Close to the Edge:
High Productivity at Plot Peripheries and the Inverse Size-Productivity Relationship

June 26, 2018

Smaller farms and plots are more productive per hectare than larger ones in most developing country data. Using unique, plot-level panel data from Uganda, we estimate the inverse size-productivity relationship and generate two important findings. First, the standard inverse relationship is a plot-level phenomenon. This renders insufficient the conventional household- or farm-level explanations related to market failure. The inverse relationship is also independent of soil quality and error in farmer-estimated plot size, eliminating the other dominant explanations in the literature. Second, the plot perimeter/area ratio, reflecting an "edge effect" discussed in the agronomy literature wherein productivity is highest around the periphery of plots, explains most or all of the inverse plot size-productivity relationship. We present suggestive evidence consistent with behavioral and biophysical mechanisms underpinning the edge effect.

Keywords: inverse relationship, productivity, behavioral, causal bounds, perceptions, edge effect

JEL: O12, O13, Q12, D91

Acknowledgements: Thanks to Clark Gray, PI on the NSF-funded survey (BCS-1226817) that collected our 2013 data, and co-PIs Ephraim Nkonya, Darrell Shultze, Kayuki Kaizzi, and Leah VanWay. Thanks to the Kampala IFPRI and HarvestPlus offices for hosting the survey, and to the entire group of surveyors. Thanks to Devon Jenkins, Rachel Hestrin, Maia Call, David Just, Kalle Hirvonen, Marc Bellemare, Kurt Lavetti, Adan Murillo, and Rachid Laajaj for helpful comments. All errors are solely our responsibility.

It has long been observed that smaller farms produce more per unit area than larger farms, ceteris paribus, across a number of developing and non-developing settings, at least over the dominant range of developing country farm sizes.[1] This has been observed in Africa (Collier, 1983; Barrett, 1996; Kimhi, 2006; Barrett, Bellemare and Hou, 2010; Larson et al., 2014), in Asia (Sen, 1962; Mazumdar, 1965; Bardhan, 1973; Carter, 1984; Heltberg, 1998; Benjamin and Brandt, 2002; Rios and Shively, 2005), in Europe (Alvarez and Arias, 2004) and in Latin America (Berry and Cline, 1979; Kagin, Taylor and Yúnez-Naude, 2015), to cite only a few studies. This inverse size-productivity relationship has long attracted the attention of development and agricultural economists because it suggests unequal marginal factor productivity across farms of different sizes, and hence inefficient resource allocation. This prospective inefficiency is important for development because small farms command 30-40 percent of farmland in the world's poorest countries (Lowder, Skoet and Raney, 2016), and support over half of the world's poor (HTF, 2003; Bezemer and Headey, 2008; Laborde Debucquet and Martin, 2018).[2]

So potential explanations for the inverse relationship have practical implications. For example, if small farms are inherently more efficient in a given setting, redistributive land reform might offer both equity and efficiency gains. If market failures drive the inverse relationship, then government interventions may be necessary. Conversely, if the inverse relationship is purely a statistical artifact attributable to measurement error or to omitted relevant variables, then the rural economy may function reasonably efficiently and interventions are more likely to do harm than good.

Using plot-level panel data from rural Uganda — to our knowledge, for the first time in this literature — we demonstrate that the inverse relationship exists at the plot level within farms, rather than at the farm level. Familiar explanations (discussed below) fail to explain the observed relationship, and so we turn to a new mechanism: if an "edge effect" repeatedly identified in the agronomy literature leads to higher marginal productivity on plot peripheries than in plot interiors, then smaller plots should be more productive than larger plots, as a larger share of small plot area falls along the perimeter. We find that the perimeter/area ratio indeed explains most or all of the observed inverse relationship at plot level, suggesting that the edge effect drives the inverse relationship in our data. While we are unable to identify whether behavioral or biophysical factors, or both, drive the edge effect, farmers do apply more labor around the more visible and accessible plot edges. Further evidence regarding misperceptions of plot size indicate that farmers' perceptions can indeed shape behavior and productivity.

Through 100 years of literature on the inverse relationship, two classes of mechanisms were commonly hypothesized. Under the first, multiple market failures cause large farms to face higher shadow factor costs than smaller farms, especially for labor. Consequently small farming families with a low opportunity cost of labor, no problems with worker supervision, and a concern for their own food security apply labor in large quantities

[1] Foster and Rosenzweig (2017) raise the good point that at some point in the upper range of the farm size distribution in developing countries, one sometimes finds productivity increases with farm size, generating a U-shaped relationship over the full range of the farm size distribution. This observation raises crucial questions that their paper tackles. It leaves unanswered the fundamental questions about factor allocation among small farms, on which most of the literature focuses, as does this paper.

[2] Laborde Debucquet and Martin calculate that 76 percent of the world's poor live in rural areas, and 54 percent are smallholder farmers.

per unit land to achieve remarkably high yields (Chayanov, 1991; Sen, 1966; Feder, 1985; Barrett, 1996).[3] By contrast, a second thread of literature explained the inverse relationship as illusory, a statistical artifact resulting from either measurement error around land size or omitted variable bias due to unobserved inputs such as soil fertility (Bhalla and Roy, 1988; Benjamin, 1995; Lamb, 2003; Assunção and Braido, 2007).

However, a wave of new evidence stemming from rich, often plot-level or panel datasets has largely invalidated the first class of mechanisms, and transformed the emphasis of the second. The inverse relationship between farm size and farm productivity persists in panel data, even when household fixed effects control for the time-invariant component of household-specific shadow prices (Henderson, 2015; Kagin, Taylor and Yúnez-Naude, 2015). Of course, this might occur if households shift participation in factor markets over time, in coordination with their farm size (Carter and Yao, 2002). But when household fixed effects are applied to plot-level data the inverse relationship remains, sometimes stronger at the plot than the farm level (Assunção and Braido, 2007; Barrett, Bellemare and Hou, 2010; Ali and Deininger, 2015). Taken together, it seems unlikely that household-specific shadow factor prices are driving the inverse relationship.

Neither omitted variable bias nor measurement around land size seems to explain the puzzle either. Including biochemical and physical soil quality controls in a plot-level dataset from Madagascar does not mitigate the inverse relationship (Barrett, Bellemare and Hou, 2010). Similarly, the inverse relationship actually increases in magnitude when farm size is based on GPS measurement rather than error-ridden farmer estimates. This is because African smallholders on very small farms tend to over-report plot size, while larger farms tend to under-report plot size (Carletto, Savastano and Zezza, 2013; Carletto, Gourlay and Winters, 2015).

While measurement error around farmer-estimated *land size* mitigates the inverse relationship, two new papers hypothesize that measurement error in farmer-recalled *production* estimates drive it. Using data from Eastern Uganda, Gourlay, Kilic and Lobell (2017) estimate an inverse relationship using conventional, farmer-recalled measures of crop production, but find no inverse relationship when yield is measured via sample crop cutting. Desiere and Jolliffe (2017) find the same in Ethiopia. In both papers, the difference between recalled yields and cut-calculated yields is inversely related to (GPS-measured) plot size. If sample cut-calculated yields are closer to "true" yields than are farmer-estimated yields, the inverse relationship may be driven by systematic patterns in farmer recall error. If, however, sampled crop cuts over-estimate yield on larger plots and under-estimate yields on smaller plots, then the disappearance rather than existence of the inverse relationship may be the statistical artifact.

We explore an entirely new mechanism, the "edge effect," and test it in plot-level, geospatially-matched panel data from rural Uganda.[4] Agronomy literature documents the existence of this edge effect, wherein the peripheral rows of a plot are more productive than interior rows (Little and Hills, 1978; Barchia and Cooper, 1996). For instance, peripheral rows of corn yielded over 30 percent more than center rows in

[3] Chayanov first observed the inverse relationship in Russian agriculture, publishing "Osnovnye idei i formy organizatsii sel'skokhozyaistvennoi kooperatsii" (The basic ideas and organizational forms of agricultural cooperation) in Moscow in 1921. The english translation was published in 1991.

[4] To the best of our knowledge, this is the first paper to use plot-level panel data for this purpose.

Argentina and the US (Verdelli, Acciaresi and Leguizamón, 2012; Ward, Roe and Batte, 2016). Cotton grown on plot edges yielded 360 percent more than cotton grown in the interior (Holman and Bednarz, 2001). Agronomic experiments suggest that plot peripheries experience higher yields due to increased sunlight exposure (Barchia and Cooper, 1996), differences in pests, biodiversity or pollination (Balagawi, Jackson and Clarke, 2014), greater nutrient uptake due to reduced competition (Watson and French, 1971) and/or greater water availability (O'Brien and Green, 1974). There might also be behavioral mechanisms underpinning the effect — plot peripheries may be more accessible to a farmer in a way that changes her management of the area, and weed growth, pest infestation or plant disease are likely more visible at the edges of a plot.

Our contribution to the inverse relationship literature is thus threefold. First, we comprehensively examine and reject previous mechanisms. The inverse relationship is at the plot rather than household level, is not driven by measurement error in land size, and is not driven by soil fertility or other, generally omitted variables. Second, we address (lack of) causality directly by estimating all results under two identification strategies, and calculating bounds around the probable, causal effect of plot size on productivity (Oster, 2016). Our results are close to identical under plot fixed effects (which hold fixed time-invariant plot characteristics like landscape or distance to household, and identify off of substantial changes in plot size, shape and yield over time) and under household-time fixed effects (which hold fixed household shadow price for each time period, and exploit within-household variation in plot characteristics).[5] Furthermore, Oster's (2016) causal bounds fit so tightly around the original estimate that under any plausible assumption, the un-estimable "causal" inverse relationship is not attenuated by omitted, relevant variables.

Third, and most importantly, we explore the edge effect as a possible explanation for the inverse relationship. Using precise plot perimeters constructed via GPS waypoints, we calculate the ratio of perimeter to area for each plot, essentially capturing the proportion of plot area that falls within the "edge." Not only does perimeter-area ratio better explain plot yields than area alone — suggesting that the inverse relationship may be thought of as misspecification error – but the inverse relationship is completely mitigated in our preferred specification by controlling for the perimeter-area ratio. We subject the edge effect to a range of robustness and placebo tests. We also show that labor intensity, which declines with plot size as do yields, also displays the edge effect, suggesting that a behavioral mechanism may be at play. Unfortunately, we cannot test for error in farmer-recalled production because we do not have crop cuts.[6]

1 Data

We use plot-level panel data from rural Uganda. The first wave of data was collected during the summer of 2003, by the International Food Policy Research Institute

[5]The similarity of these estimates suggest that neither endogeneity of plot size change (potentially biasing coefficient estimates only under plot fixed effects) nor plot-specific shadow prices (potentially biasing coefficient estimates only under household-time fixed effects) are driving the inverse relationship.

[6]Similarly, Desiere and Jolliffe (2017) cannot test the edge effect because crop cuts are gathered only within plot interiors. Gourlay, Kilic and Lobell (2017) may be able to do so with their full crop cut data, but current drafts do not show this analysis.

(IFPRI). This IFPRI survey was run in conjunction with a larger Uganda Bureau of Statistics (UBOS) survey conducted in 2002/2003. Together, the surveys collected household-level socioeconomic data, plot-level input and production data, and took plot-level soil samples for later soil analysis. Additionally, farmers estimated the size of each of their plots, and plot perimeter, plot size and plot centroid were measured via GPS. (See Appendix 1 for details on plot size calculations.) Information on the sampling strategy used in 2003 can be found in Nkonya et al. (2008).[7]

The second wave of data was collected during the summer of 2013 under a National Science Foundation (NSF) funded project. The same household- and plot-level data were collected, along with plot-level soil samples. Of the 859 households interviewed in 2003, 803 were tracked successfully and re-interviewed. Additionally, individuals who had split off from the original 2003 household to form a new household were tracked if they were still within the original parish. Appendix 2 examines attrition — households that attritted tended to live in peri-urban areas and on average were slightly younger, slightly smaller, slightly more educated and had slightly less land and fewer animals.

In each wave, soil samples were aggregated from 12-20 subsamples (based on plot size) taken in a zig-zag pattern across each plot. Samples were then analyzed for a number of biophysical and chemical characteristics at the National Agricultural Research Laboratory in Uganda using well-established protocols. Details on the soil sampling strategy as well as soil analysis can be found in Appendix 3.

In this paper, the unit of analysis is a plot of land, used to grow a single crop or multiple, mixed crops. Most, though not all, farmers have multiple plots in both 2003 and 2013.[8] While the size and shape of plots shifts substantially across the decade, we match plots across time using their geospatial location within a land parcel.[9] Physically overlapping plot polygons, drawn via waypoints at all vertices, are defined as a match. This is how we form a plot-level panel dataset — our primary dataset. Additionally, we pool all plots from both rounds into our second, pooled dataset used for analysis without plot fixed effects.[10] More details on geospatial matching are found in Appendix 1.

Change in plot size and shape over time is due almost entirely to re-organization of plots within larger, unchanged land parcels. Rotation, fallowing, and mixed cropping is common in Uganda, leading farmers to re-organize plots within farms almost every season (twice a year). This results in substantial changes in farm lay-out over a decade. No plots change ownership (such plots are dropped from our datasets), and only 8

[7]Essentially, rural households were randomly chosen within survey districts, but the survey districts themselves were chosen to represent various agro-ecological zones across Uganda. Thus, the results in this paper cannot be viewed as representative across Uganda.

[8]In 2003 and 2013, 16% and 32% percent of households, respectively had only 1 plot. Thus, 5.3% and 14.0% of plots belonged to a single-plot household in 2003 and 2013, respectively.

[9]While a plot is defined by cropping system, a parcel is a contiguous piece of land under one form of ownership, holding one or multiple plots. Fifty (57) percent of our 2003 (2013) plots come from a parcel that holds only that single plot. The average plot in 2003 (2013) is on a parcel that holds 2 (1.8) plots.

[10]Of the 2,549 plots recorded in the 2003 survey, 24 percent (605) were geospatially matched. Of the 1,656 plots recorded in the 2013 survey, 42 percent (700) were geospatially matched. We therefore have 700 plots in our plot panel dataset, each viewed in 2-4 agricultural seasons (but necessarily at least 1 agricultural season per round), totaling 2,076 observations. We have 4,205 plots in our pooled dataset, each viewed in 1-2 agricultural seasons within their respective round, totaling at 6,509 observations.

percent of plots sit on a land parcel that was sub-divided. Only 2.5 percent of plot size change is predicted by the sale or acquisition of land at the household level (Table A13). Instead, the greatest predictor of plot size change is 2003 plot size; small plots are more likely to grow (as 30 percent do) and large plots more likely to shrink (as 70 percent do), in a regression to the mean (Figure A4). Yet 2003 plot size itself is correlated with farm- and plot-level characteristics, and so change in plot size is not exogenously determined. Appendix 5 further examines the predictors of plot size change.

Because most of Uganda has two agricultural seasons, agricultural input and production data were gathered for both seasons. We therefore consider four time periods, one for each agricultural season of each round. Plot size and shape do not vary across season within round, but other plot-level characteristics such as as fertilizer use or management practices do. Plot productivity (revenue per hectare) also varies across all four time periods. To be included in the plot-level panel, a plot must be viewed in each round. Not all plots were farmed in both seasons, however. Our panel is therefore unbalanced, with plots viewed 2-4 times, depending on how many seasons they were farmed.

Table 1 summarizes key variables used in analysis in our plot-level panel dataset. (Appendix 4 summarizes the same for our pooled dataset.) Both plots and farms are shrinking over time, and at a similar rate — in 2013 the median area for either unit is about 60 percent of the median area in 2003. We measure productivity in terms of real revenue per hectare, as common in this literature, because it is difficult to evaluate the market price of the primary inputs (labor and organic matter).[11] Plots are far more productive in 2013 than in 2003, and labor intensity is higher. Soil became slightly more acidic over the decade, while organic carbon content slightly increased.[12]

Inputs, management and cropping systems also shifted slightly over the decade. Organic amendment (manure, crop residue, food residue or compost) is less likely to be applied in 2013 than in 2003. Terracing is less commonly practiced in 2013 while crop rotation is more commonly practiced. In both years the use of inorganic fertilizer is negligible, as is the use of irrigation; less than 2 percent of plots benefit from either practice in either year. Household heads appear to own and manage plots at slightly higher rates in 2013. In both years, about 50 percent of plots are under a mixed cropping system, i.e., hold multiple crops. Intercropping, defined more strictly as alternating rows of different crops, rises in prevalence between 2003 and 2013. Yet the number of plots holding each crop category (tubers, legumes, bananas, cereals and cash crops) declines between 2003 and 2013, due to a decline in the average number of crops grown per household (see Tables A3 and A5) as well as a decline in the average number of crops grown per plot.[13]

There is, of course, a selection process into the geospatially matched panel dataset. Some plots cannot be matched over time, as they have no geospatial overlap with another plot from across the decade. We drop these plots from our analysis. We also

[11] Ugandan Bureau of Statistics consumer price indices, base year 2005, are used to deflate revenue.

[12] The change in organic carbon content may be due to a change in the buffer pH within the Walkley-Black procedure used to recover organic carbon in both years. Because round fixed effects are used in all analysis, this mean shift should have no consequence for results.

[13] We do not observe fallow plots in either round, and so cannot examine changes in fallowing. Additionally, we can be sure that fallowing does not drive plot panel results; a 2003 plot fallowed in 2013 would simply disappear from the panel.

drop a handful of plots that changed household ownership over the decade. Some 2003 plots are matched to multiple 2013 plots, generally because they have been split up into smaller plots over the decade.[14] Appendix 2 shows that this selection is not random; while households selected into the panel dataset tend to be representative of the larger universe of houses, plots in the panel dataset are are larger, receive less labor per hectare, are more likely to grow bananas or cash crops, are less likely to be rotated and are more likely to receive organic amendment. These plots cannot, therefore, be viewed as representative of the larger universe of plots in our pooled data.

We also estimate all core results under household-year-season fixed effects (Appendix 4), where the identifying variation is across-plot, within-household change in plot size. This strategy was also employed by Assunção and Braido (2007). Plots within the pooled data used for this analysis *can* are representative of the larger universe of plots (Appendix 2).[15] If coefficients estimated under plot fixed effects are similar to those estimated household-time fixed effects, selection seems unlikely to be driving our results.

2 Estimation Strategy

2.1 Shadow Prices (Plot vs. Household Mechanisms)

We first estimate the inverse size-productivity ratio according to farm size, plot size, and under various fixed effects models, in order to investigate whether household-level shadow prices drive the inverse relationship, as much of the literature has long held. Let Y_{ijt} be the productivity of plot j within farm i in time period t, where productivity is defined as revenue per hectare and t takes a unique value for each year-season combination.[16] Plot area is given by A_{ijt}, and farm size is given by A_{it}.[17]

The inverse relationship can be estimated including only A_{ijt}, including only A_{it}, or including both area measures. The inverse relationship will appear as a negative and statistically significant coefficient estimate on farm size, plot size, or both. These relationships can be estimated using simple Ordinary Least Squares (OLS) with no fixed effects as in Equation 1. The relationship with plot size can additionally be estimated by including household-time fixed effects λ_{it} as in Equation 2. (Farm size can no longer be included, as it only varies by household and time.) And both relationships can be estimated by including plot fixed effects λ_{ij} and a time fixed effect λ_t as in Equation 3.

$$Y_{ijt} = \delta_1 A_{it} + \gamma_1 A_{ijt} + \varepsilon_{ijt} \tag{1}$$

$$Y_{ijt} = \gamma_2 A_{ijt} + \lambda_{it} + \varepsilon_{ijt} \tag{2}$$

[14] Of the 631 2003 plots that were geospatially matched to a 2013 plot, 70 percent were matched to exactly 1 plot from 2013, 20 percent were matched to 2 plots, 7 percent to 3 plots, 2 percent to 4 plots, and an additional 1 percent to 5-9 plots. In only 16 cases did a plot change ownership, and we drop these plots from all analysis. More information can be found in Appendix 1.

[15] This is despite the fact that households with only one plot are dropped from household-time fixed effect analysis. Household characteristics do change slightly in this smaller dataset, relative to the larger dataset, but time-invariant household characteristics are partialled out in both types of analysis.

[16] Because most plots contain multiple crops, productivity cannot be measured solely as physical yields.

[17] At this point, let plot area be measured by GPS. In the next sub-section GPS measurement is compared to farmer-recalled plot size. Farm area is given by aggregated plot area, using GPS-measured plot size when available and farmer-recalled plot size for those plots that were not visited by an enumerator.

$$Y_{ijt} = \delta_3 A_{it} + \gamma_3 A_{ijt} + \lambda_{ij} + \lambda_t + \varepsilon_{ijt} \tag{3}$$

If $\hat{\gamma}_1$ is significant and negative once A_{it} is controlled for in Equation 1, the inverse relationship must stem at least in part from phenomenon at the plot, rather than household, level. If household-level shadow prices drive the inverse relationship, then the plot-level relationship should disappear once household-time fixed effects are controlled for in Equation 2, or plot fixed effects are controlled for in Equation 3.

Because we find that the inverse relationship stems solely from plot-level, rather than household-level phenomenon (i.e., once A_{ijt} is controlled for, controlling for A_{it} offers no additional, statistically significant information), all future equations estimate γ_3 using plot and time fixed effects, as in Equation 2, but excluding A_{it}. This identification strategy, new to the inverse relationship literature, differences away all time-invariant, plot-specific shadow prices and/or characteristics that might drive the inverse relationship, related to features such as distance from the homestead, slope, or position on the toposequence. Only unobserved plot-level characteristics that vary across time with plot size may bias the estimated relationship in Equation 2.

We also, however, estimate all equations using household-time fixed effects as in Equation 2; these results appear in Appendix 4. This identification strategy controls for household-and-season-specific shadow prices and other household-level factors, both time invariant and time-varying, but allows for within-season plot-level unobservables to bias the relationship. Because these strategies identify the inverse relationship off different sources of variation, if the $\hat{\gamma}_2$ and $\hat{\gamma}_3$ estimates are statistically identical, it suggests that neither household-level nor plot-level factors are biasing the relationship — rather, plot size itself drives plot-level productivity. Otherwise, two different sources of bias must coincidentally result in the same parameter estimate.

2.2 Measurement Error

We then investigate how measurement error around plot size influences the estimated inverse relationship between plot size and productivity. Let Y^m_{ijt} be the productivity of plot j within farm i in time period t, where method m was used to measure the size of plot i. Method m may be either size reported by farmer or size measured via GPS. Similarly, let A^m_{ijt} be the area of plot i within household j, measured by method m.

$$Y^m_{ijt} = \gamma_4 A^m_{ijt} + \lambda_{ij} + \lambda_t + \varepsilon_{ijt} \tag{4}$$

If the inverse relationship is partly a statistical artifact driven by measurement error, then $\hat{\gamma}_4$ estimated under Equation 4 should be smaller in absolute value and R^2 should be smaller with the GPS plot size measure as compared to farmer recall. Alternatively, measurement error in farmer reporting of plot size might attenuate the $\hat{\gamma}_4$ estimate, as found by Carletto, Savastano and Zezza (2013).

2.3 Omitted Variables and Causality

We next consider the role of omitted variable bias. A number of plot characteristics shift between 2003 and 2013 (Table 1). If change in these characteristics (e.g., change in soil quality, crops grown) are correlated with changes in plot size and plot productivity, failing to control for them might cause a spuriously estimated inverse relationship.[18]

Equation 5 therefore controls for an exhaustive list of time-varying plot characteristics X_{ijt}, alongside plot size A_{ijt} from Equation 4, under the optimal plot size measure m.

$$Y_{ijt} = \gamma_5 A_{ijt} + \beta X_{ijt} + \lambda_{ij} + \lambda_t + \varepsilon_{ijt} \tag{5}$$

We first allow X_{ijt} to encompass a set of plot-specific soil fertility indicators, in order to explore whether omitting soil quality from Equation 4 drives a spurious correlation between plot size and plot productivity, as hypothesized in the literature. We then allow X_{ijt} to include time-varying plot characteristics relevant to plot productivity: agricultural inputs, plot ownership and management practices, and crops grown.

If the inverse relationship is robust to these controls, i.e., the coefficient estimate $\hat{\gamma}_5$ is stable with the introduction of relevant, but often omitted, variables X_{ijt}, then it is possible that the inverse relationship is at least partly causal, rather than reflecting omitted variable bias. Without further restrictive assumptions, however, it is impossible to quantify the likelihood of such causality.

Oster (2016), Krauth (2016) and Altonji, Elder and Taber (2005) develop a set of econometric techniques designed for this very purpose: to bound the causal effect of an endogenous explanatory variable under the threat of omitted relevant variable bias. The method uses restrictive though plausible assumptions regarding the relative correlations between a potentially endogenous explanatory variable (in our case, plot size) and relevant observables, and that variable and unobservables. We use the consistent estimator of bias derived by Oster (2016) to bound the causal effect of A_{ijt} on Y_{ijt}.[19]

Consider the data generating process $Y = \beta X + \psi w_1 + W_2 + \varepsilon$, where β gives the causal effect of the treatment variable X on the outcome Y, w_1 is an observable set of variables, and W_2 and the error ε are unobservable. Regressing Y on X alone results in the biased coefficient $\mathring{\beta}$ and R-squared $\mathring{R}$. Regressing Y on X and w_1 results in the (less) biased coefficient $\tilde{\beta}$ and R-squared $\tilde{R}$. The R-squared from a hypothetical but impossible regression of Y on X, w_1 and W_2 would result in R_{max}, a number which is less than 1 if measurement error or other factors prohibit the full explanation of Y.

Oster (2016) proves that with one key assumption,[20] the bias-adjusted coefficient β^* can be approximated as below, and that β^* converges in probability to the true, causal coefficient β.[21] The parameter δ gives the relative proportion of X explained by unobservables vs. observables — so δ is always > 0, and if δ is $<$ $(=)$$[>]$ 1, then X is

[18]Similarly, if plot characteristics are correlated with size and productivity across plots within time, omitting them might bias estimation under the household-time fixed effects specification.

[19]This technique will not bound the causal effect if bias is due to non-classic measurement error in Y_{ijt}. It deals strictly with omitted variable bias, with adjustment for classic measurement error only.

[20]The relative contribution of each variable within w_1 to X must be the same as the relative contribution of each variable within w_1 to Y. While unlikely to hold unless w_1 is a single variable, Oster (2016) notes that as long as deviations from this condition are not "extremely large," the calculated estimator will still provide an approximation of the consistent estimator.

