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

Leah EM Bevis*, Christopher B Barrett†

November 7, 2016

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 Uganda, we provide highly suggestive evidence of causality, for the first time in this literature, to the best of our knowledge. We then test the familiar explanations and find that they fail to explain the observed inverse relationship. We next propose and examine a new mechanism that completely explains the inverse relationship: 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, where crop status is more readily observable. We additionally present suggestive evidence that a behavioral mechanism explains the edge effect, namely that farmers 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 employ household level fixed effects to provide strong evidence that neither multiple market failures nor measurement error/omitted variables drive the inverse relationship.

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 inter-household variation in shadow prices (Assunção and Braido, 2007; Henderson, 2015; Kagin, Taylor and Yúnez-Naude, 2015). Barrett, Bellemare and Hou (2010). Carletto, Savastano and Zezza (2013) and Carletto, Gourlay and Winters (2013) all examine cross-sectional plot-level data, with multiple plots per household. Each study finds that household fixed effects attenuate only part of the inverse relationship and observes a strong, statistically significant negative relationship between plot size and plot productivity even within households. This seems to rule out the multiple market failures hypothesis as a full explanation of the phenomenon.

The statistical artifact hypothesis has not stood up well to recent empirical tests either. 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 plot size is based on 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

[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).

actually reinforces rather than explains away the inverse relationship.

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; not a single paper has yet estimated a causal relationship between plot size and plot productivity, nor proposed a plot-level mechanism to explain this inverse relationship.

We address these important gaps in knowledge using plot-level, geospatially-matched panel data to estimate the inverse relationship.[2] Because the panel spans a decade, plot shape and plot size vary over time, as do some other characteristics such as soil fertility, crops, or management practices. Plot fixed effects therefore 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.

Our contribution to the literature is three-fold. First, we validate recent findings by Barrett, Bellemare and Hou (2010) and Carletto, Savastano and Zezza (2013) indicating that household shadow prices play no role in driving the inverse relationship — the relationship is at the plot level. Measurement error around plot size also plays no role in driving the relationship. Neither does soil fertility; controlling for detailed laboratory measures of soil quality does not mitigate the inverse relationship. The relationship is also unchanged if we control for other inputs, for the crops being produced, or for patterns in ownership or management. The conventional, contending hypotheses — multiple markets failures versus statistical artifact — do not seem to explain the puzzle.

Second, while plot size is not exogenous to other time-varying plot characteristics, we use a technique proposed by Oster (2014) 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

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

Bednarz (2001) find that cotton at the edge of plots yields over three times more than cotton at the interior. 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.

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.

If plot peripheries are 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 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. 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.

It is difficult to entirely untangle the biophysical and behavioral drivers of the edge effect, however. To establish that purely behavioral phenomenon are capable of driving plot productivity, 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 quite 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).[3]

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, and associated possible selection effects. Although none of the differences were significant, 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 multiple, mixed crops. Most, though not all, farmers have multiple plots in both 2003 and 2013. 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. Because GPS waypoints were taken around the corners of all plots in both rounds of data collection, we can match plots across time using their geospatial location. This is how we form a plot-level panel dataset. Of course, some plots cannot be matched over time, as they have no geospatial overlap with another plot from across the decade. These plots are dropped

[3]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.

from our analysis. In other cases, a plot from 2003 may have split into two or more plots from 2013. In such cases the 2003 plot is matched with both of the 2013 plots.[4]

Table 1 summarizes all key variables used in analysis, for 2003 and 2013. 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/hectar/day) far higher. Soil became slightly more acidic over the decade, while organic carbon content appears to have slightly increased.[5]

Inputs, management and cropping systems also shifted 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, and mono-cropping less 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. The number of plots holding tubers, legumes, bananas and cash crops (all categories except for cereals) declined between 2003 and 2013. This is partially due to a decline in mixed cropping (i.e., less crops per plot), and partially due to a decline in the number of crops grown per household.

3 Estimation Strategy

3.1 Shadow Prices

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 profits per hectare.[6] Plot area is given by A_{ijt}, and farm size is given by A_{it}.[7]

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

[4] Of the 1,089 plots from 2003 that were successfully matched to a plot from 2013, 69 percent were matched to exactly 1 plot from 2013, 19 percent were matched to 2 plots from 2013, and the rest were matched to 3-9 plots from 2013. Of the 2003 plots that do split into multiple 2013 plots, 13 percent were under new household ownership in 2013 (usually because they were inherited by a child). The rest were split into multiple plots still owned by the same household.

[5] 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.

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

[7] 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, but farmer-recalled plot size for those plots that were not visited by an enumerator.

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 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-level fixed effects λ_{ij} as in Equation 3.

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

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

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

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-level 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-level fixed effects, and exclude A_{it}.

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 use 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.

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

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.) are correlated with both plot size and plot productivity, failing to control for them might cause a spuriously estimated inverse relationship.

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.

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

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.

Oster (2014), 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 (2014) 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 (2014) proves that with one key assumption,[8] the bias-adjusted coefficient estimate β^* can be approximated as below, and that β^* converges in probability to the true, causal coefficient β.[9] 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.

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

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

[8]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 (2014) 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.

[9]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.

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 (2014) 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 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.

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

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.[10] 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

[10]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 — one row of crops, for instance — then this rectangle will approximate the area of the plot's border.

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.

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

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 4 contains such calculations for hypothetical circular, rectangular, and triangular plots. It additionally explores the possibility that b may not be small, relative to the total size of the plot. If b is not small, then we should find that plot productivity Y_{ijt} rises with P_{ijt}/A_{ijt} and also with A_{ijt}. This would be equivalent to finding that after controlling for P_{ijt}/A_{ijt} in Equation 10, the inverse size-productivity relationship reverses, such that $\hat{\gamma}_6 > 0$.

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 a purely biophysical mechanisms, not controlled by the farmer, and would fall into the category of statistical artifact explanations of the inverse relationship.

Because we do not view these biophysical inputs, it is impossible to test for such mechanisms directly. However, Appendix 5 reports a number of indirect tests for biophysical mechanisms. While we find no evidence to suggest that soil nutrients, sunlight, or biodiversity are driving the edge effect, we certainly cannot rule these mechanisms out.

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 the interiors of their plots. Farmers might weed the edges of their plots more carefully, space crops differently around the edge of a plot, or harvest crops more assiduously around the edges of a plot, where a missed plant will be visible when later walking by. These are all behavioral mechanisms, that give rise to real inefficiencies in production due to within-plot variation in input productivity.[11]

[11] A related behavioral mechanism also seems plausible: farmers may purposely tend the edges of their plot more intensively in order to signal plot ownership to neighbors. In our data, this mechanism is indistinguishable from the spatial awareness mechanism. However, in a dataset with cross sectional or temporal variation in tenure, this mechanism might be explored. In our dataset 90 percent of plots are owned, and there is almost no variation in tenure status across time.

If this is the case, we should expect plot average labor per hectare to exhibit the same pattern as average productivity per hectare, as given by Equations 7-10. Taking average labor per plot hectare L_{ijt} as our outcome variable, we would therefore expect to estimate a negative and significant $\hat{\gamma_7}$ if only A_{ijt} is included on the right hand side of Equation 11, but to estimate an insignificant $\hat{\gamma_7}$ and a positive, significant $\hat{\theta_2}$ if P_{ijt}/A_{ijt} is also controlled for.

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

If farming families invest more labor around the edges of their plots, this suggests an increased awareness of the space within plot edges, likely to due to visibility. In order to investigate the association between "plot awareness" and plot productivity, we additionally investigate a slightly different behavioral mechanism, one purely related to farmer awareness of space and 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.

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^B reflects the average effect of over-estimating plot size on plot productivity, while 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

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. 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.

In Panel 3, plot fixed effects are introduced. (Year and season are not controlled for.) This controls for plot-specific, time-invariant characteristics or transaction costs such as distance to household, distance to road, location on the landscape, and time invariant soil characteristics. 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, as under the longstanding Chayanovian hypothesis.

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, with mean shifts in productivity across year or season also controlled for. Plots that cannot be matched across the decade are, of course, dropped. However, all results presented in this paper may instead be estimated with household-year-season fixed effects. Appendix 6 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. Explanatory power is lower for this form of variation. The coefficients estimated, however, are quantitatively (and qualitatively) the same a 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 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.[12]

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.[13]

[12]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.

