The Journal of Development Studies, 2023
Vol. 59, No. 4, 570–592, https://doi.org/10.1080/00220388.2022.2147831

Do Large Estates Benefit Smallholder Neighbours? Evidence from Malawi

JULIAN JOSEPH[*], MICHIEL VAN DIJK[*,**] & TAMÁS KRISZTIN[*]
[*]International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, [**]Wageningen Economic Research, IIASA, Laxenburg, Austria

(Original version submitted February 2022; final version accepted November 2022)

ABSTRACT *We examine spillovers from agricultural estates to Malawian smallholders within an econometric counterfactual framework. We consider economic spillovers such as income, as well as agrarian spillovers such as yields, harvests, and crop diversity. We identify long-run effects of large agricultural investments on small-scale farmers. For the location of large estates, we use a novel OpenStreetMap dataset, while data on smallholder's stems from a household survey. We provide evidence for the importance of the distance threshold for spillovers, and explore multiple thresholds. In proximity to estates we find higher groundnut and pigeon pea yields and increased crop diversity. In very close proximity, incomes are also higher. Area under cultivation in total and for maize are smaller for nearby households, while maize yields are not significantly different. Overall, our results suggest that policies should aim to leverage the increased crop diversity and groundnut yields while mitigating potential detrimental effects arising from reduced cultivated land.*

1. Introduction

Questions around the interplay of smallholder farmers and large-scale agricultural estates have received substantial attention in the past (e.g. Deininger & Byerlee, 2011). Theoretical and empirical insights contribute to the importance of these questions. On the theoretical side, it has been argued that positive spillovers from large investments to smallholders mainly occur through the transfer of technologies. Additionally, income effects and lower supply of productive land can play a role. Thus, we empirically test these main channels for spillovers from estates to smallholders in Malawi with a new data source.

Malawi has undergone substantial investments in agriculture but empirical evidence on spillovers related to those investments remains inconclusive, mainly due to of scarcity of data. The only comprehensive, country-wide study covering the long-term performance of a set of estates in Malawi was conducted by Deininger and Xia (2018) but is based on a dataset which is reported to be unreliable due to overlapping claims in land registers and insufficient coverage in the number of smallholder outcomes possibly affected by spillovers. They find significant

Correspondence Address: Julian Joseph, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. Email: joseph@iiasa.ac.at
Supplementary Materials are available for this article which can be accessed via the online version of this journal available at https://doi.org/10.1080/00220388.2022.2147831

negative impacts of distance to agricultural estates of any size and lease status, including overlapping claims, on the absolute value of agricultural output of smallholders. In the Malawian context additional empirical insights are available only for specific sugarcane estates (Herrmann & Grote, 2015). Studies from other countries in Eastern and Southern Africa (Ali, Deininger, & Harris, 2017, 2019; Deininger & Xia, 2016; Glover & Jones, 2019; Lay, Nolte, & Sipangule, 2021) reach partly contrasting results for some of the key spillover outcomes, showing the limited transferability of results from one country to another and underpinning the need for a separate analysis of the Malawian context.

In this paper, we aim to fill this research gap by examining the long-run spillovers of estates to nearby smallholders in Malawi using a novel estate dataset coupled with inverse probability weighting and propensity scoring approaches. Through aggregating existing literature, employing a new dataset of estates in Malawi and examining spillover channels through related indicators in a counterfactual framework, this study produces a detailed picture of the various ramifications of estates on the smallholders neighbouring them. Specifically, we aim to differentiate between positive spillovers, contributing to increases in productivity, production and income, and negative spillovers concerning land tenure security and increasing poverty (see e.g. De Schutter, 2011; Kleemann & Thiele, 2015).

