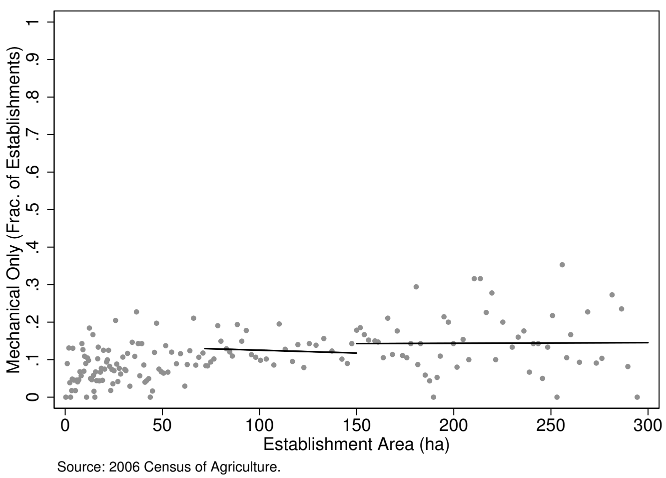


(a) Manual only



(b) Manual and mechanical

(c) Mechanical only

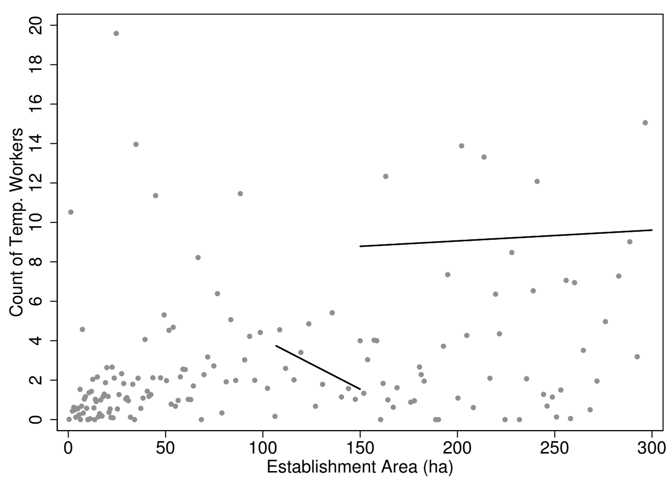


Data from 2006 Agricultural Census. Local linear estimates are plotted for $x \in [150 - h_{-}, \min\{300, 150 + h_{+}\}]$. Bin sizes set to mimic the variance of the underlying data (see Calonico et al. (2015)).

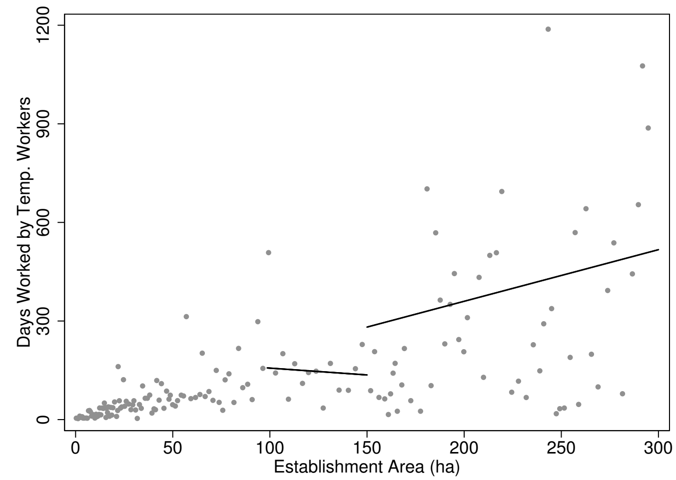
A.3.3 Continuously-varying Inputs as Outcome Variables