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

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 and plot size change are correlated with other, observable plot characteristics. Balance tests in Appendix 7 document this fact — while plot size change is unrelated to all static, 2003 plot characteristics besides 2003 plot size itself, change in plot size is predicted by other plot-level changes. In panel, 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 as relevant, observable controls are introduced (Oster, 2014). Table 4 introduces such controls, as specified in Equation 5, always controlling for plot-level 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.[14] 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 8 reports these results as a robustness check; the ratio does not change significantly across either category.

This stability of the inverse relationship coefficient estimate 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.[15] (However, to obtain comparable R-squared values we re-estimate the univariate inverse relationship, and in fact all columns of Table 4, with an identical sample across columns. These results are found in Table A12 in Appendix 7.)

If we assume $\delta = 1$ and $R_{max} = 1.3R_5$, as suggested by Oster (2014), we obtain the

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

[15] See Appendix 7 for bounds based on each set of controls in turn.

bounds [-0.658 -0.6996]. 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.)

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 A2 in Appendix 7 illustrates these possible bounds. More details can be found in Appendix 7.

4.4 The Edge Effect

Having established the inverse relationship as strongly robust, very possibly causal, and un-explained 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 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 model R^2 remains constant, confirming that plot size contributes no information on average productivity once perimeter-area ratio is known.[16,17]

A variety of additional tests, all held in Appendix 9, further suggest that the edge effect is driving the inverse relationship. The edge effect result is robust to all the controls of Table 4, and the coefficient on perimeter-area ratio is virtually indistinguishable across both plot size quantiles and perimeter-area ratio quantiles, as shown in Table A17. The inverse relationship, however, is insignificant for plots with the smallest shifts in perimeter-area ratio, and increases in magnitude as changes in perimeter-area ratio rise, as shown in Table A18. Additionally, it is worthwhile to note that perimeter-area ratio is not solely driven by plot size — twenty percent of the panel data variation in perimeter-area ratio can be explained by plot shape dummies alone, as shown in Table A19.

[16]The fact that plot size contributes no information on productivity confirms that the border area b defining plot edge is small, as explained in Appendix 4.

[17]Note that plot area and perimeter-area ratio are correlated, with a Pearson's Correlation Coefficient of -0.13. The log version of these coefficients is even more strongly correlated. But under multicollinearity we would expect large standard errors, which we do not observe. Additionally, under certain assumptions we might find that a negative correlation between plot area and perimeter-area ratio would drive significant, positive coefficients on both variables. But we do not observe this, again. The correlation between these two variables does not seem to be problematic.

4.5 Edge Effect Mechanisms

We next investigate labor as the mechanism behind the edge effect. 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, appears to be inversely correlated with plot size. In Column 2, however, it is clear that the perimeter-area ratio is actually driving the inverse relationship, as with productivity. Once the perimeter-area ratio is controlled for, the inverse relationship between plot size and labor intensity is statistically identical to zero.

Column 3 controls only for perimeter-area ratio, and as in Table 5 the R^2 remains almost constant between Columns 2 and 3; plot size contributes almost no information on average labor intensity once perimeter-area ratio is known. Additionally, the edge effect seems to affect labor intensity and productivity in exactly the same way — the third columns of Table 5 and Table 6 suggest a one-to-one increase in both plot productivity and labor intensity, respectively.[18]

Appendix 10 reports additional results for family and non-family laborers and across various types of labor — weeding labor, planting stage labor, and other labor. 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.) When it comes to various 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.

Last, we test 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 11 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. Especially if conditioned on plot area and the 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 plot productivity. Oddly, while plot productivity drops, on average, when a farmer over-estimates plot size, plot productivity rises with every marginal increase in over-estimation, and drops with marginal rises in under-estimation. Column 1 controls for log plot area and log plot perimeter-area ratio in addition to farmer perceptions, as

[18] 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.

in most other tables. Column 2 controls for area and perimeter-area ratio quadratically.[19] Column 3 controls for area and perimeter-area ratio quadratically and additionally controls for all typically "omitted" variables from Column 6 of Table 4. In all three specifications, the effect of farmer perceptions remains qualitatively and statistically the same.

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-estimate (under-estimate) of plot size would lead to inefficient allocation of resources, and higher (lower) productivity on such plots. In fact, Table A26 in Appendix 13 shows that labor intensity does respond to farmer misperceptions in plot size, though not as consistently as does productivity. (This makes sense; if farmers choose input intensity based on plot size perceptions, and do this for all inputs, we would expect a stronger relationship between resulting productivity and perceptions than we would between any given, individual input and perceptions.)

Because Appendix 11 suggest that perception error may possibly be related to crop type, Appendix 12 presents the same regression of crop productivity on plot size misperceptions by crop. While the results are qualitatively similar, they are no longer significant since sample size is cut down drastically.

It is also possible that misperceptions of plot size change, not plot productivity itself, drives 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 plot 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

We estimate the inverse size-productivity relationship at the plot level using panel data. It is clear that the inverse relationship is not explained by household shadow prices, or by other household characteristics. Additionally, we validate the findings of both Carletto, Savastano and Zezza (2013) and Barrett, Bellemare and Hou (2010) 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 input and management type, we show that the inverse relationship cannot be explained by the time-varying plot characteristics typically considered.

Moreover, the stability of the inverse relationship to this rich set of covariates is suggestive of a causal relationship. We estimate bounds around the likely causal

[19]This is because the results in Appendix 11 suggest that perceptions may respond non-linearly to area and perimeter-area ratio.

relationship between plot size and plot productivity, using the consistent estimator of bias derived by Oster (2014). 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, we show that the phenomenon is completely explained in these data by edge effects — a novel explanation in the economics literature. Plot productivity rises with perimeter-area ratio, and once perimeter-area ratio is controlled for, plot area has no remaining influence on plot productivity. This 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.[20] Therefore, in these data at least, it appears that the inverse relationship — observed at both the farm and plot level — is driven by a thin but differentially productive peripheral area around the edge of each plot, driving smaller plots to be more productive on average than larger 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 cannot test these inputs directly, but our indirect tests find no evidence of such biophysical mechanisms.

The behavioral economics literature may also help us to understand how the edge effect drives plot productivity. Just and Wansink (2014) show that food consumers use portion labeling (e.g., small, large, super-size) as objective information on portion size, and that this labeling therefore has the power to influence food consumption decisions. In a similar line of research, Wansink, Painter and North (2005) show that costumers eat more when (erroneous) visual clues suggest that they are consuming less calories than they truly are.

We hypothesize that similar, visual cues about plot size or plot space may similarly influence farmer management decisions. Farmers may be more aware of the space at the periphery of plots, due to visibility and accessibility, just as individuals are more aware of the food that they have eaten when visual cues suggest that large quantities of food are being consumed. If so, we might expect labor intensity to be higher around the edges of a plot, and perhaps other management decisions to differ between the periphery and the interior of a plot.

Our results confirm that both plot-level labor intensity and also plot-level productivity rises with perimeter-area ratio, suggesting that differential management practices drive plot peripheries, or plot edges, to be more productive than the interior region of plots. This same result can also be thought of differently — while land productivity (value added per unit of land) is falling with plot size, a pattern of increased labor intensity on smaller plots leads labor productivity (value added per unit of labor) to rise with plot

[20]If the width of the peripheral area was wide, plot productivity would respond positively to both perimeter-area ratio and also plot area.

size. Both of these patterns have been observed at the farm level in datasets across the globe (Adamopoulos and Restuccia, 2014).

Additionally, 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. Our results suggest that this may be the case. 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 drive input application rates and resulting inter-plot and intertemporal productivity differences, consistent with a behavioral explanation of the edge effects apparent in our data.

Taken together, these results suggest that farmer perceptions or awareness of plot size and plot space may significantly influence the management and productive capacity of plots in smallholder settings. This is important in its own right, and more research is merited — we might wonder, for instance, whether more accurate information on plot size or increased accessibility into the interior of plots would change farmer investment behavior and farmer productivity. Additionally, differential productivity between the interior and the periphery of plots appears to explain the long-held puzzle of inversely related plot/farm size and plot/farm productivity, in these data at least.

References

Adamopoulos, Tasso, and Diego Restuccia. 2014. “The size distribution of farms and international productivity differences.” *The American Economic Review*, 104(6): 1667–1697.

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

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. 2013. "From guesstimates to GPStimates: land area measurement and implications for agricultural analysis." *World Bank Policy Research Working Paper*, , (6550).

