Assessing misallocation in agriculture: plots versus farms *

Fernando M. Aragón Diego Restuccia Juan Pablo Rud

January 2023

Abstract

We assess the extent of misallocation in agriculture in less-developed countries comparing the analysis at the plot and farm levels. Using detailed data from Uganda, we show that the plot-level analysis leads to substantially larger estimates of reallocation gains, even after adjusting for measurement error and unobserved heterogeneity. These discrepancies arise due to greater measurement error in plot-level data and different production function estimates. Our findings suggest caution is needed when extrapolating insights obtained using plot-level analysis to results obtained at the farm level.

JEL classification: O4, O13, Q12.
Keywords: misallocation, agriculture, measurement error, distortions.

*All errors and interpretations are our own. Restuccia gratefully acknowledges the support from the Canada Research Chairs program and the Bank of Canada Fellowship. Aragon gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada, grant 435-2018-0227. Rud gratefully acknowledges the support from the Economic and Social Research Council (UK Research and Innovation), Project ES/S014438/1. The views expressed herein are not necessarily those of the Bank of Canada and are the authors alone. *Contact Information:* Aragón: Department of Economics, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada (e-mail: faragons@sfu.ca); Restuccia: Department of Economics, University of Toronto, 150 St. George Street, Toronto, ON M5S 3G7, Canada (e-mail: diego.restuccia@utoronto.ca); Rud: Department of Economics, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK (e-mail: juan.rud@rhul.ac.uk).

1 Introduction

A growing literature documents substantial dispersion in measures of marginal products of inputs across production units. This finding has been interpreted as evidence of factor misallocation (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2017). A relevant concern, however, is that the observed dispersion might reflect other factors such as overhead costs, unobserved heterogeneity, or measurement error (Bartelsman et al., 2013; Asker et al., 2014; Foster et al., 2016). There are several emerging approaches to deal with these issues. For instance, an ambitious line of research models specific sources of misallocation to identify their quantitative importance using microdata (Midrigan and Xu, 2014; David and Venkateswaran, 2019; Yang, 2021). Another method exploits the availability of panel data to purge measures of marginal products from time-invariant measurement error and overhead costs (Bils et al., 2017).

In the context of small-scale agriculture, a complementary approach emphasizes using granular data, at plot-level (Gollin and Udry, 2021; Abay et al., 2019; Desiere and Jolliffe, 2018). The intuition behind this approach is simple. If farmers can freely allocate inputs across plots within their farm operation, then the marginal productivity of inputs should be equalized across operated plots. Thus, observed within-farm dispersion in measures of marginal productivity can be attributed to other sources rather than misallocation. This approach, however, contrasts with the existing literature on misallocation in agriculture, which focuses on the household farm as the unit of analysis.

In this paper, we examine empirically whether the assessment of misallocation in agriculture (and therefore the role of mismeasurement) is affected by the level of analysis: plots or farms. Our analysis uses data and previous estimates from Uganda and, similar to recent work in the literature, assess misallocation using efficiency gains, i.e., the increase in aggregate output that could be obtained from reallocating resources across production units according to an efficient benchmark. Our main insight is that the level of data aggregation matters, and can lead to quantitatively different conclusions.

We find that efficiency gains at the plot level are extremely high, even after adjusting for measurement error. Plot-level estimates suggest efficiency gains of reallocation at the national level of more than 2,200%. After adjusting for measurement error using Gollin and Udry (2021)'s methodology, efficiency gains are still greater than 500%. These estimates imply an extent of misallocation far greater than previously documented in the literature. As a comparison, previous studies using farm-level data from China, Ethiopia and Malawi document efficiency gains ranging from 53% to 259%. Estimates using the same Ugandan dataset, but aggregated at the farm-level, also suggest more modest gains of around 175%.

The large discrepancy in assessed misallocation cast doubts on the validity of extrapolating insights obtained using plot-level analysis to results obtained at the farm level. This issue becomes apparent when assessing the role of measurement error. For example, a researcher using plot-level data to assess measurement error would observe that it explains a large fraction of the productivity dispersion. The researcher could then conclude that misallocation is not important and that previous estimates (using farm-level data) overstated its magnitude. This conclusion, however, would be misplaced: given the large initial estimates, even a substantial reduction in dispersion still leaves sizeable levels of misallocation.