[21]Under a second assumption of proportional selection, β^* can be exactly calculated under $\delta = 1$. Unwilling to make this restrictive assumption, we consider a range of δ values as well as R_{max} values.

more (equally) [less] influenced by observables than by unobservables.

$$\beta^*(R_{max}, \delta) = \tilde{\beta} - \delta\Big[\mathring{\beta} - \tilde{\beta}\Big]\frac{R_{max} - \tilde{R}}{\tilde{R} - \mathring{R}}$$

Equivalently, we calculate the bias-adjusted inverse relationship γ^* as in Equation 6, where R_4 is the R-squared obtained by estimating the univariate inverse relationship of Equation 4, and R_5 is the R-squared obtained by estimating the inverse relationship with a full set of controls as in Equation 5.

$$\gamma^*(R_{max}, \delta) = \hat{\gamma_5} - \delta\Big[\hat{\gamma_4} - \hat{\gamma_5}\Big]\frac{R_{max} - R_5}{R_5 - R_4} \tag{6}$$

Contingent on aforementioned assumptions, the causal effect of X on Y will lie within the interval $\big[\tilde{\beta}, \beta^*(R_{max}, \delta)\big]$ and Oster (2016) suggests that in most situations the causal effect will lie within the bounds of $\big[\tilde{\beta}, \beta^*(min\{1.3\tilde{R}, 1\}, 1)\big]$. We calculate an equivalent bounding interval for the inverse relationship, $\big[\hat{\gamma_5}, \gamma^*(min\{1.3R_5, 1\}, 1)\big]$, and additionally calculate bounding intervals under a range of R_{max} and δ parameters.

2.4 The Edge Effect

We next test a previously unconsidered explanation for the inverse relationship. We allow the productivity of plot j belonging to household i in time period t to be given by a combination of the productivity of the plot's interior, Y^I_{ijt}, and the productivity of the plot's periphery, Y^P_{ijt}, as suggested by the agronomy literature and shown in Equation 7. Productivity is weighted by the area of the plot's interior, A^I_{ijt} and the area of the plot's periphery, A^P_{ijt}, and the sum of these two areas gives the total area of the plot, A_{ijt}.

$$Y_{ijt} \equiv \frac{Y^I_{ijt} * A^I_{ijt} + Y^P_{ijt} * A^P_{ijt}}{A_{ijt}} \tag{7}$$

By re-arranging terms, Equation 7 can be re-written as in Equation 8.

$$Y_{ijt} \equiv \frac{Y^I_{ijt} * (A_{ijt} - A^P_{ijt}) + Y^P_{ijt} * A^P_{ijt}}{A_{ijt}} = Y^I_{ijt} + (Y^P_{ijt} - Y^I_{ijt}) * \frac{A^P_{ijt}}{A_{ijt}} \tag{8}$$

This last functional form suggests that plot productivity Y_{ijt} is an additive function of the productivity of the interior, Y^I_{ijt}, and the ratio of the plot's peripheral area A^P_{ijt} to the plot's total area A_{ijt}. However, while we view A_{ijt}, we do not view A^P_{ijt}, as we do not know the width of the peripheral area. Calculating A^P_{ijt}/A_{ijt} is therefore impossible.

We do, however, view the plot's GPS-measured perimeter, P_{ijt}. If A^P_{ijt} is roughly equivalent to $P_{ijt} * b$, where b is the width of the peripheral area, then we obtain Equation 9.[22] Figure 1 provides a schematic visual for this assumption.

$$Y_{ijt} \approx Y^I_{ijt} + (Y^P_{ijt} - Y^I_{ijt}) * b * \frac{P_{ijt}}{A_{ijt}} \tag{9}$$

[22]For intuition, we are basically assuming that the plot's periphery is thin enough to be rolled out from around the plot's perimeter as a long strip, or rectangle. This rectangle does not approximate the peripheral area if b is large relative plot length/width. But if b is small, e.g., one row of crops, then this rectangle approximates the peripheral area.

Equation 9 indicates that plot productivity should increase in P_{ijt}/A_{ijt}, given that b is a positive constant and we expect $(Y^P_{ijt} - Y^I_{ijt})$ to be positive. In this case, the inverse relationship might stem from misspecification of the true data-generating process behind average plot productivity, since plot area A_{ijt} will be inversely correlated with P_{ijt}/A_{ijt}. We can test this hypothesis by estimating Equation 10, in which γ_6 indicates the classic inverse relationship, and $\theta_1 = (Y^P_{ijt} - Y^I_{ijt}) * b$. If γ_6 becomes statistically insignificant and R^2 rises when we control for P_{ijt}/A_{ijt} in addition to A_{ijt}, then it would seem that the edge effect drives the inverse relationship.

$$Y_{ijt} = \gamma_6 A_{ijt} + \theta_1 \frac{P_{ijt}}{A_{ijt}} + \lambda_{ij} + \lambda_t + \varepsilon_{ijt} \tag{10}$$

While Equation 9 depends on the assumption that $A^P_{ijt} \approx P_{ijt} * b$ when b is small, for all plots of all shapes, this assumption can actually be quantified, and the result more rigorously shown for a variety of plot shapes. Appendix 6 contains such calculations for hypothetical circular, rectangular, and triangular plots. It additionally explores the possibility that b may not be small, relative to the total size of the plot. If b is not small, then we should find that plot productivity Y_{ijt} rises with P_{ijt}/A_{ijt} and also with A_{ijt}. This would be equivalent to finding that after controlling for P_{ijt}/A_{ijt} in Equation 10, the inverse size-productivity relationship reverses, such that $\hat{\gamma}_6 > 0$.

While Equation 10 will estimate $\hat{\theta}_1$ as an average treatment effect, we expect that the effect is truly heterogeneous; both $(Y^P_{ijt} - Y^I_{ijt})$ and b may change with adjacent plot characteristics, row organization, landscape, etc. The estimated coefficient must therefore be viewed as a context-specific average treatment effect produced by heterogeneous biophysical and behavioral mechanisms, on which more below.

2.5 Edge Effect Mechanisms

Because we find that the edge effect entirely explains the inverse relationship under plot fixed effects, we next turn to investigating the mechanisms behind the edge effect. Two broad categories of mechanisms appear plausible. First, peripheral productivity Y^P_{ijt} may be higher than interior productivity Y^I_{ijt} due to higher levels of unobserved biophysical inputs such as sunlight, water, nutrients, drainage, pest protection or other agronomic factors generally unobservable to the econometrician but clearly important to plant growth. Because we do not view biophysical inputs we cannot test for such mechanisms directly. Appendix 7 reports a number of indirect tests, all inconclusive.

The second category of edge effect mechanisms involves farmer behavior rather than biophysical inputs. It may be that Y^P_{ijt} is higher than Y^I_{ijt} because farmers tend the more-visible edges of their plots differently than they tend plot interiors. Farmers might weed plot edges more carefully, space crops differently around plot edges, or harvest crops more assiduously around plot edges, where a missed plant will most visible.[23]

If this is the case, we might expect plot labor intensity L_{ijt} to exhibit the same inverse relationship and edge effect as plot productivity. That is, we might expect a negative and significant $\hat{\gamma}_7$ if only A_{ijt} is included on the right hand side of Equation 11, but an

[23]Farmers might also tend edges more intensively to signal ownership, particularly if tenure status is questionable. We cannot investigate this mechanism due to a lack of variation in tenure status.

insignificant $\hat{\gamma}_7$ and a positive, significant $\hat{\theta}_2$ if P_{ijt}/A_{ijt} is also controlled for, suggesting that spatial patterns in labor allocation could be a mechanism behind the edge effect.

$$L_{ijt} = \gamma_7 A_{ijt} + \theta_2 \frac{P_{ijt}}{A_{ijt}} + \lambda_{ij} + \varepsilon_{ijt} \tag{11}$$

To further investigate the productivity implications of "plot awareness" we examine a second and purely behavioral mechanism, unrelated to biophysical constraints or inputs. We calculate plot size perception error as the difference between farmer-recalled and GPS-measured plot size: $e_{ijt} \equiv A^F_{ijt} - A^{GPS}_{ijt}$. If farmers apply inputs based on perceived plot size, then application rates and plot productivity should rise with perception error, signaling resource allocation inefficiencies due to farmer cognitive and behavioral error.

In Equation 12, $\mathbb{1}(e_{ijt} > 0)$ is a binary variable indicating whether the size of plot j was over-estimated by farmer i in time period t, e^O_{ijt} gives positive perception errors as a percentage of plot size (i.e., $e^O_{ijt} = e_{ijt}/A_{ijt}$ for all $e_{ijt} > 0$), and e^U_{ijt} gives negative perception errors as a percentage of plot size (i.e., $e^U_{ijt} = e_{ijt}/A_{ijt}$ for all $e_{ijt} < 0$). Thus, κ^O and κ^U give the marginal effects of over- and underestimating plot size, respectively.

$$Y_{ijt} = \gamma_8 A_{ijt} + \theta_3 \frac{P_{ijt}}{A_{ijt}} + \kappa^B \mathbb{1}(e_{ijt} > 0) + \kappa^O e^O_{ijt} + \kappa^U e^U_{ijt} + \lambda_{ij} + \varepsilon_{ijt} \tag{12}$$

If, conditional on A_{ijt} and $\frac{P_{ijt}}{A_{ijt}}$, e_{ijt} is randomly distributed, then κ^O and κ^U capture the causal effect of farmer misperceptions on per hectare productivity. We test the conditional exogeneity of e_{ijt}, and then estimate $\hat{\kappa}^O$ and $\hat{\kappa}^U$.

3 Results

3.1 Shadow Prices (Plot vs. Household Mechanisms)

Table 2 reports results for Equations 1-3. Panel 1 displays traditional, OLS estimates of the inverse relationship, as in Equation 1. Column 1 displays a significant, inverse relationship between farm size and plot productivity, while Column 2 displays a significant, inverse relationship between plot size and plot productivity. Explanatory power is greater in Column 2, however. When both farm size and plot size are controlled for in Column 3, the inverse relationship appears to exist only at the plot level.

In Panel 2 household-year-season fixed effects are introduced, controlling for household- and time-specific shadow prices as in Equation 2.[24] Variation in plot size is identified only within households within a season. (Estimating the association between farm size and productivity is therefore impossible.) The plot-level inverse relationship rises slightly, in comparison to estimates in Panel 1: a 10 percent increase in plot size is associated with a 5.6 percent decrease in plot productivity.[25]

In Panel 3 plot fixed effects are introduced alongside year and season fixed effects, as in Equation 3. Variation in plot size is therefore identified across time, stemming from the

[24] We define "time" from Section 2 as year-specific agricultural season (2 per year).

[25] This relationship is similar across years, though the estimated coefficient is larger in magnitude in 2013. See Table A21 in Appendix 8.

re-organization of plots within land parcels over a decade of agricultural seasons. Mean shifts in productivity or technologies across time are controlled for. The plot-level inverse relationship estimates in Columns 2 and 3 of this panel are slightly larger than those of Panel 2: a 10 percent increase in plot size is associated with a 6.2 percent decrease in plot productivity. As in Panel 1, farm size is superfluous once one controls for plot size. These results make it clear that the inverse relationship is a plot-level phenomenon, in these data at least, and not driven by inter-household heterogeneity in shadow prices or factor markets, as under the longstanding Chayanovian hypothesis.

The results of Table 2 are robust, as shown in Appendix 8. While 70 percent of plots grow over the decade (more information in Appendix 5), the inverse relationship exists for both shrinking and growing plots, and is unchanged in the sample of plots with a 1:1 match over the decade (Table A24).[26] The relationship is also qualitatively unchanged across functional forms, as shown in Table A25. For the remainder of this paper, main results are therefore estimated with plot fixed effects, controlling for year (2003 vs. 2013) and for season (1st vs. 2nd agricultural season) fixed effects.

All results may instead be estimated with household-year-season fixed effects. Appendix 4 reports these results, where the identifying variation is across plots, within household-year-season groups. In this case, households with only one plot in any given year-season time period are dropped (a selection process examined in Appendix 2). Explanatory power is lower when we exploit this dimension of variation. The coefficient estimates, however, are statistically (and qualitatively) the same as those estimated under plot fixed effects, with just two exceptions, both discussed in the appendix. We therefore confine discussion from here on to our preferred, plot fixed effects specification.

3.2 Measurement Error

Table 3 reports results for Equation 4, investigating whether measurement error around plot size drives or in fact mitigates the estimated inverse relationship. While the estimated relationship is statistically identical across measurement methods (GPS vs. farmer recall), R^2 is considerably higher for the GPS-measured variables in Column 2 than for the farmer-recalled variables in Column 1. This is consistent with the results of Carletto, Savastano and Zezza (2013), and counter to Lamb's (2003) hypothesis.[27]

As those authors noted, measurement error around plot size appears to weaken the relationship between plot size and productivity rather than strengthen it, at least in the Ugandan context. This is logical given that measurement error tends to be positive for smaller plots and negative for larger plots, both in our data and in the data examined by Carletto, Savastano and Zezza (2013). Figure 2 illustrates this relationship non-parametrically. The pattern makes smaller plots look less productive than they truly are, while large plots look more productive than they truly are.[28]

[26] Thirty percent of 2013 plots were matched to multiple 2003 plots (Appendix 1).

[27] Unlike the results by Carletto, Savastano and Zezza (2013), controlling for rounding in farmer-recalled plot size has no effect, and the coefficient estimate on a dummy for rounding is not significant.

[28] The bulk of plots fall within -3 and 1 on the x-axis of this figure.

3.3 Omitted Variables and Causality

If plot size (or change in plot size over the decade) was randomly distributed, we might interpret the coefficient on GPS-measured plot size in Table 3 as the causal effect of plot size on productivity. This is not the case. Plot size is correlated with observable plot characteristics, as established by regression-based balance tests in Appendix 5. In neither the pooled nor the panel setting, therefore, can plot size be considered exogenous, and omitted variable bias is a threat to causal identification.

If omitted variables drive the inverse relationship, as suggested by Lamb (2003) or Assunção and Braido (2007), we would expect the coefficient estimate on plot size to diminish in absolute value as relevant, observable controls are introduced (Oster, 2016). Table 4 introduces such controls, as specified in Equation 5, always controlling for plot fixed effects as well as year and season fixed effects. Column 1, beginning with no controls, is identical to Column 2 in Table 3 except for the loss of observations due to missing values of control variables needed in subsequent Columns.

Column 2 controls for soil characteristics. In Column 3, inputs such as labor hours, soil amendments and structures within the plot are added as controls.[29] The inverse relationship remains virtually identical in each of these specifications while the partial correlations between the controls and productivity are as one would expect. In Column 4 plot ownership and plot management is controlled for, and the inverse relationship becomes significantly larger in magnitude, as likewise occurs when crops are controlled for in Column 5. Column 6 controls for all variables simultaneously; the inverse relationship is statistically identical to the baseline estimate of Column 1.

While we control for management and crops by including them in Columns 4 and 5, one might wonder if instead the inverse relationship should be separately estimated across crop and management categories. Appendix 8 reports these results as a robustness check (Tables A22, A23); the ratio does not change significantly across either category.

This stability of the estimated inverse relationship coefficient in Table 4 is remarkable, given the richness of these time-varying control variables and the fact that plot fixed effects control for all time-invariant, plot-level and household-level characteristics. However, it is still possible that an omitted characteristic drives the inverse relationship; the stability of the association is suggestive of but not proof of causality.

In order to explore the likelihood of causality we calculate Oster's bias-adjusted estimator γ^*, as defined in Equation 6. We do this allowing X_{ijt} from Equation 5 to be the full set of controls in Column 6 of Table 4,[30] and begin by assuming $\delta = 1$ and $R_{max} = 1.3R_5$, as suggested by Oster (2016). Under these assumptions, we obtain the extremely tight bounding interval [-0.685, -0.687]. Notably, these bounds suggest that the true, causal parameter is practically identical to the estimated parameter.[31]

If we loosen these assumptions to allow a range of δ and R_{max} parameters, though still relying on Oster's restriction on intermediate coefficients, we find that *all* possible

[29] Organic and inorganic fertilizer are controlled for in a binary fashion, as few plots receive either input.
[30] See Appendix 9 for bounds based on each set of controls in turn.
[31] If anything, these bounds suggest that omitted variables bias the magnitude of the coefficient downwards, causing us to under-estimate rather than over-estimate the inverse relationship.

bounds suggest that the causal inverse relationship is almost identical to the estimated relationship, if anything slightly under-estimated. Figure A5 in Appendix 9 illustrates these possible bounds. More details can be found in Appendix 9.

3.4 The Edge Effect

Having established the inverse relationship as robust, and unexplained by measurement error in land size or omitted variables, we investigate the edge effect mechanism. Table 5 presents results for Equation 10. In Column 1, only plot size explains plot productivity, and the baseline inverse relationship is estimated. In Column 2 the perimeter-area ratio, $\frac{P_{ijt}}{A_{ijt}}$, is additionally controlled for. This ratio is strongly, positively correlated with plot productivity, as we would expect if the edges of a plot are more productive than the interior of a plot. Moreover, the inverse relationship is completely mitigated, becoming statistically identical to zero. Model R^2 also rises significantly in Column 2.

This result suggests that the inverse relationship is driven by a misspecification of the plot-level production function. In Column 3 plot size only perimeter-area ratio is controlled for; adjusted R^2 drops by only a percentage point, suggesting that plot size contributes little additional information on average productivity once perimeter-area ratio is known.[32] In fact, adjusted R-squared is higher in Column 3 than in Column 1; perimeter-area ratio alone better explains productivity than does area alone.

Because $log(P_{ijt}/A_{ijt}) = log(P_{ijt}) - log(A_{ijt})$, we can alternatively specify the regression in Column 3 controlling for perimeter and area separately. If perimeter and area predict productivity only insofar as their ratio predicts productivity, as hypothesized, then $\hat{\theta}_1$ will be estimated as the coefficient on each variable, though negative for area. Indeed, this is the case; the coefficients in Column 4 have statistically identical magnitudes.

Appendix 10 further explores evidence for the edge effect through a variety of robustness checks, placebo tests, and alternative specifications. The edge effect result is robust to all the controls of Table 4 (Table A28), can be estimated across crop and ownership/management subsets (Tables A30-A31), and is virtually indistinguishable across plot size quantiles and perimeter-area quantiles (Table A32). Oster's (2016) bias-adjusted estimator γ^* suggests that under reasonable assumptions, the edge effect is above zero (Table A29, Figure A8). While plot shape does not drive productivity it alters the marginal effect of perimeter-area ratio, which diminishes with number of plot sides (Table A33). For all shapes, however, the edge effect is far above zero.

Because plot size and perimeter-area ratio are correlated (with a Pearson's correlation coefficient of -0.133), and the log version of these variables is even more strongly correlated, collinearity is a concern.[33] Yet the edge effect is indistinguishable across correlation quintiles (Table A34). Additionally, two placebo tests replace perimeter with a second variable, similar in distribution to perimeter; while these new placebo-area ratios are highly correlated with area, just as perimeter-area ratio is, controlling for the

[32]The fact that plot size contributes little or no information on productivity suggests, according to the theory outlined in Appendix 6, that the border area b defining plot edge is small.

[33]Though plot size and perimeter-area ratio are closely related, plot shape also explains a significant portion of perimeter-area ratio. See Table A27 in Appendix 10.

placebo-area ratios does not mitigate the inverse relationship (Tables A35 and A36). These tests illustrates that collinearity cannot drive the results in Table 5.

Additionally, alternative proxies can be used to test for the edge effect, using number of sides (Table A37) or "extra/unnecessary" perimeter (Table A38), rather than perimeter, to capture the peripheral area. As we would expect if these variables are an inferior proxy for peripheral area proportion, the estimated inverse relationship is attenuated but not wholly eliminated when controlling for these proxies.

Implicit in our edge effect hypothesis is another: for plots where size and perimeter-area ratio move in the same direction over time, we should not observe an inverse relationship. Appendix 10 simulates this effect and tests it empirically. For plots where size and perimeter-area ratio both move in the same direction between 2003 and 2013, plot size is positively rather than negatively correlated with productivity (Table A39).

These results all suggest the edge effect as an important explanation of the oft-observed size-productivity relationship. We note, however, that the magnitude of the estimated effect seems large. We explore the expected elasticity of productivity with respect to both plot size and perimeter-area ratio in the last section of Appendix 10. Using interior-peripheral differentials observed in agronomy literature, we simulate an effect that is almost half of the estimated coefficient on perimeter-area ratio. While we cannot be sure why the estimated effect is twice as high, we do know that the interior-peripheral differentials discussed in agronomy literature usually stem from controlled agronomic trials, and therefore reflect biophysical mechanisms only. In farmer controlled plots, behavioral mechanisms may increase the differential between interior and peripheral productivity, increasing the edge effect. See Appendix 10 for more discussion.

While perimeter-area ratio completely mitigates the inverse relationship under plot fixed effects, Table A10 shows that under household-time fixed effects, the inverse relationship is only mitigated by about a half, more in line with theory-based expectations developed in Appendix 10. Because plot-specific shadow prices may bias estimation under household-time fixed effects, this remaining inverse relationship could be spurious correlation. However, it is also possible that while the edge effect explains much or all of the within-plot, over-time inverse relationship, other mechanisms drive the inverse relationship observed across plots within time periods. For example, Gourlay, Kilic and Lobell (2017) and Desiere and Jolliffe (2017) hypothesize that farmers over-estimate production on small plots and under-estimate it on large ones; if "small" and "large" are relative measures based on plot comparison, this mechanism might occur within time periods, but not across time within plots.

3.5 Edge Effect Mechanisms

We next explore mechanisms that might explain the edge effect. This section is necessarily more exploratory and suggestive in nature, as we cannot test for biophysical mechanisms directly. Appendix 7 reports indirect evidence regarding biophysical mechanisms. We find no evidence that sunlight or soil nutrient absorption drive the edge effect, though we cannot rule these mechanisms out. We do find that the edge effect is (insignificantly) larger in magnitude for mixed cropping and intercropped plots than for monocropped plots, as observed by Ward, Roe and Batte (2016). While other biophysical characteristics should drive heterogeneity in the edge effect, we are limited in our ability to test for such patterns. For instance, north-south row orientation mitigates the edge effect (Barchia and Cooper, 1996), but we do not observe row orientation. The characteristics of adjacent plots, also unobserved, will surely impact edge effect magnitude. A richer dataset might characterize such heterogeneity.

Because we do observe labor intensity, we can more directly examine evidence regarding a possible behavioral mechanism. Table 6 reports results for Equation 11, investigating whether the perimeter-area ratio drives labor intensity as well as plot productivity. In Column 1 labor intensity is inversely correlated with plot size. In Column 2 we see that the controlling for perimeter-area ratio mitigates the inverse relationship between plot size and labor intensity by two-thirds, and also increases adjusted R^2. Column 3 controls only for perimeter-area ratio; plot size continues to contribute information on labor intensity, though not much. As with Table 5, the relative R^2 values of Columns 1 and 3 suggest that perimeter-area ratio better captures the data generating process than does plot size. This suggests (but cannot prove) that increased labor intensity around plot edges contributes to the edge effect, and helps to drive the inverse relationship.[34,35] Appendix 7 reports additional results across various types of labor tasks and laborers; differences are not large or particularly meaningful.

The results of Table 6 are thought-provoking in one additional way. If measurement error (e.g., in land size or in production) drives the size-productivity inverse relationship or the edge effect, this measurement error would need to effect both yields and labor intensity similarly, in order for the results in Table 6 to hold. For instance, if the inverse relationship is solely driven by over-estimation of production on small plots, as suggested by Gourlay, Kilic and Lobell (2017) and Desiere and Jolliffe (2017), then labor intensity must be similarly over-estimated on these small plots. And in both cases, over-estimation is apparently better predicted by perimeter-area ratio than by plot size.

Last, we examine how farmers' misperceptions of plot area are associated with plot productivity. If perception error is exogenous to other plot conditions, then any labor or productivity response to misperceptions can be viewed as purely behavioral. Appendix 11 illustrates that, once conditioned on plot area and perimeter-area ratio, perception error appears exogenous to other time varying, plot-level characteristics, with the potential exception of crops cultivated.

Table 7 demonstrates the effect of farmer misperceptions of plot size on productivity. Plot productivity rises with over-estimation (at a diminishing rate), and drops with under-estimation (also at diminishing rate).[36] Column 1 controls for log plot area and

[34] Additionally, the edge effect seems to affect labor intensity and productivity in the same way: the third columns of Table 5 and Table 6 suggest a roughly one-to-one increase in both productivity and labor intensity. As discussed in Appendix 10, this elasticity is higher than expected, and could be driven by a few factors. Visible edges may increase application of labor farther into the plot than the peripheral area that enjoys a biophysical edge effect. Shape may also be endogenously chosen to maximize ease of labor, in a way not fully understand.

[35] It is important to note that this relationship implies that labor is applied at higher rates around plot edges, not that labor is more effective around plot edges, a possibility we cannot gauge in these data.

[36] Note that over-estimation is, on average, much larger in magnitude than under-estimation, which is bounded from below by zero. The productivity impacts of over-estimation are therefore much greater than the productivity impacts of under-estimation.

log plot perimeter-area ratio in addition to farmer perceptions, and Column 2 controls for area and perimeter-area ratio quadratically.[37] Column 3 controls for area and perimeter-area ratio quadratically and additionally controls for all typically "omitted" variables from Column 6 of Table 4. The effect of farmer perceptions remains qualitatively and statistically the same across all specifications.

The results in Table 7 suggest that farmers' misperceptions of plot size impact plot productivity. This presumably occurs through behavioral channels only, since misperceptions seem to be exogenous to plot characteristics, conditional on plot size and perimeter-area ratio. If farmers apply more (less) inputs and labor to larger (smaller) plots, as they clearly do, it seems logical that an over- (under-) estimate of plot size would lead to inefficient allocation of resources, and higher (lower) productivity. In fact, Table A45 shows that labor intensity does respond in this way to farmer over-estimation of plot size, though the results are not as strong as they are for productivity.[38]

Because Appendix 11 suggest that perception error may possibly be related to crop type, Appendix 12 presents the same regressions by crop, as well as by ownership/ management category. The results are qualitatively similar to those of Table 7, though not always significant since sample size falls drastically. It is also possible that misperceptions of plot size impact not plot productivity itself but rather *perceived* plot productivity. A farmer who believes his/her plot to be larger than it is may be more likely to over-report yields, while a farmer who under-estimates size may under-report yields. With these data we cannot differentiate such a phenomenon from the behavioral phenomenon wherein farmers actually allocated inputs according to their perceptions of plot size, and truly experience higher or lower yields as a result.