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.

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.

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. 2014. "Unobservable selection and coefficient stability: Theory and evidence." *University of Chicago Booth School of Business Working Paper.*

Pischke, Jörn-Steffen, and Hannes Schwandt. 2014. "Poorly measured confounders are more useful on the left than on the right." Working Paper, London School of Economics.

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

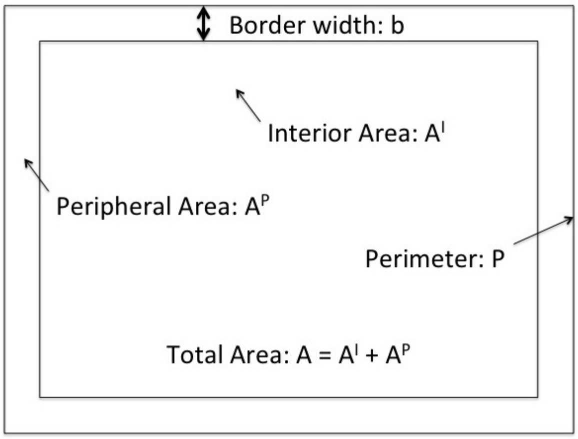
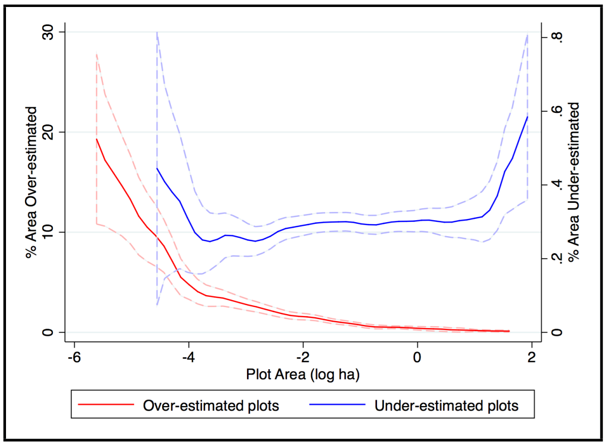


Figure 2: Misperceptions by Plot Area



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)	1.35	1.18	0.82	0.88	15.61***
Plot size (ha)	0.55	0.67	0.31	0.41	11.01***
Perimeter-area ratio (m/ha)	1,131.66	923.79	1,928.23	6,766.28	-8.98***
Plot productivity (revenue/ha)	262.25	1,327.25	1,241.49	8,637.60	-17.51***
Labor intensity (hrs/ha/day)	3.86	9.55	10.12	67.56	1.73*
Soils					
Soil pH (pH)	6.24	0.53	6.18	0.63	2.16**
Soil sand (%)	60.03	14.13	52.73	15.79	10.61***
Soil organic carbon (%)	3.46	1.62	3.71	1.85	-3.10***
Inputs					
Organic amendment (%)	20.91	40.69	12.15	32.68	5.56***
Inorganic fertilizer (%)	1.28	11.24	1.37	11.63	-0.19
Irrigation (%)	1.30	11.31	0.19	4.31	3.01***
Terracing (%)	23.01	42.11	9.33	29.10	8.84***
Management					
Head owns plot (%)	67.67	46.79	74.43	43.65	-3.49***
Head manages plot (%)	54.98	49.77	63.56	48.15	-4.10***
(Head owns)X(Head manages)	46.12	49.87	59.09	49.19	-6.13***
Crops are rotated (%)	20.83	40.63	42.83	49.51	-10.80***
Crops are mono-cropped (%)	60.55	48.90	41.46	49.29	9.10***
Mixed cropping (%)	54.06	49.86	51.96	49.98	0.98
Crops Grown					
Tubers grown (%)	41.46	49.29	24.57	43.07	8.54***
Cereals grown (%)	47.40	49.96	44.20	49.69	1.50
Legumes grown (%)	51.23	50.01	45.48	49.82	2.70***
Bananas grown (%)	51.96	49.98	27.67	44.76	11.98***
Cash crops grown (%)	32.42	46.83	18.81	39.10	7.38***

†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)	-0.425***		0.00672
	(0.0321)		(0.0360)
Plot size (log ha)		-0.552***	-0.556***
		(0.0269)	(0.0333)
Observations	2189	2189	2189
Adjusted R^2	0.095	0.206	0.206
	Panel 2 (House-year-season FE)		
Plot size (log ha)		-0.672***	
		(0.0513)	
Observations		2173	
Adjusted R^2		0.240	
	Panel 3 (Plot FE)		
Farm size (log ha)	-0.338***		0.0520
	(0.0859)		(0.0618)
Plot size (log ha)		-0.598***	-0.620***
		(0.0620)	(0.0625)
Observations	2189	2189	2189
Adjusted R^2	0.274	0.369	0.370

Panel 1: Robust standard errors
Panel 2: House-year-season clustered standard errors
Panel 3: Plot clustered standard errors
*** p<0.01, ** p<0.05, * p<0.1

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

	(1) Plot Productivity	(2) Plot Productivity
Farmer-recalled plot size (log ha)	-0.599***	
	(0.0467)	
GPS-measured plot size (log ha)		-0.598***
		(0.0620)
Observations	2183	2189
Adjusted R^2	0.223	0.369

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
*** 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)	-0.598***	-0.591***	-0.568***	-0.704***	-0.659***	-0.658***
	(0.0620)	(0.0787)	(0.0678)	(0.0682)	(0.0627)	(0.0796)
Soil pH (pH)		1.707				0.730
		(1.126)				(1.321)
Soil pH^2 (pH^2)		-0.123				-0.0360
		(0.0918)				(0.110)
Soil sand (%)		-0.00371				-0.00243
		(0.00518)				(0.00552)
Soil organic carbon (%)		-0.00924				-0.0217
		(0.0349)				(0.0429)
Labor intensity (log hrs/ha/day)			0.0931***			0.0964**
			(0.0341)			(0.0413)
Organic amendment (binary)			0.0797			0.0640
			(0.119)			(0.156)
Inorganic fertilizer (binary)			0.569**			1.228***
			(0.278)			(0.305)
Irrigation (binary)			0.409**			-0.376
			(0.202)			(0.404)
Terracing (binary)			0.279**			0.429***
			(0.110)			(0.144)
Head owns plot (binary)				-0.206		-0.196
				(0.131)		(0.152)
Head manages plot (binary)				-0.298		0.0440
				(0.190)		(0.204)
(Head owns)X(Head manages)				0.558***		0.189
				(0.211)		(0.233)
Crops are rotated (%)				-0.205**		0.0427
				(0.0951)		(0.118)
Crops are mono-cropped (%)				0.352***		0.340***
				(0.105)		(0.117)
Mixed cropping (%)				0.650***		0.581***
				(0.105)		(0.130)
Tubers grown (binary)					0.168**	-0.0474
					(0.0836)	(0.108)
Cereals grown (binary)					0.0993	-0.0102
					(0.0839)	(0.106)
Legumes grown (binary)					0.220***	0.0585
					(0.0808)	(0.109)
Bananas grown (binary)					0.270**	0.110
					(0.125)	(0.158)
Cash crops grown (binary)					0.343***	0.455***
					(0.132)	(0.158)
Observations	2189	1897	2061	1957	2189	1634
Adjusted R^2	0.369	0.364	0.373	0.411	0.384	0.423

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
*** 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
GPS-measured plot size (log ha)	-0.598***	-0.0922	
	(0.0620)	(0.121)	
Perimeter-area ratio (log m/ha)		0.928***	1.075***
		(0.230)	(0.0964)
Observations	2189	2189	2189
Adjusted R^2	0.369	0.382	0.382

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
*** 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
GPS-measured plot size (log ha)	-0.628***	-0.198	
	(0.0624)	(0.139)	
Perimeter-area ratio (log m/ha)		0.784***	1.099***
		(0.244)	(0.0956)
Observations	2080	2080	2080
Adjusted R^2	0.174	0.185	0.183

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
*** 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)	-0.496***	-0.479***	-0.313*
	(0.137)	(0.137)	(0.180)
Over-estimate (% area)	0.160***	0.156***	0.109***
	(0.0339)	(0.0335)	(0.0415)
Over-estimate squared	-0.00376***	-0.00439***	-0.00335**
	(0.00135)	(0.00127)	(0.00153)
Under-estimate (% area)	-2.659***	-2.478***	-2.046*
	(0.793)	(0.790)	(1.167)
Under-estimate squared	3.065***	2.727***	2.312
	(0.959)	(0.952)	(1.456)
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	2187	2187	1632
Adjusted R^2	0.410	0.419	0.462

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Additional plot controls are from Column 6 of Table 3
*** p<0.01, ** p<0.05, * p<0.1

Appendix 1 Plot Size and Perimeter Variables

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. In 2013 they did this by walking to each corner or flex-point of a plot and taking a waypoint. 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.