Which is the right level of analysis then? Even though plot-level analysis promises a way to address measurement error and unobserved heterogeneity, we argue that it suffers from two empirical limitations. First, granular data may actually exacerbate measurement error. We compare self-reported area of landholdings to their GPS measure and show evidence of substantial and systematic measurement error. The magnitude of this error, however, is attenuated when aggregating data at farm level.

Second, using plot-level data makes it difficult to implement estimation methods based on panel data. This can lead to substantially different production function estimates. In the

Uganda case, plot-level IV estimates suggest near constant returns to scale (0.91). These results contrast with panel-data estimates from Uganda and other Sub-Saharan countries which suggest smaller values around 0.71. These differences are economically relevant. For instance, simply changing the assumed returns to scale from 0.91 to 0.71 leads to a massive drop in efficiency gains calculated with plot-level data from 23.9-fold to 5.3-fold.

In addition to the methodological implications of using plot or farm measures, there are important implications for understanding and addressing misallocation. From a policy standpoint, the focus on the farm is relevant given the fact that land institutions in many developing countries allocate land rights at the household level (Restuccia, 2020). Also, in the context of small-scale agriculture, the presence of fixed factors shared across plots makes the household farm the appropriate production unit (De Janvry et al., 1991).

2 Does the level of microdata aggregation matter?

We start by comparing estimates of productivity dispersion and reallocation gains across plots and across farms. In our analysis, a farm is the set of plots cultivated by the household. The data comprises three waves from the Uganda Panel Survey (2009-2014), a household survey collected as part of the World Bank's Integrated Surveys of Agriculture (LSMS-ISA). Gollin and Udry (2021) and Aragón et al. (2022) provide a detailed description of the data.

2.1 Measuring misallocation

We measure the extent of misallocation by calculating efficiency gains, i.e., the ratio of efficient to actual aggregate output. This ratio quantifies the increase in aggregate output that could be obtained if resources (such as land and labor) were assigned across production units according to an efficient benchmark. Following the literature on misallocation, we use as benchmark the allocation that maximizes the aggregate output subject to the available resources,

In the special case studied in Hsieh and Klenow (2009), the efficiency gains can be summarized by the dispersion of log TFPR (a weighted average of marginal revenue products). However, in a more general case, estimating efficiency gains requires computing the output and input allocation in the efficient benchmark.

To do so, we consider an economy comprised of a given set of production units with the following Cobb-Douglas technology,

$$Y_i = s_i (L_i)^{\alpha_L} (X_i)^{\alpha_X}, \quad \alpha_L, \alpha_X > 0, \quad \alpha_L + \alpha_X < 1, \tag{1}$$

where L_i and X_i are the amounts of land and labor used in production unit i, and s_i denotes its productivity.

The efficient allocation equates marginal products of land and labor across production units. Denoting $z_i \equiv s_i^{1/(1-\alpha_L-\alpha_X)}$, we can characterize the efficient allocation as:

$$T_i^e = \frac{z_i}{\sum_i z_i} T, \qquad L_i^e = \frac{z_i}{\sum_i z_i} L, \tag{2}$$

where T and L are the aggregate amounts of land and labor. The efficiency gain, our measure of misallocation, is the ratio $\sum_i Y_i^e / \sum_i Y_i$ where Y_i^e is production-unit output associated with the efficient input allocation, and Y_i is the observed output.

There are three observations relevant to the empirical analysis. First, the efficient allocation implies a positive relation between production-unit productivity and input use. The slope of this relation is proportional to the return to scale $(\alpha_L + \alpha_X)$. Second, deviations from this positive relation are indicative of the extent of misallocation. Finally, to calculate efficiency gains we need estimates of the production-unit productivity s_i and parameters α_L and α_X.

2.2 Estimates of misallocation using plot-level data

To assess misallocation at the plot level, we rely on estimates from Gollin and Udry (2021). They use state-of-the-art methods to estimate plot-level productivity and adjust it for measurement error and unobserved heterogeneity. For our analysis, we use their two-stage least squares (2SLS) estimates. The results are, however, robust to using the alternative instrumental variables correlated random coefficients (IVCRC) estimates.