Our study contributes to the existing literature along three important dimensions. First, data on estates and their locations is difficult to obtain and use. We address this issue by introducing publicly available OpenStreetMap (OSM) as a source of data for large estate locations. In contrast to comparable studies it can contain estates of all dates of establishment. Hence it is helpful for identifying long run effects. Second, econometric methods used in this article extend those previously employed by using a counterfactual framework. Third, the variation of spillovers for varying crops is addressed and complemented with the use of an extensive set of outcome variables. Outcome variables in our model include absolute quantities harvested, yields and field sizes separate for three major crops, as well as per household indicators for income from multiple sources, assets and crop diversity. Additionally, we discuss in detail the importance of the distance threshold up to which spillovers from estates to smallholders occur. The household data we utilise allow for a previously underappreciated discussion of a sensible cutoff for this distance threshold. Similar to the long-run approach of former studies, we do not restrict the date of setup of estates as opposed to many country-level studies using panel data for very short periods of time only. While it was not possible to obtain specific dates of setup for each estate using the OSM data, a substantial part especially of the tea estates around Mulanje and Thyolo are known to date to Colonial times and reaching their current size as early as the 1930s (Chinigò, 2016).

The rest of the paper is structured as follows. Section 2 of the paper states hypothesis for spillover channels derived from theoretical and empirical literature. Section 3 presents our dataset for Malawi, including the novel OpenStreetMap data on large-scale estates. In Section 4 the econometric framework is discussed. Section 5 and 6 contain results and discussion, and Section 7 concludes.

2. Large-scale agricultural investments in Malawi and Sub-Saharan Africa

This section discusses the relevance of the estate sector in Sub-Saharan Africa (SSA) and in Malawi and potential channels through which they could cause spillovers to smallholders. Theoretical and empirical observations on channels for spillovers are considered. On the empirical side we focus on previous country-level studies of Malawi and its neighbouring countries.[1]

2.1. The relevance of estates in Malawi over time

In this research we focus on a set of larger estates of at least 50 ha because these have the largest potential to generate spillovers. Also, Deininger and Xia (2018) in their study of consequences of transfer of land rights to small estates of 10–30 ha in the 1980s and 1990s conclude that many of these are non-operational or claims in official registers are overlapping. More than two thirds of the estates in our sample are tea estates. These estates where set up in favourable high-altitude areas, often already as early as the 1930s under the colonial regime (Chinigò, 2016). Chinigo summarises that while there were attempts to provide settlements and livelihoods for estate workers starting under colonial rule, this did only in rare occasions succeed and many estate workers and surrounding households remain smallholder farmers. This has also produced conflict and other adverse effects because in many occasions estates utilise the most productive land. Hence the focus of this paper on smallholder welfare and potential channels influencing it. While agronomic conditions are somewhat different for other crops (e.g. high water requirements for sugar cane), the methodology used in this paper attempts to identify comparable smallholders for smallholders close to each of these large estates.

2.2. Transmission channels for spillovers

The economic welfare of smallholders can be impacted through various channels. Here, we focus on the two central economic outcomes, namely income and assets. This section lays out which mechanism can potentially be at work when spillovers are transmitted. In addition, the analysis considers two important intermediary outcomes that are essential to the size and volatility of income and assets. First, technology spillovers are examined by examining productivity and diversity of crops. Second, the land holdings of agricultural households are compared to analyse whether any dynamics in this realm have the potential to influence farmers' economic outcomes. In the following we discuss these potential transmission channels, their theoretical foundations, empirical evidence and how we test for them in the data for Malawian smallholders and estates.

2.2.1. Income. Theoretical implications of proximity of smallholders to estates can be derived from Kleemann and Thiele (2015) who provide a theoretical model for large-scale agricultural investments. They explicitly model the labour market for smallholders supplemented with a model of smallholder yields subject to budget constraints, with analogous equations for estates. The model and its stylised facts foundations address the most important notions for discussion of rural welfare implications of large-scale investments. In the model estate establishments affect their surrounding smallholders through two major spillover channels: compensation payments and technology spillovers, which are subdivided into improvements in farming infrastructure and knowledge transfer. The former is explicitly modelled as lower input prices. Both would induce increased income from farming activities. Income then enters wealth of the farmers which is composed of both types of incomes plus compensations for displacement. The model predicts lowered prices of inputs and subsequently rising smallholder production and income. Overall effects on smallholder economic outcomes depend on several factors: the presence and characteristics of spillovers, the level of wages, and food prices. In conclusion, if the aforementioned characteristics are met, income is increasing.