Appendix 2 Survey Attrition

Table A1: Farmer and Farm Characteristics in 2003, Tracked and Attritted Families

	(1) Tracked Mean	(2) Tracked Sd	(3) Attritted Mean	(4) Attritted Sd
Head age (years)	42	14.1	38.0	13.0
Head education (years)	4.9	3.4	5.6	3.1
Head married (%)	79.8	40.2	81.0	39.4
Household size (people)	6.0	2.8	5.1	2.9
Asset index (index)	24.1	23.6	24.2	21.3
Crops as primary income (%)	68.9	46.3	45.5	50.0
Cattle (#)	2.4	6.2	1.9	7.2
pH (pH)	6.2	0.5	6.1	0.6
Soil carbon (%)	3.2	1.5	4.1	1.8
Soil sand (%)	62.9	13.8	57.4	12.6
Farm area (ha)	1.4	3.1	0.9	1.2
Average plot area (ha)	0.6	1.4	0.4	0.8

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 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.

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

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 $A^E = A - A^I$. These area definitions can be expanded as follows.

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

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

Average productivity can therefore be re-written as below.

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

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 $A^E = A - A^I$. These area definitions can be expanded as follows.

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

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

Average productivity can therefore be re-written as below.

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

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 $A^E= A - A^I$. These area definitions can be expanded as follows.

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

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

Average productivity can therefore be re-written as below.

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

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 $b(Y^E - Y^I)[P/A]$, and falls linearly with $gb^2(Y^E - Y^I)[1/A]$, 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 5 Biophysical Mechanisms

Estimation Strategy

We can test indirectly for biophysical mechanisms in the following manner. 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 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.

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

Similarly, we do not observe sunlight as an input. However, 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. We therefore estimate Equation 10 for subsets of crops according to height, in order to test for such a pattern.

Some sources suggest that the edges of a plot are more productive because they contain greater levels of biodiversity, in terms of crops or insects that move in from outer fields. As with sunlight we are unable to observe biodiversity as an input. However, we can investigate whether the edge effect is weaker in plots that are likely to have improved levels of biodiversity. In the Ugandan context some plots contain one crop only (mono-cropping), while others contain multiple crops (mixed cropping). Farmers additionally choose to rotate crops over time for some plots, but not for all plots. We might expect mixed cropping and rotated plots to have higher levels of biodiversity. We therefore estimate Equation 10 for subsets of plots according to cropping system and rotation, in order to test for such a pattern.

Results

Table A2 below shows results for the equation specified above. Soil fertility is specified 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.

Table A3 estimates the edge effect 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 know why this would be the case. If tree crops are considered "very tall crops," however, we may again observe that sunlight does not appear to be driving the edge effect.

Table A4 again estimates the edge effect for a subset of plots, this time according to rotation status and cropping system. The edge effect appears to be stronger for plots where rotation is practiced — i.e., stronger on plots with higher biodiversity. It is also stronger for plots under a mixed cropping system (i.e., higher biodiversity) than for plots under mono-cropping. This is precisely the opposite of what might be expected if biodiversity were to drive the edge effect. (The result is also interesting in its own right; it is unclear why the edge effect would be stronger under rotation or mixed cropping.)

Table A2: Edge Effect and Soil Nutrients (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity
GPS-measured plot size (log ha)	-0.0922	-0.0208	-0.0373	-0.0448
	(0.121)	(0.140)	(0.139)	(0.141)
Perimeter-area ratio (log m/ha)	0.928***	1.172***	0.970**	1.069***
	(0.230)	(0.330)	(0.419)	(0.325)
Soil organic carbon (%)		0.319		
		(0.255)		
(Soil organic carbon)X(Perimeter-area ratio)		-0.0458		
		(0.0350)		
Soil sand (%)			-0.00734	
			(0.0458)	
(Soil sand)X(Perimeter-area ratio)			0.000538	
			(0.00650)	
Soil nitrogen (%)				1.603
				(2.937)
(Soil nitrogen)X(Perimeter-area ratio)				-0.321
				(0.403)
Observations	2189	1897	1897	1897
Adjusted R^2	0.382	0.375	0.374	0.377

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A3: 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
GPS-measured plot size (log ha)	-0.0922	-0.0305	-0.0359	-0.220
	(0.121)	(0.209)	(0.437)	(0.203)
Perimeter-area ratio (log m/ha)	0.928***	1.465***	1.508**	0.455
	(0.230)	(0.360)	(0.724)	(0.318)
Observations	2189	991	528	1290
Adjusted R^2	0.382	0.472	0.364	0.285

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A4: Edge Effect and Biodiversity (Plot Panel)

	(1) Plot Productivity All Crops	(2) Plot Productivity Rotation	(3) Plot Productivity No Rotation	(4) Plot Productivity Monocropped	(5) Plot Productivity Mixed Crops
GPS-measured plot size (log ha)	-0.0922	-0.0505	-0.329*	-0.151	0.0136
	(0.121)	(0.422)	(0.173)	(0.215)	(0.236)
Perimeter-area ratio (log m/ha)	0.928***	1.429**	0.400	0.862**	1.333***
	(0.230)	(0.642)	(0.283)	(0.348)	(0.401)
Observations	2189	604	1353	1117	1160
Adjusted R^2	0.382	0.285	0.308	0.386	0.469

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix 6 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-season-time group.

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 A7 (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 A8 (analogous to Table 6). Second, farmer perceptions no longer appear associated with plot productivity, as seen in Table A9.

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.

Table A5: Measurement Error and the Inverse Relationship (Household-Time FE)

	(1) Plot Productivity	(2) Plot Productivity
Farmer-recalled plot size (log ha)	-0.538***	
	(0.0641)	
GPS-measured plot size (log ha)		-0.678***
		(0.0510)
Observations	2183	2189
Adjusted R^2	0.126	0.245

Estimated with household-year-season fixed effects
Household-year-season-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Table A6: 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)	-0.678***	-0.643***	-0.513***	-0.663***	-0.666***	-0.430***
	(0.0510)	(0.0567)	(0.0610)	(0.0492)	(0.0522)	(0.0682)
Soil pH (pH)		-0.258				-0.627
		(1.220)				(1.363)
Soil pH^2 (pH^2)		0.0547				0.0848
		(0.0979)				(0.110)
Soil sand (%)		0.00519				0.00993*
		(0.00576)				(0.00541)
Soil organic carbon (%)		0.0576				0.0619
		(0.0436)				(0.0395)
Labor intensity (log hrs/ha/day)			0.279***			0.302***
			(0.0541)			(0.0672)
Organic amendment (binary)			0.249*			0.0806
			(0.145)			(0.157)
Inorganic fertilizer (binary)			0.575			2.146***
			(1.122)			(0.399)
Irrigation (binary)			0.308			0.00304
			(0.402)			(0.412)
Terracing (binary)			0.526***			0.279*
			(0.123)			(0.144)
Head owns plot (binary)				-0.671**		-0.569*
				(0.323)		(0.313)
Head manages plot (binary)				-0.0588		-0.267
				(0.369)		(0.277)
(Head owns)X(Head manages)				0.375		0.135
				(0.486)		(0.494)
Crops are rotated (%)				-0.622***		-0.494***
				(0.142)		(0.172)
Crops are mono-cropped (%)				0.433***		0.277
				(0.149)		(0.172)
Mixed cropping (%)				0.661***		0.338*
				(0.148)		(0.182)
Tubers grown (binary)					0.106	0.159
					(0.103)	(0.129)
Cereals grown (binary)					0.0445	-0.0941
					(0.100)	(0.119)
Legumes grown (binary)					0.164**	0.180
					(0.0823)	(0.116)
Bananas grown (binary)					0.732***	0.299**
					(0.116)	(0.145)
Cash crops grown (binary)					-0.0189	0.0262
					(0.119)	(0.128)
Observations	2189	1897	2061	1957	2189	1634
Adjusted R^2	0.245	0.279	0.307	0.294	0.304	0.450