Table 1 presents the efficiency gains and productivity dispersion assuming input reallocation at different geographic levels, and using alternative measures of productivity. Column 1 uses the baseline plot productivity. This variable is called TFPA in Gollin and Udry (2021). Column 2 uses the measure of plot productivity adjusted for measurement error and unobserved heterogeneity, called TFPB. Column 3 aggregates plot productivity (TFPA) at the farm level by calculating a weighted average.[1]

We emphasize three relevant observations from Table 1.

Observation 1: Efficiency gains estimated using plot-level data are extremely large The estimates in Column 1 imply that if the allocations of land and labor were to change to the efficient allocation at the national level, agricultural output would increase by a factor of 23.9-fold, or more than 2,200%. Efficiency gains remain large even when reallocation is limited to smaller geographical areas: 1,540% within regions and 303% within villages. These results imply a very large magnitude of factor misallocation.

These estimates are much larger than those documented in the macroeconomic literature using the farm as the unit of analysis. For example, the estimated reallocation gains (at the national level) in China, Ethiopia and Malawi are 53%, 97% and 259%, respectively (Adamopoulos et al., 2022; Chen et al., 2022; Restuccia and Santaeulalia-Llopis, 2017).

[1] Denoting s_{ij} the unadjusted productivity of plot j in farm i, and given the Cobb-Douglas technology on land and labor, aggregated farm productivity is given by $\sum_j s_{ij}(\phi_{ij}^L)^{\alpha_L}(\phi_{ij}^X)^{\alpha_X}$, where ϕ_{ij}^L and ϕ_{ij}^X are the shares of farm i's total land and labor used in plot j.

Reallocation gains are comparatively large even within narrower geographical areas. For instance, the within-village reallocation is 305% using plot-level data in Uganda whereas only 24% in China (Adamopoulos et al., 2022).

Observation 2: Estimated efficiency gains remain high even after adjusting for measurement error and unobserved heterogeneity Column 2 replicates the analysis using estimates of productivity adjusted by measurement error and unobserved heterogeneity as suggested by Gollin and Udry (2021). The adjustment dramatically reduces the dispersion of productivity by almost two-thirds. The reduction in dispersion is associated with a proportional drop in the implied efficiency gains. Nevertheless, the adjusted reallocation gains are still sizable (ranging from 147% to 568%) and well-above previous estimates of misallocation in agriculture.

This observation is particularly relevant when studying the importance of measurement error in assessments of misallocation. To see this, consider a researcher observing the relative changes in columns 1 and 2. Since measurement error seems to account for a large bulk of the efficiency gains and productivity dispersion, the researcher would conclude that misallocation in agriculture is not quantitatively important.

This conclusion, however, would be misplaced. Given the large initial estimates, even a substantial reduction in dispersion still leaves sizeable levels of efficiency gains. In this particular example, they are actually greater than estimates from previous studies.

Observation 3: The larger estimates seem to be driven by the level of analysis A possible interpretation of the results is that the plot-level analysis is picking up effectively larger misallocation in Uganda. Two pieces of evidence suggest that this is not the case, but instead that the results are driven by the level of analysis.

First, we aggregate the unadjusted plot-level estimates of productivity (TFPA) at the farm level (column 3). This simple aggregation accomplishes a sizeable reduction in productivity dispersion and implied efficiency gains. The efficiency gains are almost half the estimated gains using plot-level productivity in column 1. In the case of reallocation gains at village level, the estimates using the aggregated productivity (110%) are even smaller than after adjusting by measurement error in column 2 (147%).

Second, we replicate the analysis using estimates for Uganda from Aragón et al. (2022) (column 4). These estimates are obtained from the same dataset, but using the farm (household) as unit of analysis. This change in unit of analysis allows us not only to aggregate output and inputs at farm level, but also to use a panel data approach to estimate the production function.

The estimated reallocation gains are substantially smaller: 175% at the national level and 69% at the village level. These gains are much smaller than any estimate obtained using plot-level analysis. However, they are closer in magnitude to estimates from macroeconomic studies in other contexts.