4 Conclusion

Using plot-level panel data for the first time in this literature, we illustrate that the inverse relationship (a 6.9 percent decrease in productivity with a 10 percent increase in land size) is at the plot rather than farm level, and so cannot be driven by market failure. Neither measurement error in plot size nor oft-omitted, relevant variables drive the relationship. Acknowledging that plot size is far from exogenously determined, we estimate all results under two identification strategies. Near-identical coefficients across the two strategies suggests that the relationship is unlikely to be driven by omitted, time-varying plot characteristics (which might bias estimates under plot fixed effects, but are washed away by household-time fixed effects), or by omitted characteristics that vary across plots within a farm and time period (which might bias estimates under household-time fixed effects, but are washed away by plot fixed effects). Estimated bounds around the likely, causal relationship between plot size and productivity confirm that omitted variable bias is unlikely to drive the inverse relationship (Oster, 2016). Without alternative measure of production we cannot examine the role of error in production estimates (Gourlay, Kilic and Lobell, 2017; Desiere and Jolliffe, 2017).

While previously proposed mechanisms do not explain the inverse relationship, it might

[37] Appendix 11 suggests that perceptions may respond non-linearly to area and perimeter-area ratio.

[38] If plot size perceptions drive the intensity of multiple inputs, we expect a stronger relationship between resulting productivity and perceptions than between any individual input and perceptions.

be explained by edge effects, familiar in the agronomy literature. We show that plot productivity rises with perimeter-area ratio, and under plot fixed effects, plot area has no remaining influence on productivity once perimeter-area ratio is controlled for.[39] This does not prove that the edge effect drives the inverse relationship. However, this pattern is exactly what we expect to find if plot peripheries/edges are more productive than plot interiors, and if the width of this peripheral area is narrow. Therefore, in these data at least, particularly when it comes to variation in plot size and productivity over time, it seems likely that the inverse relationship is at least partially driven by the more productive peripheral area around the edge of each plot, which necessarily makes smaller plots more productive on average than larger plots.

The mechanism behind the edge effect is difficult to isolate; there are likely multiple mechanisms. The agronomy literature suggests that the edge effect is due to increased biophysical inputs on plot edges, such as increased sunlight exposure, access to water or nutrient uptake (Watson and French, 1971; O'Brien and Green, 1974; Barchia and Cooper, 1996; Balagawi, Jackson and Clarke, 2014). We do not observe these inputs directly, but find no indirect evidence of such biophysical mechanisms.

We do, however, observe that labor intensity rises with perimeter-area ratio, just as productivity does. This suggests a behavioral mechanism; if farmers are more aware of or can more easily access plot edges, they may weed, prune, or harvest these edges more intensely, contributing to the edge effect and to the inverse relationship. Consumer behavior responds strongly to visual cues (Wansink, Painter and North, 2005). Our results suggest that perhaps producer behavior does the same. In fact, we additionally show that farmer beliefs about plot size, independent of actual plot size, are associated with productivity; over- (under-) estimation of plot size is positively (negatively) associated with plot productivity. While the relationship is not causal, such results are consistent with cognitive error influencing input application and hence productivity, again suggesting that farmer perceptions might influence behavior and production.

Taken together, our results suggest not only that the inverse relationship is at the plot, rather than farm level, but also that both plot size and plot *shape* drive productivity. Several important implications arise from an edge effect explanation for the inverse relationship. First land transfers to smaller farmers, if they led to larger plots, would generally attenuate the productivity "advantage" experienced by smaller farms. Second, while edge effects will not be addressed by market corrections, organizational tweaks (such as recommendations regarding plot lay-out and plot shape) or behavioral nudges (such as pathways built into larger plots or reminders to weed plot interiors) might extend the benefit of edge effects to larger plots and farms. On a larger scale, whether due to biophysical or behavioral mechanisms, edge effects would seem to favor polyculture mosaics on biodiverse farms over monocultures within smallholder agriculture (Marshall and Moonen 2002, Altieri 2004, Carvalheiro et al. 2011).

To capitalize on policy possibilities, further research is needed to tease out both mechanisms behind and heterogeneity within the edge effect. Plots bordering large trees, for instance, might benefit less from the edge effect; the direction of crops rows might also influence the magnitude of the edge effect. A better understanding of yield

[39]The perimeter-area ratio is similarly positively associated with productivity under household-time fixed effects, but the inverse relationship is not fully mitigated.

heterogeneity within plots, including the implications of the edge effect, is necessary before optimal organizational or behavioral practices can be considered. This paper begins, however, by showing that the inverse relationship is a plot-level phenomenon, and that plot shape as well as plot size influences productivity.

References

Ali, Daniel Ayalew, and Klaus Deininger. 2015. "Is there a farm size–productivity relationship in African agriculture? Evidence from Rwanda." *Land Economics*, 91(2): 317–343.

Altonji, Joseph G, Todd E Elder, and Christopher R Taber. 2005. "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools." *Journal of Political Economy*, 113(1): 151–184.

Alvarez, Antonio, and Carlos Arias. 2004. "Technical efficiency and farm size: a conditional analysis." *Agricultural Economics*, 30(3): 241–250.

Assunção, Juliano J, and Luis HB Braido. 2007. "Testing household-specific explanations for the inverse productivity relationship." *American Journal of Agricultural Economics*, 89(4): 980–990.

Balagawi, Solomon, Kevin Jackson, and AR Clarke. 2014. "Resting sites, edge effects and dispersion of a polyphagous Bactrocera fruit fly within crops of different architecture." *Journal of Applied Entomology*, 138(7): 510–518.

Barchia, I, and JL Cooper. 1996. "Edge row effects on yields in an oat trial." *Animal Production Science*, 36(4): 459–461.

Bardhan, Pranab K. 1973. "Size, productivity, and returns to scale: An analysis of farm-level data in Indian agriculture." *The Journal of Political Economy*, 1370–1386.

Barrett, Christopher B. 1996. "On price risk and the inverse farm size-productivity relationship." *Journal of Development Economics*, 51(2): 193–215.

Barrett, Christopher B, Marc F Bellemare, and Janet Y Hou. 2010. "Reconsidering conventional explanations of the inverse productivity–size relationship." *World Development*, 38(1): 88–97.

Benjamin, Dwayne. 1995. "Can unobserved land quality explain the inverse productivity relationship?" *Journal of Development Economics*, 46(1): 51–84.

Benjamin, Dwayne, and Loren Brandt. 2002. "Property rights, labour markets, and efficiency in a transition economy: the case of rural China." *Canadian Journal of Economics*, 689–716.

Berry, R Albert, and William R Cline. 1979. *Agrarian structure and productivity in developing countries: a study prepared for the international labour pffice within the framework of the World Employment Programme.* Johns Hopkins University Press.

Bezemer, Dirk, and Derek Headey. 2008. "Agriculture, development, and urban bias." *World Development*, 36(8): 1342–1364.

Bhalla, Surjit S, and Prannoy Roy. 1988. "Mis-specification in farm productivity analysis: the role of land quality." *Oxford Economic Papers*, 55–73.

Carletto, Calogero, Sara Savastano, and Alberto Zezza. 2013. "Fact or artifact: the impact of measurement errors on the farm size–productivity relationship." *Journal of Development Economics*, 103: 254–261.

Carletto, Calogero, Sydney Gourlay, and Paul Winters. 2015. "From guesstimates to GPStimates: Land area measurement and implications for agricultural analysis." *Journal of African Economies*, 24(5): 593–628.

Carletto, Calogero, Sydney Gourlay, Siobhan Murray, and Alberto Zezza. 2016. "Cheaper, faster, and more than good enough: is GPS the new gold standard in land area measurement?" World Bank Policy Research Working Paper 7759.

Carter, Michael R. 1984. "Identification of the inverse relationship between farm size and productivity: an empirical analysis of peasant agricultural production." *Oxford Economic Papers*, 131–145.

Carter, Michael R, and Yang Yao. 2002. "Local versus global separability in agricultural household models: The factor price equalization effect of land transfer rights." *American Journal of Agricultural Economics*, 84(3): 702–715.

Chayanov, Alexander. 1991. *The theory of peasant co-operatives.* IB Tauris.

Collier, Paul. 1983. "Malfunctioning of African rural factor markets: Theory and a Kenyan example." *Oxford Bulletin of Economics and Statistics*, 45(2): 141–172.

Desiere, Sam, and Dean Jolliffe. 2017. "Land productivity and plot size: Is measurement error driving the inverse relationship?" *Journal of Development Economics.*

Feder, Gershon. 1985. "The relation between farm size and farm productivity: The role of family labor, supervision and credit constraints." *Journal of Development Economics*, 18(2): 297–313.

Foster, Andrew D, and Mark R Rosenzweig. 2017. "Are There Too Many Farms in the World? Labor-Market Transaction Costs, Machine Capacities and Optimal Farm Size." National Bureau of Economic Research.

Gourlay, Sydney, Talip Kilic, and David Lobell. 2017. "Could the debate be over? Errors in farmer-reported production and their implications for inverse scale-productivity relationship in Uganda." Paper Presented at the 2017 Centre for the Study of African Economies (CSAE) Conference.

Heltberg, Rasmus. 1998. "Rural market imperfections and the farm size—productivity relationship: Evidence from Pakistan." *World Development*, 26(10): 1807–1826.

Henderson, Heath. 2015. "Considering Technical and Allocative Efficiency in the Inverse Farm Size–Productivity Relationship." *Journal of Agricultural Economics*, 66(2): 442–469.

Holman, E Merritt, and Craig W Bednarz. 2001. "Alley effect on several cotton cultivars in small plot research." *Communications in Soil Science and Plant Analysis*, 32(1-2): 119–126.

HTF. 2003. "Halving hunger by 2015: a framework for action." *Hunger Task Force Interim Report, Millennium Project. New York, USA: United Nations Development Program.*

Kagin, Justin, J Edward Taylor, and Antonio Yúnez-Naude. 2015. "Inverse Productivity or Inverse Efficiency? Evidence from Mexico." *The Journal of Development Studies*, 1–16.

Kimhi, A. 2006. "Plot size and maize productivity in Zambia: The inverse relationship re-examined." *Agricultural Economics*, 35: 1–9.

Krauth, Brian. 2016. "Bounding a linear causal effect using relative correlation restrictions." *Journal of Econometric Methods*, 5(1): 117–141.

Laborde Debucquet, David, and Will Martin. 2018. "Implications of the global growth slowdown for rural poverty." *Agricultural Economics*, 49(3): 325–338.

Lamb, Russell L. 2003. "Inverse productivity: Land quality, labor markets, and measurement error." *Journal of Development Economics*, 71(1): 71–95.

Larson, Donald F, Keijiro Otsuka, Tomoya Matsumoto, and Talip Kilic. 2014. "Should African rural development strategies depend on smallholder farms? An exploration of the inverse-productivity hypothesis." *Agricultural Economics*, 45(3): 355–367.

Little, Thomas M, and Frederick Jackson Hills. 1978. *Agricultural Experimentation: Design and Analysis.* John Wiley and Sons. New York, NY, USA.

Lowder, Sarah K, Jakob Skoet, and Terri Raney. 2016. "The number, size, and distribution of farms, smallholder farms, and family farms worldwide." *World Development*, 87: 16–29.

Mazumdar, Dipak. 1965. "Size of farm and productivity: a problem of Indian peasant agriculture." *Economica*, 32(126): 161–173.

Nkonya, Ephraim, John Pender, Kayuki C Kaizzi, Edward Kato, Samuel Mugarura, Henry Ssali, and James Muwonge. 2008. *Linkages between land management, land degradation, and poverty in Sub-Saharan Africa: The case of Uganda.* Vol. 159, International Food Policy Research Institute.

O'Brien, L, and RT Green. 1974. "Plot width and cultivar interaction." *Journal of the Australian Institute of Agricultural Science*, 40: 158–159.

Oster, Emily. 2016. "Unobservable selection and coefficient stability: theory and evidence." *Journal of Business Economics and Statistics*, DOI: 10.1080/07350015.2016.1227711.

Rios, Ana R, and Gerald E Shively. 2005. "Farm size and nonparametric efficiency measurements for coffee farms in Vietnam." In *Technical Report 19159, 2005 Annual Meeting.* American Agricultural Economics Association (New Name 2008: Agricultural and Applied Economics Association).

Sen, Amartya K. 1962. "An aspect of Indian agriculture." *Economic Weekly*, 14(4-6): 243–246.

Sen, Amartya K. 1966. "Peasants and Dualism with or without Surplus Labor." *Journal of Political Economy*, 74(5): 425–450.

Verdelli, Diego, Horacio A Acciaresi, and Eduardo S Leguizamón. 2012. "Corn and soybeans in a strip intercropping system: crop growth rates, Radiation interception, and grain yield components." *International Journal of Agronomy*, 2012.

Wansink, Brian, James E Painter, and Jill North. 2005. "Bottomless Bowls: Why Visual Cues of Portion Size May Influence Intake." *Obesity Research*, 13(1): 93–100.

Ward, Barry, Brian E Roe, and Marvin T Batte. 2016. "Potential Profitability of Strip Intercropping with Corn and Soybeans." *Journal of the ASFMRA*.

Watson, DJ, and SAW French. 1971. "Interference between rows and between plants within rows of a wheat crop, and its effects on growth and yield of differently-spaced rows." *Journal of Applied Ecology*, 421–445.

Figures

Figure 1: Plot Area Schematic

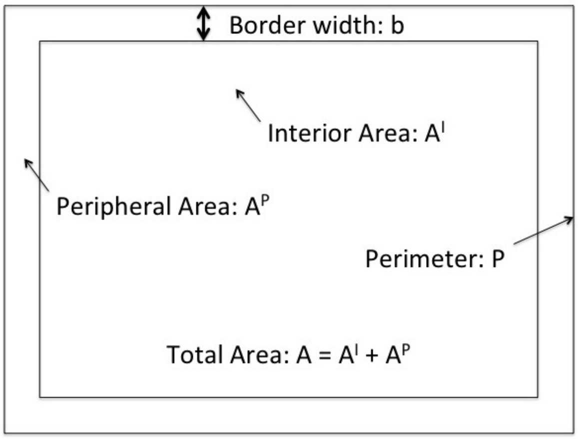
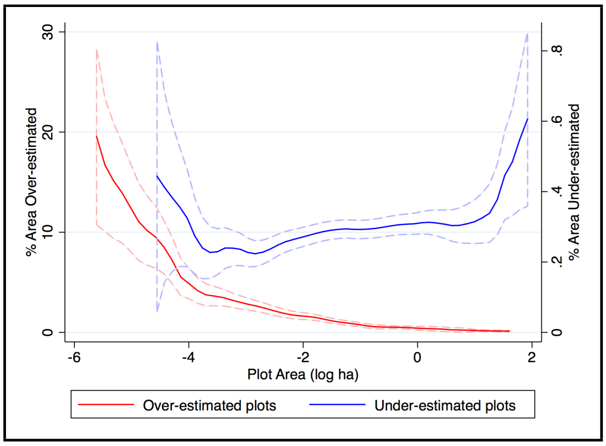


Figure 2: Misperceptions by Plot Area



Tables

Table 1: Plot Characteristics in 2003 and 2013

	2003		2013		
	Mean or Median†	Standard Deviation	Mean or Median†	Standard Deviation	T Statistic‡
Size and Productivity					
Farm size (ha)	1.03	1.13	0.60	0.89	14.33***
Plot size (ha)	0.35	0.69	0.18	0.41	11.72***
Perimeter-area ratio (m/ha)	849.45	935.42	1,137.94	6,945.52	-9.66***
Plot productivity (revenue§/ha)	102.70	1,361.44	257.63	8,867.67	-17.70***
Labor intensity (hrs/ha/day)	1.93	9.76	1.78	69.43	1.17
Soils					
Soil pH (pH)	6.22	0.51	6.18	0.63	1.60
Soil sand (%)	59.98	14.09	52.49	15.56	10.68***
Soil organic carbon (%)	3.47	1.66	3.71	1.82	-2.91***
Inputs					
Organic amendment (%)	19.94	39.98	12.04	32.56	4.94***
Inorganic fertilizer (%)	1.06	10.24	1.45	11.94	-0.79
Irrigation (%)	1.36	11.60	0.20	4.43	3.00***
Terracing (%)	23.80	42.60	9.56	29.41	8.86***
Management					
Head owns plot (%)	67.05	47.03	74.76	43.46	-3.88***
Head manages plot (%)	54.91	49.78	63.49	48.17	-3.99***
(Head owns)X(Head manages)	45.95	49.86	59.63	49.09	-6.30***
Crops are rotated (%)	21.94	41.40	42.69	49.49	-9.83***
Crops are mono-cropped (%)	43.45	49.59	36.42	48.14	3.28***
Mixed cropping (%)	54.34	49.84	52.22	49.97	0.97
Intercropping (%)	2.22	14.73	11.27	31.64	-8.36***
Crops Grown					
Tubers grown (%)	43.26	49.57	24.08	42.78	9.43***
Cereals grown (%)	49.52	50.02	43.64	49.62	2.69***
Legumes grown (%)	53.18	49.92	44.80	49.75	3.83***
Bananas grown (%)	49.42	50.02	27.94	44.89	10.30***
Cash crops grown (%)	31.02	46.28	18.98	39.23	6.39***

† The first 5 variables are all distributed log-normally, and therefore median is listed and T-statistics are generated using the log variable. For all other variables mean is listed and T-statistics are generated using the variable directly.

‡ *** $p<0.01$, ** $p<0.05$, * $p<0.1$

§ Revenue is given in real, 2005-valued dollars

Table 2: Household Shadow Prices and the Inverse Relationship (Plots Pooled)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity
		Panel 1 (No FE)	
Farm size (log ha)	-0.370***		-0.0139
	(0.0198)		(0.0228)
Plot size (log ha)		-0.508***	-0.501***
		(0.0161)	(0.0199)
Observations	5526	5526	5526
Adjusted R^2	0.066	0.168	0.168
		Panel 2 (House-year-season FE)	
Plot size (log ha)		-0.561***	
		(0.0261)	
Observations		4781	
Adjusted R^2		0.169	
		Panel 3 (Plot FE)	
Farm size (log ha)	-0.378***		0.0256
	(0.0978)		(0.0678)
Plot size (log ha)		-0.623***	-0.633***
		(0.0655)	(0.0638)
Observations	2075	2075	2075
Adjusted R^2	0.288	0.381	0.381

Dependent variable: log(revenue/hectare)
Panel 1 (Eq 1): No FE, robust standard errors
Panel 2 (Eq 2): HH-yr-ssn FE, HH-yr-ssn clustered standard errors
Panel 3 (Eq 3): Plot FE, yr and ssn FE, plot clustered standard errors
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Measurement Error and the Inverse Relationship (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity
Farmer-recalled plot size (log ha)	-0.615***	
	(0.0527)	
GPS-measured plot size (log ha)		-0.623***
		(0.0655)
Observations	2069	2075
Adjusted R^2	0.217	0.381

Col 1 dependent variable: log(revenue/farmer-recalled-hectare)
Col 2 dependent variable: log(revenue/GPS-measured-hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 4
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Omitted Variables and the Inverse Relationship (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity	(5) Plot Productivity	(6) Plot Productivity
GPS-measured plot size (log ha)	-0.684***	-0.677***	-0.616***	-0.733***	-0.743***	-0.685***
	(0.0978)	(0.101)	(0.0914)	(0.0934)	(0.0938)	(0.0876)
Soil pH (pH)		2.012				0.471
		(1.283)				(1.332)
Soil pH^2 (pH^2)		-0.143				-0.0143
		(0.106)				(0.111)
Soil sand (%)		-0.00626				-0.00795
		(0.00580)				(0.00554)
Soil organic carbon (%)		0.00262				-0.0121
		(0.0414)				(0.0463)
Labor intensity (log hrs/ha/day)			0.107**			0.102**
			(0.0439)			(0.0423)
Organic amendment (binary)			0.0905			0.128
			(0.156)			(0.163)
Inorganic fertilizer (binary)			1.479***			1.323***
			(0.249)			(0.387)
Irrigation (binary)			0.0708			-0.289
			(0.358)			(0.416)
Terracing (binary)			0.389***			0.409***
			(0.140)			(0.149)
Head owns plot (binary)				-0.109		-0.106
				(0.168)		(0.165)
Head manages plot (binary)				0.312		0.210
				(0.220)		(0.214)
(Head owns)X(Head manages)				-0.0727		-0.0363
				(0.255)		(0.245)
Crops are rotated (%)				-0.122		0.00792
				(0.121)		(0.120)
Crops are mono-cropped (%)				0.101		0.0968
				(0.275)		(0.275)
Mixed cropping (%)				0.549**		0.472*
				(0.280)		(0.274)
Tubers grown (binary)					0.126	0.0215
					(0.119)	(0.115)
Cereals grown (binary)					0.0398	-0.0444
					(0.109)	(0.106)
Legumes grown (binary)					0.208*	0.0643
					(0.107)	(0.116)
Bananas grown (binary)					0.150	0.114
					(0.176)	(0.169)
Cash crops grown (binary)					0.648***	0.492***
					(0.182)	(0.178)
Observations	1546	1546	1546	1546	1546	1546
Adjusted R^2	0.379	0.385	0.395	0.400	0.397	0.431
R^2	0.380	0.388	0.399	0.404	0.400	0.439

Dependent variable: log(revenue/hectare)
All columns estimated using only the sample for Column 6
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 5
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Edge Effect and the Inverse Relationship (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity
GPS-measured plot size (log ha)	-0.623***	-0.125		-1.024***
	(0.0655)	(0.129)		(0.130)
Perimeter-area ratio (log m/ha)		0.899***	1.095***	
		(0.234)	(0.0981)	
Perimeter (log m)				0.899***
				(0.234)
Observations	2075	2075	2075	2075
Adjusted R^2	0.381	0.393	0.393	0.393

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 10
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Edge Effect and Labor Intensity (Plot Panel)

	(1) Labor Intensity	(2) Labor Intensity	(3) Labor Intensity	(4) Labor Intensity
GPS-measured plot size (log ha)	-0.682***	-0.277**		-1.005***
	(0.0648)	(0.140)		(0.126)
Perimeter-area ratio (log m/ha)		0.727***	1.159***	
		(0.239)	(0.0970)	
Perimeter (log m)				0.727***
				(0.239)
Observations	1973	1973	1973	1973
Adjusted R^2	0.186	0.195	0.192	0.195

Dependent variable: log(hours/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 11
*** p<0.01, ** p<0.05, * p<0.1

Table 7: The Effects of Farmer Misperception of Plot Size (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity
Farmer over-estimates plot (binary)	-0.353**	-0.337**	-0.240
	(0.139)	(0.138)	(0.177)
Over-estimate (% area)	0.157***	0.153***	0.108***
	(0.0345)	(0.0338)	(0.0403)
Over-estimate squared	-0.00382***	-0.00447***	-0.00342**
	(0.00138)	(0.00126)	(0.00151)
Under-estimate (% area)	-1.996**	-1.802**	-2.129*
	(0.800)	(0.790)	(1.157)
Under-estimate squared	2.427**	2.020**	2.657*
	(0.965)	(0.947)	(1.477)
Plot Area, P-A Ratio	Yes	Yes	Yes
$(\text{Area})^2$, $(\text{P-A Ratio})^2$	No	Yes	Yes
Additional Plot Controls	No	No	Yes
Observations	2075	2075	1546
Adjusted R^2	0.413	0.425	0.458

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Additional plot controls are from Column 6 of Table 3
Table estimates Equation 12
*** p<0.01, ** p<0.05, * p<0.1

Online Appendix

Appendix 1 More on GPS: Plot Size, Perimeter, Geospatial Matching

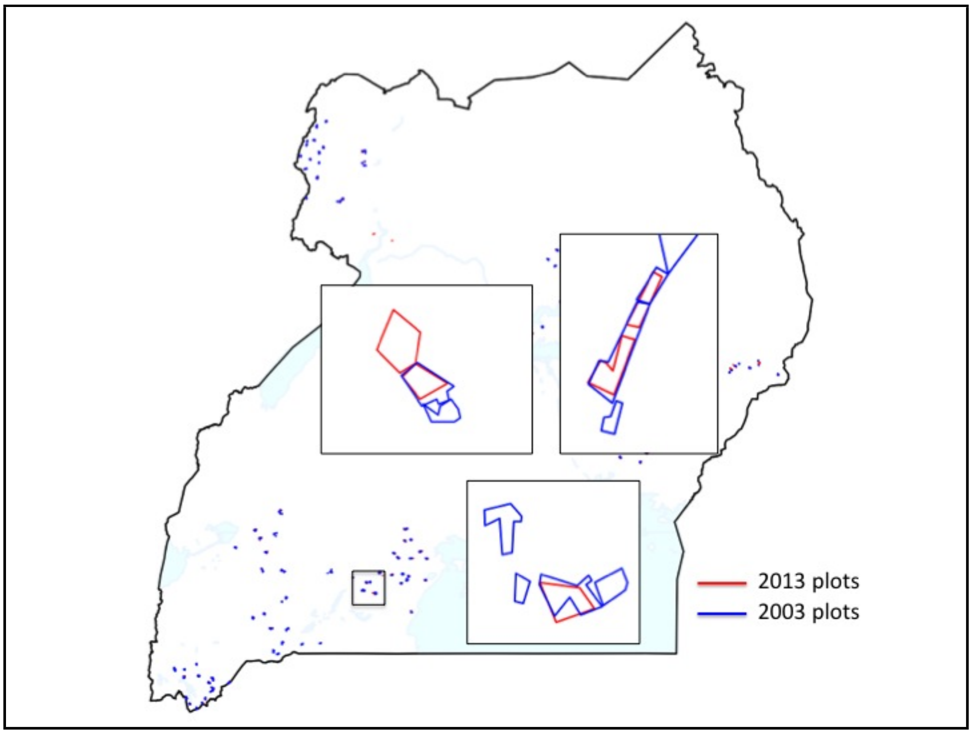
In both 2003 and 2013, enumerators collected GPS waypoints around the perimeter of each plot. In 2003 they did this by slowly walking around the plot while the GPS unit took waypoints at automated intervals, pausing at corners and flex-points in order to capture vertices. In 2013 they did this more explicitly by walking to each corner or flex-point of a plot and taking a single waypoint there. In each year they additionally took waypoints at what appeared to be the center of the plot.