Estimated with household-year-season fixed effects
Household-year-season-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Table A7: Edge Effect and the Inverse Relationship (Household-Time FE)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity
GPS-measured plot size (log ha)	-0.678***	-0.323***	
	(0.0510)	(0.107)	
Perimeter-area ratio (log m/ha)		0.646***	1.144***
		(0.168)	(0.0778)
Observations	2189	2189	2189
Adjusted R^2	0.245	0.257	0.248

Estimated with household-year-season fixed effects
Household-year-season-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Table A8: Edge Effect and Labor Intensity (Household-Time FE)

	(1) Labor Intensity	(2) Labor Intensity	(3) Labor Intensity
GPS-measured plot size (log ha)	-0.604***	-0.312***	
	(0.0432)	(0.119)	
Perimeter-area ratio (log m/ha)		0.559***	1.068***
		(0.201)	(0.0704)
Observations	2080	2080	2080
Adjusted R^2	0.261	0.272	0.261

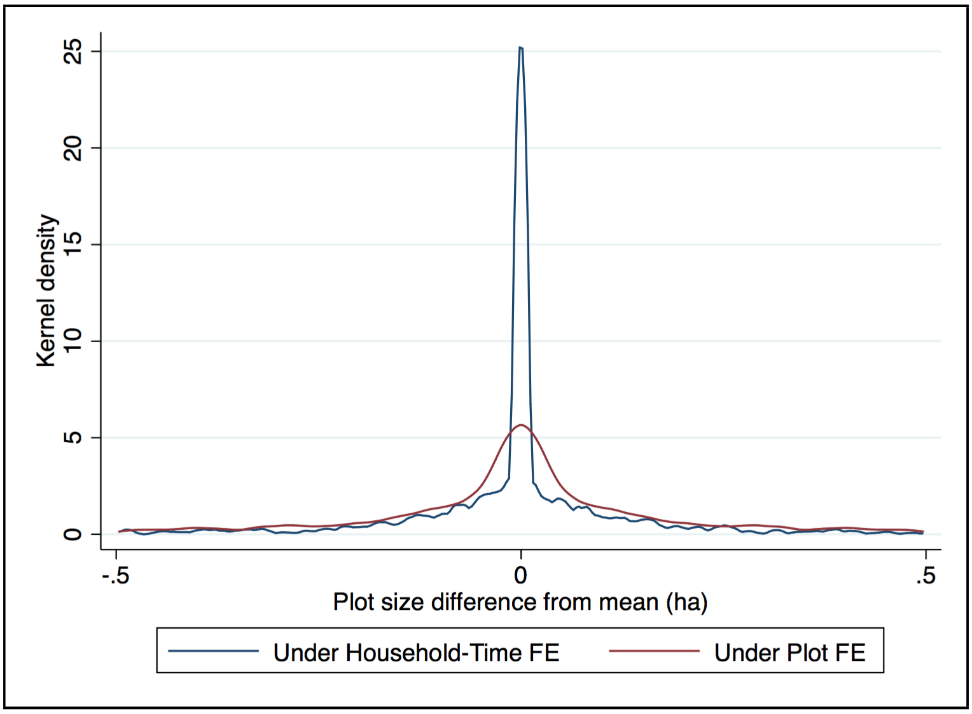
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Table A9: 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)	0.251	0.268*	0.313
	(0.154)	(0.153)	(0.220)
Over-estimate (% area)	0.0750**	0.0807**	0.00612
	(0.0369)	(0.0375)	(0.0349)
Over-estimate squared	-0.00159	-0.00204	0.000536
	(0.00116)	(0.00130)	(0.00116)
Under-estimate (% area)	0.907	0.967	1.172
	(1.043)	(1.037)	(1.411)
Under-estimate squared	-1.925	-2.009	-1.910
	(1.369)	(1.361)	(1.734)
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	2189	2189	1634
Adjusted R^2	0.277	0.278	0.469

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
Additional plot controls are from Column 6 of Table 3
p<0.01, ** p<0.05, * p<0.1

Figure A1: Identifying Variation in Plot Size



Appendix 7 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. Table A10 illustrates that *change* in plot size is conditionally exogenous with respect to round 1 data — it is highly correlated with 2003 plot size, but unrelated to other 2003 plot characteristics conditional on 2003 plot size.

However, because 2003 plot size itself is not random, change in plot size cannot be considered exogenous. Table A11 illustrates this point via a balance test — because plot size is surely measured with some error, it's endogeneity is best determined via a regression of plot size on other covariates (Pischke and Schwandt, 2014). When this regression is performed in panel, using plot fixed effects, it is clear that plot size is predicted by a number of other controls, including soil organic carbon, labor intensity, irrigation, and crop type.

We cannot, therefore, necessarily interpret the inverse relationship as causal. It is, however, remarkably robust to additional controls. Table A12 estimates the inverse relationship with an exhaustive list of additional, time-varying plot-specific controls. (This table is identical to Table 4, except that it holds sample size constant across columns, in order to obtain comparable R-squared values.) 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 A12, 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 (2014).

These bounds are displayed in Table A13. None of them contain zero. The last 3 sets of bounds — drawn from Columns 4-6 of Table A12 — suggest that the causal effect of plot size on plot productivity is actually *higher* than the estimated effect. This is because controlling for these covariates actually increases the magnitude of the estimated inverse relationship, while increasing R-squared.

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 A2 illustrates the bias-adjusted estimator γ^* calculated for every combination of $\delta \in [0, 2]$ and $R_{max} \in [.51]$, maintaining $\hat{\gamma_4} = -0.638$, $R_4 = 0.362$, $\hat{\gamma_5} = -0658$, $R_5 = 0.423$, as in the final column of Table A13, i.e., based on the full set of controls in Table A12.

Figure A2 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 the estimated inverse relationship. The bias-adjusted estimator defined by Oster (2014) will always be higher than the inverse relationship estimated under a full set of controls, and thus the causal bounds $\left[\hat{\gamma_5} \;\; \gamma^*(R_{max}, \delta)\right]$ will always suggest that omitted variable bias is mitigating, rather than inflating, the estimated inverse relationship.

Table A10: Baseline Selection into Plot Size Change (Household FE)

	(1) Plot Size Change	(2) Plot Size Change	(3) Plot Size Change	(4) Plot Size Change	(5) Plot Size Change	(6) Plot Size Change
GPS-measured plot size in '03 (log ha)	-0.622***	-0.589***	-0.621***	-0.667***	-0.611***	-0.540***
	(0.0738)	(0.0797)	(0.103)	(0.0754)	(0.0743)	(0.115)
Soil pH (pH)		3.059				3.515
		(1.984)				(2.183)
Soil pH^2 (pH^2)		-0.244				-0.281
		(0.165)				(0.180)
Soil sand (%)		0.00875*				0.00676
		(0.00511)				(0.00520)
Soil organic carbon (%)		-0.0208				-0.00830
		(0.0391)				(0.0466)
Labor intensity (log hrs/ha/day)			0.0343			0.0786
			(0.0827)			(0.0928)
Organic amendment (binary)			-0.0815			-0.189
			(0.142)			(0.177)
Inorganic fertilizer (binary)			0			0
			(.)			(.)
Irrigation (binary)			-0.118			-0.0916
			(0.429)			(0.410)
Terracing (binary)			0.128			0.156
			(0.177)			(0.242)
Head owns plot (binary)				0.425		0.371
				(0.438)		(0.468)
Head manages plot (binary)				-0.209		-0.444
				(0.299)		(0.325)
(Head owns)X(Head manages)				-0.188		0.00375
				(0.697)		(0.766)
Crops are rotated (binary)				0.0923		0.139
				(0.160)		(0.190)
Crops are mono-cropped (binary)				0.0728		0.0734
				(0.165)		(0.200)
Mixed cropping (binary)				0.219		0.268
				(0.156)		(0.179)
Tubers grown (binary)					-0.116	-0.224
					(0.144)	(0.180)
Cereals grown (binary)					-0.0430	-0.225*
					(0.116)	(0.135)
Legumes grown (binary)					0.188*	0.117
					(0.113)	(0.169)
Bananas grown (binary)					0.0794	-0.0525
					(0.134)	(0.227)
Cash crops grown (binary)					-0.00454	-0.224
					(0.137)	(0.171)
Observations	740	636	725	721	740	620
Adjusted R^2	0.201	0.188	0.209	0.212	0.204	0.217

