Close to the Edge: Do Behavioral Explanations Account for the Inverse Productivity Relationship?

Leah EM Bevis*, Christopher B Barrett†

May 29, 2017

Smaller farms and plots are more productive per hectare than larger ones. Some researchers hypothesize that it reflects household-specific shadow prices; others writers reject the relationship as spurious, invoking measurement error or omitted variables. Using unique, plot-level panel data from Uganda, we estimate the inverse relationship at the plot level and provide causal evidence that while the conventional explanations fail to explain the puzzle, the "edge effect" (productivity being highest around the periphery of plots) fully explains the phenomenon. We also find evidence that the edge effect arises due to a behavioral mechanism; labor intensity as well as productivity rises around plot edges. Productivity is also boosted by mere perceptions of plot size, illustrating that purely behavioral mechanisms can drive productivity.

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

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 and Adan Murillo for helpful comments. All errors are solely our responsibility.

*Corresponding Author. Ohio State University. Email: bevis.16@osu.edu

†Cornell University. Email: cbb2@cornell.edu

1 Introduction

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. This has been observed in Africa (Collier, 1983; Van Zyl, Binswanger and Thirtle, 1995; Barrett, 1996; Kimhi, 2006; Barrett, Bellemare and Hou, 2010; Carletto, Savastano and Zezza, 2013; Larson et al., 2014), in Asia (Sen, 1962; Mazumdar, 1965; Bardhan, 1973; Carter, 1984; Heltberg, 1998; Akram-Lodhi, 2001; 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. Of course, textbook neoclassical theory predicts equal marginal factor productivity across production units, else high marginal productivity users should purchase or rent land from low productivity users at a mutually attractive price, thereby increasing aggregate output and equalizing marginal returns. The existence of any relationship between farm productivity and farm size, either negative or positive, has therefore attracted much attention from development and agricultural economists as an important puzzle to resolve, because it suggests Pareto inefficient resource allocation.

How one explains the puzzle — the mechanism(s) that one hypothesizes generate the inverse size-productivity relationship — serves as a metaphor for how one understands the development challenge in low-income agrarian societies. Potential explanations therefore have important, practical implications. For example, if small farms are inherently more efficient in a given setting, redistributive land reform should be a source of both equity and efficiency gains. If market failures create household-specific shadow prices that drive the inverse relationship, then the fundamental welfare theorems of neoclassical economics may not hold in such settings, competitive markets do not necessarily yield Pareto optimal distributions, and government interventions in rural markets may be necessary. Conversely, if the inverse relationship is purely a statistical artifact attributable to the violation of econometric assumptions or to omitted relevant variables, then the rural economy may function as predicted by Walrasian theory and interventions may generally prove inefficient.

So how one understands the inverse size-productivity puzzle matters. Using plot-level panel data from rural Uganda we demonstrate that the inverse relationship is at the plot level, rather than at the farm level. We also provide highly suggestive evidence of causality, for the first time in this literature to the best of our knowledge. We test the familiar explanations and find that they fail to explain the observed inverse relationship. We next propose and examine a new mechanism that appears to explains the inverse relationship in our data: higher marginal productivity around the edges of plots — the "edge effect" — drives smaller plots to be more productive than larger plots, as a larger percent of small plot area falls along the plot's periphery. While we are unable to truly test whether behavioral or biophysical factors drive the edge effect, we do find that farmers seem to invest greater quantities of labor around the more highly visible and accessible plot edges.

Two classes of mechanisms are most commonly hypothesized to drive the inverse relationship. Chayanov (1991) first observed that Russian peasant farmers were more productive than larger farmers, and hypothesized that their propensity to employ

massive quantities of family labor in the farming enterprise explained this differential.[1] In seminal papers, Sen (1966) and Feder (1985) push this Chayanovian hypothesis further by theorizing that labor market failures drive high shadow prices for labor for larger farmers, who are unable to efficiently hire and/or supervise workers, and that these farm-specific shadow prices drive the inverse relationship. Barrett (1996) builds on this reasoning by showing that even in the absence of labor market failures, multiple market failures may cause smaller farmers to apply more labor per hectare, and hence be more productive, than their larger counterparts. Kagin, Taylor and Yúnez-Naude (2015) illustrate that smallholder farmers in Mexico are more technically efficient, as well as more productive, than their larger counterparts.

By contrast, a different thread of literature explains the observed inverse size-productivity relationship as illusory, a mere statistical artifact rather than causal relationship. This could result from measurement error around farm size that generates a spurious correlation (Lamb, 2003). It might also result from omitted relevant variables bias if farm size is endogenous to soil quality, with more fertile soils inducing both higher yields and denser settlement patterns leading to smaller farms (Bhalla and Roy, 1988; Benjamin, 1995; Assunção and Braido, 2007).

Resolution of the puzzle has long been complicated by the fact that few or perhaps no previous datasets were suited to provide well-identified, causal estimates of the inverse relationship. The size of a farm or even the size of a specific plot within a farm will be correlated with many unobserved factors. Nonetheless, a few recent studies examine household panel data or rich, plot-level data to provide evidence that neither multiple market failures nor measurement error/omitted variables drive the inverse relationship.

Household level fixed effects may be used in either household panel data or cross-sectional data with multiple plot observations per household. Authors using household level panel data find that the inverse relationship between farm size and farm productivity persists even when household fixed effects control for time-invariant household-specific shadow prices (Henderson, 2015; Kagin, Taylor and Yúnez-Naude, 2015). They cannot, however, rule out time-varying shadow prices, correlated with farm size, as driving the relationship. Authors using household fixed effects in plot-level data find that the inverse relationship remains and is often stronger at the plot level than it is at the farm level (Assunção and Braido, 2007; Barrett, Bellemare and Hou, 2010; Ali and Deininger, 2015).[2] These findings even more strongly indicate that household -specific shadow prices are not driving the inverse relationship. (Though it remains possible that plot-level shadow prices or observed, plot-specific omitted variables are driving the relationship.) Jointly, these household panel and plot-level studies seem to rule out multiple market failures as a full explanation of the phenomenon.

The statistical artifact hypothesis has not stood up well to recent empirical tests either.

[1] Chayanov's book was first published in English in 1991, "The Theory of Peasant Cooperatives." His original text was published in Moscow in 1921, "Osnovnye idei i formy organizatsii sel'skokhozyaistvennoi kooperatsii" (The basic ideas and organizational forms of agricultural cooperation).

[2] Assunção and Braido (2007) employ a household panel with multiple plot observations per household per round. They test the plot-level inverse relationship using both household and household-time fixed effects, with results identical across the two specifications. The later two studies employ cross-sectional, plot-level data, employing household fixed effects — which in cross sectional data, are equivalent to household-time fixed effects.

Barrett, Bellemare and Hou (2010) use cross-sectional, plot-level data from Madagascar with detailed soil quality measurements to show that including controls for soil biochemical and physical properties does not explain any part of the inverse relationship in their context. Similarly, Carletto, Savastano and Zezza (2013) show that the inverse relationship actually increases in magnitude when farm size is based on plot-specific GPS measures rather than farmer estimates that may be subject to considerable measurement error. In their data, Ugandan smallholders tend to over-report plot size while larger farmers tend to under-report plot size, so that removing measurement error around plot size actually reinforces rather than explains away the inverse relationship. Carletto, Gourlay and Winters (2015) show the same using data from Malawi, Uganda, Tanzania and Niger.

Two new papers examine a new statistical artifact hypothesis — that farmers may systematically over-estimate production on smaller plots and under-estimate it on larger plots. Using a two round household panel in a district of Eastern Uganda, Gourlay, Kilic and Lobell (2017) estimate the inverse relationship using conventional, farmer-recalled measures of crop yield, but find no inverse relationship when yields are measured via crop cutting. Desiere and Jolliffe (2017) find the same when comparing farmer-recalled yield and yields derived from crop cuts. In both papers, the difference between recalled yields and cut-calculated yields is inversely related to plot size. If cut-calculated yields are closer to "true" yields than are farmer-estimated yields, the inverse relationship is driven by systematic patterns in farmer recall error. If crop cuts over-estimate yield on larger plots and under-estimate yields on smaller plots, however, the disappearance rather than existence of the inverse relationship may be the statistical artifact.

So almost one hundred years after Chayanov first drew attention to this puzzle, it remains important and unresolved in the literature. The mechanisms previously considered most promising — the multiple markets failures and statistical artifact hypotheses — have not held up to recent, rigorous analyses. Additionally, while it is increasingly clear that the inverse relationship lies at least in part at the plot level, plot size is correlated with unobserved factors such as distance to home, crops planted, or input intensity (Tittonell et al., 2007, 2005). Thus, in the absence of random or quasi-random variation in plot size, both causality and mechanism remain unclear.

We address these important gaps in knowledge primarily by using plot-level, geospatially-matched panel data to estimate the inverse relationship.[3] While the most recent literature seems to have ruled out household-level shadow prices as driving the inverse relationship, no evidence attests to the role of plot-level shadow prices. Because plot shape and plot size vary over time (along with other plot-level characteristics), plot-level fixed effects allow us to estimate the inverse relationship while simultaneously controlling for time-invariant plot characteristics such as location within the landscape, distance to road or household, slope, elevation, and other factors that might affect plot-specific shadow prices of inputs and outputs.

However, we also estimate all results in a larger, pooled dataset of plot-level observations from both rounds, using household-time fixed effects as did Assunção and Braido (2007). The fact that this estimation method (identifying on within

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

household-time, across-plot variation) and the plot fixed effect estimation method (identifying on within-plot, over-time variation) recover almost identical inverse relationships suggests that plot size is causally associated with plot productivity. That is, in the pooled setting, only within-time-period, across-plot shadow prices or unobserved variables might bias estimation of the relationship. Yet in the geospatially-matched panel setting we cleanse all time-invariant plot-characteristics, such that only time-varying shadow prices or unobserved variables might bias estimation, and we find the same relationship. This suggests that the inverse relationship is either causal, or biased by omitted variables that change across plots within households in precisely the way that they change within plots across time.

Our contribution to the literature is therefore three-fold. First, we validate recent findings indicating that household shadow prices play no role in driving the inverse relationship — the relationship is at the plot level. We do this both by comparing the predictive value of farm size to the predictive value of plot size, and by comparing the inverse relationships estimated under various identification strategies. We additionally validate that neither measurement error around plot size, nor soil fertility, nor other often omitted, plot-specific characteristics are driving the relationship. The conventional, contending hypotheses — multiple markets failures versus statistical artifact — do not seem to explain the puzzle.

Second, we provide evidence that plot size itself is causally associated with plot productivity. We do this implicitly when comparing the inverse relationship estimated under household-time fixed effects and plot fixed effects. Additionally, thinking of causality in a more classic light, we acknowledge that plot size is not exogenous to other time-varying plot characteristics, and use a technique proposed by Oster (Forthcoming) to calculate bounds around the probable, causal effect of plot size on plot productivity. Under any plausible assumption it appears that omitted relevant variable bias slightly mitigates the point estimate on plot size, under-estimating rather than overestimating the inverse relationship. This exercise suggests that the inverse relationship is both causal and large in magnitude.

Third, we propose and test a new mechanism behind the inverse relationship, one rooted in more recent observations of the importance of both behavioral phenomena and biophysical constraints in explaining economic puzzles. The agronomy literature has long documented the "edge effect," the fact that the peripheral rows around the edge of a plot are often more productive than rows within the interior of a plot (Little and Hills, 1978; Barchia and Cooper, 1996). For instance, Verdelli, Acciaresi and Leguizamón (2012) found that the outer rows of corn in Argentinian corn and bean plots yielded 35-46 percent more than the center rows. In Illinois, border row corn yields were 37 percent higher than those of interior rows (Ward, Roe and Batte, 2016). Holman and Bednarz (2001) find that cotton at the edge of plots yields over three times more than cotton at the interior.

The peripheral area of a plot may 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) or greater water availability (O'Brien and Green, 1974). But beyond these biophysical mechanisms the periphery of a plot may

also be more visible or more accessible to a farmer in such a way that it changes his or her awareness of and thus management of this space. Behavioral economics research illustrates, for example, that individuals change food consumption behavior based on information about portion size or based on visual cues about portion size (Just and Wansink, 2014; Wansink, Painter and North, 2005). We might similarly hypothesize that farmers change crop or soil management based on visual signals of crop growth conditions. Farmers can see weed growth, pest infestation, plant disease or other yield-dampening phenomena more readily on a plot's perimeter than within its interior; and they can often reach the perimeter more easily as well.

Importantly, the edge effect is not a fixed magnitude — it varies by crop type, cropping system, plot lay-out, border direction, and likely many other factors. For instance, reviewing a couple decades of literature, Ward, Roe and Batte (2016) suggest that the edge effect is stronger in intercropped systems (common in Uganda and sub-Saharan Africa more generally) than in mono-cropped systems. Additionally, row orientation and border direction also impact the edge effect, with northern-facing border rows often having the largest productivity differential, and north-south oriented rows mitigating the edge effect Barchia and Cooper (1996). Almost certainly the properties of adjacent plots will matter — the edge effect may be larger for a plot bordering ground crops, for instance, than for a plot bordering sunlight-blocking tree crops. If farmer behavior drives the edge effect, it may be greater for plots closer to the house, or for plots with higher-value crops.

However, if plot peripheries are on average more productive than plot interiors, for any reason, then smaller plots with larger periphery-to-interior ratios will be more productive on average than larger plots, plausibly driving part or all of the oft-observed inverse relationship. In our data, controlling for the edge effect in this way (thereby estimating an average treatment effect for a likely heterogeneous process) does completely explain the inverse size- productivity relationship. Therefore, in this dataset at least, it seems that the inverse relationship is driven by the edge effect, i.e. by either differential unobserved biophysical inputs at the periphery of plots, or differential input allocation responding endogenously to behavioral cues.[4] In either case, these differential inputs lead to real inefficiency, as opposed to being merely a statistical artifact.

Having established this fact, we seek to uncover whether biophysical or behavioral phenomenon drive the edge effect. We indirectly test a number of plausible, biophysical mechanisms, and find no evidence of their driving the edge effect. We then show that labor per hectare rises with periphery-to-interior ratio, and also that this effect is only significant for family (non-hired) laborers, who may be more assiduous in tending the highly-visible periphery of a plot. These results suggest that labor and crop management patterns do indeed change around the periphery of the plot, hinting at a behavioral explanation for the edge effect.

[4]It is worth noting that we have only farmer-recalled yield data, and so cannot compare our results to those obtained using yields calculated from crop cuts, as do Gourlay, Kilic and Lobell (2017) and Desiere and Jolliffe (2017). However, if the edge effect drives the inverse relationship, yields based on crop cuts will generally fail to capture the phenomenon anyway, as crop cuts are taken in the plot interior and therefore representative only of interior productivity. (Though it is true that for a small subset of fully harvested plots, Gourlay, Kilic and Lobell (2017) find the same disappearance of the inverse relationship; in this context farmers may truly be systematically over-estimating yields on smaller plots.)

Because it is difficult to entirely untangle the biophysical and behavioral drivers of the edge effect, we then explore a second behavioral phenomenon. We show that when a plot's size is over-estimated by its attending farmer, that plot will tend to be more productive. Similarly, the under-estimation of plot size appears to lower plot productivity. Farmers seem to over-allocate inputs to plots whose size they over-estimate, and to under-allocate inputs to plots whose size they under-estimate. Because these misperceptions of plot size are purely cognitive errors, it seems that a farmer's awareness of space has the power to change his or her management patterns, and thus to drive plot productivity. While not directly related to the edge effect, this pattern makes it clear that a behavioral explanation of the edge effect, and of the inverse size-productivity relationship, is plausible.

The remainder of this paper is organized as follows. Section 2 discusses the plot-level panel data that we use in this paper. Section 3 presents, by sub-section, the equations that we later estimate. Section 4 presents our results within identical sub-sections. Section 5 concludes.

2 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 (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).[5]

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 single plot of land, used to grow a single crop or

[5]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 can not be viewed as representative across Uganda.

multiple, mixed crops. Most, though not all, farmers have multiple plots in both 2003 and 2013.[6] While the size and shape of these plots shifts across the decade, many of them are generally in the same location within a larger parcel/unit of land. Because GPS waypoints were taken around the corners of all plots in both rounds of data collection, we can match plots across rounds using their geospatial location. This is how we form a plot-level panel dataset — our primary dataset. (More details on the geospatial matching process can be found in Appendix 1.) Additionally, we pool all plots from both rounds into one dataset, which constitutes our second, pooled dataset used for analysis without plot fixed effects.[7]

Because most of Uganda has two agricultural seasons, agricultural input and production data was gathered for both seasons. We therefore consider four time periods, one for each agricultural season of each round. Plot size and shape does not vary across season within round, but other plot-level characteristics such as as fertilizer use, organic amendment or management 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 all key variables used in analysis, for 2003 and 2013, for our plot-level panel dataset. (Appendix 4 summarizes the same variables 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. Plots are also far more productive (measured in terms of revenue per hectare) in 2013 than in 2003, and labor intensity (hours/hectare/day) far higher. Soil became slightly more acidic over the decade, while organic carbon content appears to have slightly increased.[8]

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

[6] In 2003 and 2013, 16% and 31.6% 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.

[7] Of the 2,549 plots recorded in the 2003 survey, 25 percent (631) were geospatially matched. Of the 1,773 plots recorded in the 2013 survey, 42 percent (738) were geospatially matched. We therefore have 738 plots in our plot panel dataset, each viewed in 2-4 agricultural seasons (but necessarily at least 1 agricultural season per round), totaling at 2,182 observations. We have 4,322 plots in our pooled dataset, each viewed in 1-2 agricultural seasons within their respective round, totaling at 6,704 observations.