The perimeter of each plot was then created via GIS by connecting the waypoints taken around each plot. (Because in most cases plots are fairly standard sizes, generally rectangles or triangles, these perimeters are fairly accurate.) Plot size was calculated as the precise area within each perimeter. (Altitude/slope was not used, as accurate measures were not available, and additionally plots are generally too small for this to be necessary.) New plot centroids were also generated based on the GIS-determined perimeter.

While GPS measurement is not without error, Carletto et al. (2016) illustrate that this error is not a source of concern when recording plot sizes. While it seems possible that the 2003 method of recording waypoints was more subject to error, we have no way of testing this hypothesis. However, measurement error will manufacture an inverse relationship, as plots that were under-estimated in size (appearing smaller) will be over-estimated in terms of revenue per hectare, while plots that were over-estimated in size (appearing larger) will be under-estimated in terms of revenue per hectare. If measurement error was greater in 2003 than in 2013, we might expect to see a stronger inverse relationship in round 1. In fact, the opposite is observed. We have no reason to suspect, therefore, that measurement error in either round is driving any part of our results.

To match plots geospatially over time, plot shapefiles were overlaid upon one another in ArcGIS as shown by Figure A1. It was more often found that a 2003 plot geospatially overlapped with multiple 2013 plots (as in the example at the top right of A1) than the reverse, due to the general trend of plots becoming smaller over the decade. Thus, we maintain 2013 plots as the unique observation, and match each 2013 plot to one 2003 plot. In 70 percent of cases, the 2013 plot overlaps only one 2003 plot. The rest of 2013 plots overlap multiple 2003 plots — in these cases the 2003 plot of greatest area overlap is matched. In some cases multiple 2013 plots end up matched to the same 2003 plot. In fact, 26 percent of 2013 plots are matched to a 2003 plot that is also matched to a different 2013 plot — the top right zoom in of A1 shows an example of three 2013 plots that each end up matched to the same 2003 plots.

Figure A1: Overlaying 2003 and 2013 plot shapefiles



Appendix 2 Survey Attrition, Sample Selection

1. Survey Attrition 2003 - 2013

Of the 849 households interviewed in 2003, 728 households (86 percent of the original sample) was tracked for re-interview. Households were reinterviewed if they were living in the same location, or tracked to a new location and re-interviewed if they were living in the same parish. Households who had moved out of the parish were not tracked. Additionally, 10 communities were not visited in 2013 — 1 in northern Uganda, and 9 in the south-west. Fifty-eight percent of attritted households come from these 10 communities, and 42 percent are distributed across the other communities.

Table A1 examines differences between the 2003 characteristics of tracked and attritted households. Households that attritted were significantly more educated, a bit younger, and a bit smaller than tracked households. Households that attritted were also slightly further from markets, had larger plots and farms, and had better soil quality.

These differences conflate the two types of attrition, however. Table A2 differentiates between households that attritted by community and households that attritted individually. Households that attritted by community tend to be further from markets and roads, make more money in crop income, and have better soil quality than tracked households. These statistics reflect the fact that more remote communities, which were difficult and expensive to travel to, were more likely to be dropped from the 2013 survey. Households that attritted individually (i.e., had simply moved out of a community that was included in the 2013 survey) had smaller farms and made less money that tracked households, but were also younger and more educated. This likely reflects a combination of individual push and pull factors driving migration out of the local parish.

Table A1: Household-level Attrition From 2003 to 2013

	Tracked		Attritted		
	Mean	St Dev	Mean	St Dev	T-stat
Head years of education (#)	4.88	3.36	5.56	3.13	-2.08**
Head age (#)	41.97	14.08	37.96	12.97	2.92***
Household size (# people)	5.97	2.81	5.15	2.88	2.92***
Asset index (index)	13.99	0.93	14.06	0.82	-0.70
Net crop income (1,000 Ush)	494.12	1,259.58	708.30	1,750.85	-1.43
Distance to all weather road (km)	2.57	4.64	2.52	2.72	-1.65*
Distance to market (km)	3.04	3.52	3.58	3.07	-1.83*
Farm size (ha)	1.41	2.52	1.65	6.45	2.02**
Number of plots owned (#)	4.56	2.20	5.30	3.16	-2.45**
Average plot area (ha)	0.63	1.19	1.08	6.24	2.05**
Crops provide primary income (%)	68.87	46.33	45.45	50.00	5.09***
Number of cattle (#)	2.36	6.24	1.88	7.20	0.71
Soil pH (pH)	6.16	0.47	6.13	0.58	0.63
Soil carbon (%)	3.16	1.51	4.15	1.85	-6.02***
Soil sand (%)	62.92	13.78	57.38	12.64	3.87***

*** p<0.01, ** p<0.05, * p<0.1

Table A2: Household-level Attrition: Community vs. HH-level Attrittition

	Tracked		Comm Attritted			House Attritted		
	Mean	St Dev	Mean	St Dev	T-stat	Mean	St Dev	T-stat
Head years of education (#)	4.88	3.36	5.33	3.14	-1.07	5.88	3.11	-2.07**
Head age (#)	41.97	14.08	39.23	14.05	1.54	36.24	11.24	2.85***
Household size (# people)	5.97	2.81	5.10	2.80	2.47**	5.21	3.03	1.78*
Asset index (index)	13.99	0.93	14.16	0.76	-1.36	13.89	0.90	0.60
Net crop income (1,000 Ush)	494.12	1,259.58	1,104.59	2,006.55	-4.02***	164.37	1,131.06	2.50**
Distance to all weather road (km)	2.57	4.64	3.45	3.04	-3.99***	1.23	1.45	2.11**
Distance to market (km)	3.04	3.52	4.50	3.24	-3.71***	2.30	2.28	1.47
Farm size (ha)	1.41	2.52	2.06	8.23	0.30	1.09	2.41	2.92***
Number of plots owned (#)	4.56	2.20	6.10	3.64	-4.35***	4.20	1.89	1.22
Average plot area (ha)	0.63	1.19	1.35	8.12	2.29**	0.71	1.52	0.60
Crops provide primary income (%)	68.87	46.33	45.71	50.18	3.96***	45.10	50.25	3.52***
Number of cattle (#)	2.36	6.24	1.21	4.19	1.42	2.92	10.24	-0.53
Soil pH (pH)	6.16	0.47	6.11	0.65	0.73	6.15	0.45	0.13
Soil carbon (%)	3.16	1.51	4.70	1.64	-7.67***	3.32	1.85	-0.65
Soil sand (%)	62.92	13.78	55.43	12.37	4.16***	60.31	12.63	1.20

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

2. Selection into Geospatially Matched Plot Panel (Plot and Time FE)

Our primary estimations are performed in a plot panel under plot fixed effects. Plots are geospatially matched across the decade using GPS. Plots that cannot be matched across the decade are therefore dropped, and household with no geospatially matched plots are not included in the dataset. Twenty-eight percent of 2003 households and 34 percent of 2003 plots appear in the panel dataset; forty-four percent of 2013 households and 30 percent of 2013 plots appear in the panel dataset.

It seems likely that selection into the geospatially matched dataset is not random. Tables A3 and A4 therefore compare household and plot characteristics across (i) the universe of all households/plots from 2003 and 2013, and (ii) the households/plots in the geospatially matched plot-level panel dataset. In Table A3 the unit of observation is the household, while in Table A4 the unit of observation is the plot.

Table A3 suggests that 2003 households select into the panel dataset in a fairly random manner — on the whole, the datasets seem similar at the household level, though the households in the panel dataset have slightly fewer plots owned, on average, and slightly larger household sizes. The 2013 households that end up in the panel dataset have very slightly older household heads than the households in the larger 2013 dataset, and are further from roads. All other household-level characteristics are balanced across datasets.

The plot level selection is far less random, in both rounds. Table A4 shows that larger 2003 plots end up in the panel dataset; 2003 plots growing bananas or cash crops were also more likely to end up in the panel dataset. Soil quality, inputs and management also differ across the universe of all 2003 plots and the 2003 plots that end up in the panel dataset.

Table A3: Universe of All Households vs Plot FE Model Households

	2003			2013		
	Universe	Plot FE	T-stat	Universe	Plot FE	T-stat
Farm size (ha)	1.26	1.25	-0.93	0.74	0.72	0.68
Number of plots owned (#)	4.69	4.38	1.83*	4.34	4.19	1.01
Number of crops grown (#)	5.85	6.02	-1.01	3.54	3.69	-1.25
Head years of education (#)	4.94	4.88	0.23	5.21	4.94	1.08
Head age (#)	41.50	44.34	-2.78***	50.52	52.21	-1.87*
Household size (# people)	5.86	6.26	-1.83*	6.63	6.32	1.35
Asset index (index)	14.00	14.08	-1.07	13.37	13.40	-0.48
Net crop income (1,000 Ush)	527.75	537.79	-2.05**	750.58	790.32	-0.73
Distance to all weather road (km)	2.51	2.37	-0.08	5.35	4.35	1.63
Distance to market (km)	3.14	2.95	0.07	4.86	5.04	-0.51

*** p<0.01, ** p<0.05, * p<0.1

Table A4: Universe of All Plots vs Plot FE Model Plots

	2003			2013		
	Universe	Plot FE	T-stat	Universe	Plot FE	T-stat
Plot size (ha)	0.40	0.48	-5.60***	0.31	0.31	0.57
Perimeter-area ratio (m/ha)	1,375.42	1,188.29	5.19***	1,724.88	1,810.74	-0.43
Plot productivity (revenue/ha)	328.86	308.92	1.01	989.19	1,184.86	-1.40
Labor intensity (hrs/ha/day)	5.27	4.37	2.60***	7.14	9.55	-1.34
Soil pH (pH)	6.11	6.18	-2.64***	6.07	6.13	-1.97**
Soil sand (%)	60.92	60.09	1.09	55.31	53.11	2.89***
Soil organic carbon (%)	3.55	3.47	0.90	3.35	3.65	-3.36***
Organic amendment (%)	8.59	16.53	-5.85***	7.43	10.57	-2.52**
Inorganic fertilizer (%)	1.96	1.32	1.05	1.69	1.29	0.72
Irrigation (%)	1.67	1.51	0.28	0.31	0.15	0.70
Terracing (%)	14.48	23.14	-5.23***	6.59	9.46	-2.41**
Head owns plot (%)	59.87	65.79	-2.69***	74.88	74.86	0.01
Head manages plot (%)	50.22	53.72	-1.55	62.44	63.29	-0.39
(Head owns)X(Head manages)	42.45	44.96	-1.12	56.34	59.00	-1.19
Crops are rotated (%)	30.49	24.75	2.75***	52.34	45.76	2.63***
Crops are mono-cropped (%)	60.30	45.95	6.46***	47.22	39.86	3.29***
Mixed cropping (%)	35.66	51.40	-7.20***	39.55	47.57	-3.61***
Tubers grown (%)	40.21	42.48	-1.02	25.72	24.86	0.44
Cereals grown (%)	49.35	48.76	0.26	49.40	44.71	2.08**
Legumes grown (%)	49.98	52.07	-0.92	41.24	43.14	-0.85
Bananas grown (%)	24.32	42.64	-9.16***	16.36	24.43	-4.60***
Cash crops grown (%)	16.24	27.44	-6.43***	13.59	17.29	-2.32**

*** p<0.01, ** p<0.05, * p<0.1

3. Selection into Pooled Dataset (Household-Time FE)

We additionally estimate all core results using pooled data from both 2003 and 2013, via a household-year-season fixed effect model. Rather than identifying via within-plot, over-time variation, this household-time fixed effect model identifies via across-plot, within-time variation. There is also selection into this dataset, however, as all household-rounds with only 1 plot are dropped from the estimation. In 2003 and 2013, 16 percent and 32 percent percent of households, respectively, had only 1 plot. These households are thus dropped from the dataset for estimation, along with their 1 plot each — 5 and 14 percent of the total universe of plots, in 2003 and 2013 respectively.

Households with only one plot might be different than households with multiple plots. If so, the houses/plots used for household-time fixed effect analysis might not be representative of the larger universe of houses/plots in our data. As we would expect, Table A5 shows that 2003 households who make it into the household-time fixed effect analysis more plots than the larger group of 2003 households — 0.38 more plots, on a average — though also slightly smaller farms. The 2013 households who make it into this analysis also have more plots — 0.5 more plots, on average — and have larger farms. In both years, the households included in household-time fixed effect analysis have slightly higher levels of crop income than households in the larger universe of data.

Interestingly, Table A6 suggests that plot-level selection into the household-time fixed effect dataset is random. No significant difference exists between the plots that end up in this dataset and the larger universe of all pooled plots, in either 2003 or 2013.

Table A5: Universe of All Households vs HH-Time FE Model Households

	2003			2013		
	Universe	HH-time FE	T-stat	Universe	HH-time FE	T-stat
Farm size (ha)	1.26	1.21	-1.80*	0.74	0.87	-4.28***
Number of plots owned (#)	4.69	5.07	-3.10***	4.34	4.79	-3.56***
Number of crops grown (#)	5.85	6.13	-2.55**	3.54	3.93	-4.05***
Head years of education (#)	4.94	5.03	-0.50	5.21	5.35	-0.65
Head age (#)	41.50	41.37	0.18	50.52	50.17	0.46
Household size (# people)	5.86	5.97	-0.75	6.63	6.79	-0.83
Asset index (index)	14.00	14.02	-0.52	13.37	13.40	-0.54
Net crop income (1,000 Ush)	527.75	559.25	-1.84*	750.58	839.22	-2.23**
Distance to all weather road (km)	2.51	2.40	-0.00	5.35	4.89	0.37
Distance to market (km)	3.14	3.17	-0.29	4.86	4.96	0.15

*** p<0.01, ** p<0.05, * p<0.1

Table A6: Universe of All Plots vs HH-Time FE Model Plots

	2003			2013		
	Universe	HH-time FE	T-stat	Universe	HH-time FE	T-stat
Plot size (ha)	0.40	0.34	1.29	0.31	0.30	0.22
Perimeter-area ratio (m/ha)	1,375.42	1,397.59	-1.10	1,724.88	1,739.76	-0.19
Plot productivity (revenue/ha)	328.86	339.59	-0.62	989.19	1,033.93	-0.02
Labor intensity (hrs/ha/day)	5.27	5.39	-0.67	7.14	7.23	-0.32
Soil pH (pH)	6.11	6.09	0.53	6.07	6.06	0.22
Soil sand (%)	60.92	60.80	0.23	55.31	55.11	0.33
Soil organic carbon (%)	3.55	3.59	-0.48	3.35	3.34	0.14
Organic amendment (%)	8.59	8.32	0.34	7.43	6.69	0.80
Inorganic fertilizer (%)	1.96	1.86	0.25	1.69	1.55	0.30
Irrigation (%)	1.67	1.67	-0.02	0.31	0.34	-0.18
Terracing (%)	14.48	14.49	-0.02	6.59	6.55	0.05
Head owns plot (%)	59.87	59.25	0.44	74.88	74.53	0.23
Head manages plot (%)	50.22	49.86	0.25	62.44	62.43	0.00
(Head owns)X(Head manages)	42.45	42.11	0.24	56.34	56.22	0.07
Crops are rotated (%)	30.49	30.82	-0.25	52.34	53.60	-0.63
Crops are mono-cropped (%)	60.30	61.37	-0.77	47.22	48.85	-0.91
Mixed cropping (%)	35.66	34.66	0.74	39.55	38.04	0.87
Tubers grown (%)	40.21	39.01	0.87	25.72	25.47	0.16
Cereals grown (%)	49.35	48.03	0.93	49.40	49.19	0.12
Legumes grown (%)	49.98	48.99	0.70	41.24	39.59	0.94
Bananas grown (%)	24.32	23.93	0.32	16.36	16.01	0.27
Cash crops grown (%)	16.24	15.78	0.45	13.59	13.85	-0.21

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix 3 Soil Sampling and Analysis

In both survey rounds soil sampling was conducted according to standard protocols for in-field, representative soil sampling. Twelve to twenty sub-samples were taken from each plot, with a thin soil probe that reached down to 20 cm. In plots with very hard soil, occasionally an auger or a hoe was used to collect soil samples, rather than a soil probe. In such cases effort was still made to gather soil down to 20 cm.

Sub-samples were taken from randomly distributed locations around the plot, roughly following zig-zag patterns, but avoiding any "odd" patches of ground such as termite mounds or compost piles. (Soil characteristics associated with such patches may be non-representative of the plot.) After mixing all sub-samples together in a bucket, a representative quantity of 500 grams of soil was gathered for subsequent drying, grinding and analysis.

Soil samples were processed and analyzed at Uganda's National Agricultural Laboratory (NARL), in both 2003 and 2013. In each year they were air dried, ground to pass through a 2-mm sieve, and milled using aluminum or stainless steel grinders.

After grinding, soil sub-samples (roughly 0.5 grams) were analyzed for a number of characteristics. Soil pH was determined in a 2.5:1 water to soil suspension, with the pH measured in the soil suspension after a 30-minute equilibration time (Okalebo et al. 2002). Soil organic carbon was determined via the Walkely-Black method (Walkley and Black 1934). While we believe that the buffer pH changed across 2003 and 2013 for this test, round fixed effects should pick up any difference in mean extraction levels due to this methodological shift. Soil texture, including percentage sand, was determined by hydrometer method in both years, after destruction of organic matter with hydrogen peroxide and dispersion with sodium hexametaphosphate (Bouyoucos 1936; Okalebo et al. 2002).

References

Walkley, Aldous, and I Armstrong Black. 1934. "An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method." *Soil Science*, 37(1): 29-38.

Okalebo, J Robert, Kenneth W Gathua, and Paul L Woomer. 2002. *Laboratory Methods of Soil and Plant Analysis: A Working Manual.* 2 ed., Chapter: "Soil Particle Size Analysis by the Bouyoucos or Hydrometer method." Tropical Soil Biology and Fertility Programme Nairobi, Kenya.

Bouyoucos, George John. 1936. "Directions for making mechanical analyses of soils by the hydrometer method." *Soil Science*, 42(3): 225-230.

Appendix 4 Household-Time Fixed Effects

While this paper's primary results are estimated with plot fixed effects, and controlling for year and season dummies, the same results can be estimated with household-year-season fixed effects. In this case, the identifying variation comes not from within-plot, across-time changes, but rather from cross sectional variation across plots, within a household-year-season group. Figures A2 and A3 show the probability distribution and cumulative distribution, respectively, of the demeaned independent variable plot size under these two forms of identification. Plot size (in log hectares) is demeaned by household-year-season for the pooled dataset and by household-plot for the panel dataset. The distributions are similar, though a Kolmogorov-Smirnov test for equality of distributions finds them to be significantly different.

Explanatory power is lower when results are estimated via this second form of cross-sectional variation, suggesting that the plot fixed effects are a better specification. The coefficients estimated, however, are qualitatively (and quantitatively) the same as those estimated under plot fixed effects, with two exceptions. First, perimeter-area ratio explains approximately half of the inverse relationship (Table A10) rather than all of it as under plot fixed effects (Table 5). Similarly, the perimeter-area ratio explains approximately half of the inverse size-labor relationship (Table A12) rather than all of it (Table 6). Second, farmer perceptions are not as strongly associated with plot productivity in Table A12 as they are in Table 7.

What should we make of the fact that the inverse relationship remains significant, even once perimeter-area ratio is controlled for, under household-year-season fixed effects? It might be that within farms, larger plots are, on average, truly less productive than smaller plots. Alternatively, time-invariant, plot-level characteristics (washed out by plot fixed effects) may be biasing the estimation. Or, as proposed by Gourlay, Kilic and Lobell (2017) and Desiere and Jolliffe (2017), it may be that farmers over-estimate production on small plots and over-estimate production on large plots, a phenomenon that would occur across plots within farms, but is unlikely to occur across time periods. Notably, whatever the mechanism, it occurs alongside the effect of the plot perimeter-area ratio — under this specification, we seem to observe the edge effect and the inverse relationship occurring simultaneously.

Figure A2: Identifying Variation by Estimation Method (PDF)

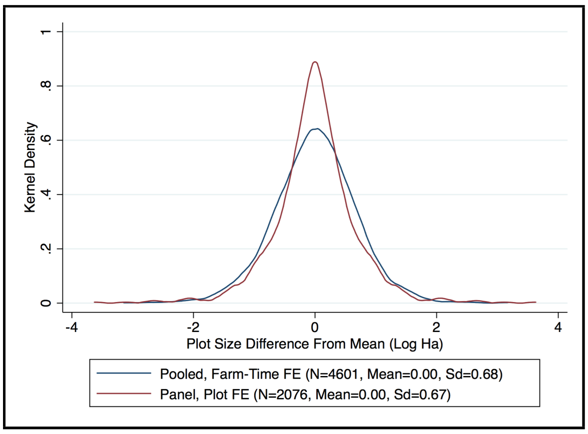


Figure A3: Identifying Variation by Estimation Method (CDF)

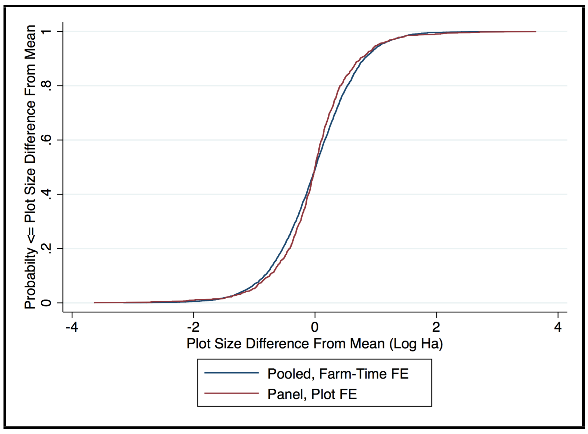


Table A7: Plot Characteristics in 2003 and 2013

	2003		2013		
	Mean or Median†	Standard Deviation	Mean or Median†	Standard Deviation	T Statistic‡
Size and Productivity					
Farm size (ha)	0.92	1.96	0.64	0.93	18.67***
Plot size (ha)	0.19	1.83	0.19	0.40	3.05***
Perimeter-area ratio (m/ha)	1,093.91	1,071.70	1,100.51	5,197.90	-2.92***
Plot productivity (revenue§/ha)	105.40	1,383.62	245.38	7,173.22	-23.28***
Labor intensity (hrs/ha/day)	2.45	15.08	1.70	55.58	9.55***
Soils					
Soil pH (pH)	6.13	0.60	6.09	0.61	1.97**
Soil sand (%)	60.29	15.11	54.42	15.47	13.15***
Soil organic carbon (%)	3.63	2.03	3.44	1.83	3.21***
Inputs					
Organic amendment (%)	10.54	30.71	8.62	28.07	2.53**
Inorganic fertilizer (%)	1.68	12.85	1.59	12.50	0.29
Irrigation (%)	1.70	12.92	0.33	5.74	4.92***
Terracing (%)	15.43	36.13	7.18	25.83	9.85***
Management					
Head owns plot (%)	61.28	48.72	73.73	44.02	-10.36***
Head manages plot (%)	50.94	50.00	62.83	48.34	-9.42***
(Head owns)X(Head manages)	42.99	49.51	56.85	49.54	-10.95***
Crops are rotated (%)	29.95	45.81	50.15	50.01	-15.52***
Crops are mono-cropped (%)	58.57	49.27	42.54	49.45	12.71***
Mixed cropping (%)	38.05	48.56	44.61	49.72	-5.24***
Intercropping (%)	3.36	18.02	12.61	33.20	-14.55***
Crops Grown					
Tubers grown (%)	40.10	49.02	25.46	43.57	12.18***
Cereals grown (%)	50.54	50.00	49.82	50.01	0.57
Legumes grown (%)	51.58	49.98	43.59	49.60	6.27***
Bananas grown (%)	29.60	45.66	19.40	39.55	9.19***
Cash crops grown (%)	18.30	38.67	15.25	35.96	3.16***

† The first 5 variables are all distributed log-normally, and therefore median is listed and T-statistics are generated using the log variable. For all other variables mean is listed and T-statistics are generated using the variable directly.