Estimated with household fixed effects
Round 1 data only, 1 observation per plot
Household-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Table A11: Balance Test for Plot Size (Plot Panel)

	(1) Plot Size	(2) Plot Size	(3) Plot Size	(4) Plot Size	(5) Plot Size	(6) Plot Size
Soil pH (pH)		0.0741				0.176
		(0.945)				(1.032)
Soil pH^2 (pH^2)		-0.00820				-0.0232
		(0.0774)				(0.0845)
Soil sand (%)		-0.000210				0.00187
		(0.00412)				(0.00408)
Soil organic carbon (%)		-0.0393				-0.0898***
		(0.0288)				(0.0264)
Labor intensity (log hrs/ha/day)			-0.265***			-0.274***
			(0.0370)			(0.0455)
Organic amendment (binary)			0.119			0.161*
			(0.0799)			(0.0881)
Inorganic fertilizer (binary)			0.314			-0.121
			(0.289)			(0.471)
Irrigation (binary)			-0.0695			-0.657**
			(0.233)			(0.277)
Terracing (binary)			0.0556			0.0982
			(0.0777)			(0.108)
Head owns plot (binary)				0.132		0.122
				(0.135)		(0.119)
Head manages plot (binary)				0.00493		-0.0270
				(0.161)		(0.165)
(Head owns)X(Head manages)				-0.0383		0.0115
				(0.189)		(0.191)
Crops are rotated (binary)				0.0546		-0.0131
				(0.0971)		(0.101)
Crops are mono-cropped (binary)				0.202**		0.198**
				(0.0973)		(0.0882)
Mixed cropping (binary)				0.432***		0.353***
				(0.0880)		(0.0880)
Tubers grown (binary)					0.386***	0.275***
					(0.0844)	(0.0982)
Cereals grown (binary)					0.294***	0.115
					(0.0719)	(0.0744)
Legumes grown (binary)					0.139**	0.0482
					(0.0663)	(0.0764)
Bananas grown (binary)					0.0195	-0.220*
					(0.0971)	(0.117)
Cash crops grown (binary)					0.585***	0.366***
					(0.105)	(0.127)
Observations	2190	1898	2062	1958	2190	1635
Adjusted R^2	0.171	0.163	0.302	0.186	0.264	0.389

Estimated with household-year-season fixed effects
Household-year-season-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Table A12: Omitted Variables and the Inverse Relationship (Plot Panel Single Sample)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity	(4) Plot Productivity	(5) Plot Productivity	(6) Plot Productivity
GPS-measured plot size (log ha)	-0.638***	-0.633***	-0.577***	-0.697***	-0.695***	-0.658***
	(0.0908)	(0.0936)	(0.0822)	(0.0874)	(0.0868)	(0.0796)
Soil pH (pH)		2.455*				0.730
		(1.296)				(1.321)
Soil pH^2 (pH^2)		-0.178*				-0.0360
		(0.106)				(0.110)
Soil sand (%)		-0.00216				-0.00243
		(0.00597)				(0.00552)
Soil organic carbon (%)		-0.00796				-0.0217
		(0.0391)				(0.0429)
Labor intensity (log hrs/ha/day)			0.109**			0.0964**
			(0.0431)			(0.0413)
Organic amendment (binary)			0.0528			0.0640
			(0.145)			(0.156)
Inorganic fertilizer (binary)			1.566***			1.228***
			(0.203)			(0.305)
Irrigation (binary)			0.0861			-0.376
			(0.356)			(0.404)
Terracing (binary)			0.388***			0.429***
			(0.137)			(0.144)
Head owns plot (binary)				-0.169		-0.196
				(0.156)		(0.152)
Head manages plot (binary)				0.0957		0.0440
				(0.201)		(0.204)
(Head owns)X(Head manages)				0.199		0.189
				(0.236)		(0.233)
Crops are rotated (binary)				-0.0813		0.0427
				(0.122)		(0.118)
Crops are mono-cropped (binary)				0.365***		0.340***
				(0.117)		(0.117)
Mixed cropping (binary)				0.664***		0.581***
				(0.122)		(0.130)
Tubers grown (binary)					0.0594	-0.0474
					(0.112)	(0.108)
Cereals grown (binary)					0.134	-0.0102
					(0.109)	(0.106)
Legumes grown (binary)					0.191*	0.0585
					(0.100)	(0.109)
Bananas grown (binary)					0.123	0.110
					(0.163)	(0.158)
Cash crops grown (binary)					0.670***	0.455***
					(0.161)	(0.158)
Observations	1634	1634	1634	1634	1634	1634
Adjusted R^2	0.362	0.370	0.381	0.391	0.384	0.423

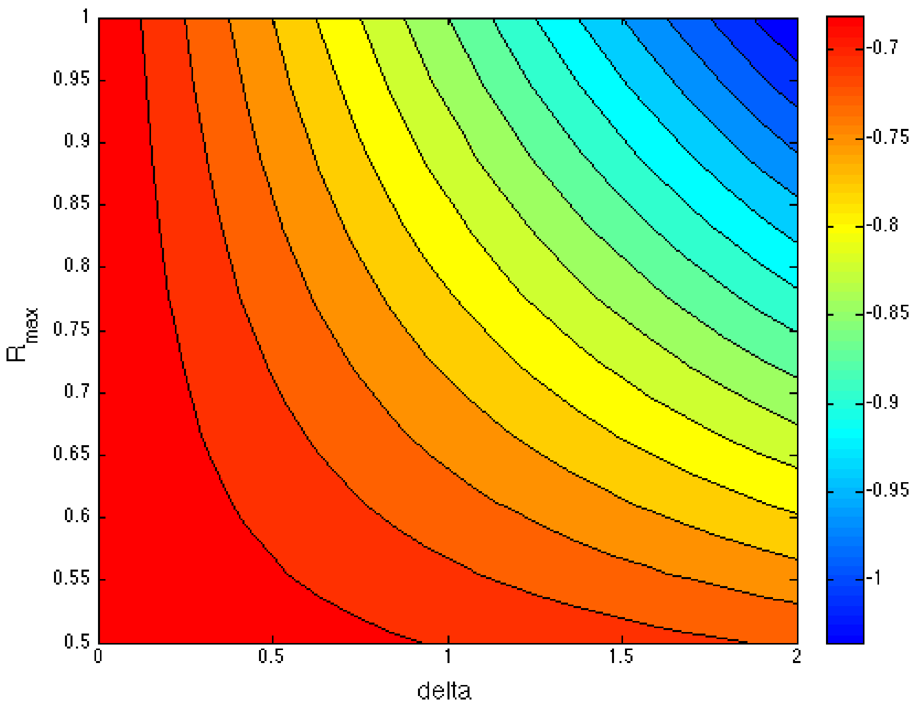
Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A13: Bounds for Potential Causal Relationship

	$R_{max} = 1.3 * R_5$, δ=1, $\hat{\gamma_4}$=-0.638, R_4=0.362				
	(1)	(2)	(3)	(4)	(5)
Column Controls	Soil	Inputs	Management	Crops	Full Set
Column R_5	0.370	0.381	0.391	0.384	0.423
Bounds $[\hat{\gamma_5}\ \gamma^*(R_{max}, \delta)]$	[-0.633 -0.564]	[-0.577 -0.210]	[-0.697 -0.936]	[-0.695 -0.994]	[-0.658 -0.6996]

Coefficients $\hat{\gamma_5}$ and R_5 from Columns 2-6 of Table A12
Coefficient $\hat{\gamma_4}$ and R_4 from Column 1 of Table A12

Figure A2: Bias-Adjusted Estimator $\gamma^*(R_{max}, \delta)$



Appendix 8 Inverse Relationship by Plot Subsets

Table A14: 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)
GPS-measured plot size (log ha)	-0.598***	-0.633***	-0.897***	-0.447***	-0.448***	-0.620***
	(0.0620)	(0.135)	(0.103)	(0.0978)	(0.0918)	(0.184)
Observations	2189	723	1002	1059	871	560
Adjusted R^2	0.369	0.292	0.450	0.316	0.283	0.307

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Table A15: 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)
GPS-measured plot size (log ha)	-0.615***	-0.466***	-0.531***	-0.800***	-0.608***	-0.788***
	(0.0945)	(0.0637)	(0.0713)	(0.186)	(0.0917)	(0.107)
Observations	1556	1297	1152	604	1117	1160
Adjusted R^2	0.381	0.327	0.330	0.267	0.377	0.451

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Appendix 9 Robustness of Edge Effect

Table A16 shows the edge effect to be robust to all controls perviously 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 (2014), using the results from Table A16. As with the inverse relationship, these bounds will contain only positive coefficients (i.e., will not include zero) for almost any assumptions about δ, R_{max}.