This substantial reduction in assessed misallocation occurs despite that dispersion of farm-productivity being greater than the dispersion of plot-productivity (after adjusting for measurement error). This finding also illustrates the limitation of using productivity dispersion alone to assess misallocation. In general, the magnitude of efficiency gains is not only a function of productivity dispersion, but also of economies of scale and of the relationship between input allocation and productivity across production units.

The main takeaway is that focusing on the plot as unit of analysis is not a useful starting point for the study of misallocation in agriculture. The large discrepancy in assessed efficiency gains casts doubts on the validity of extrapolating insights obtained using plot-level analysis to results obtained at the farm level.

Table 1: Efficiency gains and productivity dispersion in plot- and farm-level analysis

	Plot-level data			Farm-level data
	Plot productivity	Plot productivity (adjusted)	Plot productivity aggregated at farm level	Farm productivity
	(1)	(2)	(3)	(4)
A. Efficiency gains				
Nationwide	23.96	6.68	14.28	2.75
Region	16.40	5.38	8.35	2.66
Parish (Village)	4.03	2.47	2.10	1.69
B. Dispersion				
Variance of log	1.26	0.53	0.78	0.99

Notes: Efficiency gain is the ratio of aggregate output in the efficient allocation to actual output averaged over season-years. Columns 1-3 use 2SLS estimates of plot productivity from Gollin and Udry (2021). Column 1 uses the baseline (unadjusted) productivity measure (TFPA), whereas Column 2 uses the adjusted productivity measure (TFPB). Column 3 aggregates TFPA at the farm level by computing a weighed average. Column 4 uses direct estimates of farm productivity from Aragón et al. (2022).

3 What explains the different results?

We highlight two important empirical limitations of using plot-level analysis when assessing misallocation in agriculture.[2] First, plot-level analysis can exacerbate measurement error. Second, plot-level analysis may lead to different production function estimates. Together, these issues might increase the magnitude of assessed misallocation, and overstate the contribution of measurement error.

3.1 Greater measurement error in more disaggregated data

Most survey data on smallholder agriculture are based on farmers' self-reporting. This feature creates the possibility of misreporting and measurement error: farmers may round-up quantities, or simply provide guesstimates instead of actual values. In some cases, measure-

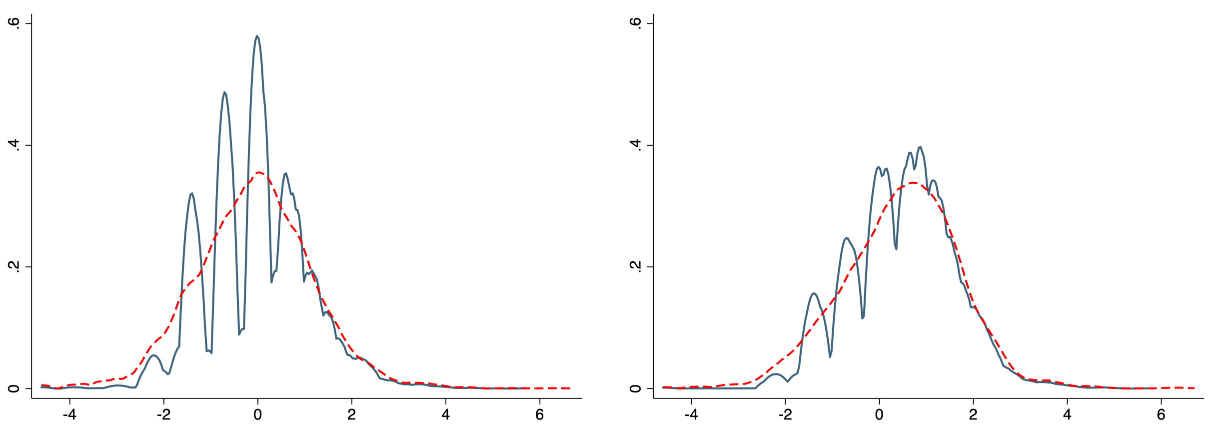
[2]Conceptually, it is not clear what it means to reallocate resources, including land, across plots since the plot is not an administrative unit of production.

ment error can also be introduced when allocating indirect costs (such as capital expenses or management labor) among particular activities or crops. This issue is a concern because measurement error can bias estimates of the production function, or be included as part of the residual often attributed to productivity.