‡ *** p<0.01, ** p<0.05, * p<0.1

§ Revenue is given in real, 2005-valued dollars

Table A8: Measurement Error and the Inverse Relationship (Household-Time FE)

	(1) Plot Productivity	(2) Plot Productivity
Farmer-recalled plot size (log ha)	-0.547***	
	(0.0262)	
GPS-measured plot size (log ha)		-0.561***
		(0.0261)
Observations	5661	4781
Adjusted R^2	0.135	0.169

Col 1 dependent variable: log(revenue/farmer-recalled-hectare)
Col 2 dependent variable: log(revenue/GPS-measured-hectare)
Estimated with household-year-season fixed effects
Household-year-season-clustered standard errors in parentheses
Table estimates Equation 4
$p<0.01$, ** $p<0.05$, * $p<0.1$

Table A9: Omitted Variables and the Inverse Relationship (Household-Time FE)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity	(5) Plot Productivity	(6) Plot Productivity
GPS-measured plot size (log ha)	-0.532***	-0.513***	-0.390***	-0.551***	-0.561***	-0.352***
	(0.0311)	(0.0310)	(0.0335)	(0.0306)	(0.0306)	(0.0347)
Soil pH (pH)		-0.846				-1.092**
		(0.558)				(0.526)
Soil pH^2 (pH^2)		0.0948**				0.109**
		(0.0469)				(0.0440)
Soil sand (%)		0.00319				0.00327
		(0.00269)				(0.00259)
Soil organic carbon (%)		0.0272				0.0310*
		(0.0196)				(0.0173)
Labor intensity (log hrs/ha/day)			0.300***			0.344***
			(0.0293)			(0.0305)
Organic amendment (binary)			0.318***			0.0656
			(0.0872)			(0.0892)
Inorganic fertilizer (binary)			1.147***			1.135***
			(0.372)			(0.366)
Irrigation (binary)			0.400			0.530**
			(0.286)			(0.250)
Terracing (binary)			0.363***			0.235**
			(0.0922)			(0.0934)
Head owns plot (binary)				-0.0991		-0.115
				(0.181)		(0.169)
Head manages plot (binary)				0.202		0.167
				(0.174)		(0.161)
(Head owns)X(Head manages)				0.0164		0.0579
				(0.211)		(0.193)
Crops are rotated (%)				-0.333***		-0.376***
				(0.0820)		(0.0843)
Crops are mono-cropped (%)				-0.0953		-0.0253
				(0.142)		(0.133)
Mixed cropping (%)				0.164		0.0599
				(0.145)		(0.136)
Tubers grown (binary)					0.277***	0.205***
					(0.0575)	(0.0561)
Cereals grown (binary)					0.00606	-0.0502
					(0.0563)	(0.0564)
Legumes grown (binary)					0.236***	0.116**
					(0.0509)	(0.0548)
Bananas grown (binary)					0.536***	0.354***
					(0.0778)	(0.0802)
Cash crops grown (binary)					0.236***	0.125*
					(0.0737)	(0.0698)
Observations	3448	3448	3448	3448	3448	3448
Adjusted R^2	0.154	0.169	0.240	0.171	0.196	0.295
R^2	0.154	0.170	0.241	0.172	0.198	0.299

Dependent variable: log(revenue/hectare)
All columns estimated using only the sample for Column 6
Estimated with household-year-season fixed effects
Household-year-season-clustered standard errors in parentheses
Table estimates Equation 5
p<0.01, ** p<0.05, * p<0.1

Table A10: Edge Effect and the Inverse Relationship (Household-Time FE)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity
GPS-measured plot size (log ha)	-0.561***	-0.288***		-0.809***
	(0.0261)	(0.0613)		(0.0564)
Perimeter-area ratio (log m/ha)		0.520***	0.993***	
		(0.106)	(0.0426)	
Perimeter (log m)				0.520***
				(0.106)
Observations	4781	4781	4781	4781
Adjusted R^2	0.169	0.175	0.169	0.175

Dependent variable: log(revenue/hectare)
Estimated with household-year-season fixed effects
Household-year-season-clustered standard errors in parentheses
Table estimates Equation 10
p<0.01, ** p<0.05, * p<0.1

Table A11: Edge Effect and Labor Intensity (Household-Time FE)

	(1) Labor Intensity	(2) Labor Intensity	(3) Labor Intensity	(4) Labor Intensity
GPS-measured plot size (log ha)	-0.550***	-0.360***		-0.733***
	(0.0229)	(0.0541)		(0.0468)
Perimeter-area ratio (log m/ha)		0.373***	0.979***	
		(0.0903)	(0.0389)	
Perimeter (log m)				0.373***
				(0.0903)
Observations	4525	4525	4525	4525
Adjusted R^2	0.220	0.224	0.211	0.224

Dependent variable: log(hours/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 11
p<0.01, ** p<0.05, * p<0.1

Table A12: The Effects of Farmer Misperception of Plot Size (Household-Time FE)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity
Farmer over-estimates plot (binary)	0.0170	0.0271	0.0796
	(0.0806)	(0.0804)	(0.0921)
Over-estimate (% area)	0.0618***	0.0653***	0.0292
	(0.0207)	(0.0210)	(0.0252)
Over-estimate squared	-0.00123*	-0.00174**	-0.00104
	(0.000684)	(0.000736)	(0.000879)
Under-estimate (% area)	-0.519	-0.490	-0.102
	(0.459)	(0.460)	(0.546)
Under-estimate squared	-0.106	-0.165	-0.283
	(0.557)	(0.562)	(0.661)
Plot Area, P-A Ratio	Yes	Yes	Yes
$(\text{Area})^2$, $(\text{P-A Ratio})^2$	No	Yes	Yes
Additional Plot Controls	No	No	Yes
Observations	4781	4781	3448
Adjusted R^2	0.188	0.190	0.301

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Additional plot controls are from Column 6 of Table 3
Table estimates Equation 12
p<0.01, ** p<0.05, * p<0.1

Appendix 5 Investigating Exogeneity of Plot Size Change (2003-2013)

Our primary identification strategy rides on variation in plot size over time, through the implementation of plot fixed effects alongside time fixed effects. Between 2003 and 2013, 30 percent of plots grew and 70 percent of plots shrank. Expanding plots increased by an average (median) 0.17 (0.07) hectares, while shrinking plots decreased by an average (median) 0.45 (0.22) hectares.

It is therefore important to examine the primary predictors of plot size change between 2003 and 2013. For instance, one might hypothesize that plot size change is driven by farm- or parcel-level changes (farm or parcel sub-divisions, land sales, ownership changes, etc.). However, plots within our sample do **not** change ownership — while the 2013 survey did track plots passed from father to son, or plots sold or exchanged between neighbors, we drop these plots from both the panel and pooled datasets. Additionally, only 8 percent of plots sit on land parcels that experience sub-division. Hence, the majority of plot size change is due to re-arrangement of plots within existing, unchanged land parcels.

Table A13 examines potential correlates of plot size change via regression. The primary predictor is plot size in 2003, which explains 36 percent of variation in plot size change over the decade (Column 1). Larger plots are more likely to shrink over time, and smaller plots more likely to grow, in a pattern of regression to the mean (Figure A4). Farm- and parcel-level changes over the decade are also associated with plot size change, but explain only 2.5 percent of variation (Column 2), and contribute almost no explanatory once 2003 plot size is controlled for (Column 3).

Table A14 examines 2003 correlates with eventual plot size change. Again, 2003 plot size is clearly the primary predictor. None-the-less, other 2003 characteristics are joint, significantly predictors of plot size change, as indicated by the F-statistic at the bottom of Table A14. Table A15 conducts a more formal, regression-based balance test with the full panel dataset, regressing plot size on all covariates under plot-level fixed effects and year and season fixed effects, as in our primary specifications. The F-statistics of joint covariate significance again reject exogeneity.

Tables A13, A14, and A15 indicate that plot size change over the decade is not exogenous to other plot-level or household-level characteristics. Change in size is largely driven by regression to the mean over time, and correlated with other plot-level characteristics. Thus, the coefficient on plot size, in regressions capturing the inverse relationship or otherwise, therefore cannot be viewed as causal. It is for precisely this reason that we estimate Emily Oster's causal bounds around the coefficient that captures the inverse relationship (Appendix 9) and the coefficient that captures the edge effect (Appendix 10).

Table A13: Predictors of Plot Size Change (OLS)

	(1) Plot Size Change	(2) Plot Size Change	(3) Plot Size Change
GPS-measured plot size '03 (log ha)	-0.298***		-0.298***
	(0.0338)		(0.0329)
Houshold lost a parcel (binary)		0.0573	-0.0598
		(0.0574)	(0.0371)
Household aquired a parcel (binary)		0.124*	0.129**
		(0.0675)	(0.0576)
Plot is on a sub-divided parcel (binary)		-0.341*	-0.102
		(0.189)	(0.141)
Observations	700	656	656
Adjusted R^2	0.355	0.0208	0.358
R^2	0.356	0.0253	0.362

Dependent variable: ([2013 revenue/hectare] - [2003 revenue/hectare])
Household-clustered standard errors in parentheses for all columns
*** p<0.01, ** p<0.05, * p<0.1

Figure A4: Plot Size Regression to Mean

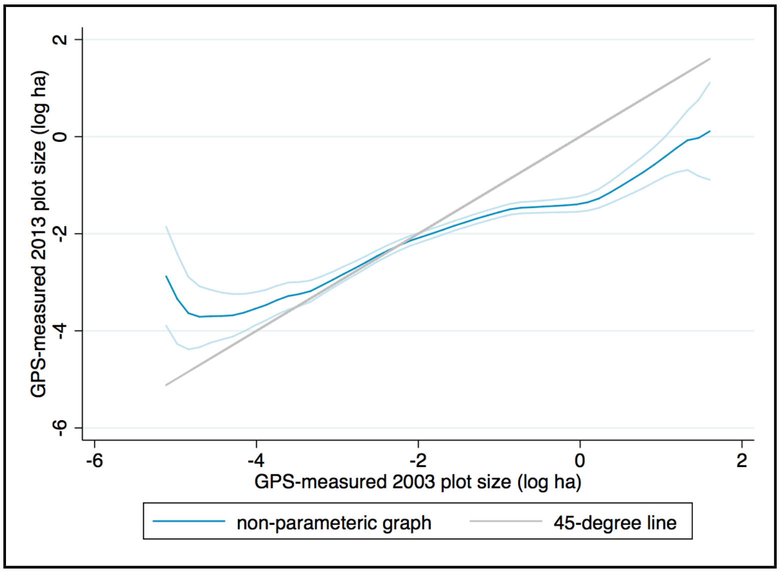


Table A14: Correlates of Plot Size Change in Round 1 (OLS)

	(1) Plot Size Change	(2) Plot Size Change	(3) Plot Size Change	(4) Plot Size Change	(5) Plot Size Change	(6) Plot Size Change
GPS-measured plot size '03 (log ha)	-0.295***	-0.315***	-0.314***	-0.296***	-0.273***	-0.311***
	(0.0375)	(0.0398)	(0.0400)	(0.0361)	(0.0373)	(0.0416)
Soil pH (pH)		-0.317				-0.312
		(0.420)				(0.447)
Soil pH^2 (pH^2)		0.0319				0.0300
		(0.0363)				(0.0393)
Soil sand (%)		0.00281*				0.00275*
		(0.00154)				(0.00157)
Soil organic carbon (%)		-0.0163				-0.0145
		(0.0114)				(0.0116)
Labor intensity (log hrs/ha/day)			-0.0377*			-0.0195
			(0.0217)			(0.0236)
Organic amendment (binary)			0.0308			0.0576
			(0.0491)			(0.0631)
Inorganic fertilizer (binary)			-0.0349			-0.104
			(0.146)			(0.169)
Irrigation (binary)			-0.164			-0.136
			(0.106)			(0.105)
Terracing (binary)			0.0855			0.0749
			(0.0665)			(0.0784)
Head owns plot (binary)				0.157***		0.162***
				(0.0581)		(0.0576)
Head manages plot (binary)				0.116*		0.119*
				(0.0664)		(0.0700)
(Head owns)X(Head manages)				-0.314***		-0.324***
				(0.0911)		(0.0915)
Crops are rotated (%)				-0.0348		-0.0118
				(0.0565)		(0.0649)
Crops are mono-cropped (%)				-0.0634		-0.0573
				(0.0642)		(0.0645)
Mixed cropping (%)				-0.0938		-0.0588
				(0.0586)		(0.0670)
Tubers grown (binary)					-0.134**	-0.130**
					(0.0530)	(0.0556)
Cereals grown (binary)					-0.0280	-0.0148
					(0.0516)	(0.0564)
Legumes grown (binary)					0.00471	0.0600
					(0.0460)	(0.0635)
Bananas grown (binary)					-0.0731	-0.0942
					(0.0570)	(0.0725)
Cash crops grown (binary)					-0.0978	-0.0855
					(0.0611)	(0.0619)
Observations	588	588	588	588	588	588
Adjusted R^2	0.343	0.350	0.346	0.356	0.354	0.374
R^2	0.344	0.356	0.352	0.364	0.360	0.396
Joint test (F-stat) p-value		0.0704	0.306	0.00705	0.201	0.00179

Dependent variable: ([2013 revenue/hectare] - [2003 revenue/hectare])
All covariates are from round 1 (R1) of the geospatially matched panel dataset
Household-clustered standard errors in parentheses for all columns
F-statistic p-value indicates join significance of all covariates except 2003 plot size
*** p<0.01, ** p<0.05, * p<0.1

Table A15: Plot Size Balance Test (Plot Panel, Plot FE)

	(1) Plot Size	(2) Plot Size	(3) Plot Size	(4) Plot Size	(5) Plot Size
Soil pH (pH)	-0.177				-0.255
	(1.141)				(1.048)
Soil pH^2 (pH^2)	0.00654				0.00991
	(0.0947)				(0.0861)
Soil sand (%)	-0.000591				0.00187
	(0.00500)				(0.00436)
Soil organic carbon (%)	-0.0347				-0.0523*
	(0.0352)				(0.0287)
Labor intensity (log hrs/ha/day)		-0.263***			-0.262***
		(0.0480)			(0.0453)
Organic amendment (binary)		0.207**			0.242***
		(0.0918)			(0.0833)
Inorganic fertilizer (binary)		-0.106			-0.374
		(0.655)			(0.646)
Irrigation (binary)		-0.350			-0.656**
		(0.337)			(0.263)
Terracing (binary)		0.0447			0.0274
		(0.101)			(0.104)
Head owns plot (binary)			0.0309		0.102
			(0.127)		(0.116)
Head manages plot (binary)			-0.0386		0.0305
			(0.178)		(0.152)
(Head owns)X(Head manages)			0.0199		-0.0888
			(0.204)		(0.173)
Crops are rotated (%)			-0.0645		-0.0800
			(0.102)		(0.0923)
Crops are mono-cropped (%)			0.273		0.249*
			(0.206)		(0.148)
Mixed cropping (%)			0.550***		0.436***
			(0.196)		(0.145)
Tubers grown (binary)				0.402***	0.352***
				(0.118)	(0.0940)
Cereals grown (binary)				0.176**	0.138**
				(0.0801)	(0.0692)
Legumes grown (binary)				0.0943	0.0970
				(0.0804)	(0.0736)
Bananas grown (binary)				0.0758	-0.103
				(0.132)	(0.118)
Cash crops grown (binary)				0.402***	0.327**
				(0.148)	(0.132)
Observations	1547	1547	1547	1547	1547
Adjusted R^2	0.213	0.353	0.238	0.270	0.434
R^2	0.216	0.356	0.242	0.274	0.442
Joint test (F-stat) p-value	0.760	0.000000571	0.00358	0.00000372	2.19e-10

Dependent variable: log(hectares)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
F-statistic p-value indicates join significance of all covariates, excluding fixed effects
*** p<0.01, ** p<0.05, * p<0.1

Appendix 6 Perimeter-Area Ratio by Plot Shape

Rather than assuming a generically shaped plot, we can assume plots of various, specific plot shapes in order to show more quantitatively that, with a small border width b, plot productivity Y_{ij} will always increase in P_{ijt}/A_{ij}, where P_{ijt} is the perimeter of the plot and A_{ijt} is the area of the plot. For the following calculations, we drop the ij subscript for all variables, for simplicity in notation. In all cases therefore, we define average productivity of the plot in question as below, exactly as in Equation 7.

$$Y \equiv \frac{Y^I * A^I + Y^P * A^P}{A}$$

Circle

Assume a circular plot with radius R, diameter D, border width b, perimeter $P = 2\pi R$ and total area $A = \pi R^2$. The interior of the plot has area $A^I = \pi(R-b)^2$, and the periphery, or border area, of the plot has area $A^P = A - A^I$. These area definitions can be expanded as follows.

$$A^I = \pi(R^2 - 2bR + b^2) = \pi R^2 - 2\pi bR + \pi b^2$$

$$A^P = (\pi R^2) - (\pi R^2 - 2\pi bR + \pi b^2) = 2\pi bR - \pi b^2$$

Average productivity can therefore be re-written as below.

$$\begin{aligned}
Y &= \frac{1}{A}(\pi R^2 - 2\pi bR + \pi b^2)Y^I + \frac{1}{A}(2\pi bR - \pi b^2)Y^P \\
&= \frac{1}{A}(2\pi bR - \pi b^2)(Y^P - Y^I) + \frac{1}{A}(\pi R^2)Y^I \\
&= \frac{1}{A}(2\pi bR)(Y^P - Y^I) - \frac{1}{A}(\pi b^2)(Y^P - Y^I) + \frac{1}{A}(A)Y^I \\
&= \frac{1}{A}bP(Y^P - Y^I) - \frac{1}{A}(\pi b^2)(Y^P - Y^I) + Y^I \\
&= \left[Y^I\right] + b(Y^P - Y^I)\left[\frac{P}{A}\right] - (\pi b^2)(Y^P - Y^I)\left[\frac{1}{A}\right]
\end{aligned}$$

Average productivity is therefore expected to rise with $\frac{P}{A}$ and to fall with $\frac{1}{A}$. However, if we assume that b is so small that b^2/A is close to zero, the last term of the equation disappears, and we expect productivity to rise only with $\frac{P}{A}$.

Rectangle

Assume a rectangular plot with length L and width W, border width b, perimeter $P = 2L + 2W$ and total area $A = WL$. The interior of the plot has area $A^I = (W - 2b)(L - 2b)$, and the periphery of the plot has area $A^P = A - A^I$. These area definitions can be expanded as follows.

$$A^I = WL - 2Wb - 2Lb + 4b^2$$

$$A^P = WL - (WL - 2Wb - 2Lb + 4b^2) = 2Wb + 2Lb - 4b^2$$

Average productivity can therefore be re-written as below.

$$
\begin{aligned}
Y &= \frac{1}{A}(WL - 2Wb - 2Lb + 4b^2)Y^I + \frac{1}{A}(2Wb + 2Lb - 4b^2)Y^P \\
&= \frac{1}{A}(2Wb + 2Lb - 4b^2)(Y^P - Y^I) + \frac{1}{A}(WL)Y^I \\
&= \frac{1}{A}(2Wb + 2Lb)(Y^P - Y^I) - \frac{1}{A}(4b^2)(Y^P - Y^I) + \frac{1}{A}(A)Y^I \\
&= \frac{1}{A}b(P)(Y^P - Y^I) - \frac{1}{A}(4b^2)(Y^P - Y^I) + Y^I \\
&= \left[Y^I\right] + b(Y^P - Y^I)\left[\frac{P}{A}\right] - (4b^2)(Y^P - Y^I)\left[\frac{1}{A}\right]
\end{aligned}
$$

Average productivity is therefore expected to rise with $\frac{P}{A}$ and to fall with $\frac{1}{A}$. However, if we assume that d is so small that b^2/A is close to zero, the last term of the equation disappears, and we expect productivity to rise only with $\frac{P}{A}$.

Triangle

Assume an equilateral triangular plot with each side being length S, border width b, perimeter P= $3S$ and total area A= $\frac{\sqrt{3}}{4}S^2$. The interior of the plot has area $A^I = \frac{\sqrt{3}}{4}(S - 2\sqrt{3}b)^2$, and the periphery of the plot has area A^P= $A - A^I$. These area definitions can be expanded as follows.

$$
A^I = \frac{\sqrt{3}}{4}\left[S^2 - 4\sqrt{3}bS + 12b^2\right] = \frac{\sqrt{3}}{4}S^2 - 3bS + 3\sqrt{3}b^2
$$

$$
A^P = \frac{\sqrt{3}}{4}S^2 - \left[\frac{\sqrt{3}}{4}S^2 - 3bS + 3\sqrt{3}b^2\right] = 3bS - 3\sqrt{3}b^2
$$

Average productivity can therefore be re-written as below.

$$
\begin{aligned}
Y &= \frac{1}{A}\left[\frac{\sqrt{3}}{4}S^2 - 3bS + 3\sqrt{3}b^2\right]Y^I + \frac{1}{A}\left[3bS - 3\sqrt{3}b^2\right]Y^P \\
&= \frac{1}{A}\left[3bS - 3\sqrt{3}b^2\right](Y^P - Y^I) + \frac{1}{A}\left[\frac{\sqrt{3}}{4}S^2\right]Y^I \\
&= \frac{1}{A}\left[3bS\right](Y^P - Y^I) - \frac{1}{A}\left[3\sqrt{3}b^2\right](Y^P - Y^I) + \frac{1}{A}(A)Y^I \\
&= \frac{1}{A}b(P)(Y^P - Y^I) - \frac{1}{A}\left[3\sqrt{3}b^2\right](Y^P - Y^I) + Y^I \\
&= \left[Y^I\right] + b(Y^P - Y^I)\left[\frac{P}{A}\right] - (3\sqrt{3}b^2)(Y^P - Y^I)\left[\frac{1}{A}\right]
\end{aligned}
$$

Average productivity is therefore expected to rise with $\frac{P}{A}$ and to fall with $\frac{1}{A}$. However, if we assume that d is so small that b^2/A is close to zero, the last term of the equation disappears, and we expect productivity to rise only with $\frac{P}{A}$.

What if the periphery is wide?

In all three of these specifications, we see that Y rises linearly with $b(Y^P - Y^I)[P/A]$, and falls linearly with $gb^2(Y^P - Y^I)[1/A]$, where g is a scaling factor that varies by plot shape. For circles, $g = \pi \approx 3.142$, for rectangles $g = 4$, and for equilateral triangles $g = 3\sqrt{3} \approx 5.196$. (So, it appears that g falls as the number of sides increases.)

Therefore, under each of these three specifications, if the width of the periphery/border length b is so very narrow that b^2/A is close to zero, we would expect to find that average plot productivity Y rises only with P/A. If the periphery length b is wide, however, we expect that average plot productivity rises with P/A and also falls with $1/A$.

However, when we do these regressions in practice, we regress the log form of these variables — so we would regress $log(Y)$ on $log(P/A)$ and $log(1/A)$. But $log(1/A)$ $= log(A^{-1}) = -1 * log(A)$. So if the periphery length b is wide, we expect that average plot productivity both rises only with $log(P/A)$ and also rises with $log(A)$. Increasing with $log(A)$ is equivalent to decreasing with $log(1/A)$.

In Table 5, however, we find that plot area has no additional, explanatory power after perimeter-area ratio is controlled for. This suggests that, indeed, periphery length b is narrow enough that in most cases we can assume that b^2/A is close to zero.

Appendix 7 More on Edge Effect Mechanisms

It seems feasible that both biophysical and behavioral mechanisms drive the edge effect, but we have limited ability to test either hypothesis. Table 6 provides some evidence that farmers provide more labor to plots with a higher ratio of peripheral area, suggesting that labor allocation may play into the edge effect. Below, we provide some further analysis regarding labor inputs as a mechanism. Second, we indirectly test for biophysical mechanisms as best as possible given that we have no data on biophysical inputs.

Behavioral Mechanisms

Table A16 illustrates that the edge effect is statistically identical across family and non-family labor, though it becomes slightly smaller in magnitude and insignificant for non-family labor. Likely, however, this is due to a reduced sample size. (In Uganda, both hired labor and exchange labor are relatively rare.) Table A17 estimates the edge effect by various labor tasks. It appears that edge effect most strongly drives weeding and planting labor. The result is difficult to interpret, however, because the third category of "other labor" includes labor allocated towards a litany of other tasks, none of which account for any significant proportion of total labor across households. All in all, little can be gleaned in these data about how the edge effect might vary by types of labor or laborers.

Biophysical Mechanisms

First, if soil nutrients are more plentiful at the edges of a plot, therefore driving these edges to be more productive, then we might expect the edge effect to function most strongly in nutrient-constrained settings. We therefore modify Equation 10 to control for plot-specific soil quality S_{ijt} and interactions between soil quality and the perimeter-area ratio, as in Equation 13 below. If $\hat{\theta}$ reduces in magnitude and $\hat{\eta}$ is significant and negative, this indicates that the edge effect is particular strong in nutrient-constrained settings.

$$Y_{ijt} = \gamma A_{ijt} + \theta \frac{P_{ijt}}{A_{ijt}} + \zeta S_{ijt} + \eta \left[\frac{P_{ijt}}{A_{ijt}} * S_{ijt} \right] \tag{13}$$

Table A18 below shows results for Equation 13, specifying soil fertility in three ways — by soil organic matter, soil sand content, and soil nitrogen. In all cases, the coefficient on the perimeter-area ratio is unchanged, and the coefficient on the interaction between soil fertility and the ratio is insignificant. While this result does not, of course, prove that soil fertility gradients do *not* drive the gradient, it also does not support nutrient availability as an edge effect mechanism.

Second, if differential access to sunlight drives the edges of a plot to be more productive, then we might expect the edge effect to function most strongly with taller plants such as maize, millet, or sesame, where the plants around plot edges likely block sunlight from the plants in the interior. For crops grown close to the ground, such as groundnuts or potatoes, we might expect the edge effect to be weaker. Table A19 therefore estimates Equation 10 for subsets of plots according to crop height. The edge effect is identical for tall crops (Column 2) and for ground crops (Column 3), a finding that does not support sunlight as the mechanism behind the edge effect. Interestingly, the edge effect is smaller and insignificant for tree crops (bananas, cassava and coffee). Because tree crops differ from seasonal crops in terms of management, labor, biophysical inputs and more customary inputs, it is difficult to interpret this result.

Third, Ward, Roe and Batte (2016) suggest that the edge effect is stronger in intercropped systems than in monocropped systems, and imply that this effect is due to spacing and light. Table A20 therefore estimates Equation 10 for all plots in Column 1, for plots that are monocropped (according to the farmer) in Column 2, and for plots that are intercropped or contain mixed crops in Column 3. While the magnitude of the coefficient on perimeters-area ratio does rise for mixed cropped and intercropped plots, the difference is not statistically significant.

Table A16: Edge Effect and Labor Intensity by Labor Type (Plot Panel)

	(1) Labor Intensity (All)	(2) Labor Intensity (Family)	(3) Labor Intensity (Non-Family)
Perimeter-area ratio (log m/ha)	1.159***	1.129***	0.726***
	(0.0970)	(0.111)	(0.184)
Observations	1973	1938	747
Adjusted R^2	0.192	0.171	0.085

Dependent variable: log(hours/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 11 for data subsets
p<0.01, ** p<0.05, * p<0.1

Table A17: Edge Effect and Labor Intensity by Labor Type (Plot Panel)

	(1) Labor Intensity (All)	(2) Labor Intensity (Weeding)	(3) Labor Intensity (Planting)	(4) Labor Intensity (Other)
Perimeter-area ratio (log m/ha)	1.159***	1.202***	1.445***	0.795***
	(0.0970)	(0.105)	(0.105)	(0.150)
Observations	1973	1780	1340	1298
Adjusted R^2	0.192	0.234	0.283	0.072

Dependent variable: log(hours/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 11 for data subsets
p<0.01, ** p<0.05, * p<0.1

Table A18: Edge Effect and Soil Nutrients (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity
Perimeter-area ratio (log m/ha)	1.095***	1.217***	1.225***	1.150***
	(0.0981)	(0.186)	(0.416)	(0.175)
Soil organic carbon (%)		0.298		
		(0.247)		
(Soil organic carbon)X(Perimeter-area ratio)		-0.0413		
		(0.0341)		
Soil sand (%)			0.00873	
			(0.0487)	
(Soil sand)X(Perimeter-area ratio)			-0.00217	
			(0.00690)	
Soil nitrogen (%)				1.252
				(2.908)
(Soil nitrogen)X(Perimeter-area ratio)				-0.249
				(0.399)
Observations	2075	1794	1794	1794
Adjusted R^2	0.393	0.388	0.388	0.388

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 13
*** p<0.01, ** p<0.05, * p<0.1

Table A19: Edge Effect and Sunlight (Plot Panel)

	(1) Plot Productivity (All Crops)	(2) Plot Productivity (High Crops)	(3) Plot Productivity (Low Crops)	(4) Plot Productivity (Tree Crops)
Perimeter-area ratio (log m/ha)	1.095***	1.463***	1.471***	0.755***
	(0.0981)	(0.132)	(0.166)	(0.129)
Observations	2075	955	505	1214
Adjusted R^2	0.393	0.471	0.358	0.278

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 3 for plot subsets
*** p<0.01, ** p<0.05, * p<0.1

Table A20: Edge Effect and Biodiversity (Plot Panel)

	(1) Plot Productivity (All Plots)	(2) Plot Productivity (Monocropped)	(3) Plot Productivity (Mixed or Intercropped)
Perimeter-area ratio (log m/ha)	1.095***	0.981***	1.275***
	(0.0981)	(0.205)	(0.129)
Observations	2075	829	1245
Adjusted R^2	0.393	0.353	0.436

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 3 for plot subsets
*** p<0.01, ** p<0.05, * p<0.1

Appendix 8 Inverse Relationship Robustness

Several robustness checks confirm that the plot-level inverse relationship holds across rounds, across data-subsets, and (qualitatively) across functional forms. Table A21 estimates the relationship with household-year-season fixed effects for round 1 only (Column 1), for Round 2 only (column 2) and for both rounds, as in Panel 2 of Table 2. The inverse relationship is larger in magnitude for round 2, but in both years the relationship is strongly, statistically significant and of a magnitude comparable to previous studies. Tables A22 and A23 illustrate that the inverse relationship is fairly stable in magnitude across crop subsets and across managers and agricultural management styles.