Table A16 illustrates the edge effect to be linear across both perimeter-area ratio and plot size. Perimeter-area and plot size quantiles are created in panel; they may change across time for a given plot.

Conversely, Table A18 shows that the inverse relationship is strongest for those plots that experience large shifts in perimeter-area ratio over time, and weakest for those plots that experience very small shifts in perimeter-area ratio over time. In this table, quantiles reflect plot-specific change across time, rather than plot- and time-specific size. The interactions between these quantiles and plot size is therefore akin to estimating the inverse relationship separately, by groups defined by absolute size in the perimeter-area ratio shift over the decade.

Table A19 illustrates the explanatory power of plot shape vs. plot area, when it comes to explaining perimeter-area ratio. Of course, plot shape cannot be fully controlled for, as there is no continuous variable or set of categorical variables that can fully capture plot shape — every single plot is shaped differently. But three dummies for triangular, rectangular, and multi-sided plots capture 20 percent of the variation in perimeter-area ratio.

Table A16: Robustness of Edge Effect (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)	-0.0823	-0.0399	-0.107	-0.274**	-0.212*	-0.183
	(0.122)	(0.141)	(0.133)	(0.136)	(0.127)	(0.162)
Perimeter-area ratio (log m/ha)	0.943***	0.998***	0.855***	0.763***	0.805***	0.859***
	(0.232)	(0.280)	(0.248)	(0.250)	(0.229)	(0.291)
Km to Residence	Yes	No	No	No	No	Yes
Soil Controls	No	Yes	No	No	No	Yes
Input Controls	No	No	Yes	No	No	Yes
Management Controls	No	No	No	Yes	No	Yes
Crop Controls	No	No	No	No	Yes	Yes
Observations	2178	1897	2061	1957	2189	1625
Adjusted R^2	0.383	0.381	0.386	0.422	0.395	0.443

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A17: Edge Effect by Quantiles (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity	(3) Plot Productivity
Perimeter-area ratio (log m/ha)	1.075***		
	(0.0964)		
P-A ratio quantiles=1 × Perimeter-area ratio (log m/ha)		1.331***	
		(0.201)	
P-A ratio quantiles=2 × Perimeter-area ratio (log m/ha)		1.306***	
		(0.187)	
P-A ratio quantiles=3 × Perimeter-area ratio (log m/ha)		1.241***	
		(0.179)	
P-A ratio quantiles=4 × Perimeter-area ratio (log m/ha)		1.247***	
		(0.171)	
P-A ratio quantiles=5 × Perimeter-area ratio (log m/ha)		1.256***	
		(0.157)	
Plot size quantiles=1 × Perimeter-area ratio (log m/ha)			1.117***
			(0.153)
Plot size quantiles=2 × Perimeter-area ratio (log m/ha)			1.101***
			(0.164)
Plot size quantiles=3 × Perimeter-area ratio (log m/ha)			1.077***
			(0.172)
Plot size quantiles=4 × Perimeter-area ratio (log m/ha)			1.117***
			(0.179)
Plot size quantiles=5 × Perimeter-area ratio (log m/ha)			1.136***
			(0.188)
Observations	2189	2189	2189
Adjusted R^2	0.382	0.392	0.387

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A18: Inverse Relationship by Perimeter-Area Change Quantiles (Plot Panel)

	(1) Plot Productivity	(2) Plot Productivity
GPS-measured plot size (log ha)	-0.598***	
	(0.0620)	
(GPS-measured plot size)x(P-A ratio abs. change Q1)		0.215
		(0.235)
(GPS-measured plot size)x(P-A ratio abs. change Q2)		-0.338**
		(0.166)
(GPS-measured plot size)x(P-A ratio abs. change Q3)		-0.431***
		(0.107)
(GPS-measured plot size)x(P-A ratio abs. change Q4)		-0.653***
		(0.0683)
Observations	2187	2187
Adjusted R^2	0.369	0.377

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A19: Perimeter-Area Ratio Explained by Shape and Size (Plot Panel)

	(1) P-A Ratio	(2) P-A Ratio	(3) P-A Ratio	(4) P-A Ratio
Plot is triangle (binary)	1.035***		0.108	-0.671***
	(0.297)		(0.106)	(0.164)
Plot is rectangle (binary)	0.0815		-0.174**	-0.259***
	(0.138)		(0.0705)	(0.0935)
Plot has > 4 sides (binary)	-0.251*		-0.0854	-0.136
	(0.139)		(0.0710)	(0.0935)
GPS-measured plot size (log ha)		-0.545***	-0.546***	-0.416***
		(0.0173)	(0.0152)	(0.0695)
(Plot is triangle)x(GPS-measured plot size)				-0.333***
				(0.0799)
(Plot is rectangle)x(GPS-measured plot size)				-0.129*
				(0.0693)
(Plot has > s4 sides)x(GPS-measured plot size)				-0.0991
				(0.0691)
Observations	2190	2190	2190	2190
Adjusted R^2	0.195	0.883	0.893	0.902

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 10 Edge Effect and Labor Intensity by Labor Subsets

Table A20: Edge Effect and Labor Intensity by Labor Type (Plot Panel)

	(1) Labor Intensity (All)	(2) Labor Intensity (Family)	(3) Labor Intensity (Non-Family)
GPS-measured plot size (log ha)	-0.198	-0.240	-0.0802
	(0.139)	(0.153)	(0.285)
Perimeter-area ratio (log m/ha)	0.784***	0.684**	0.528
	(0.244)	(0.268)	(0.445)
Observations	2080	2044	789
Adjusted R^2	0.185	0.165	0.075

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Table A21: 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)
GPS-measured plot size (log ha)	-0.198	-0.391**	-0.423**	-0.216
	(0.139)	(0.158)	(0.171)	(0.307)
Perimeter-area ratio (log m/ha)	0.784***	0.555**	0.733***	0.355
	(0.244)	(0.275)	(0.278)	(0.476)
Observations	2080	1873	1401	1368
Adjusted R^2	0.185	0.239	0.291	0.058

Estimated with plot, year and season fixed effects
Plot-clustered standard errors in parentheses
p<0.01, ** p<0.05, * p<0.1

Appendix 11 Exogeneity of Perception Error

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

Figure A3: Plot Size Perception Error over Plot Size

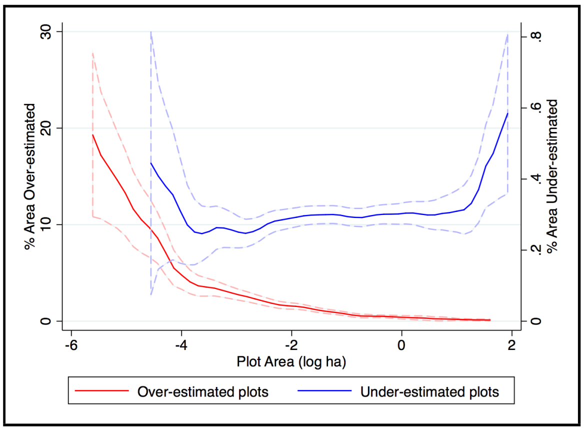
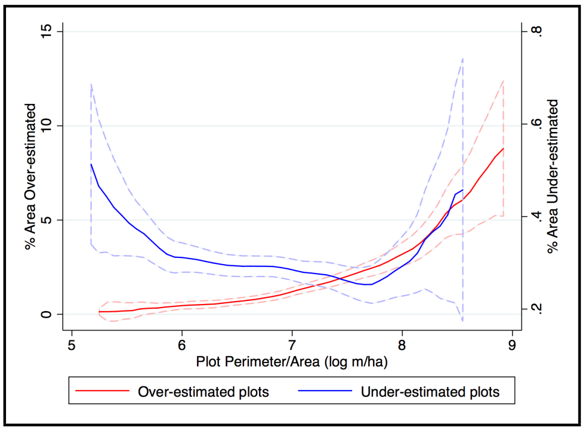