If growers were shifting large amounts of area from manual to mechanical harvesting, we would expect to see similarly large changes in a variety of inputs that are measured by the Census of Agriculture. Repeating the estimation procedure from Section 2.3.2, I consider three labor-related inputs (days paid to temporary workers, the number of temporary workers, and the total number of workers) and five machine-related inputs (expenditure on contracting services, fuel expenditure, the number of harvesting machines, machine rental expenditure, the value of all vehicles). Figures 18 and 19 show binned scatterplots of the data along with regression lines. In Table 4, I present estimates from the same regression discontinuity procedure as above to see if regulated growers used different inputs than their unregulated counterparts. The estimates are noisy so, in general, the confidence intervals do not exclude large effects of regulation. However, only one of the point estimates are statistically significant and some have counterintuitive signs. The number of employees is higher among regulated establishments, and significant at the 5 percent level, while fuel expenditure and the number of harvesting machines is lower. Finally, the graphs provide no visual evidence of input changes at or near the threshold.

Figure 18: Binned Scatter Plots and Local Linear Estimates for Manual Harvesting Inputs

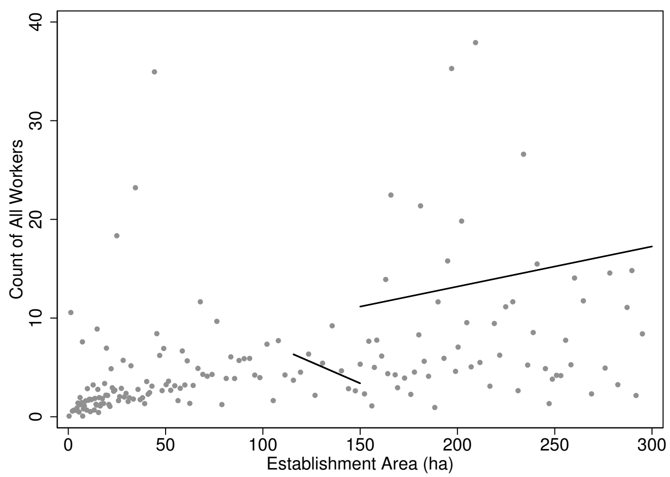


(a) Count of Temporary Workers



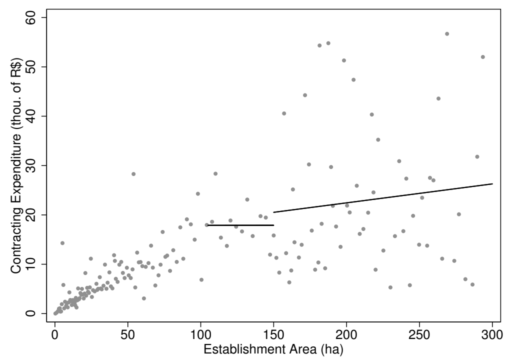
(b) Days Worked by Temporary Workers

(c) Count of All Workers

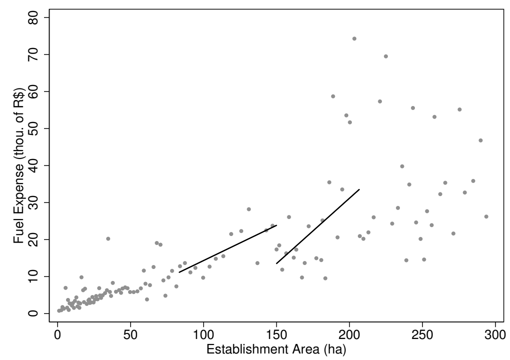


Data from 2006 Agricultural Census. Local linear estimates are plotted for $x \in [150 - h_{-}, \min\{300, 150 + h_{+}\}]$. Bin sizes set to mimic the variance of the underlying data (see Calonico et al. (2015)).

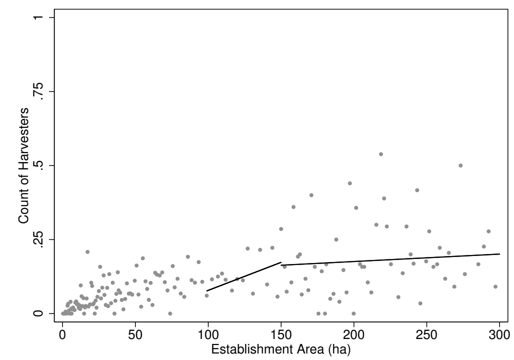
Figure 19: Binned Scatter Plots and Local Linear Estimates for Mechanical Harvesting Inputs