Nolte and Ostermeier (2017) put forth an additional theoretical source of spillovers. In the case of labour creation, which can be assumed to significantly affect a number of outcomes such as income, assets and possibly productivity, spillovers might occur when labour is indirectly created through the operation of an estate but not the estate itself. The presence and size of spillovers at the same time depend on the crops grown by the investor, as well as their strategic intentions.

Empirical research has shown that indeed, the income channel is affected differently in different settings. Both positive and negative overall income effects of recent large-scale investments on smallholders have been found in the literature for Malawi and neighbouring countries (Herrmann, 2017; Herrmann & Grote, 2015; Osabuohien, Efobi, Herrmann, & Gitau, 2019). The overall effects were found to vary according to smallholder characteristics such as gender and according to the influence of estates on income components such as agricultural production. Wages were found to increase in some settings in estate proximity but not necessarily for individuals directly employed by them. An important lever to increase income in this respect has been found to be outgrower contracts (Herrmann, Jumbe, Bruentrup, & Osabuohien, 2018; Herrmann & Grote, 2015). Wage labour on estates on the contrary can be associated with lower income than producing export crops reducing income for those who shift from the former to the latter activity (Schuenemann, Thurlow, & Zeller, 2017). Overall however, empirical evidence from SSA suggest only very limited effects of estates on overall labour demand in a region (Ali et al., 2017, 2019), hence the importance of the wage labour channel is also limited.

The empirical section thus analyzes treatment effects on two types of income: wage income and total smallholder income. Wages could be influenced either directly or indirectly as outlined above. Total income is a composite of wages, agricultural income and any other type. Given the mixed previous evidence we rely on the cited theoretical model for the tested hypothesis, namely increased income of smallholders. As an indicator for past income, treatment effects for assets are analysed. While this method is a first attempt to look into estimates beyond the cross-sectional data, more research will be needed to quantify these effects in more detail.

2.2.2. Technology spillovers. An appealing feature of estate investments and foreign direct investments for policymakers in host and origin countries and other stakeholders is their potential for knowledge transfers. If these materialise from the estates to surrounding farmers technology could spread, thereby raising productivity.

On the macro side Dhahri and Omri (2020) found positive effects of FDI inflows on agricultural production in a panel of 50 developing countries. Ben Slimane, Huchet-Bourdon, and Zitouna (2016) however estimates zero effects of FDI in agriculture on agricultural production but instead pointed at a significant increase in food security from these agricultural investments in a sample of 55 developing countries. The drivers and the effects of large-scale farming investments have been shown to vary across countries in the region in a study in Ethiopia, Ghana and Tanzania by Cotula et al. (2014) calling for micro-level evidence. We therefore examine spillovers to yields which are an indicator for productivity of smallholders.

2.2.2.1. Yields. In case knowledge spillovers occur, these would entail the transfer of technology which can lead to increased productivity. For two of the three major crops included in the empirical analysis, groundnuts and pigeon peas, agricultural extension is sometimes delivered through or in collaboration with estates in Malawi (Tsusaka, Msere, Siambi, Mazvimavi, & Okori, 2016). We therefore hypothesise that especially in these crops productivity is higher for smallholders closer to estates. Increased productivity could in turn directly contribute to higher incomes if prices and cultivated areas do not fall.

Empirical evidence from neighbouring Mozambique suggests no short-run productivity spillovers (Deininger & Xia, 2016). In the longer run, these are however more likely to occur through observed increased access to improved inputs delivered by some estates (Glover & Jones, 2019). In Zambia, a recent study (Lay et al., 2021) finds significantly higher in maize yields for smallholders close to estates. A potential reason for this differential, however, could be not a rise in productivity but a less steep decrease in these areas as opposed to areas further away from estates as smallholders expand area for maize cultivation. Their results are in line with the proposition that land-rich households are found to be more often positively affected than land-poor (Herrmann, 2017; West & Haug, 2017).