[8] This change in organic carbon content may be due to a slight change in analysis technique rather than a true change in soil organic organic. In both years, soil organic matter was obtained via the Walkley-Black test. However, the buffer pH changed across years, potentially causing more organic matter to be extracted from samples in the 2013 test. Because round fixed effects are used in all analysis, this mean shift should have no consequence for our results.

and 2003, due to a decline in the average number of crops grown per household (see Tables A9 and A11) as well as a decline in the average number of crops grown per plot.

In our primary analysis, the inverse relationship is driven by within-plot change in size over time (controlling for year and season fixed effects), and Appendix 5 holds detailed information on the determinants of this plot size change. Thirty-two percent of plots grow over the decade while sixty-eight percent of plots shrink over the decade. These changes in size are not random: the primary predictor of plot size change is 2003 plot size. Larger plots are both more likely to shrink and shrink more, on average. Plots are also more likely to shrink if they are located on parcels of land that have been subdivided between 2003 and 2013. Generally, however, these are the larger plots, and starting plot size has more explanatory power than parcel division, when it comes to predicting changes in size across all plots. (Also note that the vast majority of geospatially matched plots, all but 16, remain under the same ownership over the decade. Changes in ownership are not driving plot size change or productivity changes.)

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. These plots are dropped from our analysis. Some 2003 plots are matched to multiple 2013 plots, generally because they have been split up into smaller plots over the decade.[9]

Because this selection process may not be random, Appendix 6 compares household and plot characteristics in the plot panel dataset to household and plot characteristics within the larger universe of all pooled plots from 2003 and 2013. While household-level characteristics differ only slightly across these two datasets, plot-level characteristics differ quite a lot — plots within the panel dataset 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 than plots within the larger dataset. These plots cannot, therefore, be viewed as representative of the larger universe of plots in our pooled dataset. However, we also estimate all core results under household-time (specifically, household-year-season) fixed effects. Plots within the household-time fixed effect analysis can be viewed as representative of the larger universe of plots (also shown in Appendix 6).[10] If results stemming from plot fixed effect estimation are similar to results stemming household-time fixed effect estimation, it therefore seems unlikely that selection is driving/biasing these results.

[9] Of the 631 plots from 2003 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 were matched to 3 plots, 2 percent were matched to 4 plots, and an additional 1 percent were 5-9 plots. In only 16 cases (1.8 percent of the plots in our panel), part of a 2003 plot is split off to be moved under new management in 2013. (Generally this means the plot was inherited by a child.) No other plots change ownership. More details can be found in Appendix 1.

[10] 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 pooled dataset, but time-invariant household characteristics are partialled out in both types of analysis anyway.

3 Estimation Strategy

3.1 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. 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.[11] (Uganda has two agricultural seasons.) Plot area is given by A_{ijt}, and farm size is given by A_{it}.[12]

The inverse relationship can be estimated including only A_{ijt}, including only A_{it}, or including both areas. The inverse relationship will appear as a negative and statistically significant coefficient on farm size, a negative and statistically significant coefficient on 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 again be estimated by including plot fixed effects λ_{ij} and a time fixed effect λ_t as in Equation 3.

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

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

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

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 are driving the inverse relationship, the plot-level relationship should cease to exist once household-time fixed effects are controlled for in Equation 2, or plot fixed effects are controlled for in Equation 3.

Because we do 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, and exclude A_{it}, as in Equation 2. This identification strategy, new to the inverse relationship literature, partials out all time-invariant, plot-specific shadow prices and/or characteristics as driving the inverse relationship. Only plot-level characteristics that vary across time with plot size may bias the estimated relationship.

We also, however, estimate all equations using household-time fixed effects as in Equation 2; these results are found in Appendix 4. This identification strategy partials out all household-specific shadow prices, both time invariant and time-varying, but allows for within-time plot-level factors to bias the relationship. If the inverse

[11] Because a variety of crops are being grown across and within plots, productivity cannot be measured solely as physical yields.

[12] 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.

relationship is identical across these two identification strategies, it suggests that neither household-level nor plot-level factors are biasing the relationship, and that plot size itself truly drives plot productivity. (Though it is also feasible that two different sources of bias might coincidentally result in the same inverse relationship.)

3.2 Measurement Error

We then investigate how measurement error around plot size influences the estimated inverse relationship between plot size and plot productivity. Let Y_{ijt}^m 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_{ijt}^m be the area of plot i within household j, measured by method m.

$$Y_{ijt}^m = \gamma_4 A_{ijt}^m + \lambda_{ij} + \lambda_t + \varepsilon_{ijt} \tag{4}$$

If the inverse relationship is in part a statistical artifact driven by measurement error, the $\hat{\gamma}_4$ estimated under Equation 4 should be weaker and R^2 should be smaller when plot size is measured by GPS rather than being recalled by farmers. If the opposite is true, then measurement error instead attenuates $\hat{\gamma}_4$, as found by Carletto, Savastano and Zezza (2013).

3.3 Omitted Variables and Causality

Having set the issue of measurement error aside, and chosen a statistically preferable method m for defining plot size, we explore the issue of causal identification and omitted variable bias. A number of plots characteristics shift between 2003 and 2013, as evidenced in Table 1. If some of these characteristics (soil quality, crops grown, etc.) change across rounds in a way that is related to changes in plot size and plot productivity, failing to control for them might cause a spuriously estimated inverse relationship.[13]

We therefore control for an exhaustive list of time-varying plot characteristics X_{ijt}, as in Equation 5, alongside plot size A_{ijt} from Equation 4.

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

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. We then allow X_{ijt} to include other time-varying plot characteristics relevant to plot productivity: agricultural inputs, plot ownership and management style, and crops grown.

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

[13] Similarly, if these characteristics are correlated across plots within time, with plot size and plot productivity, failing to control for them might bias the estimated relationship under the household-time fixed effect specification.

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

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 (Forthcoming) proves that with one key assumption,[14] the bias-adjusted coefficient estimate β^* can be approximated as below, and that β^* converges in probability to the true, causal coefficient β.[15] The parameter δ gives the relative proportion of X explained by unobservables vs. observables — i.e., if δ is $< (=)[>]$ 1, then X is 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.

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

The causal effect of X on Y will lie within the bounding box $\big[\tilde{\beta}, \beta^*(R_{max}, \delta)\big]$, and Oster (Forthcoming) 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 box for the inverse relationship, $\big[\hat{\gamma_5}, \gamma^*(min\{1.3R_5, 1\}, 1)\big]$, and additionally calculate bounding boxes under even more restrictive R_{max} and δ parameters.

3.4 The Edge Effect

We next propose and test a new, previously unconsidered mechanism behind the inverse relationship. We allow the productivity of plot j belonging to household i in time

[14] 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 (Forthcoming) 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.

[15] Under a second assumption of proportional selection — that X is equally related to w_1 and W_2 — β^* can be exactly calculated, using the same equation and letting $\delta = 1$. This approximation, however, allows a range of δ values to be considered.

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 we assume that A^P_{ijt} is roughly equivalent to $P_{ijt} * b$, where b is the width of the peripheral area, then we can can rewrite Equation 8 as in Equation 9.[16] Figure 1 provides a schematic visual for this assumption.

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

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. If this is the case, the inverse relationship could stem from mis-specification 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 this equation, γ_6 indicates the classic inverse relationship, and $\theta_1 = (Y^P_{ijt} - Y^I_{ijt}) * b$. If γ_6 becomes 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 7 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

[16] For intuition, we are basically assuming that the plot's periphery is thin enough that it could be rolled out from around the plot's perimeter in the form of a rectangle. Such a rectangle would not approximate the area of the plot's periphery if the b was quite large with respect to the plot edges. But if b is small — e.g., one row of crops — then this rectangle approximates the peripheral area.

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 may truly be heterogeneous — both $(Y^P_{ijt} - Y^I_{ijt})$ and b may change with plot lay-out, adjacent plot characteristics, crop type, location on the landscape, etc. The estimated coefficient must therefore be viewed as a context-specific average.

3.5 Edge Effect Mechanisms

Because we do find that the edge effect entirely explains the inverse relationship, we next turn to investigating the mechanisms behind the edge effect. Two 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 biophysical inputs such as sunlight, water, nutrients or biodiversity, factors generally ignored by the econometrician. These are purely biophysical mechanisms, not controlled by the farmer. Because we do not view biophysical inputs we cannot test for such mechanisms directly; Appendix 8 reports a number of indirect tests, all inconclusive.

The second category of 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 highly-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 be visible when walking by.[17]

If this is the case, we might expect average labor per hectare L_{ijt} to exhibit the same pattern as average productivity per hectare. 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 insignificant $\hat{\gamma}_7$ and a positive, significant $\hat{\theta}_2$ if P_{ijt}/A_{ijt} is also controlled for. (Of course, this is suggestive evidence only that labor serves as a mechanism behind the edge effect.)

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

If farmers invest more labor around the edges of their plots, this could reflect an increased awareness of the space around plot edges. To further investigate the productivity implications of "plot awareness" we examine a second and purely behavioral mechanism, one wholly unrelated to biophysical constraints or inputs.

Because we have data on farmer-recalled plot size and also GPS-measured plot size, we can calculate plot size perception error as the difference between these two sizes: $e_{ijt} \equiv A^F_{ijt} - A^{GPS}_{ijt}$. That is, we can calculate how much a farmer under-estimates or over-estimates his or her plot relative to the GPS-measured plot size. If farmers apply inputs based on perceived plot size, then productivity should rise with perception error, signaling real resource allocation inefficiencies due to farmer error.

[17] A related mechanism also seems plausible: farmers may tend the edges of their plot more intensively in order to signal plot ownership to neighbors, particularly if tenure status is questionable. We are unable to investigate this mechanism due to a lack of variation in tenure status.

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^B, κ^O and κ^U capture the causal effect of farmer awareness of plot space on per hectare productivity. We test the conditional exogeneity of e_{ijt}, and then estimate $\hat{\kappa^B}$, $\hat{\kappa^O}$ and $\hat{\kappa^U}$.

4 Results

4.1 Shadow Prices (Plot vs. Household Mechanisms)

Table 2 reports results for Equations 1-3. Panel 1 displays traditional, OLS estimates of the inverse relationship. 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, and 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. That is, a fixed effect is included for every unique combination of household, year, and agricultural season.[18] This effectively controls for household- and time-specific shadow prices — variation in plot size is identified only within households and within a single time period. Estimating the association between farm size and plot productivity is therefore impossible, of course. However, the plot-level inverse relationship found in Column 2 of this panel is actually significantly greater in magnitude than the plot-level inverse relationship in Panel 1. A 10 percent increase in plot size appears to drive a 6.7 percent decrease in plot productivity. (This relationship is similar across years, though the estimated coefficient is slightly larger in magnitude in 2013. See Table A18 in Appendix 9.)

In Panel 3, plot fixed effects are introduced. Year and season are additionally controlled for. This estimation method therefore controls for plot-specific, time-invariant characteristics, shadow prices or transaction costs such as distance to household, distance to road, location on the landscape, and time invariant soil characteristics. It also controls for any mean shifts in productivity or production technology across time. The plot-level inverse relationships in Columns 2 and 3 of this panel are strong and statistically equivalent to the plot-level inverse relationship in Panel 2. 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.

[18] Most of Uganda has 2 agricultural seasons per year, and so we define "time" from Section 3 as year-season.

For the remainder of this paper, results are therefore estimated with plot fixed effects, controlling for year (2003 vs. 2013) and for season (1st vs. 2nd agricultural season) as dummies. The identifying variation is therefore within-plot, across-time variation in size, shape, and other characteristics. Plots that cannot be matched across the decade are, of course, dropped, and Appendix 6 examines this selection process. Appendix 5 examines the non-random processes driving changes in plot size over time; essentially, larger plots are more likely to shrink, and smaller plots are more likely to grow. However, the inverse relationship exists for both shrinking and growing plots, as illustrated in Table A21 of Appendix 9. The same table shows the inverse relationship to be unchanged if estimated using only those plots for which one 2003 plot was matched to one 2013 plot (rather than one 2003 plot matching to multiple 2013 plots). The relationship is qualitatively unchanged across functional forms, as shown in Table A22.

All results presented in this paper 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 again examined in Appendix 6. Explanatory power is lower for this form of variation. The coefficients estimated, however, are quantitatively (and qualitatively) the same as those estimated under plot fixed effects, with just two exceptions, both discussed in the appendix.

4.2 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 identical across measurement methods (GPS-measured vs. farmer-recalled), 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 found by Carletto, Savastano and Zezza (2013), and counter to Lamb's (2003) hypothesis.[19]

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.[20]

4.3 Omitted Variables and Causality

If plot size (or change in plot size over the decade) was randomly distributed, we could interpret the coefficient on GPS-measured plot size in Table 3 as the causal effect of plot size on plot productivity. This is not the case, however. Plot size is correlated with other, observable plot characteristics under any identification strategy — Table A23 in

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

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

Appendix 9 illustrates this fact via regression-based balance tests under OLS, household-time fixed effects, and plot fixed effects.[21] In both the pooled and panel setting, therefore, plot size cannot 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 on plot size to diminish in absolute value as relevant, observable controls are introduced (Oster, Forthcoming). 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 the coefficient on GPS-measured plot size in Table 3.

In Column 2, soil characteristics are controlled for. In Column 3, inputs such as labor hours, soil amendments and structures within the plot are controlled for.[22] The inverse relationship remains virtually identical in each of these specifications. In Column 4 plot ownership and plot management is controlled for, and the inverse relationship becomes significantly larger in magnitude. In Column 5 crops are controlled for. In Column 6 all variables are simultaneously controlled for, and the inverse relationship is statistically identical to (though actually slightly large in magnitude than) 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 9 reports these results as a robustness check; 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 possible that some other, still-omitted, time-varying plot characteristic drives the inverse relationship. The stability/robustness 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,[23] and begin by assuming $\delta = 1$ and $R_{max} = 1.3R_5$, as suggested by Oster (Forthcoming). Under these assumptions, we obtain the bounds [-0.658 -0.696].

Notably, these bounds suggest that the causal inverse relationship is actually *higher* than the estimated inverse relationship. This is because controlling for a full set of covariates actually increases the magnitude of the estimated inverse relationship, while increasing R-squared. (I.e., it appears that omitted relevant variables downwardly bias, rather than upwardly bias, the inverse relationship.)

[21] If 2003 plot size is included in the third column, almost all associations with other covariates become insignificant. That is, plot size change over the decade is primarily associated with other covariates through it's association with 2003 plot size. None-the-less, as we do not control for 2003 plot size in our primary regressions, plot size change cannot be considered exogenous for our purposes.

[22] Organic and inorganic fertilizer are controlled for in a binary fashion, as few plots receive either input.

[23] See Appendix 10 for bounds based on each set of controls in turn.

If we loosen these assumptions to allow a range of δ and R_{max} parameters, we again find that *all* possible bounds suggest the causal inverse relationship to be higher than the estimated inverse relationship. Figure A6 in Appendix 10 illustrates these possible bounds. More details can be found in Appendix 10.

4.4 The Edge Effect

Having established the inverse relationship as strongly robust, very possibly causal, and unexplained by any previously considered mechanism, we now investigate the newly proposed edge effect mechanism. Columns 1-3 of Table 5 presents results for Equation 10, which specifies plot productivity as a function of the perimeter-area ratio, $\frac{P_{ijt}}{A_{ijt}}$. In Column 1, only plot size explains plot productivity, and the baseline inverse relationship is estimated. In Column 2 the perimeter-area ratio 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 — 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 is dropped so that only the perimeter-area ratio is controlled for, and adjusted R^2 drops by only a percentage point, suggesting that plot size contributes very little additional information on average productivity once perimeter-area ratio is known.[24] In fact, adjusted R-squared is higher in Column 3 than in Column 1 — perimeter-area ratio better explains plot 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 of Table 5 as in Column 4, controlling for perimeter and area separately. If Equation 10 specifies the true functional form, i.e., perimeter and area only predict productivity inso far as their ratio predicts productivity, then the same $\hat{\theta}_1$ coefficient will be estimated as the coefficient on each variable. Indeed, this is the case in Column 4.

Appendix 11 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 A26), can be estimated across crop and ownership/ management subsets (Tables A28-A29), and is virtually indistinguishable across both plot size quantiles and perimeter-area ratio quantiles (Table A30). Oster's bias-adjusted estimator γ^* again suggests that under any reasonable assumption, the edge effect is far above zero (Table A27 and Figure A9). The effect does vary by plot shape — Table A31 illustrates that while plot shape does not directly drive productivity, it alters the marginal effect of perimeter-area ratio. For all shapes, however, the marginal effect is far above zero.

Because plot size and perimeter-area ratio are so closely related (with a Pearson's Correlation Coefficient of -0.133), and the log version of these variables is even more strongly correlated (with a correlation coefficient of -0.939), multicollinearity is a

[24] The fact that plot size contributes little or no information on productivity suggests, according to the theory outlined in Appendix 7, that the border area b defining plot edge is small.

concern.[25] Yet the edge effect is indistinguishable across correlation quintiles (Table A32). Additionally, two placebo tests replace perimeter with a second variable, and while these new placebo-area ratios are highly correlated with area, just as perimeter-area ratio is, controlling for the ratios does not mitigate the inverse relationship (Tables A33 and A34). These tests suggest that multicollinearity is not driving the results in Table 5.