(a) Contracting Expenditure

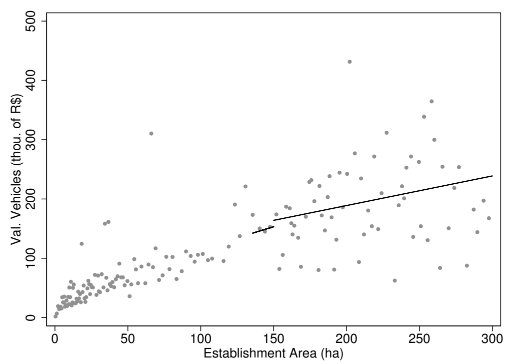


(b) Fuel Expenditure

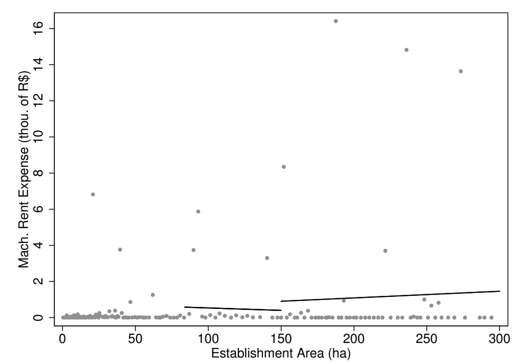


(c) Count of Harvesting Machines

(d) Value of Vehicles



(e) Machine Rental Expenditure



Data from 2006 Agricultural Census. Local linear estimates are plotted for $x \in [150 - h_-, \min\{300, 150 + h_+\}]$. Bin sizes set to mimic the variance of the underlying data (see Calonico et al. (2015)).

Table 4: The Effect of Regulation on Input Use (RD)

Input	$\hat{\tau}$	95% CI	$\bar{Y}$	h_-	h_+	N_-	N_+
Temp. days paid	146	[-167; 305]	197	51	1,214	1,047	2,590
Temp. employ.	7.25	[-2.03; 15.7]	3.6	43	443	869	2,097
Tot. employ.	7.77	[1.23; 21]	5.22	34	340	707	1,946
Cont. Exp.	2.63	[-7.09; 13.9]	14.9	45	466	906	2,139
Fuel Exp.	-10.3	[-12.7; 8.9]	.	67	57	978	429
Harvesters	-.00976	[-.149; .0911]	.118	51	392	1,040	2,031
Mach. Rent Exp.	506	[-2,827; 2,396]	491	66	269	1,555	1,767
Val. Vehicles	10.7	[-63.9; 172]	132	14	556	235	2,239

A.4 Difference-in-Difference Estimates Using the Size Threshold

The area threshold in the São Paulo regulation also enables me to estimate the effect of regulation via difference-in-differences. Recall that São Paulo was the only regulated state at the time of the 2006 Census of Agriculture and that establishments below 150 hectares were exempted from the regulation. I compare the difference between São Paulo establishments above and below the 150 hectare threshold to the same difference among non-São Paulo establishments. Specifically, I estimate:

$$\begin{aligned} Y_i = &\beta_0 + \beta_1 \mathbb{1}(\mathrm{T}_i > 150 \text{ ha}) + \beta_2 \mathbb{1}(\mathrm{State}_i = \text{ SP}) \\ &+ \beta_3 \big[\mathbb{1}(\mathrm{T}_i > 150 \text{ ha}) \times \mathbb{1}(\mathrm{State}_i = \text{ SP})\big] + \varepsilon_i \end{aligned} \tag{23}$$

where β_3 is the effect of the regulation. The unit of observation is an agricultural establishment, indexed by i. The outcome Y_i may be one of three variables: indicators for i) establishments that use manual harvesting only, ii) establishments that use mechanical harvesting only, or iii) establishments that use both. In the reported specifications, I add state fixed effects and a polynomial in area.