We focus on measurement error on land. To assess whether the level of analysis matter, we compare two measures of size of land holdings: self-reported by the farmer, and a GPS measure collected by enumerators. Albeit not exempt from measurement error, this last measure is arguably more precise and less prone to farmer's misreporting (Carletto et al., 2017). For each variable, we aggregate the data at both farm-level and at parcel-level. A parcel is a set of plots and thus less dissagregated than plot-level analysis. The GPS measures, however, are only available at this level of resolution.

Figure 1 displays the distribution of land holdings using both measures. Panel (a) shows the distribution of the original data at the parcel level, while panel (b) shows the distribution of the data aggregated at the farm level. There are three relevant observations.

Figure 1: Distribution of landholding size, self-reported and GPS measured



(a) Parcel level (b) Farm level

Notes: Distribution of the log area of landholdings at the parcel level (panel a) and aggregated to the farm level (panel b). Solid lines represent self-reported values, while red-dashed lines represent GPS measures.

First, there are evident discrepancies between both measures. This discrepancy has been documented by other studies and interpreted as evidence of measurement error in self-reported values (see, for instance, Judge and Schechter, 2009; Carletto et al., 2015; Gourlay et al., 2019; Abay et al., 2021). Interestingly, the GPS measure follows a smooth bell-shaped distribution, while the self-reported measures are heaped around certain values. The observed 'heaping' is indicative of respondents (or enumerators) rounding the reported size (Abay et al., 2019, 2021; Carletto et al., 2013).

Second, the discrepancy between both measures is more pronounced among smaller units, on the left side of the distribution. This evidence suggests that the measurement error is not classical, but correlated to unit size. This pattern has been documented in other studies. For instance, Abay et al. (2021) reports a negative correlation between plot size and measurement error in land in four Sub-Saharan African countries.

Finally, the differences between self-reported and GPS measures are greater when using more granular data. This observation suggests that using parcel-level data can exacerbate measurement error. To quantitatively assess this issue, we measure the distance between self-reported and GPS measures using their absolute relative difference (ARD).[3] The average ARD using farm-level data is 0.678. However, this value increases to 0.891 when using parcel-level data. This corresponds to an increase of almost a third in this measure of distance.

3.2 Different production function estimates

Production function estimates are key inputs to assess factor misallocation. The main econometric challenge in estimating the production function parameters is the presence of determinants of production, such as productivity shocks, that are unobserved to the econometrician but observed by the farmer.

Panel data offer a way to address this endogeneity problem. For instance, if the unobserved productivity shocks are time-invariant (such as location, soil quality, or farming

[3] Formally, we define $ARD = abs[\frac{self_i - GPS_i}{GPS_i}]$, where i is the unit of observation (farm or parcel).

ability), the production function can be consistently estimated using a panel data model with fixed effects (Ackerberg et al., 2015). In contexts with imperfect input markets, a simplified dynamic panel model can also address auto-regressive, time-varying, productivity shocks (Shenoy, 2017, 2020).[4]

Plot-level analysis, however, limits the use of panel data methods. This occurs because the available plot-level data is mostly cross-sectional.[5] In contrast, there are several agricultural surveys, such as the World Banks' Integrated Surveys of Agriculture, that already include a panel of households. For instance, in the case of Uganda, Gollin and Udry (2021) are unable to use panel data methods with their plot-level data. Instead, they rely on a clever instrumental variable approach using self-reported productivity shocks on nearby plots. In contrast, Aragón et al. (2022) use the same dataset aggregated at the household level and estimate a panel-data model with fixed effects.

These methodological differences can generate substantially different estimates of returns to scale and productivity. To see this, we display the production function estimates from these two studies (see Table 2). Note that are sizable differences in the contributions of land and labor, and in the implied returns to scale. In particular, the IV estimates using plot-level data suggest a larger contribution of land, and returns to scale closer to unity. This is consistent with the farm being an aggregation of constant-return-to-scales' plots with some fixed factors.

These differences in production function estimates matter for the empirical assessment of misallocation. We illustrate this point in two ways. First, we re-calculate the efficiency gains (obtained in column 1 of Table 1) using the same estimates of plot-level productivity from

[4]Proxy variables methods such as Olley and Pakes (1996), Levinsohn and Petrin (2003) and Ackerberg et al. (2015) also exploit panel data. However, they are of limited use in contexts with suspect factor misallocation because, as shown by Shenoy (2020), their identification assumptions fail when market frictions distort input choices.