Additionally, an alternative proxy can be used to test for the edge effect, using number of sides rather than perimeter to capture the peripheral area (Table A35). As we would expect if this variable was an inferior proxy for peripheral area proportion, the inverse relationship is mitigated when number of sides per unit area is controlled for, though does not become zero. Both this result and Table 5 clearly show that both plot shape and plot size predict productivity, and suggest the edge effect as the mechanism, or at least a mechanism. It must be noted, however, that the magnitude of the effect is surprising; we explore the expected elasticity of productivity with respect to both plot size and perimeter-area ratio in the last section of Appendix 11, and find that our theoretical edge effect can explain only approximately a third of the estimated coefficient on perimeter-area ratio, and only approximately a fourth of the estimated inverse relationship.

While perimeter-area ratio completely mitigates the inverse relationship under plot fixed effects, Table A5 shows that under household-time fixed effects, the inverse relationship is only mitigated by about a third. 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 instance, 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.

4.5 Edge Effect Mechanisms

We next explore mechanisms behind the edge effect. While we cannot test for biophysical mechanisms directly, Appendix 8 holds indirect evidence regarding biophysical mechanisms. We find no evidence that sunlight or soil nutrient absorption drive the edge effect, though 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).

Because we view labor inputs, we can more directly examine evidence regarding a 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, measured in labor hours per hectare, is inversely correlated with plot size. In Column 2, however, it appears that the perimeter-area ratio drives much of the

[25] Though plot size and perimeter-area ratio are closely related, plot shape also explains a significant portion of perimeter-area ratio, especially within plot fixed effects. See Table A25 in Appendix 11.

inverse relationship, as with productivity. Once the perimeter-area ratio is controlled for, the inverse relationship between plot size and labor intensity reduces by two-thirds, and is significant only at the 10 percent level. Additionally, controlling for perimeter-area ratio increases R^2 relative to Column 1, as in Table 5.

Column 3 controls only for perimeter-area ratio. Adjusted R^2 drops slightly relative to Column 2 — this fact and the marginal significance on plot size in Column 2 suggest that plot size does contribute information on labor intensity, conditional on perimeter-area ratio, though not very much. The fact that the coefficients on plot size and perimeter in Column 4 differ slightly also suggests that one of these variables may be associated with productivity outside of the association seen with their ratio. However, like the results from 5, the relative R^2 values of Columns 1 and 3 show that, if each variable is considered alone, 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 hence helps to drive the inverse relationship.[26,27] Appendix 8 reports additional results for family and non-family laborers and across various types of labor; differences in effect are not large or particularly meaningful.

Last, we examine how farmer awareness of plot area impacts plot productivity, via a slightly different route. Farmer perception error around plot size is captured through the difference between farmer-recalled plot size and GPS-measured plot size. If perception error is exogenous to other plot conditions, then any labor response or productivity response to perception error can be viewed as a purely behavioral mechanism. Appendix 12 investigates this exogeneity. While perception error is clearly related to plot size and the perimeter-area ratio when all data are pooled, under a plot fixed effects model it appears largely unrelated to plot conditions. If conditioned on plot area and perimeter-area ratio, perception error appears exogenous to all other time varying, plot-level characteristics, with the potential exception of crop choice, which is in any case a product of farmer behavior rather than a biophysical characteristic.

Table 7 demonstrates the effect of farmer misperceptions of plot size on productivity. Plot productivity rises with over-estimation (with diminishing returns), and drops with under-estimation (with diminishing returns).[28] Column 1 controls for log plot area and log plot perimeter-area ratio in addition to farmer perceptions, and Column 2 controls for area and perimeter-area ratio quadratically.[29] 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.

[26] Additionally, the 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, respectively. As discussed in the last section of Appendix 11, this elasticity seem too high to be realistic.

[27] It is important to note here that this relationship implies that labor is being applied at higher rates around the edges of plots — not that labor is more effective around the edge of plots, a possibility we cannot gauge in these data.

[28] 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.

[29] Appendix 12 suggests that perceptions may respond non-linearly to area and perimeter-area ratio.

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 A41 in Appendix 13 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. (This makes sense; if farmers choose input intensity based on plot size perceptions, for all inputs, we would expect a stronger relationship between resulting productivity and perceptions than between any given, individual input and perceptions.)

Because Appendix 12 suggest that perception error may possibly be related to crop type, Appendix 13 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 is cut drastically.

It is also possible that misperceptions of plot size impact, not plot productivity itself, but *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.

5 Conclusion

Using plot-level panel data for the first time in this literature, we estimate the inverse size-productivity relationship under plot fixed effects. We show that the inverse relationship is at the plot level, not the farm level, and thus cannot be explained by household shadow prices or by other, omitted household characteristics. Additionally, the inverse relationship estimated under plot fixed effects (partialling out time invariant, plot-specific shadow prices but vulnerable to shadow prices that change over time) is identical to that estimated under household-time fixed effects (partialling out household-time-specific shadow prices but vulnerable to cross-sectional, plot-specific shadow prices). This stability across two identification methods, each with different, potentially biasing factors, suggests that the relationship may be causal.

Additionally, we validate the findings of Carletto, Savastano and Zezza (2013), Barrett, Bellemare and Hou (2010) and Carletto, Gourlay and Winters (2015) by showing that neither soil fertility nor measurement error around plot size drives a spurious, inverse relationship between plot size and plot productivity. By controlling for a rich set of plot characteristics such as inputs and management, we show that the inverse relationship cannot be explained by the time-varying plot characteristics typically considered.

The stability of the inverse relationship to this rich set of covariates is again suggestive of a causal relationship. We estimate bounds around the likely causal relationship between plot size and plot productivity, using the consistent estimator of bias derived by Oster (Forthcoming). This exercise suggests that the causal relationship is

significant and large in magnitude; the inverse relationship may be slightly greater, in fact, than the estimated elasticity, which suggests that a 10 percent increase in plot size decreases plot productivity by 6.6 percent.

While previously proposed mechanisms do not explain the inverse relationship, our analysis suggests that the phenomenon may be explained by edge effects — a novel explanation in the economics literature. 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. (Perimeter-area ratio is similarly associated with productivity under household-time fixed effects, but the inverse relationship is not fully mitigated.) 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 a thin but differentially productive peripheral area around the edge of each plot, which makes smaller plots more productive on average than large plots.

The mechanism behind the edge effect is difficult to isolate; in all likelihood there are multiple mechanisms. The agronomy literature suggests that the periphery of a plot is often more productive than the interior of a plot due to increased inputs such as 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 plot-level labor intensity rises with perimeter-area ratio, just as productivity does. This suggests a behavioral mechanism — if farmers are more aware of plot edges, and/or can more easily access plot edges, they may tend these edges more intensely in the form of weeding, pruning, or even harvesting. This could contribute to the edge effect, and to the inverse relationship. Literature from behavioral economics shows that consumption responds strongly to visual cues (Wansink, Painter and North, 2005). Our results suggest that production does the same.

In fact, we additionally show that farmer beliefs about plot size — independent of actual plot size — may influence investments in plots and therefore plot productivity, just as portion labeling influences consumer reference frames for portion size, and therefore influences food consumption choices (Just and Wansink, 2014). Farmer misperceptions of plot size, exogenous to other plot characteristics conditional on plot size and perimeter-area ratio, appear to drive plot productivity. Marginal increases in plot size over-estimation drive up plot productivity, while marginal increases in plot size under-estimation drive down plot productivity. These results suggest that farmer cognitive error drives input application rates and resulting productivity, consistent with the possibility of a labor-driven edge effect.

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 plot productivity. Robustness checks suggest that a number of plot shape characteristics influence productivity, all in line with the edge effect. The fact that labor intensity as

well as productivity rises with increased peripheral area further suggests that the edge effect is driving the inverse relationship, and that a behavioral mechanism may be at play. Further research might do better at teasing 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 should also influence magnitude. With this knowledge, optimal layout and organization of plots might be considered. This paper shows with clarity however, that the inverse relationship is a plot-level phenomenon, and that farmer choices regarding plot shape as well as plot size influence productivity.

References

Akram-Lodhi, Haroon. 2001. "Vietnam's agriculture: is there an inverse relationship?" *ISS Working Paper Series/General Series*, 348: 1–36.

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.

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

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

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.

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?" Paper Presented at the 2017 Centre for the Study of African Economies (CSAE) Conference.

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.

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.

Just, David R, and Brian Wansink. 2014. "One man's tall is another man's small: How the framing of portion size influences food choice." *Health Economics*, 23(7): 776–791.

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.

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.

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.

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.

Oster, Emily. Forthcoming. "Unobservable selection and coefficient stability: Theory and validation." *Journal of Business Economics and Statistics.*

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.

Tittonell, PABLO, Bernard Vanlauwe, N De Ridder, and Ken E Giller. 2007. "Heterogeneity of crop productivity and resource use efficiency within smallholder Kenyan farms: Soil fertility gradients or management intensity gradients?" *Agricultural Systems*, 94(2): 376–390.

Tittonell, P, Bernard Vanlauwe, PA Leffelaar, Keith D Shepherd, and Ken E Giller. 2005. "Exploring diversity in soil fertility management of smallholder farms in western Kenya: II. Within-farm variability in resource allocation, nutrient flows and soil fertility status." *Agriculture, Ecosystems & Environment*, 110(3): 166–184.

Van Zyl, Johan, Hans Binswanger, and Colin Thirtle. 1995. "The relationship between farm size and efficiency in South African agriculture." World Bank.

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.

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.

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

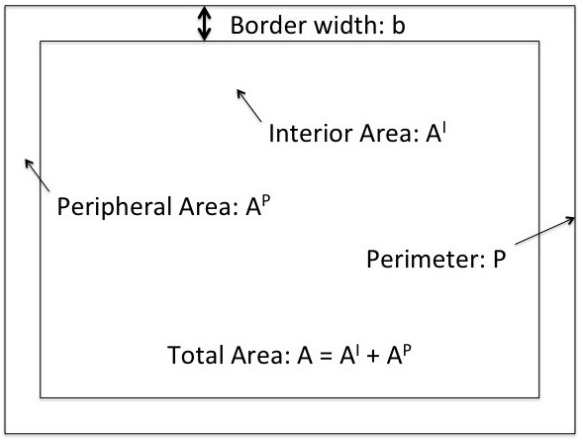
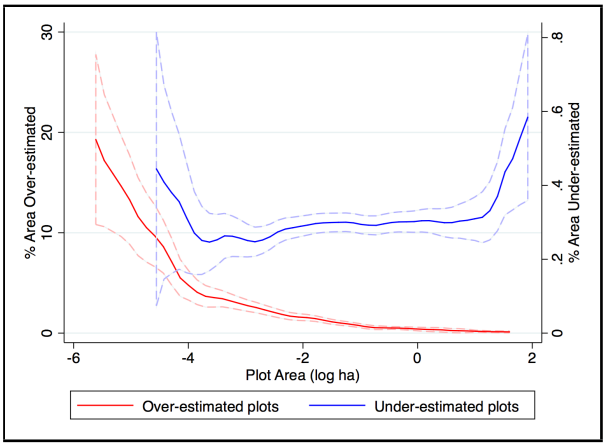


Figure 2: Misperceptions by Plot Area



Tables

Table 1: Plot Characteristics in 2003 and 2013

	2003		2013		
	Mean or Median†	Standard Deviation	Mean or Median†	Standard Deviation	T Statistic‡
Size, Productivity, Labor					
Farm size (ha)	1.04	1.19	0.57	0.88	15.83***
Plot size (ha)	0.35	0.68	0.18	0.40	12.21***
Perimeter-area ratio (m/ha)	849.45	925.11	1,146.56	6,778.42	-10.08***
Plot productivity (revenue§/ha)	103.06	1,328.82	256.12	8,653.08	-18.06***
Labor intensity (hrs/ha/day)	1.94	9.61	1.78	67.70	1.49
Soils					
Soil pH (pH)	6.23	0.52	6.18	0.63	1.96*
Soil sand (%)	60.12	14.05	52.65	15.82	10.82***
Soil organic carbon (%)	3.45	1.65	3.72	1.87	-3.27***
Inputs					
Organic amendment (%)	19.89	39.94	12.01	32.52	5.06***
Inorganic fertilizer (%)	1.19	10.86	1.47	12.03	-0.56
Irrigation (%)	1.30	11.33	0.19	4.32	3.01***
Terracing (%)	23.19	42.22	9.37	29.15	8.89***
Management					
Head manages plot (%)	55.36	49.73	63.70	48.11	-3.98***
(Head owns)X(Head manages)	46.47	49.90	59.30	49.15	-6.05***
Crops are rotated (%)	21.75	41.27	42.87	49.52	-10.26***
Crops are mono-cropped (%)	42.99	49.53	36.94	48.29	2.89***
Mixed cropping (%)	54.54	49.82	52.06	49.98	1.16
Intercropping (%)	2.47	15.54	10.91	31.19	-7.99***
Crops Grown					
Tubers grown (%)	43.35	49.58	24.84	43.23	9.30***
Cereals grown (%)	49.50	50.02	43.63	49.62	2.75***
Legumes grown (%)	53.07	49.93	45.10	49.78	3.74***
Bananas grown (%)	48.67	50.01	27.59	44.72	10.38***
Cash crops grown (%)	30.98	46.26	18.97	39.23	6.54***

†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)	-0.378***		-0.0358*
	(0.0188)		(0.0217)
Plot size (log ha)		-0.516***	-0.496***
		(0.0159)	(0.0198)
Observations	5709	5709	5709
Adjusted R^2	0.073	0.171	0.172
	Panel 2 (House-year-season FE)		
Plot size (log ha)		-0.565***	
		(0.0259)	
Observations		4845	
Adjusted R^2		0.170	
	Panel 3 (Plot FE)		
Farm size (log ha)	-0.321***		0.0634
	(0.0875)		(0.0618)
Plot size (log ha)		-0.621***	-0.648***
		(0.0636)	(0.0637)
Observations	2181	2181	2181
Adjusted R^2	0.282	0.381	0.381

Dependent variable: log(revenue/hectare)
Panel 1: No FE, robust standard errors
Panel 2: House-year-season FE, house-year-season clustered standard errors
Panel 3: Plot FE, year and season FE, plot clustered standard errors
Panels 1, 2, and 3 estimate Equations 1, 2, and 3
*** 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)	-0.629***	
	(0.0496)	
GPS-measured plot size (log ha)		-0.621***
		(0.0636)
Observations	2175	2181
Adjusted R^2	0.229	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)	-0.670***	-0.662***	-0.603***	-0.712***	-0.728***	-0.667***
	(0.0965)	(0.0994)	(0.0900)	(0.0925)	(0.0921)	(0.0864)
Soil pH (pH)		2.239*				0.711
		(1.306)				(1.351)
Soil pH^2 (pH^2)		-0.159				-0.0329
		(0.107)				(0.112)
Soil sand (%)		-0.00374				-0.00468
		(0.00593)				(0.00560)
Soil organic carbon (%)		-0.00942				-0.0218
		(0.0403)				(0.0447)
Labor intensity (log hrs/ha/day)			0.106**			0.102**
			(0.0425)			(0.0416)
Organic amendment (binary)			0.0497			0.0750
			(0.148)			(0.156)
Inorganic fertilizer (binary)			1.559***			1.299***
			(0.204)			(0.315)
Irrigation (binary)			0.0637			-0.283
			(0.372)			(0.432)
Terracing (binary)			0.423***			0.465***
			(0.137)			(0.145)
Head owns plot (binary)				-0.111		-0.104
				(0.164)		(0.160)
Head manages plot (binary)				0.268		0.119
				(0.207)		(0.206)
(Head owns)X(Head manages)				-0.0558		0.000317
				(0.241)		(0.236)
Crops are rotated (%)				-0.0969		0.0434
				(0.118)		(0.116)
Crops are mono-cropped (%)				0.0776		0.0557
				(0.263)		(0.264)
Mixed cropping (%)				0.507*		0.414
				(0.266)		(0.262)
Tubers grown (binary)					0.101	0.00536
					(0.117)	(0.114)
Cereals grown (binary)					0.0836	-0.00553
					(0.109)	(0.105)
Legumes grown (binary)					0.202**	0.0617
					(0.102)	(0.113)
Bananas grown (binary)					0.148	0.117
					(0.168)	(0.163)
Cash crops grown (binary)					0.676***	0.522***
					(0.177)	(0.173)
Observations	1623	1623	1623	1623	1623	1623
Adjusted R^2	0.370	0.378	0.390	0.390	0.390	0.425
R^2	0.371	0.380	0.393	0.393	0.393	0.433

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)	-0.621***	-0.124		-1.027***
	(0.0636)	(0.126)		(0.128)
Perimeter-area ratio (log m/ha)		0.903***	1.099***	
		(0.231)	(0.0961)	
Perimeter (log m)				0.903***
				(0.231)
Observations	2181	2181	2181	2181
Adjusted R^2	0.381	0.393	0.392	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***	-0.264*		-1.019***
	(0.0626)	(0.141)		(0.126)
Perimeter-area ratio (log m/ha)		0.755***	1.169***	
		(0.242)	(0.0939)	
Perimeter (log m)				0.755***
				(0.242)
Observations	2076	2076	2076	2076
Adjusted R^2	0.186	0.196	0.193	0.196

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)	-0.405***	-0.400***	-0.329*
	(0.136)	(0.136)	(0.177)
Over-estimate (% area)	0.157***	0.154***	0.110***
	(0.0337)	(0.0332)	(0.0404)
Over-estimate squared	-0.00378***	-0.00444***	-0.00343**
	(0.00135)	(0.00124)	(0.00151)
Under-estimate (% area)	-2.242***	-2.156***	-2.726**
	(0.767)	(0.761)	(1.147)
Under-estimate squared	2.718***	2.464***	3.324**
	(0.917)	(0.904)	(1.455)
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	2181	2181	1623
Adjusted R^2	0.414	0.424	0.455