Note that this identification strategy yields a different parameter than the RD strategy in Section 2.3.2 even though they both use the same 150 hectare threshold. The difference-in-differences estimate can be interpreted as an average treatment effect on the treated so, in principle, it may be more sensitive to changes away from 150 hectare threshold (see, e.g., Athey & Imbens (2006)). That said, most establishments are near or below the threshold so this procedure may also fail to detect an effect on the small number of large establishments.

The estimates, shown in Tables 5, 6, and 7, argue that regulation encouraged growers to move towards mechanical harvesting. The estimates of β_3 indicate the average effect of regulation on harvesting practices for regulation establishments. Regulated establishments are about 10 percentage points less likely to use manually only. Regulated establishments are 3 percentage points more likely to use mechanical only and 7 percentage points more likely to use both techniques. The estimated effect of the regulation is significant in all specifications.

While these effects seem large compared to the means, they are within the confidence intervals of the RD estimates in Section 2.3.2 and they are small compared to the intended effect of the regulation and the eventual outcome of complete mechanization. Recall that, above the threshold, more than 60 percent of establishments report manual harvesting only. According to the regulation, none of them should rely exclusively on manual harvesting. Moreover, establishments over 150 hectares are almost completely mechanized by 2014. If the regulation causes 10 percent of establishments to use some mechanized harvesting, then

Table 5: Likelihood of Harvesting Only Manually

	(1)	(2)	(3)	(4)
Area > 150ha (β_1)	0.084***	0.098***	-0.044	-0.044***
	0.020	0.030	0.037	0.010
São Paulo (β_2)	-0.188	-0.078	-0.090	-0.090***
	0.021	0.061	0.063	0.005
Regulated (β_3)	-0.104***	-0.134***	-0.159***	-0.159***
	0.024	0.041	0.041	0.013
$\bar{Y}$	0.781	0.781	0.781	0.781
σ_Y	0.413	0.413	0.413	0.413
Clust. SE	Y	Y	Y	
Area poly	Y	Y		
State FE	Y			
N	30,423	30,423	30,423	30,423

Agricultural establishments are the unit of observation. Removed observations in the top and bottom 1% by area. The outcome variable is an indicator for Harvesting Only Manually. Assignment variable is Establishment Area. In the indicated columns, SE estimates are clustered by county.

Table 6: Likelihood of Harvesting Only Mechanically

	(1)	(2)	(3)	(4)
Area > 150ha (β_1)	-0.019**	-0.021**	0.019	0.019***
	0.009	0.009	0.014	0.006
São Paulo (β_2)	0.063	0.043**	0.047***	0.047***
	0.014	0.017	0.018	0.003
Regulated (β_3)	0.030*	0.034*	0.040**	0.040***
	0.016	0.018	0.018	0.008
$\bar{Y}$	0.067	0.067	0.067	0.067
σ_Y	0.249	0.249	0.249	0.249
Clust. SE	Y	Y	Y	
Area poly	Y	Y		
State FE	Y			
N	30,423	30,423	30,423	30,423

Agricultural establishments are the unit of observation. Removed observations in the top and bottom 1% by area. The outcome variable is an indicator for Harvesting Only Mechanically. Assignment variable is Establishment Area. In the indicated columns, SE estimates are clustered by county.

Table 7: Likelihood of Harvesting Manually and Mechanically

	(1)	(2)	(3)	(4)
Area > 150ha (β_1)	-0.064***	-0.077***	0.025	0.025***
	0.016	0.025	0.029	0.008
São Paulo (β_2)	0.125	0.035	0.042	0.042***
	0.015	0.050	0.051	0.004
Regulated (β_3)	0.074***	0.100***	0.119***	0.119***
	0.019	0.032	0.032	0.011
$\bar{Y}$	0.152	0.152	0.152	0.152
σ_Y	0.359	0.359	0.359	0.359
Clust. SE	Y	Y	Y	
Area poly	Y	Y		
State FE	Y			
N	30,423	30,423	30,423	30,423

Agricultural establishments are the unit of observation. Removed observations in the top and bottom 1% by area. The outcome variable is an indicator for Harvesting Manually and Mechanically. Assignment variable is Establishment Area. In the indicated columns, SE estimates are clustered by county.

the regulation is certainly insufficient to explain the change in harvesting techniques.