The inverse relationship hypothesis (i.e. the negative relationship of farm size and productivity) has been confirmed for Malawian smallholders (Julien, Bravo-Ureta, & Rada, 2019). This implies limited possibility for smallholders with small field sizes to increase yields because no yield gap to be closed exists between smallholders and estates (Ali et al., 2019). The varying contributions of the different crops grown are important and are considered in our empirical analysis. Similarly productivity and poverty have been shown to be inversely related at a low magnitude in Malawian agriculture (Darko, Palacios-Lopez, Kilic, & Ricker-Gilbert, 2018). Regarding income and assets this would lead to an observable reduction in poverty in our empirical results only if large productivity increases for a significant number of crops is observed.

Given the dynamics and the studies on the channels influencing yields we test the hypothesis that yields are on average significantly higher in proximity to estates. For cash crops this could be more pronounced than for the Malawian staple crop maize where results might depend more strongly on cultivated area.

2.2.2.2. Crop diversity. If increased private extension via estates and subsequent productivity gains appear, these can lead to more knowledge of a wide range of crops and corresponding cultivation techniques. In the opposite direction depending on the impact of estates on smallholders' land availability, crop diversity could decrease if farmers shift from cash crops or vegetables to staple crops to maintain food security.

One of the few studies discussing shifts in the crop portfolio is Lay et al. (2021). The authors argue that resource allocation could be shifted from cultivation of other crops to maize. In line with the previous argument this would reduce crop diversity. Hence, the expected treatment effect for the impact of estate proximity on crop diversity is negative.

2.2.3. Area under cultivation. The final outcome under investigation, area under cultivation by smallholders is interrelated with all previously described channels. The hypothesis for total cultivated area is a lowered field size close to estate given the lower supply of land after accounting for land holdings by the estate. This relates directly to the debate on loss of access to land and land rights which was one of the most controversial themes in past research (Oberlack, Tejada, Messerli, Rist, & Giger, 2016).

Given the large size of estates under study, tenure security and availability of land with sufficiently high quality are two of the main driving factors of area cultivated by smallholders and potentially influenced by estate proximity. The establishment of large farms could negatively impact small-scale agriculture in Malawi because land rights are often insecure. The exclusion from land (rental) markets caused by the insecurity of land tenure tends to harm the most vulnerable land poor households (Deininger, Savastano, & Xia, 2017; Deininger, Xia, & Holden, 2019) which would reduce their cultivated area since they rent less land (Ali & Deininger, 2022). Given the nature of the registration of customary land encompassed in the 2016 Malawi Customary Act the detrimental effects of insecure land tenure could be reinforced as Zuka (2019) found. Additionally, the reduced perceived and possibly actual tenure insecurity in areas where estates are established leads to lower conservation efforts especially for soils which is likely to lead to lowered productivity in the long run (Lovo, 2016).

An important interaction between area, income and yields is via the wages and compensation payments. In the theoretical model of Kleemann and Thiele (2015), farmers give up their farms if compensation payments or wages exceed a minimum threshold. Importantly, this threshold is higher than the pure compensation payments for foregone farming profits if the land was given up. In addition, in cases where wages on estates go beyond a defined threshold it also becomes more profitable for the smallholder to shift to wage labour and cultivate less land.

3. Data

Reliable data often remain the most limiting factor in the study of agricultural investments in SSA. This study uses a novel way of identifying large agricultural estates by using data from OpenStreetMap (OSM) and combines it with one of the largest available surveys for individual farmers, the World Bank's Living Standard Measurement Surveys - Integrated Surveys on Agriculture (LSMS-ISA).

3.1. Smallholder farmer data

The LSMS-ISA is a large survey specifically targeted at smallholder farmers in Malawi which is implemented as part of the Integrated Household Survey (IHS) of the Malawi National Statistics Office with support from the World Bank and other international donors. For the purpose of studying the long-run effects of estates we used the cross-section survey data from the 2016/2017 growing season which includes 12,480 households. The resulting data is representative at district, regional and national level when used as a whole but cannot be guaranteed to remain representative when it is split for instances close and far away from estates. It includes geo-locations of all households interviewed with a maximum offset of 5 km which is important to consider when calculating the distance from estates.