[5]Collecting a panel dataset at the plot level does not necessarily address this concern, given the endogeneity of plot formation, which would likely introduce additional biases.

Table 2: Production function estimates at the plot and farm levels

	IV (2SLS) (1)	Panel data with fixed effects (2)
Land contribution (α_L)	0.69	0.37
Labor contribution (α_X)	0.22	0.34
Returns to scale ($\alpha_L + \alpha_X$)	0.91	0.71
Aggregation level	Plot	Household

Notes: Column 1 displays 2SLS estimates reported in Table 9 (column 3) in Gollin and Udry (2021). Column 2 display estimates reported in Table A.1 (column 1) in Aragón et al. (2022).

Gollin and Udry (2021) but using the land and labor contributions estimated using panel-data methods in Aragón et al. (2022). This change implies a reduction in returns to scale from 0.91 to 0.71. This last value is similar to returns to scale documented by recent studies using farm-level data such as Shenoy (2017), Restuccia and Santaeulalia-Llopis (2017) and Manysheva (2021). This small change leads to a substantial reduction in the efficiency gains at the national level from 23.9-fold to 5.3-fold. A similar pattern is found for gains at the regional level (from 16.4 to 4.6-fold) and the parish (village) level (from 4 to 2.3-fold).

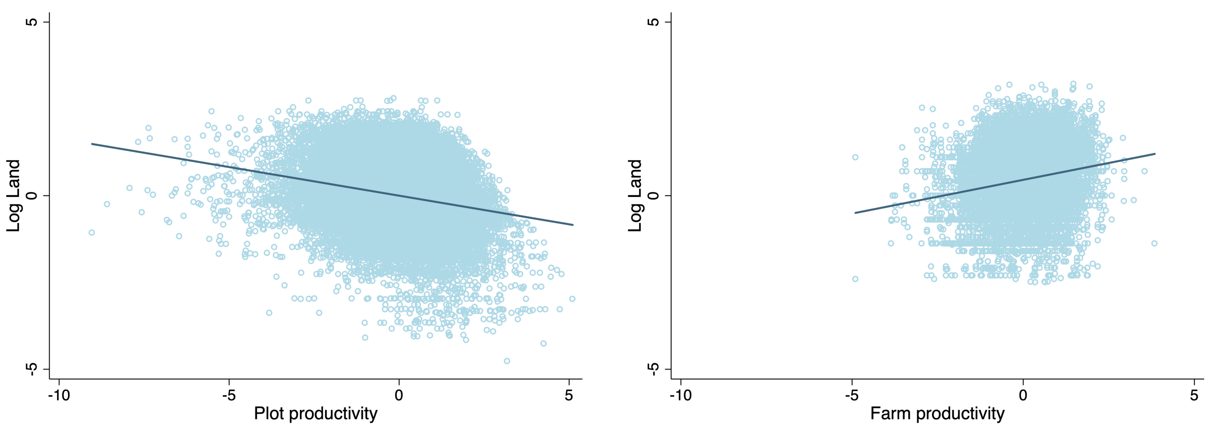
Second, we report the relationship between land use and productivity across production units using both plot-level and farm-level analysis (see Figure 2). As shown in equation (2), the efficient allocation requires a strong positive relationship between productivity and input use. Deviations from this benchmark would be indicative of the extent of misallocation and have been the focus of an expanding literature in development and agricultural economics (Adamopoulos and Restuccia, 2014; Restuccia, 2020).

Based on the estimated returns to scale in Table 2, the slope of the (log) land-productivity relationship in the efficient allocation should be 11.1 across plots and 3.4 across farms. In both cases, the main observation from Figure 2 is that the slope of the relationship between productivity and land input is much smaller than that required in the efficient allocation.

While the farm estimates (Panel b) show a weak but positive relationship, i.e. a slope of 0.26, the plot-level estimates (Panel a) actually show a negative relationship, i.e. an elasticity of -0.16.