Columns 2 and 3 of Table A24 shows that while some plots are growing across time and some plots are shrinking across time, the inverse relationship is estimated for both categories of change. Column 4 shows that while some 2013 plots are matched to multiple 2003 plots, the inverse relationship is statistically identical if we restrict the sample to only those plots that are matched one-to-one, across the decade.

Table A25 illustrates that the inverse relationship, typically estimated via logged variables since both land size (hectares) and productivity (revenue per hectare) are distributed log normally, can also be estimated via other functional forms. In Panel 1, log revenue per hectare is regressed on non-log versions of land size — as with the traditional version of this regression, we see that plot size drives the inverse relationship, not farm size. Because taking the log of a variable is a non-linear transformation, Panel 2 runs the same regression but including squared terms. These terms increase explanatory power (though not to the level achieved by logged variables), and results again confirm that plot size drives the inverse relationship rather than farm size. Panels 3 and 4 hold the same covariates, but use revenue per hectare rather than logged revenue per hectare as the dependent variable. Though statistical significance is lost on all variables, the inverse relationships still holds, and more importantly farm size again adds no additional information, conditional on knowing plot size.

Table A21: The Inverse Relationship by Round (HH-Time FE)

	(1) Plot Productivity (2003 only)	(2) Plot Productivity (2013 only)	(3) Plot Productivity (2003 & 2013)
Plot size (log ha)	-0.476***	-0.679***	-0.561***
	(0.0351)	(0.0362)	(0.0261)
Observations	2804	1977	4781
Adjusted R^2	0.118	0.257	0.169

Dependent variable: log(revenue/hectare)
Estimated with household-year-season fixed effects
Household-year-season-clustered standard errors in parentheses
Table estimates Equation 2 for data subsets by round
p<0.01, ** p<0.05, * p<0.1

Table A22: The Inverse Relationship by Crop (Plot Panel)

	(1) Plot Productivity (All Plots)	(2) Plot Productivity (Tubers)	(3) Plot Productivity (Cereal)	(4) Plot Productivity (Legumes)	(5) Plot Productivity (Banana)	(6) Plot Productivity (Cash Crops)
Plot size (log ha)	-0.623***	-0.611***	-0.897***	-0.492***	-0.459***	-0.536***
	(0.0655)	(0.139)	(0.103)	(0.107)	(0.0944)	(0.202)
Observations	2075	699	966	1017	802	518
Adjusted R^2	0.381	0.306	0.456	0.336	0.290	0.297

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 2 for data subsets by crop
p<0.01, ** p<0.05, * p<0.1

Table A23: The Inverse Relationship by Ownership/Management (Plot Panel)

	(1) Plot Productivity (Head Owns)	(2) Plot Productivity (Head Manages)	(3) Plot Productivity (Owns & Manages)	(4) Plot Productivity (Crops Rotated)	(5) Plot Productivity (Mono-Cropped)	(6) Plot Productivity (Mixed Cropping)
Plot size (log ha)	-0.625***	-0.476***	-0.511***	-0.739***	-0.612***	-0.813***
	(0.101)	(0.0713)	(0.0770)	(0.214)	(0.132)	(0.111)
Observations	1472	1228	1096	581	829	1105
Adjusted R^2	0.379	0.324	0.321	0.251	0.361	0.448

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 2 for data subsets by ownership/management
p<0.01, ** p<0.05, * p<0.1

Table A24: The Inverse Relationship by Plot Change Categories (Plot Panel)

	(1) Plot Productivity (All)	(2) Plot Productivity (Shrunk)	(3) Plot Productivity (Grew)	(4) Plot Productivity (Single-Matched)
Plot size (log ha)	-0.623***	-0.755***	-0.528**	-0.665***
	(0.0655)	(0.0861)	(0.227)	(0.0749)
Observations	2075	1443	632	1428
Adjusted R^2	0.381	0.461	0.117	0.441

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 2 for data subsets
p<0.01, ** p<0.05, * p<0.1

Table A25: The Inverse Relationship under Various Functional Forms (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity
	Panel 1: Log Productivity, Linear		
Farm size (ha)	-0.161**		0.00473
	(0.0764)		(0.0682)
Plot size (ha)		-0.577***	-0.580***
		(0.121)	(0.129)
Observations	2075	2075	2075
Adjusted R^2	0.264	0.285	0.285
	Panel 2: Log Productivity, Non-Linear		
Farm size (ha)	-0.423***		-0.0561
	(0.137)		(0.133)
(Farm size)^2	0.0467**		0.00919
	(0.0197)		(0.0194)
Plot size (ha)		-1.256***	-1.225***
		(0.279)	(0.287)
(Plot size)^2		0.211**	0.204**
		(0.0839)	(0.0854)
Observations	2075	2075	2075
Adjusted R^2	0.268	0.298	0.298
	Panel 3: Productivity, Linear		
Farm size (ha)	-126.7		132.3
	(126.0)		(88.30)
Plot size (ha)		-810.0	-907.8
		(571.9)	(616.1)
Observations	2076	2076	2076
Adjusted R^2	0.012	0.014	0.014
	Panel 4: Productivity, Non-Linear		
Farm size (ha)	-1016.6		-240.4
	(753.0)		(286.6)
(Farm size)^2	158.8		60.46
	(119.2)		(53.89)
Plot size (ha)		-3021.6	-2900.9
		(2025.1)	(1923.6)
(Plot size)^2		688.5	637.3
		(468.0)	(434.4)
Observations	2076	2076	2076
Adjusted R^2	0.014	0.021	0.021

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot clustered standard errors in parentheses
Table estimates equations similar to Equation 2
*** p<0.01, ** p<0.05, * p<0.1

Appendix 9 Causal Bounds around the IR

If plot size was randomly distributed, or if the 2003-2013 change in plot size was randomly distributed, the inverse relationship estimated under plot fixed effects would be causal. Appendix 5 indicates that this is not the case; plot size change over the decade is correlated with other plot-level characteristics.

Yet, the results of Table 4 show the relationship to be remarkably robust to additional controls. The coefficient on plot size, which we interpret as the inverse relationship, is statistically indistinguishable across columns, and actually rises in magnitude between Column 1 (univariate regression) and Column 6 (full controls).

In order to explore the likelihood of causality we calculate Oster's bias-adjusted estimator γ^*, as defined in Equation 6. We do this five times, allowing X_{ijt} from Equation 5 to take the value of each set of controls in Columns 2-5 of Table 4, as well as the full set of controls in Column 6. We assume $\delta = 1$ and $R_{max} = 1.3R_5$, as suggested by Oster (2016). (The bounds also rely on the relative contribution of each observed variable to plot size vs. productivity; Oster notes that if deviations from this restriction are not large, the estimated bounds will approximate the true bounds.)

These bounds are displayed in Table A26. None of them come close to containing zero. The last set of bounds — drawn from Column 6 of Table 4, controlling for all covariates — suggest that the causal effect of plot size on plot productivity is virtually identical to (if anything, higher than) the original, estimated effect. This is because controlling for these covariates increases R-squared while changing the estimated inverse relationship very little, in the end actually increasing the estimated relationship very slightly once all covariates are included.

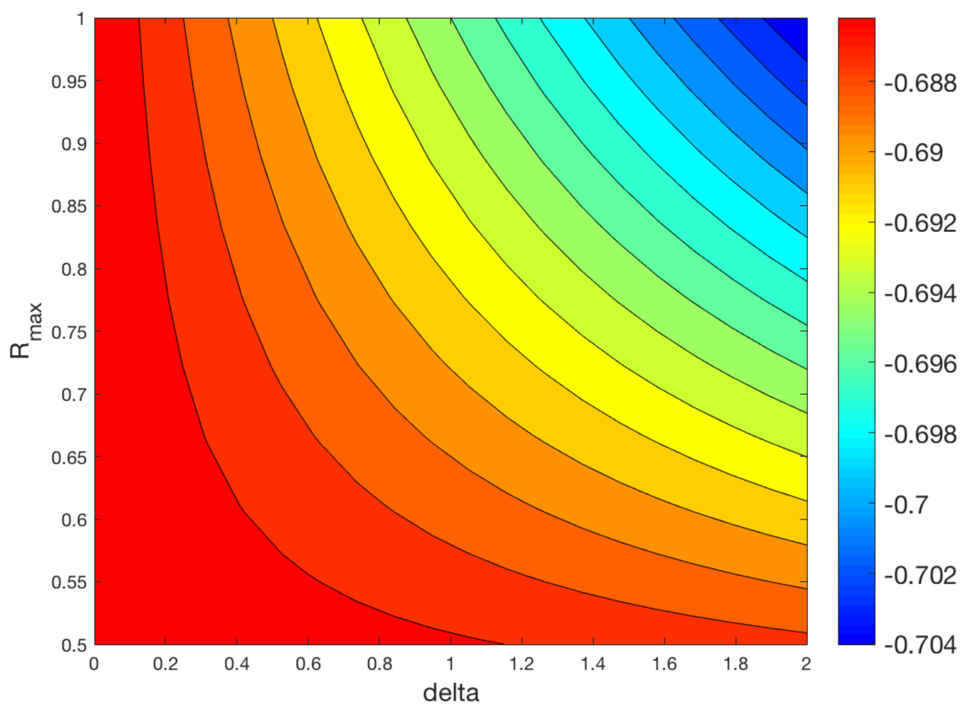
We can also alter the assumptions of $\delta = 1$ and $R_{max} = 1.3R_5$, to examine the range of bounds that are possible under various δ and R_{max} parameters. Figure A5 illustrates the bias-adjusted estimator γ^* calculated for every combination of $\delta \in [0, 2]$ and $R_{max} \in [0.5, 1]$, maintaining $\hat{\gamma_4} = -0.684$, $R_4 = 0.380$, $\hat{\gamma_5} = -0.685$, $R_5 = 0.439$, as in the final column of Table A26, i.e., based on the full set of controls in Table 4. Figure A5 illustrates that, if we consider the full set of time-varying plot controls as our observables, there is no feasible combination of δ and R_{max} parameters that suggests the causal inverse relationship to be over-estimated. All estimates suggest that, if anything, the relationship may be slightly under-estimated, hovering at some point slightly below -0.7.

Table A26: Bounds for Potential Causal Relationship t

	$R_{max} = 1.3 * R_5$, δ=1, $\hat{\gamma_4}$=-0.684, R_4=0.380				
	(1)	(2)	(3)	(4)	(5)
Column Controls	Soil	Inputs	Management	Crops	Full Set
R_5	0.388	0.399	0.404	0.400	0.439
Bounds $[\hat{\gamma_5}, \gamma^*(R_{max}, \delta)]$	[-0.677, -0.575]	[-0.616, -0.188]	[-0.733, -0.980]	[-0.743, -0.097]	[-0.685, -0.687]

Coefficient $\hat{\gamma_4}$ and R_4 from Column 1 of Table 4
Coefficients $\hat{\gamma_5}$ and R_5 from Columns 2-6 of Table 4

Figure A5: Bias-Adjusted Estimator $\gamma^*(R_{max}, \delta)$
($\hat{\gamma_4} = -0.684$, $R_4 = 0.380$, $\hat{\gamma_5} = -0.685$, $R_5 = 0.439$)



Appendix 10 More on the Edge Effect

Explaining Perimeter-Area Variation

Figure A6 illustrates a remarkably tight, linear, non-parametric relationship between plot size and perimeter-area ratio. It also displays kernel density distributions for both plot size and perimeter-area ratio. Neither distribution displays long tails on either side; medium-sized plots provide the bulk of variation both for plot size and perimeter-area ratio.

Because our primary results identify coefficients using plot fixed effects, Figure A7 displays the same non-parametric relationship and kernel density distributions, except for change in plot size over time, and change in perimeter-area ratio over time. Again, neither relationship displays a particularly long tail; the bulk of variation in both variables comes from the same, medium-change plots.

Table A27 examines the predictors of the perimeter-area ratio in our panel dataset via a regression framework. The first three columns are estimated via OLS, and so explain both cross-sectional variation and variation within plots over time. Columns 4-6 are estimated via plot fixed effects, and so explain only variation in perimeter-area ratio within plots over time. Column 1 shows that plots with interior vertices (that is, plots for which a corner cuts into the main area, perhaps tracing around an object) have higher perimeter area relationships, as we would expect. The perimeter-area ratio of a polygon theoretically decreases with number of sides — a triangle has the highest proportion of area around the periphery, and a circle the lowest. This appears in the data; in Column 1 we see that perimeter-area ratio is highest for triangular plots, and decreases at a diminishing rate with number of sides.

While Column 1 of Table A27 accounts for only plot shape, Column 2 of Table A27 accounts for only size. Shape variables explain 21 percent of variation in perimeter-area ratio, while plot size explains 88 percent of variation. This make it clear that plot size is the primary factor driving perimeter-area relationship. Yet plot shape variables still contribute to explanatory power in Column 3, even conditional on plot size. Roughly the same is seen under plot fixed effects, in Columns 4-6.

Causal Bounds Around the Edge Effect

Table A28 shows the edge effect to be robust to all controls previously considered — just as the inverse relationship was robust to these controls. The stability of the coefficient on perimeter-area ratio is again strongly suggestive of causality, and causal bounds can similarly be estimated along the lines suggested by Oster (2016).

As we did with the inverse relationship, we can calculate the univariate relationship between perimeter-area ratio and productivity, as in Equation 14. We can then add in controls X_{ijt}, as in Equation 15.

$$Y_{ijt} = \theta_4 \frac{P_{ijt}}{A_{ijt}} + \lambda_{ij} + \lambda_t + \varepsilon_{ijt} \tag{14}$$

$$Y_{ijt} = \theta_5 \frac{P_{ijt}}{A_{ijt}} + \beta X_{ijt} + \lambda_{ij} + \lambda_t + \varepsilon_{ijt} \tag{15}$$

Equation 14 is estimated in Column 1 of Table A28, and Equation 15 is estimated in Columns 2-6 of Table A28, allowing X_{ijt} to hold a different set of controls in each column. (These controls are identical to those considered in Table 5.)

We calculate the bias-adjusted perimeter-area coefficient θ^* as in Equation 16, where R_4 is the R-squared obtained by estimating the univariate relationship of Equation 14 (given in Column 1 of Table A28), and R_5 is the R-squared obtained by estimating the same relationship with controls as in Equation 15 (given in Columns 2-6 of Table A28).

$$\theta^*(R_{max}, \delta) = \hat{\theta}_5 - \delta\Big[\hat{\theta}_4 - \hat{\theta}_5\Big]\frac{R_{max} - R_5}{R_5 - R_4} \tag{16}$$

As done in Appendix 9 for the inverse relationship, we can then calculate bounding intervals for the coefficient on perimeter-area ratio, $\big[\hat{\theta}_5, \theta^*(min\{1.3R_5, 1\}, 1)\big]$, which follows Oster's suggested starting assumption of $\delta = 1$ and $R_{max} = 1.3R_5$.

These bounds are displayed in Table A29. They suggest that the causal parameter is somewhere around 1 or hovering above 1. Column 5 uses all controls to estimate the causal interval, and suggests that the causal perimeter lies within $[1.127, 0.975]$.

Additionally, using the full set of controls such that $\hat{\theta}_4 = 1.188$, $R_4 = 0.387$, $\hat{\theta}_5 = 1.127$, $R_5 = 0.440$, we calculate θ^* under the full possible range of R_{max} values and all plausible δ values. Figure A8 illustrates for almost all feasible combinations of δ and R_{max}, the bounding interval around a theoretically causal parameter will not contain zero. Only once $R_{max} > 9.44$ and $\delta > 1.68$ does the bounding interval contain zero.

Robustness of Edge Effect and the Role of Plot Shape

Tables A30 and A31 illustrate that the edge effect holds across crop subsets and ownership/management subsets, as does the inverse relationship. Additionally, it holds across plot size and perimeter-area ration quantiles, as illustrated by Table A32. (That is, like the inverse relationship, the effect is linear, rather than driven by extreme values of plot size or perimeter-area relationship.)

It is possible that the coefficient on perimeter-area ratio captures something about plot shape unrelated to peripheral productivity. It might be that plots with more perimeter per area (i.e., triangular plots, or 4-sided plots with acute angles) are more productive for reasons unrelated to the edge effect. If so, we would expect that controlling for plot shape directly, or for number of sides, would mitigate or eliminate the coefficient on perimeter-area ratio.

Columns 2 and 3 of Table A33 show that this is not the case — there is no direct, significant impact of shape on productivity, whether shape is quantified as number of plot sides (Column 2) or categorized into triangular, 4-sided, more than 4 sides (Column 3). The coefficient on perimeter-area ratio is unchanged between the base specification in Column 1, the specification controlling for plot sides in Column 2, and the within-plot-shape specification of Column 3. Figure A9 similarly illustrates that while productivity is, on average, highest for triangles and lowest for multi-sided plots, the difference is not significant.

However, Columns 4 and 5 of Table A33 demonstrate that the marginal impact of perimeter-area ratio does change with plot shape. For each additional plot side, the marginal impact of perimeter-area ratio decreases by 0.0762. For triangular plots, a ten percent increase in perimeter-area ratio results in a 14.4 percent increase in productivity; for plots with four sides this figure is 10.6 percent, and for plots with more than 4 sides (the omitted category) this figure is 7.9 percent. Figure A10, graphing productivity demeaned by year and season over perimeter-area ratio, illustrates that this differential can be seen in the (almost) raw data. Figure A11 again illustrates the differential by graphing productivity predicted by the regression model of Column 5, Table A33.

Investigating Multicollinearity

It is important to note that plot size and perimeter-area ratio are highly correlated, with a Pearson's Correlation Coefficient of -0.133. The log version of these coefficients is even more strongly correlated, with a correlation coefficient of -0.939. Essentially, perimeter-area ratio is a non-linear transform of plot size. It is logical, therefore, to be concerned that multicollinearity may in some way effect the coefficients estimated in Column 2 of Table 5, where both variables are included simultaneously as coefficients. We address this concern in a few ways.

First, Table A34 estimates the coefficient on perimeter-area ratio according to quintiles of correlation between plot size and perimeter-area relationship. (More specifically, log plot size is regressed on log perimeter-area ratio, and correlation quintiles are defined according to the residual.) The coefficient on perimeter-area ratio is stable across these quintiles.

Second, Table A35 displays a placebo test. We replace perimeter in the perimeter-area ratio with a new, placebo variable. This placebo variable has an identical distribution to perimeter but is randomly generated, and then divided by area in order to simulate a placebo version of the perimeter-area ratio. The logged versions of area and placebo-area ratio are highly correlated, with a correlation coefficient of -0.886 — close to the correlation of the true variables. Yet Table A35 illustrates that while placebo-area ratio has some explanatory power (Column 3), it is lower than the explanatory power of area alone (Column 1). This is in contrast to the true variable, which has higher explanatory power than area alone (Table 5). When both variables are controlled for simultaneously, the inverse relationship is unchanged. Again, this is in contrast to the results of Table 5, where the inverse relationship becomes zero once perimeter-area ratio is controlled for.

Third, Table A36 goes even further by replacing perimeter in perimeter-area ratio with a placebo variable that is not only identically distributed to perimeter, but is also similarly correlated with area. The logged versions of true perimeter and plot size are correlated with a coefficient of 0.934; the logged versions of this new placebo perimeter and plot size are correlated with a correlation coefficient of 0.932. Logged versions of placebo-area and plot size are correlated with a correlation coefficient of 0.939, just as with the true variables.

The results of this second placebo tests are identical to the first; the placebo-area variable has lower explanatory power than true area, and controlling for this variable in no way mitigates the inverse relationship. Together, Tables A34-A36 suggest that multicollinearity is unlikely to be driving the results in Table 5.

Alternate Indicators for Edge Effect

We hypothesize that perimeter-area ratio explains plot productivity because it is a proxy for the proportion of plot area that lies around the plot periphery. Essentially, the perimeter-area ratio specifies quantity perimeter per unit area. This suggests a less perfect proxy for the same concept — the number of plot sides per unit area. Number of sides and perimeter are highly related, with a correlation coefficient of 0.614. If quantity of perimeter per unit area drives up productivity, it would make sense for numbers of sides per unit area to also drive up productivity, and that controlling for this ratio would mitigate the inverse relationship.

This new ratio is actually more highly correlated with plot size than is perimeter-area ratio; while the logged version of plot size and perimeter-area ratio are correlated with a coefficient of -0.939, the logged versions of plot size and sides-area ratio are correlated with a coefficient of -0.952. (This is because there is less variation in number of sides than in perimeter.)

Table A37 estimates the same regressions as Table 5, but replacing perimeter with number of sides. Controlling for this ratio in Column 2 mitigates the inverse relationship, as we would expect if sides-area ratio is an imperfect proxy for proportion of plot area around the periphery. It does so less, however, than controlling for perimeter-area ratio. This makes sense, if perimeter-area ratio is a better proxy for the true variable of interest. (Additionally, it further suggests that multicollinearity does not drive the results in Table 5, since this new proxy is actually more collinear with plot size.) Column 3 shows that sides-area ratio does almost as well as plot size in explaining plot productivity — this was not true for the placebo tests in Tables A35 and A36.

An additional proxy can be found for the concept of perimeter, or perimeter-area ratio. We propose that any plot, "extra" or "unnecessary" perimeter can be quantified as the ratio between true perimeter and the perimeter that would be necessary for the same area, if the plot was a perfect square. While perimeter is correlated with area (r=0.85), the extra perimeter ratio is necessarily uncorrelated (r=0.04). We would expect extra perimeter to increase plot productivity, particularly when the extra perimeter is large with respect to plot area. Extra perimeter can therefore serve as a second, imperfect proxy for perimeter (one uncorrelated with area), and extra perimeter per hectare can serve as a proxy for perimeter-area ratio.

Table A38 estimates these regressions, similarly to Table A37. We see that indeed, the inverse relationship is mitigated again by half, and this time becomes insignificant once extra perimeter per hectare is controlled for. Column 3 implies that for every percentage increase in extra perimeter per hectare of plot, productivity increases by 0.6 percent.

Last, recall that we hypothesize that the inverse relationship exists largely because (across the entire sample) plot size is inversely correlated with perimeter-area ratio. And in fact, for about 90 percent of observations, if plot size goes up (down) between 2003 and 2013, perimeter-area ratio goes down (up). For a small selection of observations, however, perimeter-area ratio increases as plot size increases, or perimeter-area ratio decreases as plot size decreases. If the edge effect truly drives the inverse relationship, we would not expect to observe an inverse relationship for these observations. Rather, we would expect a positive relationship between plot size and productivity, as a decrease in perimeter-area ratio (and coincidentally in plot size) would drive a decrease in productivity, and an increase in perimeter-area ratio (and coincidentally in plot size) would drive a increase in productivity.

Table A39 tests this hypothesis. Column 1 estimates the inverse relationship for all observations for which plot size and perimeter-area ratio move in opposite directions over time. The estimated coefficient on plot size (-0.62) is almost identical to the estimated coefficient in full sample (-0.621). Column 2 displays results for the same regression, but for the sample of plots for which plot size and perimeter-area ratio move in the same direction. As hypothesized, for this sample of 234 plots there exists a positive relationship between plot size and productivity. Columns 3-5 run this same regression on random sub-samples of plots, in order to ensure that the result is not merely an artifact of reduced sample size. For all three columns, an inverse relationship is observed. (And while some sub-samples will result in a reduction of statistical significance or a reduction of coefficient magnitude, we have never observed a random sub-sample for which the estimated coefficient was positive.)

Magnitude of the Edge Effect

We estimate a large edge effect in our data, and it is logical to ask whether this effect, or the accompanying inverse relationship, can be simulated under reasonable assumptions. We walk through such a simulation in Figures A12 - A23.

Figures A12 and A13 illustrate the change in perimeter-area ratio that results from changes in plot shape (holding area constant) and changes in plot size (holding shape constant). Perimeter-area ratio increases as angles become more extreme (moving to the right in Figure A12) and decreases as size increases (moving to the right in Figure A13).

Figures A14 and A15 illustrate the accompanying changes in plot productivity (created as an area-weighted average of interior and peripheral productivity), given the simulated shape and size shifts. Peripheral area is taken to be 2 times interior productivity, a typically observed differential in controlled, agricultural trials, and the border area is taken to be a set length across all shapes (precisely, a tenth of the length of the squares's side). It is visually clear that productivity moves closely with perimeter area ratio — though not in an exactly linear fashion, given that perimeter-area ratio is merely a proxy for (peripheral area) / (total area).