Figure A4: Plot Size Perception Error over Perimeter-Area Ratio



However, Column 1 of Table A22 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 A23 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 A24 models the binary indicator for over-estimation of plot size, using the same covariates again. As with Tables A22 and A23, 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 A22: 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)	0.253	0.230	0.127	-0.454	-0.0641
	(0.680)	(0.747)	(0.717)	(1.047)	(0.843)
(GPS-measured plot size)^2	0.191	0.122	0.159	0.116	0.180
	(0.246)	(0.285)	(0.274)	(0.344)	(0.233)
Perimeter-area ratio (log m/ha)	-17.16	-18.53	-19.03	-21.63	-15.13
	(13.35)	(12.98)	(13.92)	(16.19)	(12.62)
(Perimeter-area ratio)^2	1.303	1.386	1.396	1.542	1.153
	(0.934)	(0.914)	(0.985)	(1.126)	(0.871)
Soil pH (pH)		10.01*			
		(5.951)			
Soil pH^2 (pH^2)		-0.781*			
		(0.470)			
Soil sand (%)		0.00333			
		(0.0215)			
Soil organic carbon (%)		0.152			
		(0.234)			
Labor intensity (log hrs/ha/day)			0.165		
			(0.133)		
Organic amendment (binary)			0.400		
			(0.529)		
Inorganic fertilizer (binary)			0.608		
			(0.938)		
Irrigation (binary)			0.0497		
			(0.560)		
Terracing (binary)			0.718		
			(0.486)		
Head owns plot (binary)				0.0130	
				(0.905)	
Head manages plot (binary)				-2.368	
				(1.551)	
(Head owns)X(Head manages)				1.909	
				(1.671)	
Crops are rotated (binary)				0.0281	
				(0.372)	
Crops are mono-cropped (binary)				-0.0514	
				(0.707)	
Mixed cropping (binary)				0.731	
				(0.606)	
Tubers grown (binary)					-0.198
					(0.516)
Cereals grown (binary)					0.472
					(0.518)
Legumes grown (binary)					0.912**
					(0.458)
Bananas grown (binary)					1.641**
					(0.690)
Cash crops grown (binary)					0.967
					(0.588)
Observations	867	767	823	766	867
Adjusted R^2	0.281	0.300	0.278	0.295	0.308

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 A23: 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)	-0.0996	-0.0589	-0.117	-0.108	-0.0863
	(0.0870)	(0.130)	(0.107)	(0.117)	(0.0850)
(GPS-measured plot size)^2	-0.0396	-0.0937**	-0.0597	-0.0310	-0.0466
	(0.0399)	(0.0410)	(0.0446)	(0.0405)	(0.0452)
Perimeter-area ratio (log m/ha)	-3.872**	-4.606**	-4.788**	-3.701**	-3.815**
	(1.809)	(2.049)	(2.218)	(1.842)	(1.854)
(Perimeter-area ratio)^2	0.277**	0.353**	0.342**	0.259*	0.272*
	(0.141)	(0.154)	(0.170)	(0.142)	(0.145)
Soil pH (pH)		0.807			
		(0.550)			
Soil pH^2 (pH^2)		-0.0762			
		(0.0475)			
Soil sand (%)		-0.000710			
		(0.00284)			
Soil organic carbon (%)		-0.0147			
		(0.0226)			
Labor intensity (log hrs/ha/day)			0.00441		
			(0.0237)		
Organic amendment (binary)			0.0361		
			(0.0970)		
Inorganic fertilizer (binary)			0.0801		
			(0.123)		
Irrigation (binary)			-0.338**		
			(0.164)		
Terracing (binary)			0.0461		
			(0.0683)		
Head owns plot (binary)				0.155**	
				(0.0719)	
Head manages plot (binary)				0.0806	
				(0.107)	
(Head owns)X(Head manages)				-0.0603	
				(0.124)	
Crops are rotated (binary)				0.00385	
				(0.0506)	
Crops are mono-cropped (binary)				-0.0199	
				(0.0670)	
Mixed cropping (binary)				0.0270	
				(0.0579)	
Tubers grown (binary)					-0.0736
					(0.0478)
Cereals grown (binary)					-0.0765
					(0.0490)
Legumes grown (binary)					-0.0428
					(0.0485)
Bananas grown (binary)					0.0503
					(0.0817)
Cash crops grown (binary)					-0.0842
					(0.0812)
Observations	607	507	577	551	607
Adjusted R^2	0.088	0.130	0.125	0.176	0.130

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 A24: 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)	-0.0202	0.0197	0.0138	-0.0507	-0.0644
	(0.0697)	(0.0839)	(0.0812)	(0.0917)	(0.0736)
(GPS-measured plot size)^2	0.0295*	0.0305	0.0386**	0.0237	0.0240
	(0.0165)	(0.0187)	(0.0183)	(0.0210)	(0.0167)
Perimeter-area ratio (log m/ha)	1.013**	1.322***	1.442***	0.979*	0.900**
	(0.412)	(0.464)	(0.460)	(0.534)	(0.422)
(Perimeter-area ratio)^2	-0.0610**	-0.0765***	-0.0864***	-0.0585*	-0.0548**
	(0.0265)	(0.0293)	(0.0298)	(0.0345)	(0.0272)
Soil pH (pH)		-0.293			
		(0.452)			
Soil pH^2 (pH^2)		0.0240			
		(0.0369)			
Soil sand (%)		-0.00344			
		(0.00210)			
Soil organic carbon (%)		0.00945			
		(0.0141)			
Labor intensity (log hrs/ha/day)			-0.0374**		
			(0.0155)		
Organic amendment (binary)			-0.0190		
			(0.0543)		
Inorganic fertilizer (binary)			-0.122		
			(0.135)		
Irrigation (binary)			-0.00136		
			(0.214)		
Terracing (binary)			0.0841*		
			(0.0498)		
Head owns plot (binary)				0.00457	
				(0.0629)	
Head manages plot (binary)				-0.117	
				(0.0825)	
(Head owns)X(Head manages)				0.122	
				(0.0964)	
Crops are rotated (binary)				0.0109	
				(0.0436)	
Crops are mono-cropped (binary)				0.0734	
				(0.0536)	
Mixed cropping (binary)				0.0501	
				(0.0524)	
Tubers grown (binary)					-0.00447
					(0.0402)
Cereals grown (binary)					0.00898
					(0.0396)
Legumes grown (binary)					0.0718*
					(0.0367)
Bananas grown (binary)					-0.0149
					(0.0550)
Cash crops grown (binary)					0.116**
					(0.0550)
Observations	1478	1278	1403	1320	1478
Adjusted R^2	0.159	0.150	0.174	0.163	0.165

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

Appendix 12 Perception Error by Crop Type

Table A25: The Effects of Farmer Misperception of Plot Size (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)	-0.699*	-0.399*	-0.403*	-0.198	-0.313
	(0.366)	(0.213)	(0.230)	(0.275)	(0.330)
Over-estimate (% area)	0.161	0.0369	0.135*	0.212***	0.136*
	(0.0976)	(0.0765)	(0.0731)	(0.0557)	(0.0821)
Over-estimate squared	-0.00354	-0.00212	-0.00344	-0.00563**	-0.00284
	(0.00333)	(0.00267)	(0.00238)	(0.00232)	(0.00255)
Under-estimate (% area)	-6.241**	-2.665**	-3.934***	1.484	-1.348
	(3.009)	(1.231)	(1.446)	(1.671)	(2.073)
Under-estimate squared	8.783*	2.615*	5.027***	-1.109	1.619
	(4.838)	(1.393)	(1.697)	(1.985)	(2.626)
Observations	721	1000	1058	871	560
Adjusted R^2	0.325	0.503	0.375	0.324	0.362

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

Appendix 13 Labor Intensity and Perception Error

Table A26: 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)	-0.270*	-0.263	0.0721
	(0.162)	(0.162)	(0.201)
Over-estimate (% area)	0.0917**	0.103**	0.0942**
	(0.0398)	(0.0408)	(0.0447)
Over-estimate squared	-0.00183	-0.00252	-0.00168
	(0.00149)	(0.00154)	(0.00154)
Under-estimate (% area)	0.466	0.479	2.422**
	(0.993)	(1.008)	(1.174)
Under-estimate squared	-0.996	-1.039	-3.836***
	(1.265)	(1.295)	(1.411)
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	2078	2078	1633
Adjusted R^2	0.207	0.211	0.266

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