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

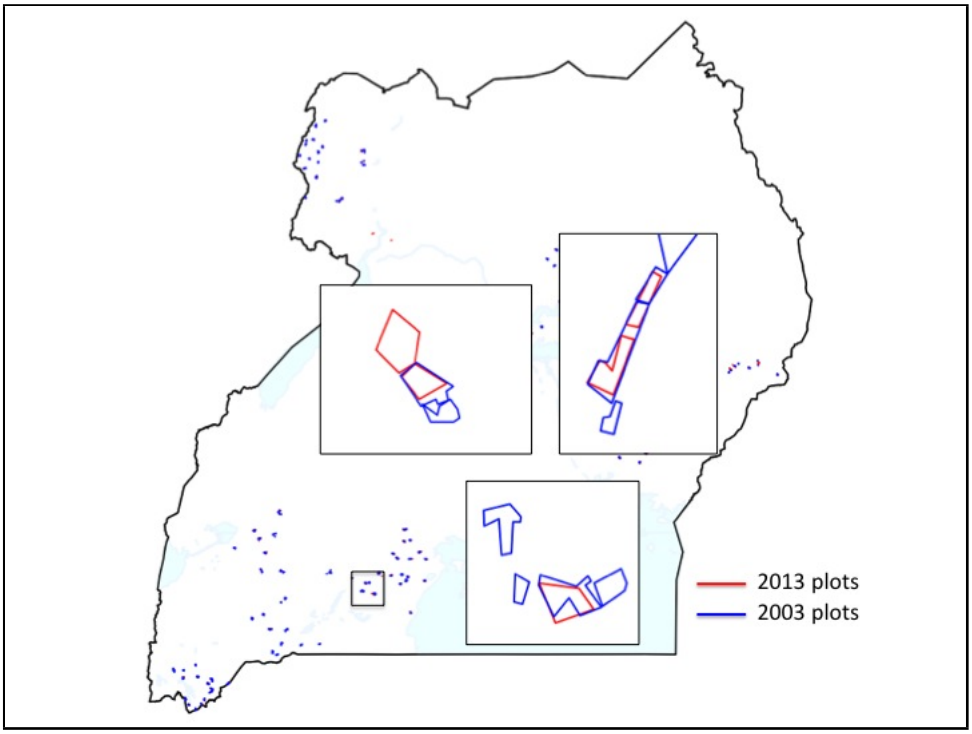
In both 2003 and 2013, enumerators collected GPS waypoints around the perimeter of each plot. In 2003 they did this by slowly walking around the plot while the GPS unit took waypoints at automated intervals — the GPS data suggests that they were instructed to pause at corners and flex-points, in order 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. 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

Table A1: Household-level Attrition From 2003 to 2013

	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***

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, Gathua and Woomer, 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, Gathua and Woomer, 2002).

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. Fig A2 shows the distribution of the demeaned independent variable plot size, under these two forms of identification. The pooled data variable is given by log plot size, demeaned by household-year-season categories. The panel data variables is given by plot size, demeaned by household-plot categories. The two distributions are clearly quite 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, implying that the plot level 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, the perimeter-area ratio explains approximately half, but not 100 percent, of the inverse size-productivity relationship, as seen in Tables A5 (analogous to Table 5). Similarly, and the perimeter-area ratio explains approximately half, but not 100 percent, of the inverse size-labor relationship, as seen in Table A6 (analogous to Table 6). Second, farmer perceptions no longer appear associated with plot productivity, as seen in Table A7.

The fact that the inverse relationship remains under household-year-season fixed effects, even once perimeter-area ratio is controlled for, suggests that larger plots are on average less productive than smaller plots within households, but due to time-invariant, plot-level characteristics that are merely associated with plot size. This phenomenon appears to be in addition to the (causal) effect of plot perimeter-area ratio, which drives the inverse relationship under plot fixed effects. Controlling for distance to residence does not mitigate the inverse relationship, however. Controlling for additional, time-varying plot characteristics such as those explored in Table 4 mitigates the relationship only sightly. (These regressions are available upon request.)

An alternate explanation might revolve around the scale of variation in plot size under these two model specifications. Difference from mean plot size is larger under the plot fixed effect specification than under the household-year-season fixed effect specification; Figure A1 illustrates this fact. Pooling positive and negative deviations from mean plot size, the mean (median) deviation under plot fixed effects is 0.18 (0.08), whereas the mean (median) deviation under household-year-season fixed effects is 0.09 (0.02). (This is largely because plots and farms shrank significantly over the decade, as shown in the Table 1.) The perimeter-area ratio only proxies for the edge effect when the area of the "edge" of a plot is approximately given by (perimeter)*(border length), and this works best with larger plots. With smaller plots, (perimeter)*(border length) will over-estimate the area at the edge of a plot. It is possible, therefore, that with less variation in plot size the perimeter-area ratio becomes a poor proxy for the edge effect. This might explain why it does less to mitigate the inverse relationship under the household-year-season fixed effects.

Figure A2: Identifying Variation in Plot Size

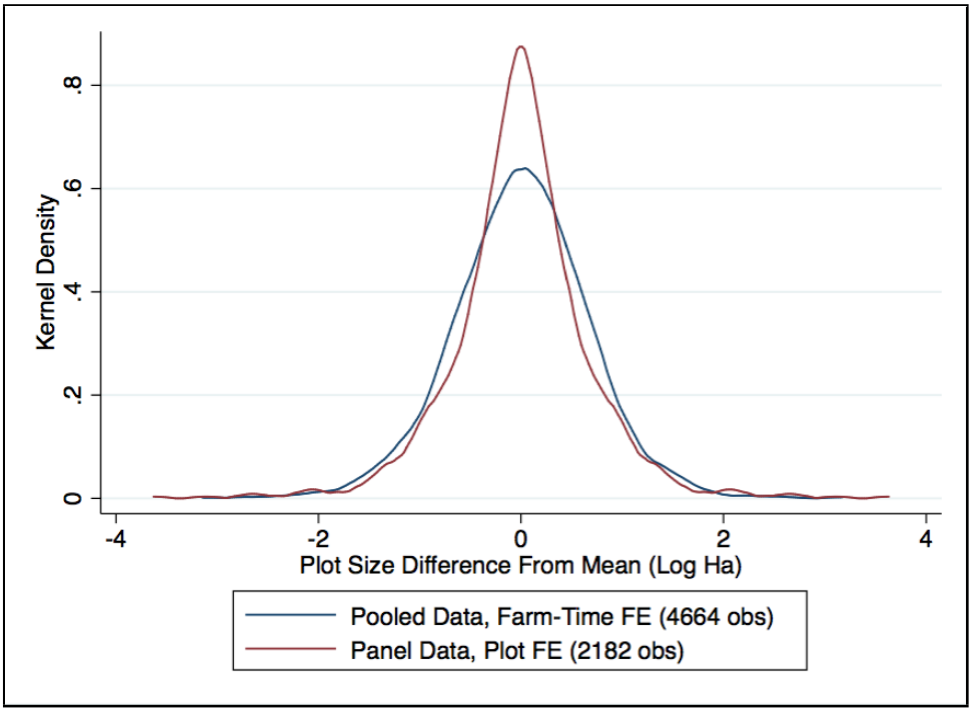


Table A2: Plot Characteristics in 2003 and 2013

	2003		2013		
	Mean or Median†	Standard Deviation	Mean or Median†	Standard Deviation	T Statistic‡
Size, Productivity, Labor					
Farm size (ha)	0.92	1.96	0.61	0.91	22.09***
Plot size (ha)	0.19	1.83	0.18	0.39	4.02***
Perimeter-area ratio (m/ha)	1,093.91	1,071.70	1,117.90	5,141.13	-4.14***
Plot productivity (revenue§/ha)	105.40	1,383.62	247.99	6,948.77	-24.11***
Labor intensity (hrs/ha/day)	2.45	15.08	1.73	53.60	9.02***
Soils					
Soil pH (pH)	6.13	0.60	6.09	0.60	2.27**
Soil sand (%)	60.29	15.11	54.60	15.52	12.96***
Soil organic carbon (%)	3.63	2.03	3.43	1.84	3.58***
Inputs					
Organic amendment (%)	10.54	30.71	8.44	27.80	2.85***
Inorganic fertilizer (%)	1.68	12.85	1.66	12.77	0.07
Irrigation (%)	1.70	12.92	0.31	5.52	5.21***
Terracing (%)	15.43	36.13	6.81	25.19	10.67***
Management					
Head owns plot (%)	61.28	48.72	74.19	43.77	-11.04***
Head manages plot (%)	50.94	50.00	64.39	47.89	-10.96***
(Head owns)X(Head manages)	42.99	49.51	57.76	49.40	-11.96***
Crops are rotated (%)	29.95	45.81	49.40	50.01	-15.30***
Crops are mono-cropped (%)	58.57	49.27	42.43	49.43	13.10***
Mixed cropping (%)	38.05	48.56	44.20	49.67	-5.02***
Intercropping (%)	3.36	18.02	12.96	33.59	-15.17***
Crops Grown					
Tubers grown (%)	40.10	49.02	26.07	43.91	11.93***
Cereals grown (%)	50.54	50.00	50.45	50.01	0.07
Legumes grown (%)	51.58	49.98	44.01	49.65	6.08***
Bananas grown (%)	29.60	45.66	19.07	39.29	9.76***
Cash crops grown (%)	18.30	38.67	15.34	36.04	3.15***

†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 A3: Measurement Error and the Inverse Relationship (Household-Time FE)

	(1) Plot Productivity	(2) Plot Productivity
Farmer-recalled plot size (log ha)	-0.547***	
	(0.0260)	
GPS-measured plot size (log ha)		-0.565***
		(0.0259)
Observations	5730	4845
Adjusted R^2	0.136	0.170

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 A4: 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)	-0.533***	-0.514***	-0.390***	-0.551***	-0.561***	-0.352***
	(0.0309)	(0.0309)	(0.0334)	(0.0305)	(0.0305)	(0.0346)
Soil pH (pH)		-0.856				-1.097**
		(0.557)				(0.526)
Soil pH^2 (pH^2)		0.0955**				0.110**
		(0.0469)				(0.0440)
Soil sand (%)		0.00318				0.00327
		(0.00269)				(0.00259)
Soil organic carbon (%)		0.0270				0.0309*
		(0.0196)				(0.0173)
Labor intensity (log hrs/ha/day)			0.299***			0.344***
			(0.0292)			(0.0304)
Organic amendment (binary)			0.318***			0.0661
			(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***			0.234**
			(0.0922)			(0.0934)
Head owns plot (binary)				-0.0985		-0.115
				(0.181)		(0.169)
Head manages plot (binary)				0.202		0.166
				(0.174)		(0.161)
(Head owns)X(Head manages)				0.0161		0.0584
				(0.211)		(0.193)
Crops are rotated (%)				-0.333***		-0.375***
				(0.0820)		(0.0843)
Crops are mono-cropped (%)				-0.104		-0.0370
				(0.141)		(0.133)
Mixed cropping (%)				0.153		0.0482
				(0.145)		(0.135)
Tubers grown (binary)					0.277***	0.205***
					(0.0572)	(0.0559)
Cereals grown (binary)					0.00661	-0.0488
					(0.0559)	(0.0560)
Legumes grown (binary)					0.234***	0.114**
					(0.0507)	(0.0546)
Bananas grown (binary)					0.535***	0.354***
					(0.0777)	(0.0801)
Cash crops grown (binary)					0.236***	0.125*
					(0.0737)	(0.0698)
Observations	3472	3472	3472	3472	3472	3472
Adjusted R^2	0.154	0.170	0.241	0.171	0.197	0.295
R^2	0.154	0.171	0.242	0.173	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 A5: 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)	-0.565***	-0.304***		-0.802***
	(0.0259)	(0.0604)		(0.0556)
Perimeter-area ratio (log m/ha)		0.499***	0.995***	
		(0.105)	(0.0424)	
Perimeter (log m)				0.499***
				(0.105)
Observations	4845	4845	4845	4845
Adjusted R^2	0.170	0.176	0.169	0.176

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 A6: 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)	-0.551***	-0.373***		-0.722***
	(0.0228)	(0.0538)		(0.0463)
Perimeter-area ratio (log m/ha)		0.349***	0.976***	
		(0.0894)	(0.0387)	
Perimeter (log m)				0.349***
				(0.0894)
Observations	4575	4575	4575	4575
Adjusted R^2	0.219	0.223	0.209	0.223

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 A7: 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)	0.0337	0.0435	0.0779
	(0.0804)	(0.0803)	(0.0917)
Over-estimate (% area)	0.0704***	0.0739***	0.0320
	(0.0204)	(0.0208)	(0.0249)
Over-estimate squared	-0.00151**	-0.00203***	-0.00115
	(0.000672)	(0.000734)	(0.000869)
Under-estimate (% area)	-0.365	-0.336	-0.106
	(0.460)	(0.461)	(0.544)
Under-estimate squared	-0.329	-0.390	-0.284
	(0.560)	(0.565)	(0.659)
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	4845	4845	3472
Adjusted R^2	0.190	0.192	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 Plot Size Change Over Time

Plots, defined as an area of land with a continuous cropping system, are situated on parcels. A parcel is a contiguous piece of land under one form of ownership, and in our pooled dataset of all plots from all time periods, parcels have between 1 and 10 plots on them. Fifty percent of our 2003 plots come from a parcel that holds only that single plot — this figure is 57 percent in 2013. The average plot in 2003 is on a parcel that holds 2 plots, and the average plot in 2013 is on a parcel that holds 1.8 plots.

Between 2003 and 2013, 32 percent of plots grew and 68 percent of plots shrank — but the majority of this variation is not driven by parcel-level divisions or changes. Indeed, in our full dataset of 2013 plots, only 7 percent sit on a parcel that experienced sub-division between 2003 and 2013. This is despite the fact that 21 percent of houses did sub-divide at least 1 parcel over the decade, and 40 percent of households either lost an entire parcel over the decade or lost a sub-division of a parcel. (Seventy percent of land disposition, including both entire parcels being lost and parcel sub-divisions being lost, is done for the purpose of land sale or land gift/bequeathment.)

Rather, the primary predictor of plot size change (change in plot hectares between 2003 and 2013) is starting plot size in 2003. This is illustrated by Figure A3. Plots that were larger in 2003 are more likely to shrink over the decade. Contingent on shrinking, they also shrink more. (Though conversely, for plots that grow over time, 2003 size is positively correlated with how much they grow.) While it seems plausible that plots belonging to households who lost a parcel, or plots situated on parcels that experienced sub-division, might shrink more than other plots, Figure A4 suggests that this may not be the case. The univariate, non-parametric relationship between starting plot size and plot size change is the same across (i) all pots, (ii) plots belonging to households who lost a parcel, and (iii) plots situated on parcels that experienced sub-division. (However, this figure also illustrates that plots on parcels that experience subdivision generally started out larger in 2003 than other plots.)

Table A8 examines correlates with plot size change — the logarithm of the difference between plot size in 2013 and plot size in 2003 — via a multiple regression framework. In Columns 1-4 plot size change is estimated as a function of 2003 covariates only, one observation per plot, by simple OLS. (For continuous covariates observed in both seasons of 2003 an average is used, and for binary covariates observed in both seasons of 2003 the maximum value is used.) In Columns 5-8 plot size change is again estimated as a function of 2003 covariates, but including household fixed effects. Under both identification strategies (i.e., whether viewing variation across all plots or across plots within households), the strongest predictor of plot size change is 2003 plot size. Neither parcel subdivision nor the household-level loss of a parcel is significantly associated with plot size change, as suggested by Figure A4.

Yet, in Column 7 we do observe a negative effect of both parcel sub-division and parcel sub-division interacted with plot size. Figure A5 displays a similar effect graphically — the variable being graphed is plot size change demeaned by household, and while 2003 plot size is negatively associated with (demeaned) plot size change for all plots, this is particularly true for those plots on subdivided parcels. The confidence intervals also make it clear, however, why the effect is not statistically significant.