Figures A16 and A17 illustrate the elasticity of productivity with respect to perimeter-area ratio, given the simulated shape and size shifts. These elasticities change across shapes, but for a peripheral-interior differential of 2, they hover between 0.2 and 0.27, meaning that a one percent increase in perimeter-area ratio — whether driven by changing shape or by changing size — increases productivity by around 0.2-0.27 percent. Notably, this elasticity is far less than edge effect parameter estimated in our results, which implied a one to one increase in productivity with perimeter-area ratio. Figures A18 and A19 again illustrate the elasticity of productivity with respect to perimeter-area ratio, but now assuming that peripheral area is 3.6 times interior productivity, as observe by Holman and Bednarz (2001). This assumption increases the elasticity considerably, now hovering between 0.37 and 0.46, though it still falls short of the one to one ratio observed in our estimations.

Why is the edge effect observed in our data so much greater than the simulated edge effect? We hazard that two factors likely play in. First, we are simulating a clear border region with differentially high productivity, and an interior region with homogenous productivity. Many papers instead note that productivity is higher around plot edges, and degrades slowly as one enters the plot (Holman and Bednarz, 2001). In such a case, there is not one peripheral area, but rather productivity is a degrading function of distance from plot edges. This would increase observed edge effect, though we do not simulate it as we have very little idea of what that function would look like.

Second, we are simulating productivity differentials observed in experimental agronomy trials — but such controlled trials will almost always observed a edge effect driven purely by biophysical factors. In farmer plots, where labor, input trends, harvesting intensity, etc., are not tightly controlled, behavioral mechanisms driven by farmer choices/behavior may increase the differential beyond that observed in controlled trials. This, too, would increase the edge effect.

Third, and even more speculatively, it is possible that plot shape drives productivity in some way that we do not understand. We take the association between perimeter-area ratio and productivity as strongly indicative of the edge effect. However, one might conservatively conclude that plot shape clearly plays into driving plot productivity, but that we do not fully understand why plot shape is so important.

Figures A20 - A23 illustrate the elasticity of productivity with respect to area. Figures A20 and A21 illustrate this elasticity while increasing plot size and holding shape constant. As with the edge effect, the inverse relationship observed is far lower than the inverse relationship estimated in our data. We simulate an inverse relationship of -0.07 to -0.17, while we estimate an inverse relationship of -0.6.

Figures A22 and A23 illustrate the same elasticity, but under simulated, simultaneous shifts in both plot shape and plot size. In Figure A22, the effect of the shape shifts dominates the effect of the size shifts, and perimeter-area ratio rises. Hence, productivity increases, and so we do not observe the inverse relationship. In Figure A23, the effect of the size shifts dominate the effect of the shape shifts, and perimeter-area ratio decreases. Hence, productivity decreases, and we observe an inverse relationship. This simulates the effect we found empirically in Table A39: for plots where size increases while perimeter-area ratio also increases, or plots where size decreases while perimeter-area ratio also decreases, we do not expect to find an inverse relationship. Instead, we expect to estimate a positive relationship between plot size and productivity for these plots.

Figure A6: Plot Size vs. Perimeter-Area Ratio

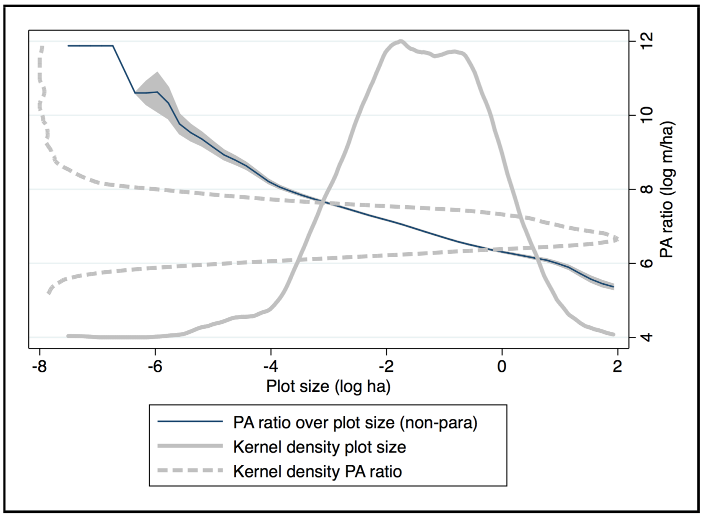


Figure A7: Plot Size vs. Perimeter-Area Ratio: Change between 2003 and 2013

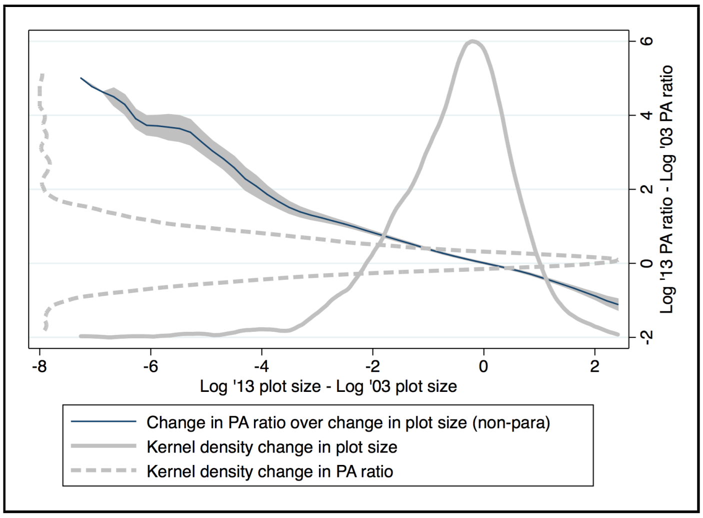


Table A27: Perimeter-Area Ratio Explained by Shape and Size (Plot Panel)

	OLS			Plot Fixed Effects		
	(1) P-A Ratio	(2) P-A Ratio	(3) P-A Ratio	(4) P-A Ratio	(5) P-A Ratio	(6) P-A Ratio
Plot has inner vertices (binary)	0.193***		0.0503***	0.0239		0.0604***
	(0.0464)		(0.0158)	(0.0462)		(0.0163)
Plot has 3 sides (binary)	1.127***		0.398***	1.065***		0.299***
	(0.242)		(0.103)	(0.283)		(0.0787)
Plot has 4 sides (binary)	0.0221		0.0107	0.172**		0.00569
	(0.0606)		(0.0177)	(0.0745)		(0.0230)
Number of sides (#)	-0.214***		-0.00810	-0.0331		0.0142
	(0.0277)		(0.0110)	(0.0451)		(0.0167)
(Number of sides)^2	0.00717***		0.00204***	-0.000914		0.000765
	(0.00112)		(0.000499)	(0.00226)		(0.000846)
GPS-measured plot size (log ha)		-0.512***	-0.531***		-0.533***	-0.581***
		(0.0108)	(0.00902)		(0.0176)	(0.0154)
Observations	2064	2076	2064	2064	2076	2064
Adjusted R^2	0.213	0.882	0.906	0.245	0.879	0.910

Dependent variable: log(perimeter/hectare)
Cols 4-6: Estimated via OLS
Cols 4-6: Estimated with plot fixed effects
Plot-clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A28: Robustness of Edge Effect (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity	(5) Plot Productivity	(6) Plot Productivity
Perimeter-area ratio (log m/ha)	1.203***	1.189***	1.097***	1.248***	1.255***	1.143***
	(0.140)	(0.145)	(0.138)	(0.131)	(0.132)	(0.127)
Soils	No	Yes	No	No	No	Yes
Inputs	No	No	Yes	No	No	Yes
Management	No	No	No	Yes	No	Yes
Crops	No	No	No	No	Yes	Yes
Observations	1546	1546	1546	1546	1546	1546
Adjusted R^2	0.394	0.400	0.409	0.410	0.407	0.437
R^2	0.395	0.402	0.412	0.414	0.410	0.445

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 10 with various controls
*** p<0.01, ** p<0.05, * p<0.1

Table A29: Bounds for Potential Causal Relationship

	$R_{max} = 1.3 * R_5$, δ=1, $\hat{\theta}_4$=1.203, R_4=0.395				
	(1)	(2)	(3)	(4)	(5)
Column Controls	Soil	Inputs	Management	Crops	Full Set
R_5	0.402	0.412	0.414	0.410	0.445
Bounds $[\hat{\theta}_5 \ \theta^*(R_{max}, \delta)]$	[1.189, 0.948]	[1.097, 0.326]	[1.248, 1.542]	[1.255, 1.681]	[1.143, 0.982]

Coefficient $\hat{\theta}_4$ and R_4 from Column 1 of Table A28
Coefficients $\hat{\theta}_5$ and R_5 from Columns 2-6 of Table A28

Figure A8: Bias-Adjusted Estimator $\theta^*(R_{max}, \delta)$
($\hat{\theta}_4 = 1.203$, $R_4 = 0.395$, $\hat{\theta}_5 = 1.143$, $R_5 = 0.445$)

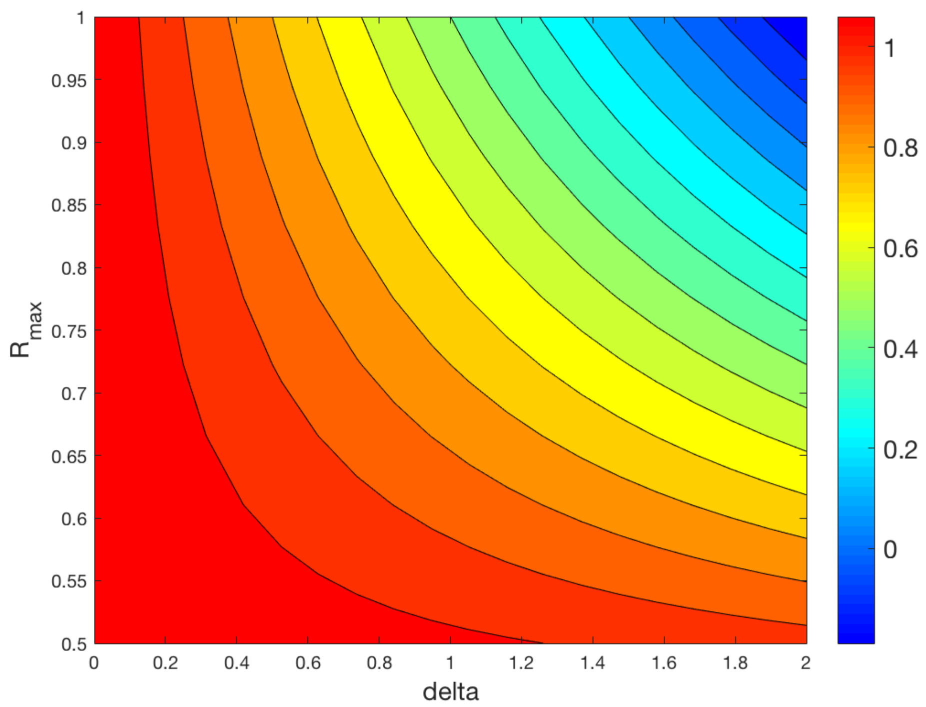


Table A30: The Edge Effect by Crop (Plot Panel)

	(1) Plot Productivity (All Plots)	(2) Plot Productivity (Tubers)	(3) Plot Productivity (Cereal)	(4) Plot Productivity (Legumes)	(5) Plot Productivity (Banana)	(6) Plot Productivity (Cash Crops)
Perimeter-area ratio (log m/ha)	1.095***	1.172***	1.481***	0.878***	0.741***	1.010***
	(0.0981)	(0.197)	(0.125)	(0.168)	(0.139)	(0.330)
Observations	2075	699	966	1017	802	518
Adjusted R^2	0.393	0.312	0.478	0.340	0.284	0.304

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 10 for data subsets by crop
p<0.01, ** p<0.05, * p<0.1

Table A31: The Edge Effect by Ownership/Management (Plot Panel)

	(1) Plot Productivity (Head Owns)	(2) Plot Productivity (Head Manages)	(3) Plot Productivity (Owns & Manages)	(4) Plot Productivity (Crops Rotated)	(5) Plot Productivity (Mono-Cropped)	(6) Plot Productivity (Mixed Cropping)
Perimeter-area ratio (log m/ha)	1.183***	0.862***	0.987***	1.408***	0.981***	1.324***
	(0.148)	(0.136)	(0.147)	(0.299)	(0.205)	(0.135)
Observations	1472	1228	1096	581	829	1105
Adjusted R^2	0.403	0.318	0.324	0.271	0.353	0.463

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 10 for data subsets by ownership/management
p<0.01, ** p<0.05, * p<0.1

Table A32: Edge Effect by Quantiles (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity
Perimeter-area ratio (log m/ha)	1.095***		
	(0.0981)		
P-A ratio quantiles=1 × Perimeter-area ratio (log m/ha)		1.355***	
		(0.192)	
P-A ratio quantiles=2 × Perimeter-area ratio (log m/ha)		1.334***	
		(0.180)	
P-A ratio quantiles=3 × Perimeter-area ratio (log m/ha)		1.257***	
		(0.171)	
P-A ratio quantiles=4 × Perimeter-area ratio (log m/ha)		1.262***	
		(0.163)	
P-A ratio quantiles=5 × Perimeter-area ratio (log m/ha)		1.277***	
		(0.150)	
Plot size quantiles=1 × Perimeter-area ratio (log m/ha)			1.112***
			(0.153)
Plot size quantiles=2 × Perimeter-area ratio (log m/ha)			1.092***
			(0.164)
Plot size quantiles=3 × Perimeter-area ratio (log m/ha)			1.054***
			(0.172)
Plot size quantiles=4 × Perimeter-area ratio (log m/ha)			1.113***
			(0.180)
Plot size quantiles=5 × Perimeter-area ratio (log m/ha)			1.119***
			(0.190)
Observations	2075	2075	2075
Adjusted R^2	0.393	0.405	0.400

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Table estimates Equation 10 for data subsets
*** p<0.01, ** p<0.05, * p<0.1

Table A33: Edge Effect by Plot Shape (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity	(5) Plot Productivity
Perimeter-area ratio (log m/ha)	1.095*** (0.0981)	1.110*** (0.102)	1.109*** (0.0975)	1.485*** (0.164)	0.756*** (0.125)
Number of sides (#)		0.00929 (0.0152)		0.565*** (0.178)	
Plot has 3 sides (binary)			-0.0121 (0.300)		-8.185*** (1.907)
Plot has 4 sides (binary)			-0.124 (0.0963)		-2.510** (1.000)
(Perimeter-area ratio)x(Number of sides)				-0.0844*** (0.0269)	
(Perimeter-area ratio)x(Plot has 3 sides)					1.042*** (0.242)
(Perimeter-area ratio)x(Plot has 4 sides)					0.348** (0.144)
Observations	2075	2063	2075	2063	2075
Adjusted R^2	0.393	0.396	0.393	0.403	0.407

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Figure A9: Productivity by Shape

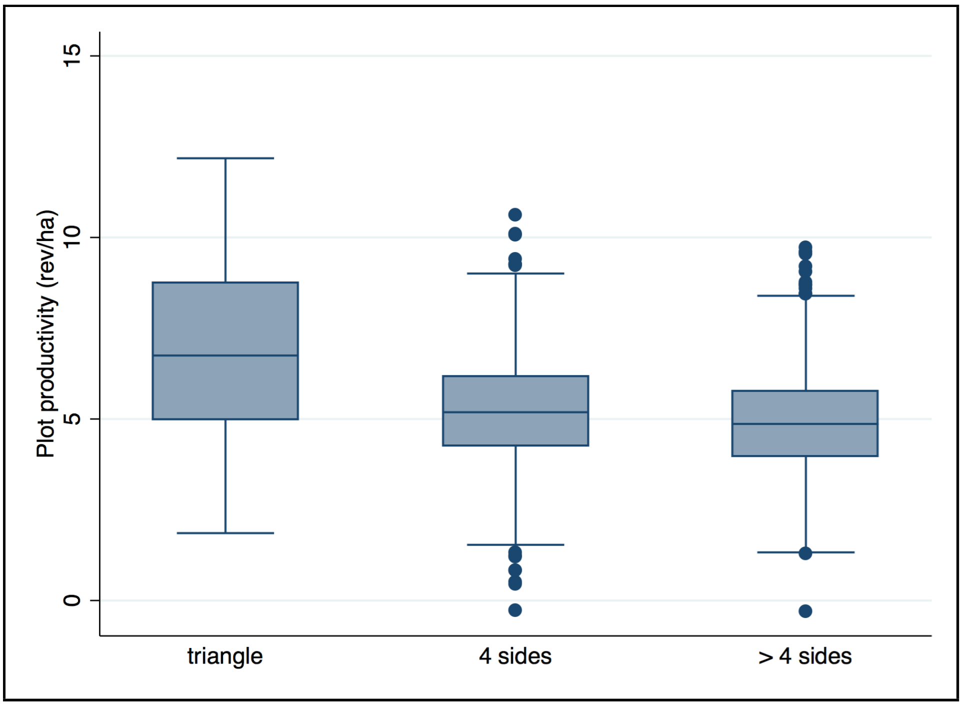


Figure A10: Productivity Demeaned by Round and Season

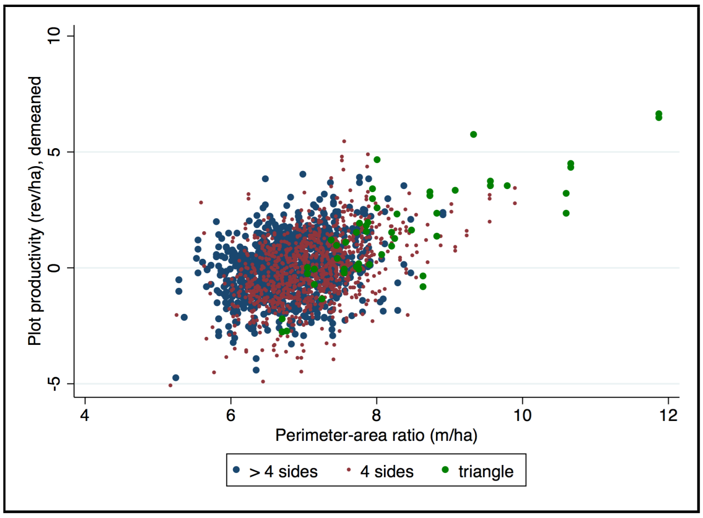


Figure A11: Productivity Prediction from Table A33 Col 5

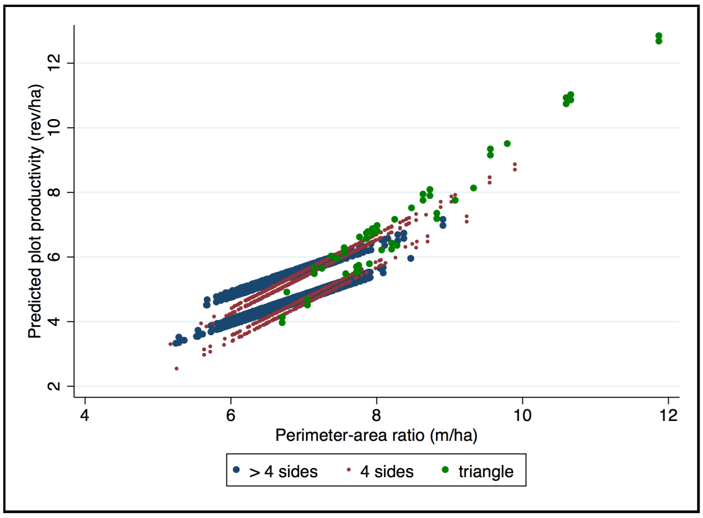


Table A34: Edge Effect by Area Correlation with Perimeter-Area Ratio (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity
GPS-measured plot size (log ha)	-0.0672	
	(0.139)	
Correlated Q1 × Perimeter-area ratio (log m/ha)	1.002***	1.111***
	(0.259)	(0.100)
Correlated Q2 × Perimeter-area ratio (log m/ha)	1.047***	1.156***
	(0.260)	(0.100)
Correlated Q3 × Perimeter-area ratio (log m/ha)	1.024***	1.133***
	(0.260)	(0.101)
Correlated Q4 × Perimeter-area ratio (log m/ha)	1.021***	1.130***
	(0.258)	(0.100)
Correlated Q5 × Perimeter-area ratio (log m/ha)	1.002***	1.108***
	(0.251)	(0.0974)
Observations	2075	2075
Adjusted R^2	0.397	0.397

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Table A35: Edge Effect Placebo Test (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity
GPS-measured plot size (log ha)	-0.623***	-0.634***		-0.623***
	(0.0655)	(0.0713)		(0.0655)
Placebo-area ratio (log m/ha)		-0.0118	0.348***	
		(0.0437)	(0.0541)	
Placebo (log m)				-0.0118
				(0.0437)
Observations	2075	2075	2075	2075
Adjusted R^2	0.381	0.381	0.327	0.381

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Table A36: Edge Effect Placebo Test with Multicollinearity (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity
GPS-measured plot size (log ha)	-0.623***	-0.627***		-0.618***
	(0.0655)	(0.0937)		(0.0873)
Placebo-area ratio (log m/ha)		-0.00963	0.907***	
		(0.125)	(0.115)	
Placebo (log m)				-0.00963
				(0.125)
Observations	2075	2075	2075	2075
Adjusted R^2	0.381	0.381	0.348	0.381

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Table A37: Suggestive Edge Effect: Number of Sides (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity
GPS-measured plot size (log ha)	-0.623***	-0.352***		-0.680***
	(0.0655)	(0.113)		(0.0714)
Sides-area ratio (log #/ha)		0.328**	0.689***	
		(0.129)	(0.0734)	
Sides (log #)				0.328**
				(0.129)
Observations	2075	2063	2063	2063
Adjusted R^2	0.381	0.388	0.382	0.388

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Table A38: Suggestive Edge Effect: Extra Perimeter (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity
GPS-measured plot size (log ha)	-0.623***	0.324		-0.526***
	(0.0655)	(0.240)		(0.0512)
Extra perimeter† / area (log ratio/ha)		0.899***	0.604***	
		(0.234)	(0.0567)	
Extra perimeter† (ratio)				0.175***
				(0.0302)
Observations	2075	2075	2075	2075
Adjusted R^2	0.381	0.393	0.392	0.402

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
†Extra perimeter = (true perimeter) / (perimeter if square)
p<0.01, ** p<0.05, * p<0.1

Table A39: Inverse Relationship Relies on PA Ratio (Plot Panel)

	(1) Plot Productivity (Same Direction)	(2) Plot Productivity (Opposite Directions)	(3) Plot Productivity (Random Sample)	(4) Plot Productivity (Random Sample)	(5) Plot Productivity (Random Sample)
GPS-measured plot size (log ha)	-0.621***	0.428	-0.810***	-0.260	-0.643***
	(0.0667)	(0.283)	(0.263)	(0.167)	(0.168)
Observations	1845	230	208	206	210
Adjusted R^2	0.400	0.137	0.424	0.312	0.340

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Figure A12: Perimeter-area Ratio Changing Shape