This result implies an even larger deviation from the efficient benchmark in the plot-level analysis, and thus a greater implied factor misallocation. Note that the same pattern between land input and productivity arises when adjusting plot-productivity by the within-farm dispersion across plots as in Gollin and Udry (2021) since the adjustment in this approach amounts to a scalar reduction in variance which is proportional to the productivity variance across plots within farm households.

Figure 2: Land size and productivity across production units



(a) IV estimates (plots) (b) Panel estimates (farms)

Notes: Both panels display the scatter-plot of size of production unit (measured by area planted) and productivity, and a fitted linear regression. Panel (a) uses plot-level measures of size and productivity from 2SLS estimates in Gollin and Udry (2021). Panel (b) uses farm-level measures of size and productivity from a panel data model in Aragón et al. (2022).

4 Measurement error and misallocation

If plot-level data is problematic, how can we assess the extent of measurement error in measures of misallocation? A growing literature have instead turned to approaches exploiting panel data. We follow Bils et al. (2017) in exploiting time-variation in the data to address measurement error. The extent to which variation over time in inputs is not reflected in variation in output, which varies across levels of measured distortions, provides a metric of the extent of measurement error.

For comparison, Gollin and Udry (2021) argue that measurement error and other sources of unobserved heterogeneity play a substantial role in accounting for the apparent misallocation in agriculture. Using plot-level data, they show that adjusting for measurement error by the within-farm dispersion in productivity, reduces estimates of misallocation by almost two thirds. As discussed earlier, the implied large reduction in misallocation is due in great part to the exceptionally large estimates of misallocation at the plot level.

This method to identify mismeasurement, however, is not applicable to the farm level since it relies on the assumption of efficient within-farm allocation of resources. An alternative approach, proposed by Bils et al. (2017), exploits panel data to quantify the extent to which misallocation reflects additive measurement error. The starting point is the observation that the ratio of first differences (i.e., the change in revenue divided by the change in inputs) is a measure of marginal product purged from constant measurement error. Based on this insight, they develop a metric, λ, that captures the fraction of the dispersion in revenue productivity (TFPR) that is due to true variation in distortions.

Using the panel household-farm data for Uganda from Aragón et al. (2022), we find that the estimate of λ is fairly high (0.926), implying that only about 7.4% of the variation in misallocation can be ascribed to measurement error. This result is consistent with the findings in Adamopoulos et al. (2022) using Chinese panel data, where this method detects only 4% measurement error in farm-level measures and 10% in cross sectional farm-level data.

The extent of farm-level measurement error is substantially smaller than that implied by an analysis at the plot level. It is also much smaller than that in the manufacturing sector analyzed in Bils et al. (2017) for India and the United States, which suggests caution is needed when making comparisons of measurement error across sectors, countries, and applications.

5 Conclusion

Does exploiting granularity in micro data provide a better assessment of misallocation in developing economies? We address this question using a common dataset for Uganda analyzed at two levels of aggregation: plots versus farms.

We show that the plot-level analysis produces larger estimates of misallocation, even after controlling for unobserved heterogeneity and measurement error. The large discrepancy in assessed efficiency gains suggest that caution is needed when extrapolating insights obtained from plot-level analysis, such as the importance of measurement error, to results obtained at the farm level. We trace the differential results between the plot and the farm level analyses to greater measurement error in dissagregated data, and differences in the estimates of production function parameters.

Overall, our analysis suggest that, despite their potential advantages to purge data from measurement error and unobserved heterogeneity, plot-level analysis might not be appropriate to assess misallocation in agriculture. Instead, focusing on the farm provides a more accurate assessment, particularly in the context of small-scale farming in low-income countries.

References

Abay, Kibrom A., Gashaw T. Abate, Christopher B. Barrett, and Tanguy Bernard, "Correlated non-classical measurement errors, 'Second best' policy inference, and the inverse size-productivity relationship in agriculture," *Journal of Development Economics*, 2019, *139*, 171 – 184.

Abay, Kibrom A, Leah EM Bevis, and Christopher B Barrett, "Measurement Error Mechanisms Matter: Agricultural intensification with farmer misperceptions and misreporting," *American Journal of Agricultural Economics*, 2021, *103* (2), 498–522.

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer, "Identification properties of recent production function estimators," *Econometrica*, 2015, *83* (6), 2411–2451.