Figure A3: Plot Size Change by 2003 Plot Size

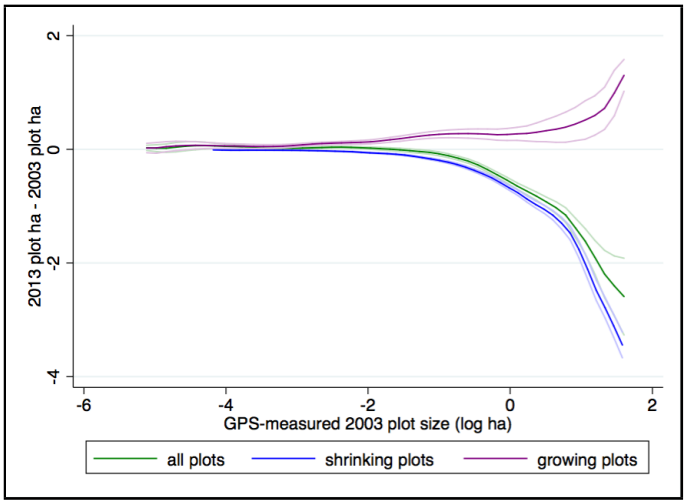


Figure A4: Plot Size Change by Parcel Occurrences

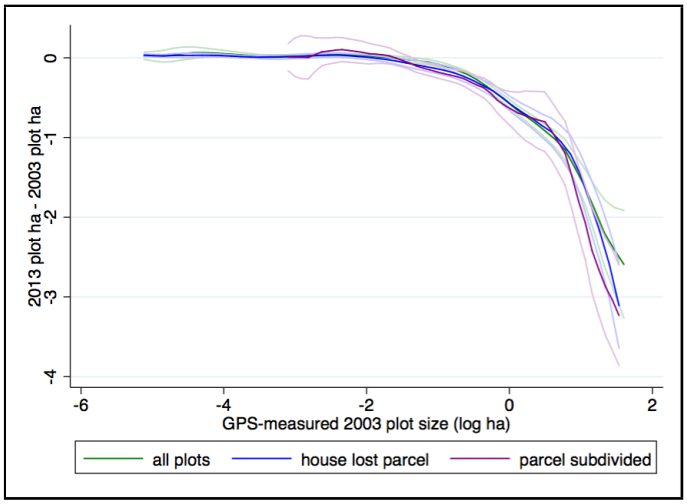


Figure A5: Plot Size Change by 2003 Plot Size (Deviation from Household Mean)

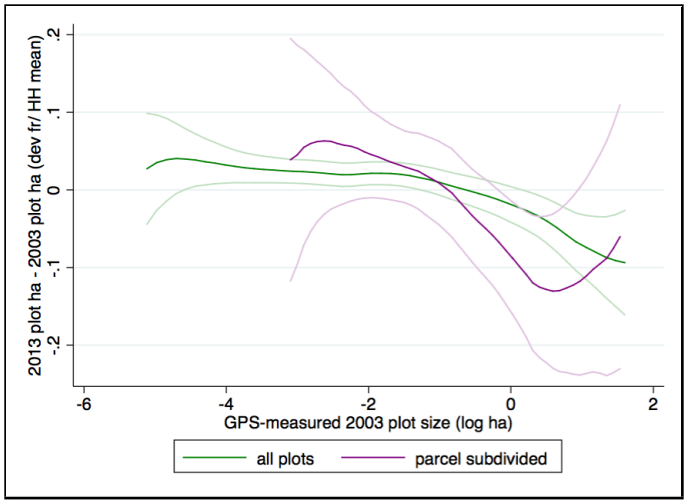


Table A8: Correlates of Plot Size Change in Geospatially-Matched Plots (OLS)

	OLS				Household Fixed Effects			
	(1) Plot Size Change	(2) Plot Size Change	(3) Plot Size Change	(4) Plot Size Change	(5) Plot Size Change	(6) Plot Size Change	(7) Plot Size Change	(8) Plot Size Change
GPS-measured plot size '03 (log ha)	-0.498***		-0.496***	-0.502***	-0.600***		-0.529***	-0.422***
	(0.0441)		(0.0705)	(0.0724)	(0.0702)		(0.0961)	(0.126)
Plot on a sub-divided parcel (binary)		-0.391	-0.0697	0.350		-0.777***	-0.234	0.672
		(0.253)	(0.322)	(0.236)		(0.218)	(0.359)	(0.502)
(Plot size '03)x(Lost parcel/sub-division)			-0.0856	0.0627			0.0853	0.390*
			(0.179)	(0.127)			(0.174)	(0.210)
At least 1 parcel/division was lost (binary)		0.0240	-0.139	-0.204				
		(0.132)	(0.192)	(0.197)				
(Plot size '03)x(Sub-divided parcel)			0.0206	-0.00560			-0.0917	-0.102
			(0.0911)	(0.0930)			(0.158)	(0.163)
Soil pH (pH)				2.775**				3.877*
				(1.170)				(1.986)
Soil pH^2 (pH^2)				-0.224**				-0.301*
				(0.0991)				(0.167)
Soil sand (%)				-0.00204				0.00656
				(0.00353)				(0.00601)
Soil organic carbon (%)				0.0298				-0.0512
				(0.0273)				(0.0616)
Labor intensity (log hrs/ha/day)				-0.0731				0.107
				(0.0487)				(0.0878)
Organic amendment (binary)				-0.0998				-0.145
				(0.127)				(0.184)
Inorganic fertilizer (binary)				-1.638				0
				(1.224)				(.)
Irrigation (binary)				0.298				-0.0694
				(0.279)				(0.369)
Terracing (binary)				0.254*				0.0794
				(0.130)				(0.245)
Head owns plot (binary)				0.234*				0.173
				(0.137)				(0.567)
Head manages plot (binary)				0.248				-0.572*
				(0.173)				(0.311)
(Head owns)X(Head manages)				-0.523**				0.428
				(0.211)				(0.824)
Crops are rotated (%)				0.0596				0.0730
				(0.119)				(0.179)
Crops are mono-cropped (%)				-0.168				-0.271
				(0.169)				(0.219)
Mixed cropping (%)				-0.0989				0.126
				(0.182)				(0.207)
Tubers grown (binary)				-0.309***				-0.107
				(0.113)				(0.175)
Cereals grown (binary)				-0.148				-0.221
				(0.108)				(0.144)
Legumes grown (binary)				0.105				0.0432
				(0.127)				(0.171)
Bananas grown (binary)				-0.109				-0.124
				(0.160)				(0.203)
Cash crops grown (binary)				-0.135				-0.235
				(0.125)				(0.176)
Observations	738	656	656	552	738	657	656	552
Adjusted R^2	0.260	0.005	0.253	0.315	0.197	0.004	0.175	0.194

Dependent variable: log([2013 revenue/hectare] - [2003 revenue/hectare])
All covariates are from round 1 (R1) of the geospatially matched panel dataset
Columns 1-4 estimated via OLS; Columns 4-5 estimated via household fixed effects
Household-clustered standard errors in parentheses for all columns
*** p<0.01, ** p<0.05, * p<0.1

Appendix 6 Selection into Datasets

1. Selection into Geospatially-Matched Plot Panel

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 A9 and A10 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 A9 the unit of observation is the household, while in Table A10 the unit of observation is the plot.

Table A9 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 A10 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 A9: 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	1.26	-0.68	0.69	0.66	0.15
Number of plots owned (#)	4.69	4.41	1.64	4.15	4.11	0.33
Number of crops grown (#)	5.85	5.99	-0.83	3.41	3.51	-0.90
Head years of education (#)	4.94	4.84	0.39	5.45	5.23	0.92
Head age (#)	41.50	43.39	-1.85*	49.53	51.37	-2.04**
Household size (# people)	5.86	6.32	-2.13**	6.46	6.08	1.77*
Asset index (index)	14.00	14.14	-1.88*	13.34	13.42	-1.22
Net crop income (1,000 Ush)	527.75	529.16	-2.07**	746.48	820.83	-0.95
Distance to all weather road (km)	2.51	2.42	0.20	5.66	4.82	2.04**
Distance to market (km)	3.14	2.83	1.04	4.68	5.13	-0.58

Table A10: 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	-5.56***	0.30	0.31	0.29
Perimeter-area ratio (m/ha)	1,375.42	1,190.42	5.14***	1,759.60	1,796.13	0.02
Plot productivity (revenue/ha)	324.62	291.09	1.40	1,006.35	1,139.03	-0.68
Labor intensity (hrs/ha/day)	5.19	4.34	2.49**	6.79	9.07	-1.09
Soil pH (pH)	6.11	6.19	-2.95***	6.06	6.12	-2.03**
Soil sand (%)	60.92	60.30	0.83	55.52	53.37	2.89***
Soil organic carbon (%)	3.55	3.44	1.16	3.33	3.64	-3.63***
Organic amendment (%)	8.94	17.12	-6.01***	7.28	10.16	-2.41**
Inorganic fertilizer (%)	1.69	1.27	0.75	1.80	1.63	0.31
Irrigation (%)	1.67	1.44	0.39	0.29	0.14	0.68
Terracing (%)	14.48	22.66	-5.03***	6.27	9.24	-2.63***
Head owns plot (%)	59.87	66.09	-2.87***	75.30	74.93	0.19
Head manages plot (%)	50.22	54.20	-1.79*	63.79	63.69	0.05
(Head owns)X(Head manages)	42.45	45.48	-1.38	57.13	58.94	-0.84
Crops are rotated (%)	30.49	24.51	2.91***	51.60	45.73	2.41**
Crops are mono-cropped (%)	59.95	47.23	5.81***	47.15	41.06	2.80***
Mixed cropping (%)	36.05	49.76	-6.37***	39.59	47.56	-3.69***
Tubers grown (%)	40.21	42.79	-1.18	26.34	25.61	0.38
Cereals grown (%)	49.35	48.65	0.31	49.86	44.72	2.35**
Legumes grown (%)	49.98	51.98	-0.90	41.68	43.63	-0.90
Bananas grown (%)	24.32	42.16	-9.06***	16.13	24.12	-4.72***
Cash crops grown (%)	16.24	27.42	-6.52***	13.71	17.21	-2.26**

2. Selection into Pooled Dataset under Household-Time Fixed Effects
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, in any number of ways. If so, the houses/plots used for household-time fixed effect analysis might not be representative of the larger universe of houses/plots that we view in our data. As we would expect, Table A11 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 A12 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 A11: 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*	0.69	0.85	-5.82***
Number of plots owned (#)	4.69	5.07	-3.10***	4.15	4.75	-4.91***
Number of crops grown (#)	5.85	6.13	-2.55**	3.41	3.90	-5.26***
Head years of education (#)	4.94	5.03	-0.50	5.45	5.41	0.21
Head age (#)	41.50	41.37	0.18	49.53	49.96	-0.56
Household size (# people)	5.86	5.97	-0.75	6.46	6.72	-1.41
Asset index (index)	14.00	14.02	-0.52	13.34	13.38	-0.80
Net crop income (1,000 Ush)	527.75	559.25	-1.84*	746.48	826.69	-2.37**
Distance to all weather road (km)	2.51	2.40	-0.00	5.66	5.14	0.69
Distance to market (km)	3.14	3.17	-0.29	4.68	4.88	-0.38

Table A12: 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	0.30	0.30	-0.08
Perimeter-area ratio (m/ha)	1,375.42	1,397.59	-1.10	1,759.60	1,768.43	0.16
Plot productivity (revenue/ha)	324.62	334.36	-0.49	1,006.35	1,061.54	-0.14
Labor intensity (hrs/ha/day)	5.19	5.30	-0.55	6.79	7.06	0.06
Soil pH (pH)	6.11	6.09	0.53	6.06	6.06	0.16
Soil sand (%)	60.92	60.80	0.23	55.52	55.32	0.34
Soil organic carbon (%)	3.55	3.59	-0.48	3.33	3.32	0.17
Organic amendment (%)	8.94	8.61	0.41	7.28	6.56	0.81
Inorganic fertilizer (%)	1.69	1.53	0.43	1.80	1.57	0.51
Irrigation (%)	1.67	1.67	-0.02	0.29	0.33	-0.24
Terracing (%)	14.48	14.49	-0.02	6.27	6.43	-0.18
Head owns plot (%)	59.87	59.25	0.44	75.30	75.08	0.14
Head manages plot (%)	50.22	49.86	0.25	63.79	63.28	0.30
(Head owns)X(Head manages)	42.45	42.11	0.24	57.13	57.05	0.05
Crops are rotated (%)	30.49	30.82	-0.25	51.60	52.80	-0.62
Crops are mono-cropped (%)	59.95	61.08	-0.81	47.15	48.92	-1.01
Mixed cropping (%)	36.05	35.03	0.75	39.59	37.77	1.07
Tubers grown (%)	40.21	39.01	0.87	26.34	26.03	0.20
Cereals grown (%)	49.35	48.03	0.93	49.86	49.31	0.31
Legumes grown (%)	49.98	48.99	0.70	41.68	39.21	1.44
Bananas grown (%)	24.32	23.93	0.32	16.13	15.93	0.15
Cash crops grown (%)	16.24	15.78	0.45	13.71	13.90	-0.16

Appendix 7 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 8 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 A13 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 A14 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 A15 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 simsim, 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 A16 therefore estimates Equation 10 for subsets of plots according to crop height. The edge effect is identical for tall crops (Column 2) and for low to the 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 A17 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 A13: 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)	1.098***	1.066***	0.652***
	(0.0957)	(0.107)	(0.182)
Observations	2080	2044	789
Adjusted R^2	0.183	0.162	0.076

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 A14: 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)	1.098***	1.178***	1.388***	0.685***
	(0.0957)	(0.101)	(0.101)	(0.164)
Observations	2080	1873	1401	1368
Adjusted R^2	0.183	0.231	0.284	0.056

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 A15: 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)	1.076***	1.205***	1.024***	1.139***
	(0.0963)	(0.182)	(0.388)	(0.169)
Soil organic carbon (%)		0.324		
		(0.249)		
(Soil organic carbon)X(Perimeter-area ratio)		-0.0465		
		(0.0342)		
Soil sand (%)			-0.00759	
			(0.0454)	
(Soil sand)X(Perimeter-area ratio)			0.000578	
			(0.00645)	
Soil nitrogen (%)				1.642
				(2.807)
(Soil nitrogen)X(Perimeter-area ratio)				-0.326
				(0.386)
Observations	2189	1898	1898	1898
Adjusted R^2	0.383	0.376	0.374	0.377

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 A16: 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)	1.076***	1.512***	1.567***	0.809***
	(0.0963)	(0.123)	(0.178)	(0.116)
Observations	2189	991	528	1290
Adjusted R^2	0.383	0.472	0.365	0.284

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 A17: 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)	1.076***	1.103***	1.244***
	(0.0963)	(0.134)	(0.131)
Observations	2189	1118	1308
Adjusted R^2	0.383	0.386	0.435

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 9 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 A18 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 A19 and A20 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 A21 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 A22 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 esteemed 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 A18: 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***	-0.686***	-0.565***
	(0.0351)	(0.0358)	(0.0259)
Observations	2804	2041	4845
Adjusted R^2	0.118	0.259	0.170

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 A19: 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)	-0.621***	-0.621***	-0.892***	-0.495***	-0.447***	-0.572***
	(0.0636)	(0.134)	(0.102)	(0.104)	(0.0940)	(0.196)
Observations	2181	744	1015	1071	831	544
Adjusted R^2	0.381	0.299	0.457	0.335	0.283	0.300

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 A20: 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)	-0.632***	-0.482***	-0.536***	-0.766***	-0.645***	-0.814***
	(0.0971)	(0.0688)	(0.0736)	(0.208)	(0.131)	(0.108)
Observations	1551	1298	1154	610	872	1162
Adjusted R^2	0.388	0.330	0.336	0.260	0.360	0.451

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 A21: 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)	-0.621***	-0.770***	-0.429**	-0.660***
	(0.0636)	(0.0828)	(0.215)	(0.0731)
Observations	2181	1531	650	1508
Adjusted R^2	0.381	0.463	0.110	0.437

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 A22: 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)	-0.120*		0.0246
	(0.0667)		(0.0543)
Plot size (ha)		-0.603***	-0.621***
		(0.120)	(0.125)
Observations	2181	2181	2181
Adjusted R^2	0.262	0.287	0.286
	Panel 2: Log Productivity, Non-Linear		
Farm size (ha)	-0.442***		-0.0862
	(0.115)		(0.115)
(Farm size)^2	0.0498***		0.0157
	(0.0128)		(0.0123)
Plot size (ha)		-1.292***	-1.242***
		(0.269)	(0.279)
(Plot size)^2		0.218***	0.205**
		(0.0825)	(0.0832)
Observations	2181	2181	2181
Adjusted R^2	0.269	0.300	0.300
	Panel 3: Productivity, Linear		
Farm size (ha)	-77.93		128.5
	(89.99)		(78.28)
Plot size (ha)		-794.9	-890.0
		(557.4)	(601.5)
Observations	2182	2182	2182
Adjusted R^2	0.011	0.014	0.014
	Panel 4: Productivity, Non-Linear		
Farm size (ha)	-735.4		-65.11
	(529.5)		(168.4)
(Farm size)^2	101.5		26.84
	(71.12)		(25.52)
Plot size (ha)		-2876.8	-2849.0
		(1921.0)	(1889.4)
(Plot size)^2		657.8	634.2
		(445.7)	(431.2)
Observations	2182	2182	2182
Adjusted R^2	0.013	0.021	0.020

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 10 Potential Causality of the IR

If plot size was randomly distributed, or if change in plot size was a random treatment between 2003 and 2013, the inverse relationship estimated in plot panel data would be causal. Yet Table A23 illustrates that this is not the case. Column 1 regresses plot size on all covariates from Table 4 using ordinary least squares, and Columns 2 and 3 do the same using household-year-season fixed effects (as in Appendix 4) and plot fixed effects (as in Tables 3-7), respectively. In all three cases, it is clear that plot size is predicted by other covariates.

Column 4 of A23 explains change in plot size — the logarithm of the difference between plot size in 2013 and plot size in 2003 — as a function of 2003 covariates only, using household fixed effects. (For continuous covariates observed in both seasons of 2003 an average is used, and for binary covariates observed in both seasons of 2003 the maximum value is used.) Column 5 does the same, but additionally controls for starting plot size in 2003. This column suggests that conditional on starting plot size, plot size change is exogenous. (Results from Appendix 5 suggest the same.) However, we cannot condition on 2003 plot size in panel estimations of the IR, and so for our purposes, plot size cannot be viewed as exogenous.

We cannot, therefore, interpret the inverse relationship as causal. 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 (Forthcoming).

These bounds are displayed in Table A24. None of them contain 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 the original, estimated effect. This is because controlling for these covariates increases R-squared while changing the estimated inverses relationship very little.