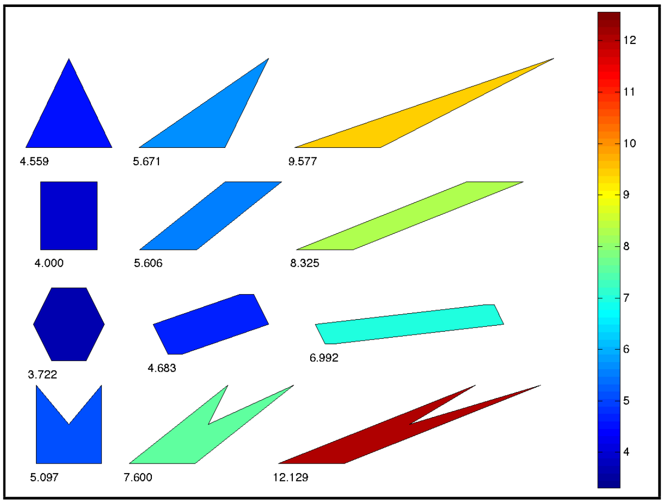


Figure A13: Perimeter-area Ratio Changing Size

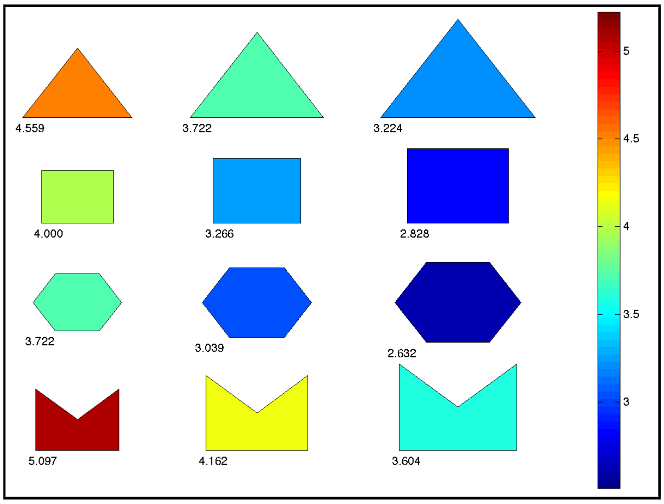


Figure A14: Productivity ($Y^P = 2Y^I$) Changing Shape

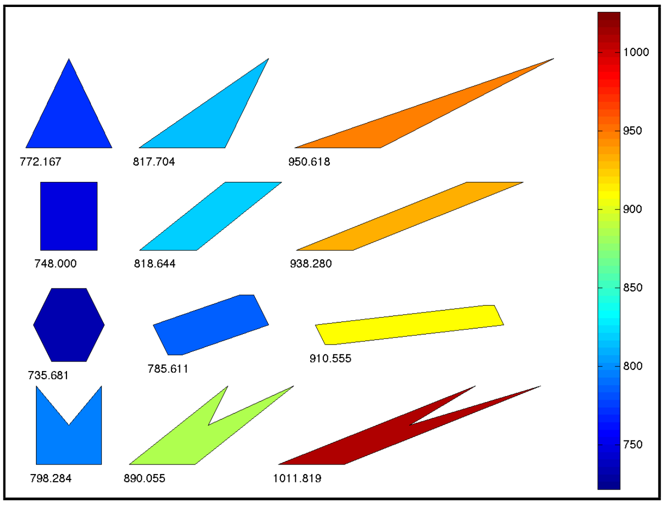


Figure A15: Productivity ($Y^P = 2Y^I$) Changing Size

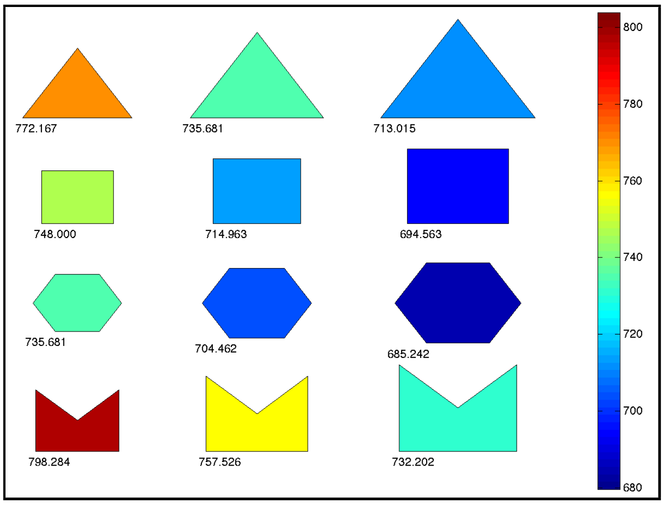


Figure A16: Elasticity wrt PA ratio ($Y^P = 2Y^I$): Changing Shape

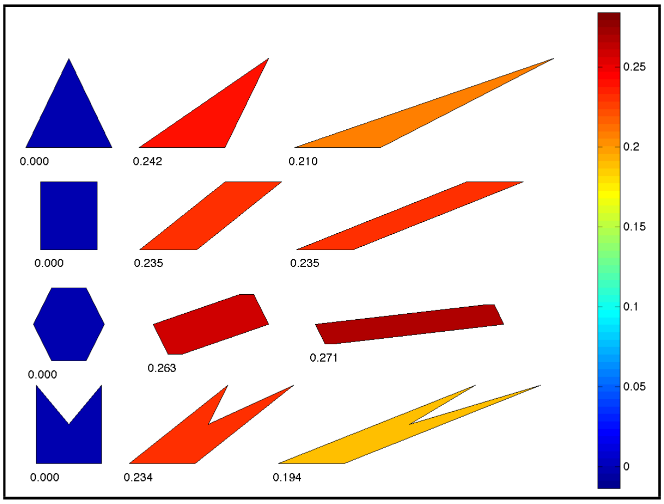


Figure A17: Elasticity wrt PA ratio ($Y^P = 2Y^I$): Changing Size

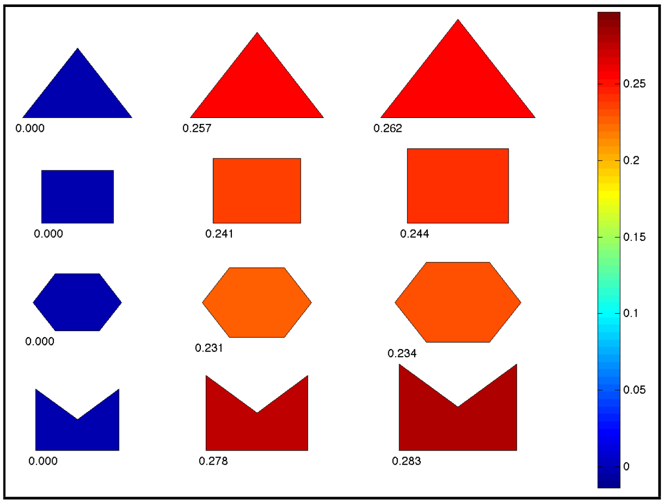


Figure A18: Elasticity wrt PA ratio ($Y^P = 3.6Y^I$): Changing Shape

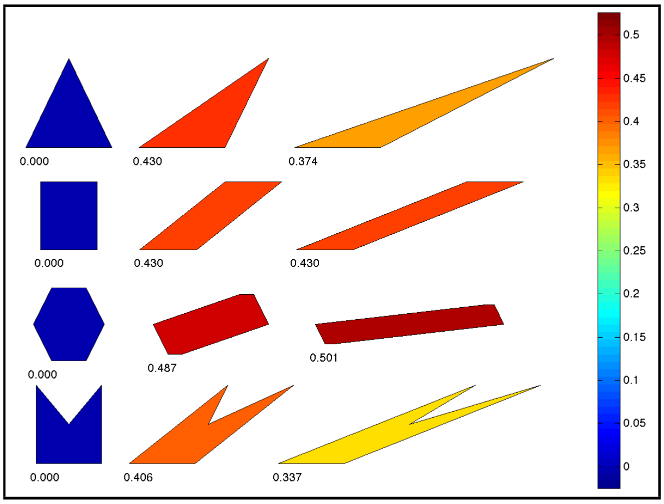


Figure A19: Elasticity wrt PA ratio ($Y^P = 3.6Y^I$): Changing Size

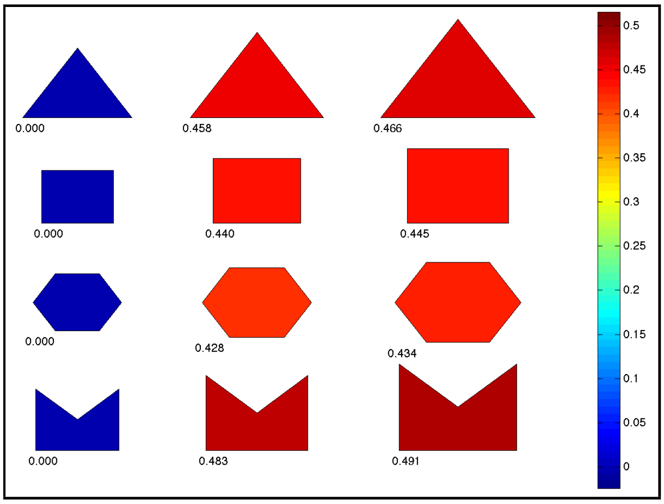


Figure A20: Elasticity wrt Area ($Y^P = 2Y^I$): Changing Size

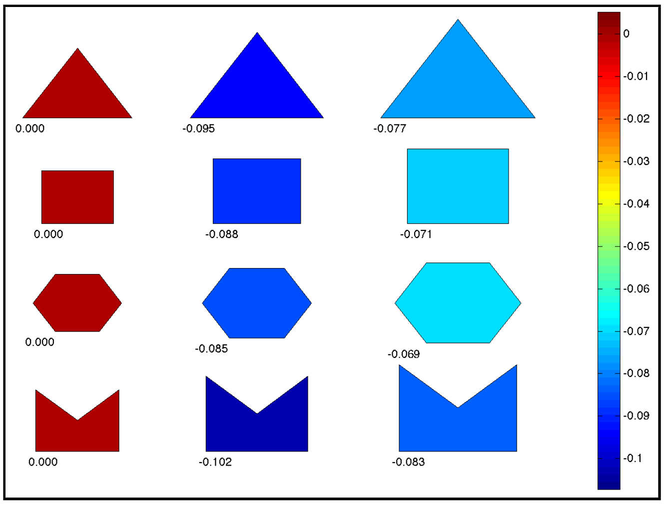


Figure A21: Elasticity wrt Area ($Y^P = 3.6Y^I$): Changing Size

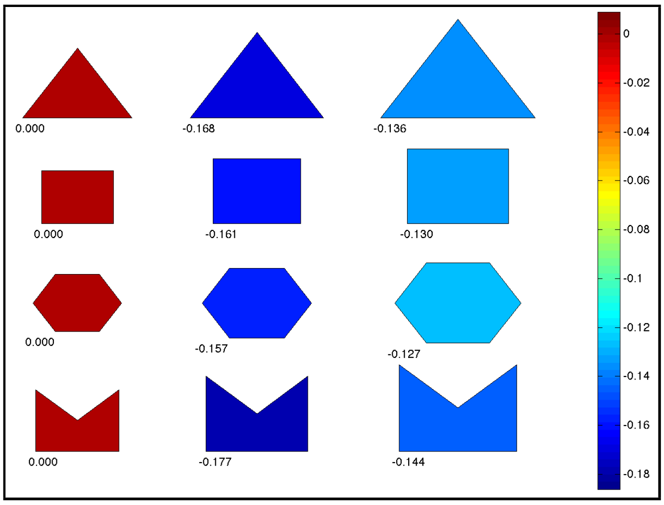


Figure A22: Elasticity wrt Area ($Y^P = 2Y^I$): Changing Shape and Size

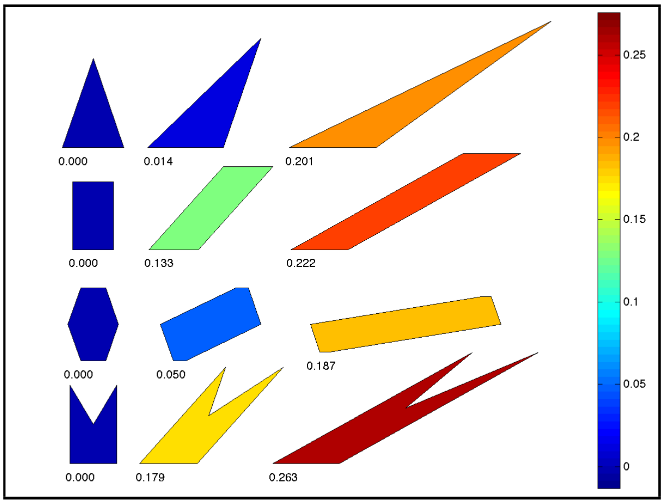
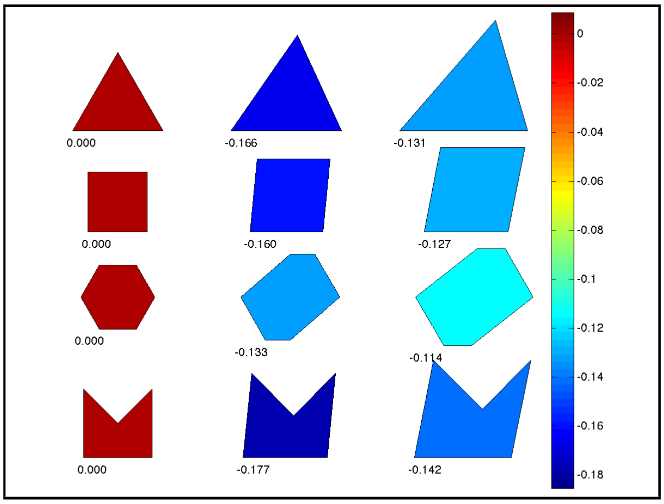


Figure A23: Elasticity wrt Area ($Y^P = 3.6Y^I$): Changing Shape and Size



Appendix 11 Exogeneity of Perception Error

When pooling data across plots and rounds, Figures A2 and A3 illustrate a clear, non-parametric relationship between perception error, plot size and the perimeter-area ratio. Over-estimation is negatively correlated with plot area and positively correlated with the perimeter-area ratio. Under-estimation moves in the opposite direction, though with a slightly noisier relationship. (Far more plots are over-estimated than under-estimated, and so the noise around under-estimation may be due to small sample size.) In both cases, perception error is measured in absolute terms, as a percent of the GSP-measured plot area.

Figure A24: Plot Size Perception Error over Plot Size

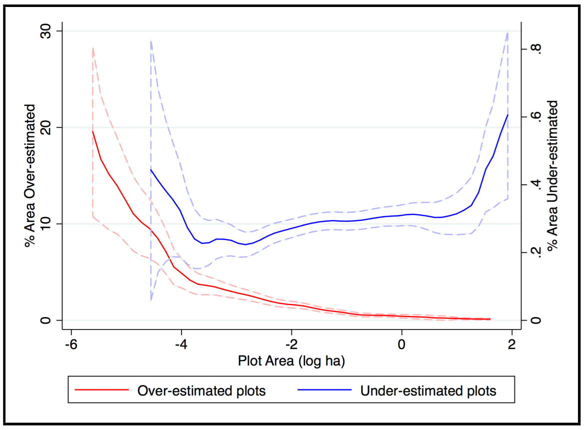
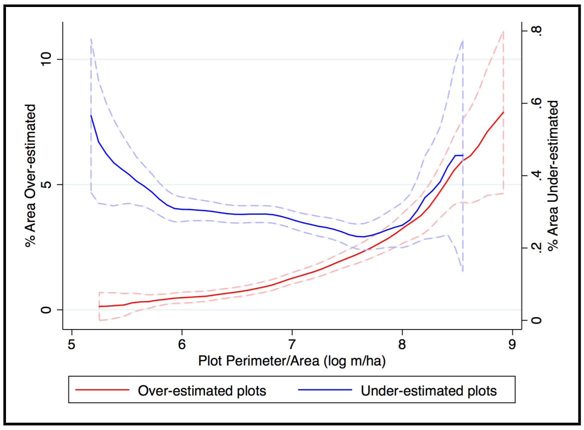


Figure A25: Plot Size Perception Error over Perimeter-Area Ratio



However, Column 1 of Table A40 shows that under plot fixed effects, neither plot area nor the perimeter-area ratio predict the over-estimation of plot size. (The same is true if plot area and the perimeter-area ratio are controlled for in a linear fashion. We choose quadratic controls due to the shape of the relationships in Figures 1 and 2.) Columns 2-5 control for other plot conditions — the same covariates that were considered as potential omitted variables in Table 4. The majority of these variables are also unrelated to over-estimation; only a few coefficients are significant, no more than one might expect by chance.

Column 1 of Table A41 similarly models the under-estimation of plot size as a quadratic function of plot area and the perimeter-area ratio, under a plot fixed effect model. In this case, it does appear that the perimeter-area ratio is weakly, negatively related to under-estimation. Yet conditional on plot area and the perimeter-area ratio, most other variables are unrelated to over-estimation. Only irrigation and tubers grown are significantly related, but only 6 observations are used to pick up irrigation variation under the plot fixed effect model, making this a volatile coefficient.

Last, Table A42 models the binary indicator for over-estimation of plot size, using the same covariates again. As with Tables A40 and A41, this binary variables appears unrelated to plot conditions under the fixed effect model, once conditioned on plot area and plot perimeter-area ratio.

By and large, it appears likely that perception error, under the fixed effect model, is exogenous to plot conditions, perhaps with the exception of crop choice.

Table A40: Exogeneity of Continuous Over-Estimation (Plot Panel)

	(1) Over- Estimation	(2) Over- Estimation	(3) Over- Estimation	(4) Over- Estimation	(5) Over- Estimation
GPS-measured plot size (log ha)	0.209 (0.635)	0.431 (0.634)	0.0779 (0.674)	-0.0340 (0.812)	-0.0841 (0.798)
(GPS-measured plot size)^2	0.135 (0.257)	0.114 (0.301)	0.105 (0.294)	0.167 (0.343)	0.133 (0.243)
Perimeter-area ratio (log m/ha)	-16.08 (13.00)	-16.13 (12.24)	-17.47 (13.50)	-17.90 (14.82)	-14.55 (12.62)
(Perimeter-area ratio)^2	1.268 (0.918)	1.269 (0.875)	1.335 (0.968)	1.341 (1.053)	1.147 (0.877)
Soil pH (pH)		8.262* (4.476)			
Soil pH^2 (pH^2)		-0.681* (0.382)			
Soil sand (%)		0.0102 (0.0219)			
Soil organic carbon (%)		0.224 (0.215)			
Labor intensity (log hrs/ha/day)			0.0920 (0.135)		
Organic amendment (binary)			0.634 (0.567)		
Inorganic fertilizer (binary)			0.827 (1.196)		
Irrigation (binary)			-0.216 (0.574)		
Terracing (binary)			0.716 (0.515)		
Head owns plot (binary)				-0.00514 (0.960)	
Head manages plot (binary)				-1.239 (1.276)	
(Head owns)X(Head manages)				0.731 (1.387)	
Crops are rotated (binary)				-0.172 (0.316)	
Crops are mono-cropped (binary)				-1.194 (0.957)	
Mixed cropping (binary)				-0.511 (0.849)	
Tubers grown (binary)					0.0267 (0.478)
Cereals grown (binary)					0.120 (0.436)
Legumes grown (binary)					1.103** (0.455)
Bananas grown (binary)					1.130*** (0.417)
Cash crops grown (binary)					0.828 (0.504)
Observations	824	726	780	726	824
Adjusted R^2	0.310	0.338	0.310	0.332	0.332

Estimated with plot and year fixed effects; Plot-clustered standard errors in parentheses
Observations include plots that were over-estimated; 1 observation per year
p<0.01, ** p<0.05, * p<0.1

Table A41: Exogeneity of Continuous Under-Estimation (Plot Panel)

	(1) Under-Estimation	(2) Under-Estimation	(3) Under-Estimation	(4) Under-Estimation	(5) Under-Estimation
GPS-measured plot size (log ha)	-0.174*	-0.178	-0.154	-0.189	-0.153*
	(0.0932)	(0.149)	(0.118)	(0.118)	(0.0897)
(GPS-measured plot size)^2	-0.0609*	-0.0735*	-0.0514	-0.0272	-0.0634
	(0.0368)	(0.0427)	(0.0441)	(0.0390)	(0.0417)
Perimeter-area ratio (log m/ha)	-4.800***	-3.846**	-4.469**	-3.524**	-4.491***
	(1.726)	(1.938)	(2.183)	(1.703)	(1.735)
(Perimeter-area ratio)^2	0.335**	0.271*	0.308*	0.229*	0.311**
	(0.133)	(0.145)	(0.166)	(0.127)	(0.134)
Soil pH (pH)		1.291**			
		(0.534)			
Soil pH^2 (pH^2)		-0.117**			
		(0.0464)			
Soil sand (%)		-0.00153			
		(0.00306)			
Soil organic carbon (%)		-0.0154			
		(0.0231)			
Labor intensity (log hrs/ha/day)			0.00799		
			(0.0257)		
Organic amendment (binary)			0.0802		
			(0.0906)		
Inorganic fertilizer (binary)			0.148		
			(0.103)		
Irrigation (binary)			-0.356*		
			(0.185)		
Terracing (binary)			0.0636		
			(0.0670)		
Head owns plot (binary)				0.156**	
				(0.0711)	
Head manages plot (binary)				0.164	
				(0.115)	
(Head owns)X(Head manages)				-0.130	
				(0.139)	
Crops are rotated (binary)				0.00349	
				(0.0551)	
Crops are mono-cropped (binary)				-0.0102	
				(0.0774)	
Mixed cropping (binary)				0.0242	
				(0.0725)	
Tubers grown (binary)					-0.0916*
					(0.0500)
Cereals grown (binary)					-0.0536
					(0.0460)
Legumes grown (binary)					-0.0687
					(0.0486)
Bananas grown (binary)					0.0613
					(0.0793)
Cash crops grown (binary)					-0.138*
					(0.0737)
Observations	572	479	547	521	572
Adjusted R^2	0.136	0.162	0.169	0.195	0.198

Estimated with plot and year fixed effects; Plot-clustered standard errors in parentheses
Observations include plots that were under-estimated; 1 observation per year
p<0.01, ** p<0.05, * p<0.1

Table A42: Exogeneity of Binary Over-Estimation (Plot Panel)

	(1) Over-Est Binary	(2) Over-Est Binary	(3) Over-Est Binary	(4) Over-Est Binary	(5) Over-Est Binary
GPS-measured plot size (log ha)	-0.0282	0.0124	0.0305	-0.0384	-0.0831
	(0.0712)	(0.0883)	(0.0826)	(0.0925)	(0.0760)
(GPS-measured plot size)^2	0.0333**	0.0329*	0.0439**	0.0245	0.0259
	(0.0166)	(0.0199)	(0.0186)	(0.0214)	(0.0169)
Perimeter-area ratio (log m/ha)	1.141***	1.413***	1.584***	1.124**	0.971**
	(0.417)	(0.488)	(0.474)	(0.530)	(0.427)
(Perimeter-area ratio)^2	-0.0716***	-0.0837***	-0.0967***	-0.0679**	-0.0621**
	(0.0267)	(0.0309)	(0.0307)	(0.0343)	(0.0272)
Soil pH (pH)		-0.164			
		(0.467)			
Soil pH^2 (pH^2)		0.0137			
		(0.0383)			
Soil sand (%)		-0.00181			
		(0.00212)			
Soil organic carbon (%)		0.000945			
		(0.0139)			
Labor intensity (log hrs/ha/day)			-0.0344**		
			(0.0158)		
Organic amendment (binary)			-0.0321		
			(0.0539)		
Inorganic fertilizer (binary)			0.0604		
			(0.151)		
Irrigation (binary)			-0.0104		
			(0.220)		
Terracing (binary)			0.0783		
			(0.0499)		
Head owns plot (binary)				-0.0506	
				(0.0628)	
Head manages plot (binary)				-0.0990	
				(0.0855)	
(Head owns)X(Head manages)				0.132	
				(0.0995)	
Crops are rotated (binary)				0.0236	
				(0.0432)	
Crops are mono-cropped (binary)				-0.0186	
				(0.0632)	
Mixed cropping (binary)				0.0247	
				(0.0636)	
Tubers grown (binary)					-0.00422
					(0.0407)
Cereals grown (binary)					0.0210
					(0.0399)
Legumes grown (binary)					0.0783**
					(0.0373)
Bananas grown (binary)					0.0154
					(0.0555)
Cash crops grown (binary)					0.116**
					(0.0569)
Observations	1400	1209	1330	1250	1400
Adjusted R^2	0.170	0.152	0.176	0.160	0.177

Estimated with plot and year fixed effects; Plot-clustered standard errors in parentheses
Observations include all plots; 1 observation per year
p<0.01, ** p<0.05, * p<0.1

Appendix 12 More on Perception Error

Tables A43 and A44 indicate that the productivity impacts of plot size misperception are qualitatively unchanged across crop and ownership/management categories, though significance is lost on many coefficients. Over-estimation of plot size is associated with higher productivity (with diminishing returns) for all crops, and under-estimation of plot size is associated with lower productivity (with diminishing returns) for all crops but banana. Over-estimation of plot size is associated with higher productivity (with diminishing returns) for plots under all categories of ownership and management, though the effect of under-estimation is lost.

Additionally, Table A45 shows that labor intensity increases with over-estimation of plot size, significantly with diminishing returns, just as productivity does. The effect of under-estimation is lost, perhaps because there is far less variation in under-estimation than in over-estimation, making these coefficients difficult to estimate if the effect is weak.

Table A43: The Effects of Farmer Misperception of Plot Size by Crop (Plot Panel)

	(1) Plot Productivity (Tubers)	(2) Plot Productivity (Cereal)	(3) Plot Productivity (Legumes)	(4) Plot Productivity (Banana)	(5) Plot Productivity (Cash Crops)
Farmer over-estimates plot (binary)	-0.758**	-0.234	-0.223	-0.160	-0.185
	(0.342)	(0.213)	(0.215)	(0.287)	(0.345)
Over-estimate (% area)	0.0277	0.00575	0.139*	0.209***	0.122
	(0.102)	(0.0742)	(0.0732)	(0.0557)	(0.0830)
Over-estimate squared	0.000819	-0.000976	-0.00387	-0.00601***	-0.00302
	(0.00338)	(0.00264)	(0.00236)	(0.00224)	(0.00264)
Under-estimate (% area)	-7.953***	-2.059	-3.328**	2.201	-0.892
	(2.496)	(1.277)	(1.393)	(1.724)	(2.276)
Under-estimate squared	11.80***	2.015	4.195**	-1.601	1.695
	(3.570)	(1.468)	(1.813)	(1.985)	(2.929)
Observations	699	966	1017	802	518
Adjusted R^2	0.346	0.501	0.390	0.335	0.360

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Plot area and area-perimeter ratio are controlled for quadratically in all columns
Table estimates Equation 12 for data subsets
p<0.01, ** p<0.05, * p<0.1

Table A44: The Effects of Farmer Misperception of Plot Size by Ownership/Management (Plot Panel)

	(1) Plot Productivity (Head Owns)	(2) Plot Productivity (Head Manages)	(3) Plot Productivity (Owns & Manages)	(4) Plot Productivity (Crops Rotated)	(5) Plot Productivity (Mono-Cropped)	(6) Plot Productivity (Mixed Cropping)
Farmer over-estimates plot (binary)	-0.0540	0.296	0.441*	-1.227*	-0.558*	0.0930
	(0.202)	(0.224)	(0.259)	(0.638)	(0.338)	(0.227)
Over-estimate (% area)	0.139***	0.130**	0.0728	0.205**	0.206***	0.198***
	(0.0534)	(0.0636)	(0.0735)	(0.103)	(0.0754)	(0.0518)
Over-estimate squared	-0.00553*	-0.00374	-0.000854	-0.00730	-0.00773**	-0.00605***
	(0.00292)	(0.00243)	(0.00314)	(0.00473)	(0.00320)	(0.00185)
Under-estimate (% area)	0.537	1.418	2.332	-6.843*	-1.544	0.463
	(1.071)	(1.303)	(1.508)	(3.530)	(2.402)	(1.533)
Under-estimate squared	-0.315	-1.884	-2.887	8.289*	1.312	-0.369
	(1.260)	(1.643)	(1.874)	(4.259)	(3.448)	(2.066)
Observations	1472	1228	1096	581	829	1105
Adjusted R^2	0.426	0.356	0.350	0.298	0.394	0.510

Dependent variable: log(revenue/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Plot area and area-perimeter ratio are controlled for quadratically in all columns
Table estimates Equation 12 for data subsets
p<0.01, ** p<0.05, * p<0.1

Table A45: Labor Intensity Effects of Farmer Misperception of Plot Size (Plot Panel)

	(1) Labor Intensity	(2) Labor Intensity	(3) Labor Intensity
Farmer over-estimates plot (binary)	-0.211	-0.202	0.106
	(0.175)	(0.176)	(0.209)
Over-estimate (% area)	0.0689*	0.0772*	0.0629
	(0.0408)	(0.0418)	(0.0478)
Over-estimate squared	-0.00124	-0.00177	-0.000758
	(0.00149)	(0.00153)	(0.00160)
Under-estimate (% area)	0.530	0.569	2.487**
	(1.093)	(1.114)	(1.217)
Under-estimate squared	-1.087	-1.183	-3.828**
	(1.414)	(1.456)	(1.483)
Plot Area, P-A Ratio	Yes	Yes	Yes
$(\text{Area})^2$, $(\text{P-A Ratio})^2$	No	Yes	Yes
Additional Plot Controls	No	No	Yes
Observations	1973	1973	1547
Adjusted R^2	0.207	0.210	0.251

Dependent variable: log(hours/hectare)
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Additional plot controls are from Column 6 of Table 3, excluding labor intensity
Table estimates Equation 12 for labor intensity, rather than for productivity
*** p<0.01, ** p<0.05, * p<0.1