Adamopoulos, Tasso and Diego Restuccia, "The Size Distribution of Farms and International Productivity Differences," *American Economic Review*, June 2014, *104* (6), 1667–97.

_ , Loren Brandt, Jessica Leight, and Diego Restuccia, "Misallocation, selection, and productivity: A quantitative analysis with panel data from china," *Econometrica*, 2022, *90* (3), 1261–1282.

Aragón, Fernando M., Diego Restuccia, and Juan Pablo Rud, "Are small farms really more productive than large farms?," *Food Policy*, 2022, *106*, 102168.

Asker, John, Allan Collard-Wexler, and Jan De Loecker, "Dynamic inputs and resource (mis) allocation," *Journal of Political Economy*, 2014, *122* (5), 1013–1063.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta, "Cross-country differences in productivity: The role of allocation and selection," *American Economic Review*, 2013, *103* (1), 305–34.

Bils, Mark, Peter J Klenow, and Cian Ruane, "Misallocation or mismeasurement?," *Working paper No. 599, Center for International Development, Stanford University*, March 2017.

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

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

_ , _ , Siobhan Murray, and Alberto Zezza, "Cheaper, faster, and more than good enough: is GPS the new gold standard in land area measurement?," in "Survey Research Methods," Vol. 11 2017, pp. 235–265.

Chen, Chaoran, Diego Restuccia, and Raül Santaeulàlia-Llopis, "The effects of land markets on resource allocation and agricultural productivity," *Review of Economic Dynamics*, 2022, *45*, 41–54.

David, Joel M and Venky Venkateswaran, "The sources of capital misallocation," *American Economic Review*, 2019, *109* (7), 2531–67.

Desiere, Sam and Dean Jolliffe, "Land productivity and plot size: Is measurement error driving the inverse relationship?," *Journal of Development Economics*, 2018, *130*, 84–98.

Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf, "Firm-level dispersion in productivity: is the devil in the details?," *American Economic Review*, 2016, *106* (5), 95–98.

Gollin, Douglas and Christopher Udry, "Heterogeneity, measurement error, and misallocation: Evidence from african agriculture," *Journal of Political Economy*, 2021, *129* (1), 000–000.

Gourlay, Sydney, Talip Kilic, and David B Lobell, "A new spin on an old debate: Errors in farmer-reported production and their implications for inverse scale-Productivity relationship in Uganda," *Journal of Development Economics*, 2019, *141*, 102376.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, November 2009, *124* (4), 1403–1448.

Janvry, Alain De, Marcel Fafchamps, and Elisabeth Sadoulet, "Peasant household behaviour with missing markets: some paradoxes explained," *The Economic Journal*, 1991, *101* (409), 1400–1417.

Judge, George and Laura Schechter, "Detecting problems in survey data using Benfordâs Law," *Journal of Human Resources*, 2009, *44* (1), 1–24.

Levinsohn, James and Amil Petrin, "Estimating production functions using inputs to control for unobservables," *The Review of Economic Studies*, 2003, *70* (2), 317–341.

Manysheva, Kristina, "Land Property Rights, Financial Frictions, and Resource Allocation in Developing Countries," *mimeo*, 2021.

Midrigan, Virgiliu and Daniel Yi Xu, "Finance and Misallocation: Evidence from Plant-Level Data," *American Economic Review*, February 2014, *104* (2), 422–58.

Olley, G Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64* (6), 1263–1297.

Restuccia, Diego, "The Impact of Land Institutions and Misallocation on Agricultural Productivity," Technical Report, National Bureau of Economic Research March 2020.

_ **and Raul Santaeulalia-Llopis**, "Land misallocation and productivity," Technical Report, National Bureau of Economic Research 2017.

_ **and Richard Rogerson**, "The causes and costs of misallocation," *Journal of Economic Perspectives*, 2017, *31* (3), 151–74.

Shenoy, Ajay, "Market failures and misallocation," *Journal of Development Economics*, 2017, *128*, 65–80.

_ , "Estimating the Production Function Under Input Market Frictions," *Review of Economics and Statistics*, 2020, pp. 1–45.

Yang, Mu-Jeung, "Micro-level Misallocation and Selection," *American Economic Journal: Macroeconomics*, October 2021, *13* (4), 341–68.