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 A6 illustrates the bias-adjusted estimator γ^* calculated for every combination of $\delta \in [0, 2]$ and $R_{max} \in [.51]$, maintaining $\hat{\gamma_4} = -0.670$, $R_4 = 0.371$, $\hat{\gamma_5} = -0658$, $R_5 = 0.423$, as in the final column of Table A24, i.e., based on the full set of controls in Table 4. Figure A6 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 suggest the causal inverse relationship to be lower than -0.6.

Table A23: Balance Test for Plot Size (OLS, Household-Time FE, Plot FE)

	(1) Plot Size	(2) Plot Size	(3) Plot Size	(4) Change in Size	(5) Change in Size
GPS-measured plot size '03 (log ha)					-0.502***
					(0.105)
Soil pH (pH)	1.900***	-0.113	-0.112	4.530**	3.977*
	(0.352)	(0.324)	(1.056)	(2.269)	(2.089)
Soil pH^2 (pH^2)	-0.149***	-0.00209	-0.000869	-0.349*	-0.320*
	(0.0295)	(0.0272)	(0.0872)	(0.191)	(0.173)
Soil sand (%)	0.00722***	-0.000908	0.0104***	0.00378	0.00834
	(0.00132)	(0.00182)	(0.00385)	(0.00628)	(0.00570)
Soil organic carbon (%)	-0.0468***	-0.0315**	-0.0733***	-0.00838	-0.00672
	(0.0105)	(0.0127)	(0.0280)	(0.0527)	(0.0489)
Labor intensity (log hrs/ha/day)	-0.401***	-0.447***	-0.254***	0.281***	0.0899
	(0.0154)	(0.0215)	(0.0444)	(0.0736)	(0.0825)
Organic amendment (binary)	0.281***	0.458***	0.273***	-0.397**	-0.0812
	(0.0594)	(0.0670)	(0.0886)	(0.178)	(0.174)
Inorganic fertilizer (binary)	-0.245	0.466***	-0.00592	0	0
	(0.273)	(0.163)	(0.526)	(.)	(.)
Irrigation (binary)	-0.214	0.143	-0.664***	-0.302	-0.105
	(0.170)	(0.128)	(0.241)	(0.221)	(0.371)
Terracing (binary)	0.0901*	0.289***	0.134	0.0226	0.137
	(0.0533)	(0.0644)	(0.109)	(0.262)	(0.243)
Head owns plot (binary)	-0.0244	-0.0275	0.138	0.0706	0.310
	(0.0514)	(0.116)	(0.120)	(0.423)	(0.464)
Head manages plot (binary)	0.0221	0.351***	0.157	-0.638	-0.468
	(0.0615)	(0.121)	(0.155)	(0.402)	(0.330)
(Head owns)X(Head manages)	-0.0405	-0.151	-0.292	0.448	-0.00681
	(0.0757)	(0.147)	(0.180)	(0.808)	(0.739)
Crops are rotated (%)	-0.0889**	0.142**	-0.211**	0.113	0.0845
	(0.0378)	(0.0575)	(0.0928)	(0.177)	(0.176)
Crops are mono-cropped (%)	0.0666	0.0123	0.380***	-0.0867	-0.148
	(0.0710)	(0.0840)	(0.141)	(0.254)	(0.210)
Mixed cropping (%)	0.273***	0.194**	0.466***	0.263	0.186
	(0.0722)	(0.0875)	(0.136)	(0.218)	(0.195)
Tubers grown (binary)	0.247***	0.212***	0.476***	-0.388**	-0.185
	(0.0343)	(0.0369)	(0.0984)	(0.182)	(0.170)
Cereals grown (binary)	0.351***	0.315***	0.208***	-0.446***	-0.253*
	(0.0337)	(0.0368)	(0.0710)	(0.144)	(0.131)
Legumes grown (binary)	0.244***	0.264***	0.177**	0.0145	0.0954
	(0.0357)	(0.0379)	(0.0737)	(0.170)	(0.167)
Bananas grown (binary)	-0.00194	0.00802	0.0676	-0.125	-0.144
	(0.0476)	(0.0569)	(0.113)	(0.206)	(0.195)
Cash crops grown (binary)	0.410***	0.230***	0.460***	-0.383**	-0.254
	(0.0466)	(0.0495)	(0.130)	(0.189)	(0.174)
Sample	Pooled	Pooled	Panel	Panel R1	Panel R1
Observations	4075	3476	1624	616	616
R^2	0.360	0.364	0.381	0.170	0.244

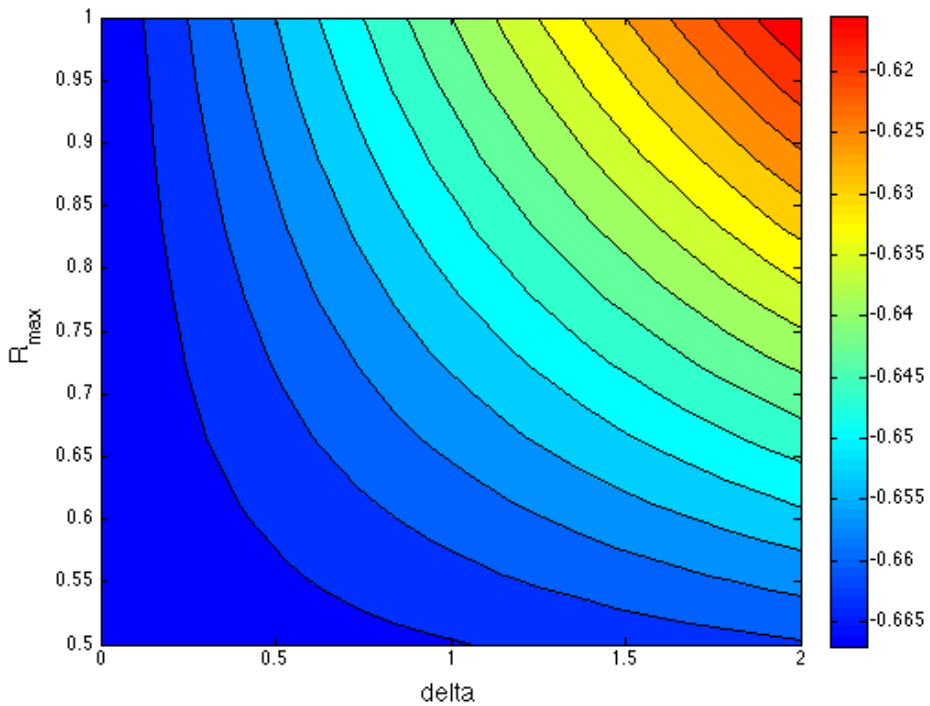
Dependent variable Cols 1-3: log(revenue/hectare)
Dependent variable Cols 4-5: log([2013 revenue/hectare] - [2003 revenue/hectare])
Column 1: Estimated with no fixed effects
Column 2: Estimated with household-year-season fixed effects
Columns 1, 2: Household-year-season-clustered standard errors in parentheses
Column 3: Estimated with plot fixed effects, plot-clustered standard errors in parentheses
Columns 4, 5: Estimated with household fixed effects, household-clustered standard errors in parentheses
Columns 4, 5: All covariates are from round 1 (R1), and the dependent variable is change in plot size between 2003 and 2013
p<0.01, ** p<0.05, * p<0.1

Table A24: Bounds for Potential Causal Relationship

	$R_{max} = 1.3 * R_5$, δ=1, $\hat{\gamma_4}$=-0.670, R_4=0.371				
	(1)	(2)	(3)	(4)	(5)
Column Controls	Soil	Inputs	Management	Crops	Full Set
Column R_2	0.380	0.393	0.393	0.393	0.433
Bounds $\left[\hat{\gamma_5}\ \gamma^*(R_{max}, \delta)\right]$	[-0.662, -0.561]	[-0.603, -0.244]	[-0.712, -0.937]	[-0.728, -1.039]	[-0.667, -0.661]

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

Figure A6: Bias-Adjusted Estimator $\gamma^*(R_{max}, \delta)$



Appendix 11 More on the Edge Effect

Explaining Perimeter-Area Variation

Figure A7 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 A8 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 A25 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 bares out in the data; in Column 1 we see that perimeter-area ratio is highest for triangular plots, and decreases with number of sides.

While Column 1 of Table A25 accounts for only plot shape, Column 2 of Table A25 accounts for only size. Shape variables explain 21 percent of variation, 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, except that plot shape variables in Column 4 now explain 40 percent of variation. This is almost half of the variation explained by plot size — so under plot fixed effects, both plot shape and plot size contribute substantially to variation in perimeter-area ratio.

Causality of Edge Effect

Table A26 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 (Forthcoming), using the results from Table A26.

Table A27 holds these causal bounds, again allowing X_{ijt} from Equation 5 to take the value of each set of controls in Columns 2-5 of Table A26, and then considering the full set of controls displayed in Column 6 of Table A26. We assume $\delta = 1$ and

$R_{max} = 1.3R_5$, as suggested by Oster (Forthcoming). None of the resulting bounds in A27 contain zero — in fact, the bounds contain only values over 0.97.

As in Appendix 10, we can examine the range of bounds that are possible under various δ and R_{max} parameters. Figure A9 illustrates the bias-adjusted estimator γ^* calculated for every combination of $\delta \in [0, 2]$ and $R_{max} \in [.51]$, maintaining $\hat{\gamma_4} = 1.162$, $R_4 = 0.382$, $\hat{\gamma_5} = 1.126$, $R_5 = .439$, as in the final column of Table A27. Even under very conservative assumptions ($R_{max} = 1$, $\delta = 2$), the range of possibly causal bounds contains only values substantially over zero.

Robustness of Edge Effect and the Role of Plot Shape

Tables A28 and A29 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 A30. (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 A31 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 A10 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 A31 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 A11, graphing productivity demeaned by year and season over perimeter-area ratio, illustrates that this differential can be seen in the (almost) raw data. Figure A12 again illustrates the differential by graphing productivity predicted by the regression model of Column 5, Table A31.

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 A32 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 A33 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 A33 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 A34 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 A32-A34 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 A35 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 A33 and A34.

Magnitude of the Edge Effect

The magnitude of the edge effect theoretically depends on two factors — border width b and the differential between interior and peripheral productivity $(Y^I - Y^P)$. Additionally, the estimated edge effect relies on the approximation $\frac{A^P_{ijt}}{A_{ijt}} \approx \frac{b*P_{ijt}}{A_{ijt}}$, and this approximation will vary by plot shape. Therefore, we expect the coefficient estimated on $\frac{P_{ijt}}{A_{ijt}}$ to also vary by shape, as well as perhaps by data subset if b or $(Y^I - Y^P)$ vary by data subset.

Tables A28-A30 suggest that the coefficient on $\frac{P_{ijt}}{A_{ijt}}$ is stable across data subsets; a 10 percent increase in perimeter-area ratio is associated with a 10 percent increase productivity. Table A31 suggests that this relationship varies slightly by plot shape; the marginal effect of perimeter-area ratio is larger for triangular plots and smaller for plots with more than 4 sides.

It is logical to ask whether these coefficients square with the theoretical productivity increases that we would expect under the edge effect. Below, Figure A13 illustrates how perimeter-area ratio changes within plot shape type (by row) as sides and angles shift (by column). We assume a fixed border width b, chosen such that for a square plot of 10 rows, 1 meter per row, the border will be defined as 1 row. Figure A14 illustrates the productivity of each plot, under this border width assumption and assuming that peripheral productivity is twice that of interior productivity. Notably, productivity does *not* move in a one to one relationship with perimeter-area ratio, as we estimated empirically.

In fact, Figure A15 displays calculated production elasticity with respect to perimeter-area ratio. (This is simply the observed percentage change in productivity that comes with the observed percentage change in perimeter-area ratio, change being defined relative to the first column of shapes.) A 10 percent increase in perimeter-area ratio is expected to increase production by 1.9-2.7 percent, about a fifth of the elasticity we estimate in data. If we assume a higher interior-periphery differential this elasticity increases — Figure A16 displays calculated production elasticities when we assume the peripheral rows are three times as productive as interior rows. Under this assumption a 10 percent increase in perimeter-area ratio is expected to increase production by 3.0-4.3

percent, which still falls far below the one to one change we observe in the data. Changing the border width again changes calculated elasticities, but elasticities never move above 0.5 under reasonable assumptions.

Figures A17-A20 display the same statistics as Figures A13-A16, but examine pure shifts in plot area (holding shape constant) rather than pure shifts in plot shape (holding area constant). In the second column area is increased by 50 percent relative to the first column, and in the third column area is increased by 100 percent, the area of column one. The elasticities obtained under these shifts are similar to the elasticities obtained under perimeters-area ratio shifts — a 10 percent increase in perimeter-area ratio returns a 2.3-2.9 percent increase in productivity if peripheral productivity is twice interior productivity, and returns a 3.7-4.4 percent increase if peripheral productivity is thrice interior productivity.

Last, Figures A21 and A22 display elasticities of productivity with respect to area (i.e. the inverse size-productivity relationship), rather than with respect to perimeter-area ratio. Again, these theoretical elasticities are far lower than the estimated coefficients we actually observe on plot area in multivariate regressions. We estimate an inverse relationship of around 0.7, implying that a 10 percent increase in plot size would decrease productivity by around 7 percent. Under reasonable assumptions border with and interior-peripheral productivity assumptions, the edge effect would only drive a decrease in productivity of about 0.5-1.5 percent, for a 10 percent increase in size.

Figure A7: Plot Size vs. Perimeter-Area Ratio

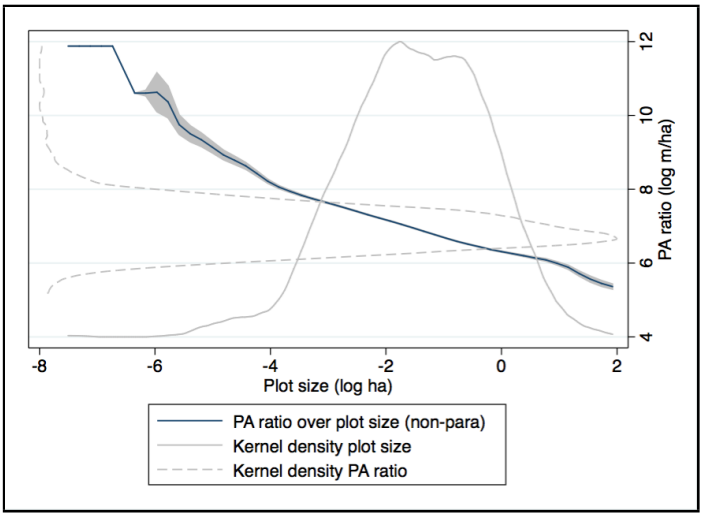


Figure A8: Plot Size vs. Perimeter-Area Ratio: Change over Time

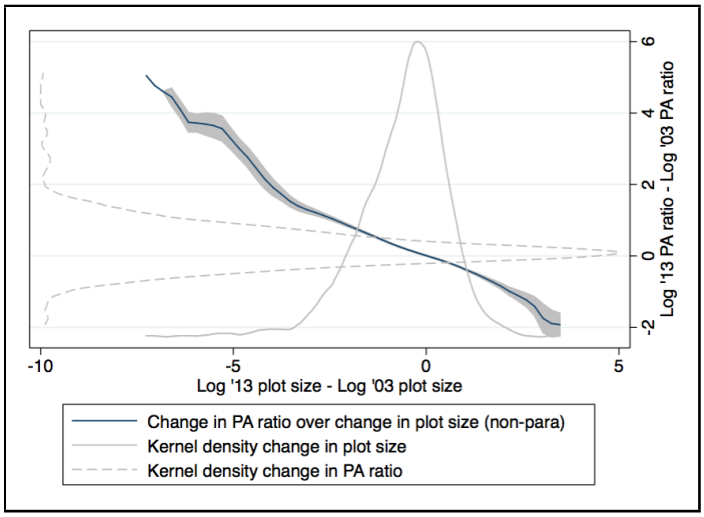


Table A25: Perimeter-Area Ratio Explained by Shape and Size (Plot Panel)

	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)	0.191***		0.0530***	0.0264		0.0589***
	(0.0457)		(0.0156)	(0.0456)		(0.0160)
Plot has 3 sides (binary)	1.103***		0.403***	1.060***		0.301***
	(0.238)		(0.101)	(0.277)		(0.0782)
Plot has 4 sides (binary)	0.0246		0.0146	0.181**		0.00739
	(0.0593)		(0.0173)	(0.0719)		(0.0223)
Number of sides (#)	-0.211***		-0.00771	-0.0329		0.0142
	(0.0272)		(0.0109)	(0.0444)		(0.0165)
(Number of sides)^2	0.00702***		0.00200***	-0.000909		0.000725
	(0.00110)		(0.000499)	(0.00224)		(0.000843)
GPS-measured plot size (log ha)		-0.511***	-0.529***		-0.530***	-0.576***
		(0.0103)	(0.00865)		(0.0169)	(0.0149)
Observations	2170	2182	2170	2170	2182	2170
Adjusted R^2	0.206	0.883	0.907	0.244	0.881	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 A26: 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)	1.188***	1.174***	1.081***	1.229***	1.238***	1.127***
	(0.140)	(0.145)	(0.137)	(0.131)	(0.132)	(0.126)
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	1623	1623	1623	1623	1623	1623
Adjusted R^2	0.386	0.393	0.404	0.401	0.399	0.432
R^2	0.387	0.396	0.407	0.405	0.402	0.440

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 A27: Bounds for Potential Causal Relationship

	$R_{max} = 1.3 * R_5$, δ=1, $\hat{\gamma_4}$=1.188, R_4=0.387				
	(1)	(2)	(3)	(4)	(5)
Column Controls	Soil	Inputs	Management	Crops	Full Set
Column R_2	0.396	0.407	0.405	0.402	0.440
Bounds $\left[\hat{\gamma_5}\ \gamma^*(R_{max}, \delta)\right]$	[1.174, 0.989]	[1.081, 0.428]	[1.229, 1.506]	[1.238, 1.640]	[1.127, 0.975]

Coefficients $\hat{\gamma_5}$ and R_5 from Columns 2-6 of Table A25
Coefficient $\hat{\gamma_4}$ and R_4 from Column 1 of Table A25

Figure A9: Bias-Adjusted Estimator $\gamma^*(R_{max}, \delta)$

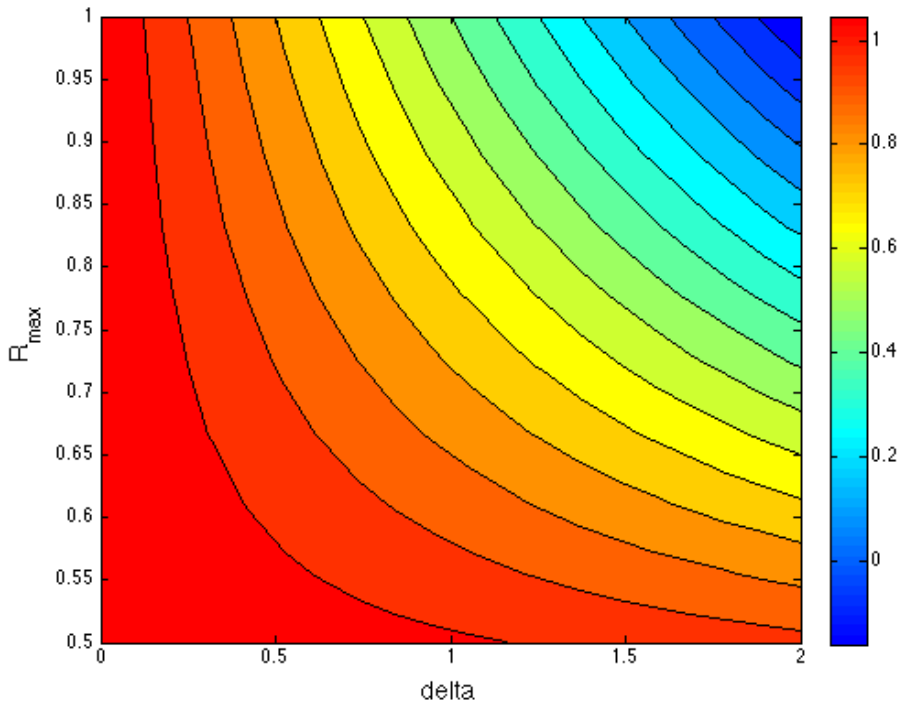


Table A28: The Inverse Relationship by Crop (Plot Panel)

	(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)
Perimeter-area ratio (log m/ha)	1.099***	1.190***	1.484***	0.887***	0.732***	1.063***
	(0.0961)	(0.193)	(0.123)	(0.163)	(0.138)	(0.326)
Observations	2181	744	1015	1071	831	544
Adjusted R^2	0.392	0.303	0.480	0.340	0.278	0.306

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 A29: 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)
Perimeter-area ratio (log m/ha)	1.197***	0.881***	1.034***	1.435***	1.038***	1.330***
	(0.142)	(0.133)	(0.144)	(0.285)	(0.206)	(0.132)
Observations	1551	1298	1154	610	872	1162
Adjusted R^2	0.411	0.325	0.339	0.281	0.351	0.466

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 A30: Edge Effect by Quantiles (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity
Perimeter-area ratio (log m/ha)	1.099***		
	(0.0961)		
P-A ratio quantiles=1 × Perimeter-area ratio (log m/ha)		1.364***	
		(0.192)	
P-A ratio quantiles=2 × Perimeter-area ratio (log m/ha)		1.342***	
		(0.180)	
P-A ratio quantiles=3 × Perimeter-area ratio (log m/ha)		1.273***	
		(0.171)	
P-A ratio quantiles=4 × Perimeter-area ratio (log m/ha)		1.275***	
		(0.163)	
P-A ratio quantiles=5 × Perimeter-area ratio (log m/ha)		1.287***	
		(0.150)	
Plot size quantiles=1 × Perimeter-area ratio (log m/ha)			1.116***
			(0.150)
Plot size quantiles=2 × Perimeter-area ratio (log m/ha)			1.096***
			(0.161)
Plot size quantiles=3 × Perimeter-area ratio (log m/ha)			1.063***
			(0.169)
Plot size quantiles=4 × Perimeter-area ratio (log m/ha)			1.113***
			(0.176)
Plot size quantiles=5 × Perimeter-area ratio (log m/ha)			1.122***
			(0.186)
Observations	2181	2181	2181
Adjusted R^2	0.392	0.402	0.398

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 A31: 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)	1.099***	1.112***	1.107***	1.463***	0.785***
	(0.0961)	(0.100)	(0.0956)	(0.162)	(0.121)
Number of sides (#)		0.00837		0.526***	
		(0.0148)		(0.175)	
Plot has 3 sides (binary)			0.0341		-7.674***
			(0.298)		(1.909)
Plot has 4 sides (binary)			-0.115		-2.302**
			(0.0928)		(0.961)
(Perimeter-area ratio)x(Number of sides)				-0.0785***	
				(0.0265)	
(Perimeter-area ratio)x(Plot has 3 sides)					0.982***
					(0.242)
(Perimeter-area ratio)x(Plot has 4 sides)					0.319**
					(0.138)
Observations	2181	2169	2181	2169	2181
Adjusted R^2	0.392	0.396	0.393	0.401	0.404

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 A10: Productivity by Shape

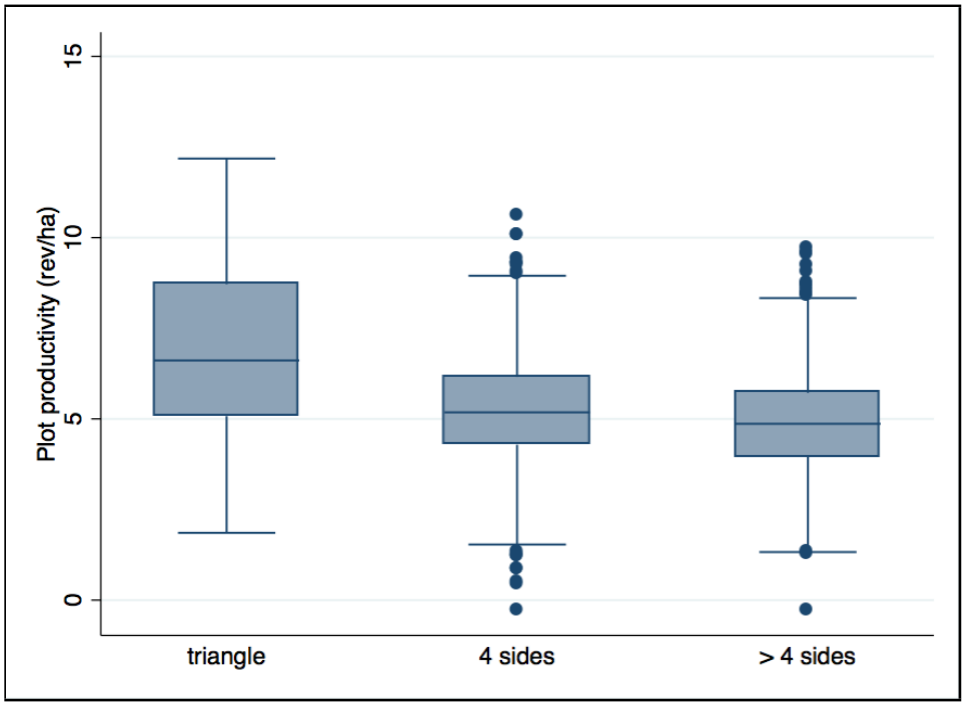


Figure A11: Productivity Demeaned by Round and Season

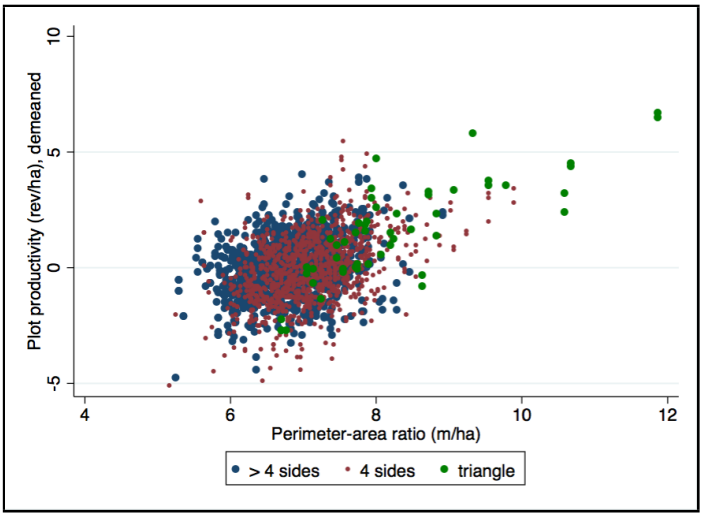


Figure A12: Productivity Prediction from Table A31 Col 5

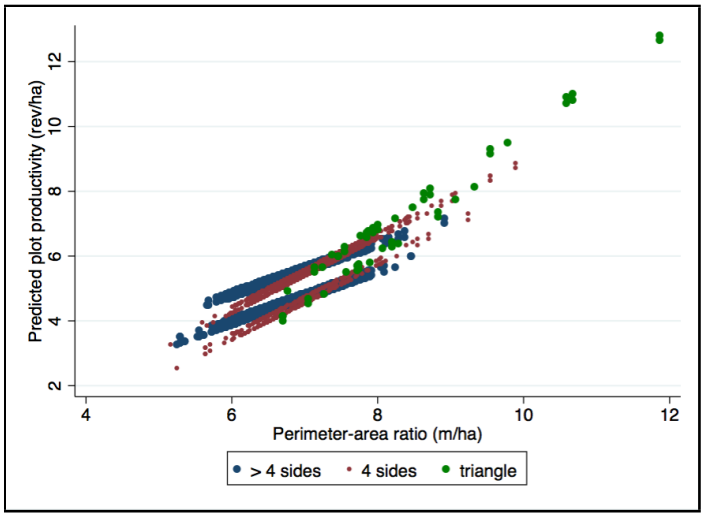


Table A32: Edge Effect by Area Correlation with Perimeter-Area Ratio (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity
GPS-measured plot size (log ha)	-0.0664	
	(0.136)	
Correlated Q1 × Perimeter-area ratio (log m/ha)	1.007***	1.115***
	(0.253)	(0.0973)
Correlated Q2 × Perimeter-area ratio (log m/ha)	1.049***	1.156***
	(0.254)	(0.0979)
Correlated Q3 × Perimeter-area ratio (log m/ha)	1.033***	1.141***
	(0.254)	(0.0984)
Correlated Q4 × Perimeter-area ratio (log m/ha)	1.027***	1.135***
	(0.252)	(0.0975)
Correlated Q5 × Perimeter-area ratio (log m/ha)	1.008***	1.113***
	(0.246)	(0.0952)
Observations	2181	2181
Adjusted R^2	0.396	0.396

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 A33: Edge Effect Placebo Test (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity
GPS-measured plot size (log ha)	-0.621***	-0.589***		-0.622***
	(0.0636)	(0.0741)		(0.0637)
Placebo-area ratio (log m/ha)		0.0332	0.370***	
		(0.0412)	(0.0495)	
Placebo (log m)				0.0332
				(0.0412)
Observations	2181	2181	2181	2181
Adjusted R^2	0.381	0.381	0.332	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 A34: 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)	-0.621***	-0.708***		-0.535***
	(0.0636)	(0.0818)		(0.0910)
Placebo-area ratio (log m/ha)		-0.173	0.853***	
		(0.117)	(0.116)	
Placebo (log m)				-0.173
				(0.117)
Observations	2181	2181	2181	2181
Adjusted R^2	0.381	0.382	0.339	0.382

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: Suggestive Edge Effect (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity
GPS-measured plot size (log ha)	-0.621***	-0.365***		-0.674***
	(0.0636)	(0.110)		(0.0694)
Sides-area ratio (log #/ha)		0.309**	0.682***	
		(0.125)	(0.0714)	
Sides (log #)				0.309**
				(0.125)
Observations	2181	2169	2169	2169
Adjusted R^2	0.381	0.387	0.380	0.387

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 A13: Perimeter-area Ratio Changing Shape

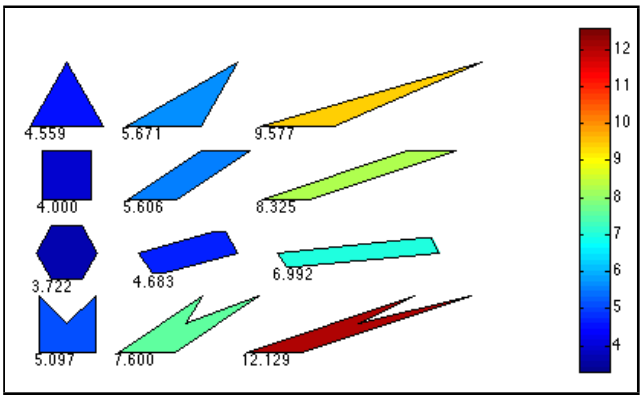


Figure A14: Productivity ($Y^P = 2Y^I$) Changing Shape

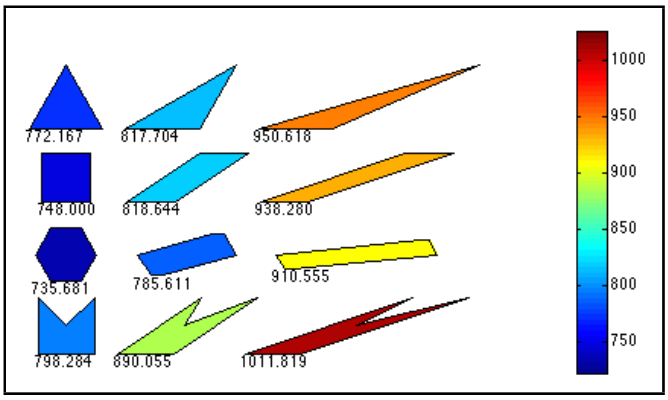


Figure A15: Elasticity wrt PA ratio ($Y^P = 2Y^I$): Changing Shape

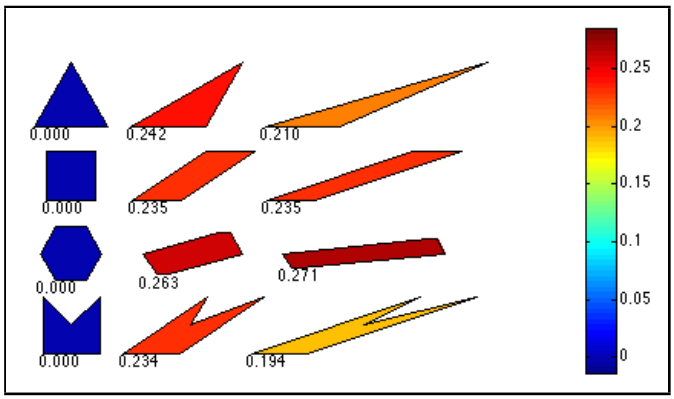


Figure A16: Elasticity wrt PA ratio ($Y^P = 3Y^I$): Changing Shape

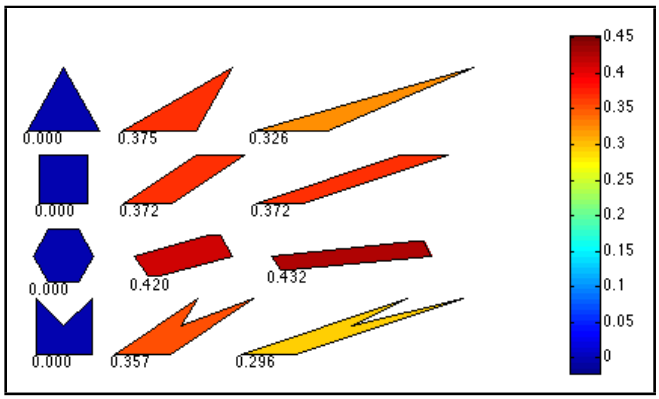


Figure A17: Perimeter-area Ratio Changing Size

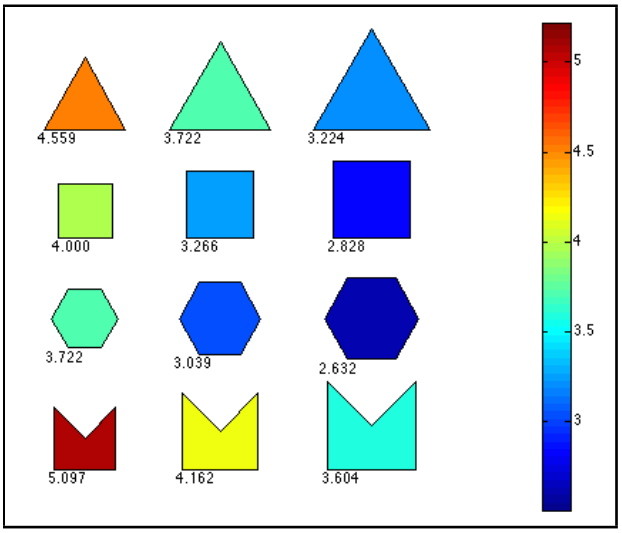


Figure A18: Productivity ($Y^P = 2Y^I$) Changing Size

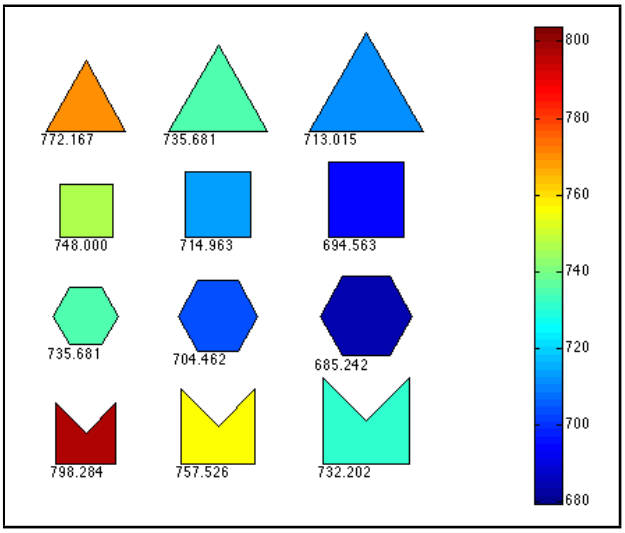


Figure A19: Elasticity wrt PA ratio ($Y^P = 2Y^I$): Changing Size

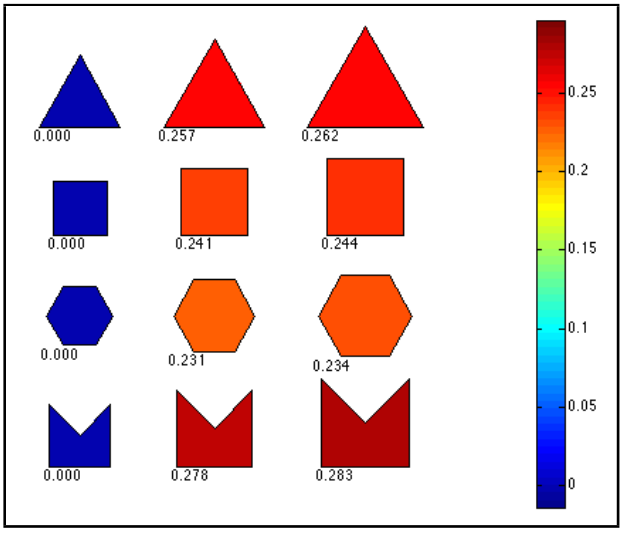


Figure A20: Elasticity wrt PA ratio ($Y^P = 3Y^I$): Changing Size

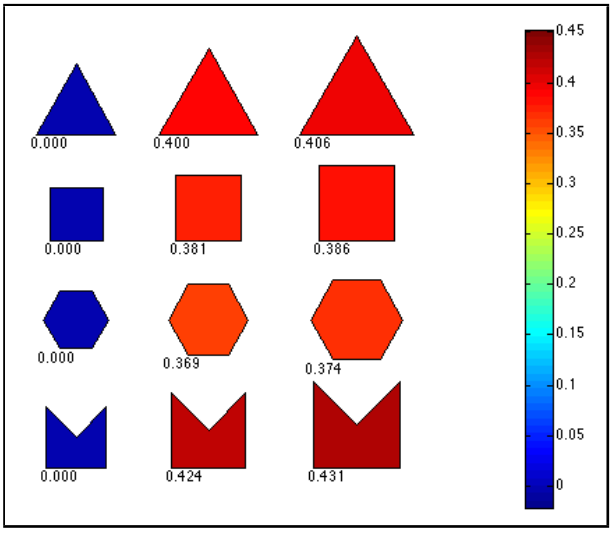


Figure A21: Elasticity wrt Area ($Y^P = 2Y^I$): Changing Size

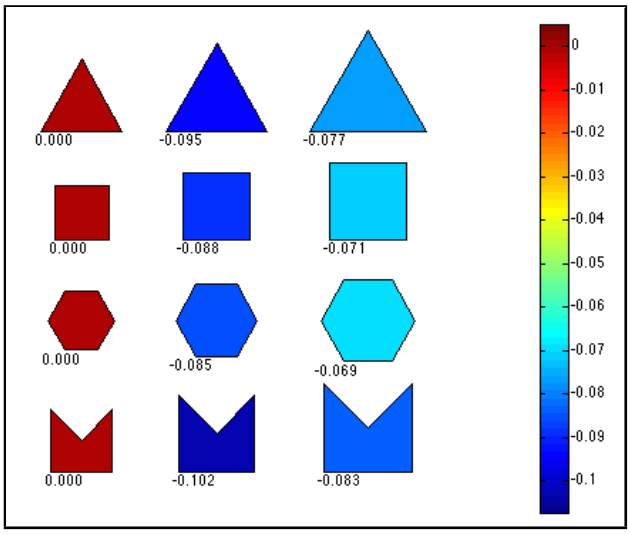
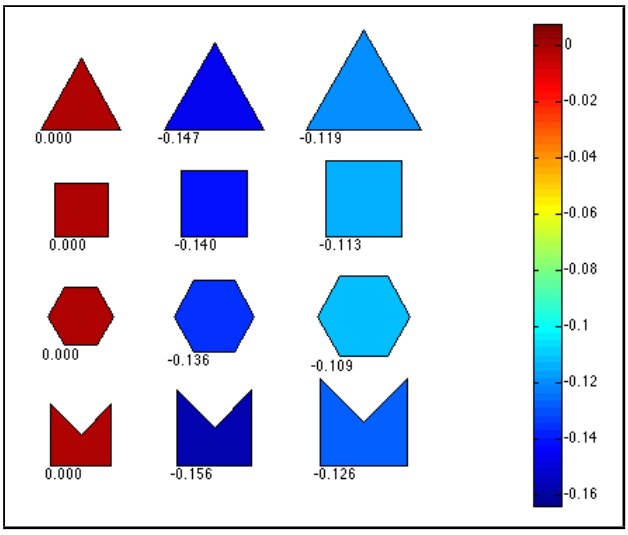


Figure A22: Elasticity wrt Area ($Y^P = 3Y^I$): Changing Size



Appendix 12 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 A23: Plot Size Perception Error over Plot Size

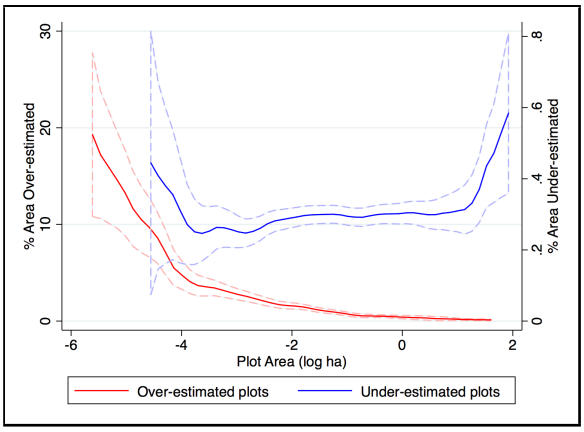
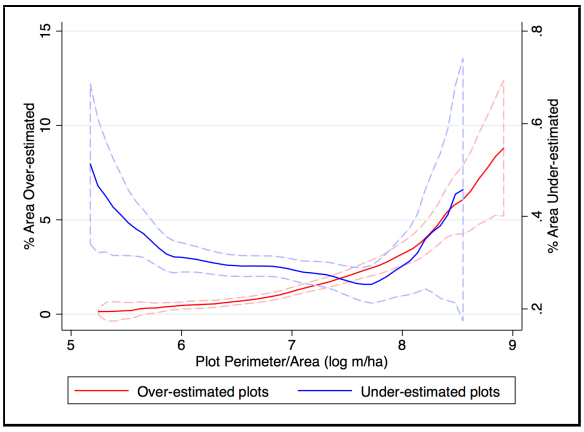


Figure A24: Plot Size Perception Error over Perimeter-Area Ratio



However, Column 1 of Table A36 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 A37 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 A38 models the binary indicator for over-estimation of plot size, using the same covariates again. As with Tables A36 and A37, 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 A36: 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)	0.417	0.632	0.315	-0.0211	0.116
	(0.631)	(0.638)	(0.654)	(0.790)	(0.795)
(GPS-measured plot size)^2	0.198	0.186	0.172	0.177	0.199
	(0.242)	(0.276)	(0.271)	(0.330)	(0.233)
Perimeter-area ratio (log m/ha)	-15.37	-15.53	-16.76	-17.78	-14.11
	(12.80)	(12.12)	(13.22)	(14.65)	(12.35)
(Perimeter-area ratio)^2	1.219	1.224	1.285	1.330	1.113
	(0.903)	(0.863)	(0.946)	(1.042)	(0.857)
Soil pH (pH)		7.621*			
		(4.430)			
Soil pH^2 (pH^2)		-0.631*			
		(0.378)			
Soil sand (%)		0.00816			
		(0.0214)			
Soil organic carbon (%)		0.216			
		(0.216)			
Labor intensity (log hrs/ha/day)			0.122		
			(0.137)		
Organic amendment (binary)			0.409		
			(0.529)		
Inorganic fertilizer (binary)			0.544		
			(1.161)		
Irrigation (binary)			-0.140		
			(0.548)		
Terracing (binary)			0.646		
			(0.503)		
Head owns plot (binary)				0.0128	
				(0.897)	
Head manages plot (binary)				-1.166	
				(1.190)	
(Head owns)X(Head manages)				0.642	
				(1.310)	
Crops are rotated (binary)				-0.140	
				(0.310)	
Crops are mono-cropped (binary)				-1.236	
				(0.917)	
Mixed cropping (binary)				-0.563	
				(0.803)	
Tubers grown (binary)					0.0409
					(0.454)
Cereals grown (binary)					0.206
					(0.431)
Legumes grown (binary)					1.060**
					(0.451)
Bananas grown (binary)					1.044**
					(0.407)
Cash crops grown (binary)					0.705
					(0.497)
Observations	860	758	816	758	860
Adjusted R^2	0.325	0.351	0.325	0.333	0.344

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 A37: 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)	-0.143	-0.173	-0.150	-0.156	-0.130
	(0.0873)	(0.131)	(0.110)	(0.108)	(0.0866)
(GPS-measured plot size)^2	-0.0335	-0.0776*	-0.0520	-0.0222	-0.0463
	(0.0382)	(0.0431)	(0.0440)	(0.0385)	(0.0442)
Perimeter-area ratio (log m/ha)	-3.783**	-4.331**	-4.569**	-3.258*	-3.948**
	(1.682)	(1.924)	(2.132)	(1.703)	(1.768)
(Perimeter-area ratio)^2	0.262**	0.310**	0.318*	0.212	0.275**
	(0.130)	(0.145)	(0.163)	(0.129)	(0.137)
Soil pH (pH)		1.242**			
		(0.520)			
Soil pH^2 (pH^2)		-0.113**			
		(0.0446)			
Soil sand (%)		-0.00180			
		(0.00292)			
Soil organic carbon (%)		-0.0138			
		(0.0228)			
Labor intensity (log hrs/ha/day)			0.00214		
			(0.0224)		
Organic amendment (binary)			0.0822		
			(0.0910)		
Inorganic fertilizer (binary)			0.148		
			(0.101)		
Irrigation (binary)			-0.344*		
			(0.180)		
Terracing (binary)			0.0608		
			(0.0649)		
Head owns plot (binary)				0.145**	
				(0.0689)	
Head manages plot (binary)				0.164	
				(0.105)	
(Head owns)X(Head manages)				-0.131	
				(0.127)	
Crops are rotated (binary)				0.0105	
				(0.0504)	
Crops are mono-cropped (binary)				-0.00590	
				(0.0731)	
Mixed cropping (binary)				0.0175	
				(0.0684)	
Tubers grown (binary)					-0.0757
					(0.0473)
Cereals grown (binary)					-0.0534
					(0.0442)
Legumes grown (binary)					-0.0613
					(0.0438)
Bananas grown (binary)					0.0871
					(0.0771)
Cash crops grown (binary)					-0.121*
					(0.0735)
Observations	612	509	582	555	612
Adjusted R^2	0.107	0.163	0.172	0.186	0.164

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 A38: 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)	-0.0227	0.0164	0.0177	-0.0396	-0.0606
	(0.0695)	(0.0847)	(0.0799)	(0.0894)	(0.0743)
(GPS-measured plot size)^2	0.0312**	0.0295	0.0404**	0.0273	0.0270*
	(0.0158)	(0.0185)	(0.0176)	(0.0204)	(0.0161)
Perimeter-area ratio (log m/ha)	1.079***	1.391***	1.519***	1.149**	0.986**
	(0.402)	(0.460)	(0.453)	(0.510)	(0.414)
(Perimeter-area ratio)^2	-0.0666***	-0.0807***	-0.0919***	-0.0705**	-0.0618**
	(0.0257)	(0.0289)	(0.0292)	(0.0329)	(0.0263)
Soil pH (pH)		-0.242			
		(0.463)			
Soil pH^2 (pH^2)		0.0200			
		(0.0379)			
Soil sand (%)		-0.00249			
		(0.00205)			
Soil organic carbon (%)		0.00618			
		(0.0138)			
Labor intensity (log hrs/ha/day)			-0.0377**		
			(0.0153)		
Organic amendment (binary)			-0.0405		
			(0.0524)		
Inorganic fertilizer (binary)			-0.0276		
			(0.146)		
Irrigation (binary)			0.000518		
			(0.216)		
Terracing (binary)			0.0828*		
			(0.0492)		
Head owns plot (binary)				-0.0261	
				(0.0615)	
Head manages plot (binary)				-0.134*	
				(0.0811)	
(Head owns)X(Head manages)				0.148	
				(0.0948)	
Crops are rotated (binary)				0.0229	
				(0.0419)	
Crops are mono-cropped (binary)				-0.0185	
				(0.0615)	
Mixed cropping (binary)				0.0112	
				(0.0616)	
Tubers grown (binary)					-0.0203
					(0.0402)
Cereals grown (binary)					0.00921
					(0.0397)
Legumes grown (binary)					0.0780**
					(0.0363)
Bananas grown (binary)					-0.0361
					(0.0556)
Cash crops grown (binary)					0.114**
					(0.0561)
Observations	1476	1271	1401	1316	1476
Adjusted R^2	0.165	0.152	0.180	0.167	0.172

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 13 More on Perception Error

Tables A39 and A40 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 A41 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 A39: 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)	-0.736**	-0.251	-0.313	-0.200	-0.202
	(0.331)	(0.207)	(0.228)	(0.286)	(0.329)
Over-estimate (% area)	0.0549	-0.00297	0.127*	0.220***	0.133
	(0.107)	(0.0729)	(0.0739)	(0.0554)	(0.0811)
Over-estimate squared	0.00000489	-0.000686	-0.00361	-0.00629***	-0.00346
	(0.00352)	(0.00258)	(0.00235)	(0.00222)	(0.00261)
Under-estimate (% area)	-7.762***	-2.222*	-3.752***	2.009	-1.013
	(2.329)	(1.216)	(1.424)	(1.718)	(2.187)
Under-estimate squared	11.69***	2.130	4.699***	-1.449	1.852
	(3.374)	(1.382)	(1.799)	(1.981)	(2.841)
Observations	744	1015	1071	831	544
Adjusted R^2	0.340	0.503	0.387	0.330	0.359

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 A40: 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)	-0.0731	0.199	0.360	-1.175*	-0.524	0.0632
	(0.194)	(0.208)	(0.243)	(0.631)	(0.326)	(0.219)
Over-estimate (% area)	0.139***	0.115*	0.0624	0.198*	0.198***	0.202***
	(0.0518)	(0.0625)	(0.0711)	(0.102)	(0.0748)	(0.0504)
Over-estimate squared	-0.00542*	-0.00310	-0.000551	-0.00717	-0.00722**	-0.00620***
	(0.00287)	(0.00243)	(0.00300)	(0.00465)	(0.00318)	(0.00181)
Under-estimate (% area)	0.327	0.631	1.750	-6.727*	-1.568	0.405
	(0.997)	(1.181)	(1.357)	(3.537)	(2.215)	(1.488)
Under-estimate squared	0.0251	-0.722	-1.969	8.185*	1.303	-0.281
	(1.158)	(1.443)	(1.614)	(4.273)	(3.097)	(2.011)
Observations	1551	1298	1154	610	872	1162
Adjusted R^2	0.433	0.354	0.361	0.306	0.393	0.512

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 A41: 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)	-0.308*	-0.304*	0.00615
	(0.168)	(0.168)	(0.201)
Over-estimate (% area)	0.0763*	0.0870**	0.0701
	(0.0404)	(0.0417)	(0.0477)
Over-estimate squared	-0.00144	-0.00205	-0.000877
	(0.00149)	(0.00154)	(0.00160)
Under-estimate (% area)	-0.0234	-0.0301	1.951*
	(1.015)	(1.029)	(1.179)
Under-estimate squared	-0.421	-0.444	-3.217**
	(1.294)	(1.324)	(1.446)
Plot Area, P-A Ratio	Yes	Yes	Yes
(Area)^2, (P-A Ratio)^2	No	Yes	Yes
Additional Plot Controls	No	No	Yes
Observations	2076	2076	1624
Adjusted R^2	0.209	0.212	0.256

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