All outcomes assessed in the empirical part of this paper are calculated from the LSMS-ISA sample. Harvests in kg are calculated by converting harvested volumes for each crop and each measurement unit into kg. Field sizes are calculated by transforming all other units into hectares for each crop. To compute the area for each crop, total area of a plot is multiplied with the share of that plot under the specific crop. Farm size or total area of the farm is the sum of fields for all crops. Yields are harvested quantities in kg divided by the hectarage for a specific crop. Wages are summed up over the whole year for daily, weekly or monthly payments. Total income in addition includes wage income, casual 'ganyu' labour income and value of agricultural production over the course of a whole year. Assets are measured in tropical livestock units (TLU) as is common in the region because most smallholders hold the majority of their assets in livestock (Sauer, Mason, Maredia, & Mofya-Mukuka, 2018; Smale & Mason, 2014). We supplement the analysis with measures for the number and concentration of crops grown. This includes a count of the number of different crops grown and a concentration index. It is calculated for each smallholder as a Shannon index.[2]

Inter-group differences in descriptive statistics (supplemental Tables A3 and A2 in appendix) in crop yields and other household level variables of close-by and distant to estate groups are not easily attributable only to the presence of the estate investor. These descriptive results have only limited meaning because estates are likely to be located in superior areas where smallholders could then not maintain their farming operations. Therefore outcomes need to be analyzed in a setup accounting for several control variables which is described in the next section.

3.2. Large-farm data

For this study we used a novel approach to create a database with the location of large estates based on information from the OpenStreetMap (OSM) project. OSM was founded in the United Kingdom in 2004 and aims to create a free, worldwide geographic dataset (www.openstreetmap.org). It mainly relies on data collected by volunteers, which are entered into the central database with support of specialised editors. To create the database, we processed raw OSM data and combined it with several other sources for cross-referencing and validation (detailed description in Annex).

We used OSM to identify the location and main crop of large farms in Malawi. This is a novel data source, that hitherto has not been used by other studies to analyse spillovers from large farms to smallholders. After processing, cross-referencing and validation, OSM data provide detailed and open information on the location of estates that resembles high-resolution land use maps in Google Earth. Such information is largely unavailable for most countries, which makes OSM a potential interesting source for the study of large farms.

Three different data sources were used to study the behaviour of large firms in the literature. First, an alternative open data source that resembles OSM is the Land Matrix (https://landmatrix.org), which also offers information on commercial farms. However, it mostly contains data on foreign direct investment and in most cases the geodata is limited to regions or centroids of investment sites.

A second source of information is nationally representative large-farmer surveys, in most cases organised by the national statistical agency. Such surveys do not exist for Malawi but were analyzed by Ali et al. (2019) for Ethiopia and by Deininger and Xia (2016) for Mozambique, and offer a very comprehensive and detailed picture of the large-farm sector in both countries. Nonetheless, a severe weakness of these datasets is that they are not in the public domain and therefore cannot be accessed and validated by most researchers.

Finally, official land registers are yet another source that were explored to investigate large-farm behaviour. Unfortunately, also this type of data is not publicly accessible, which makes it of limited use for scientific research. Moreover, several studies have questioned the reliability of land registers in developing countries. Deininger and Xia (2018) extensively discuss the use of agricultural estate data from official land registers in the case of Malawi but concluded that there was a substantial overlap in many of the registered land claims and that there was high uncertainty in regard to whether the land claims are actually used for farming.

OSM data offers an interesting fourth type of information that can be used to identify the location of large estates and assess spillovers from these estates to smallholders. The main advantage of OSM over several other sources is that access to it is open and free.

In line with comparable literature (Deininger & Xia, 2016; e.g. Glover & Jones, 2019) we assume estates to have a minimum size of 50 ha. This resulted in a database with in total 59 estates distributed over all three regions of Malawi (Figure 2). A large share (41) of the estates are tea estates followed by 3 maize, 5 sugarcane and 5 tobacco estates. A small number of estates produce coffee, soybeans and wheat. Most farms are in the Southern Region and include nearly all tea estates and some of the largest sugar growers. A much smaller number of farms are located in the Central and Northern regions, including a variety of crops. A comparison between satellite imagery from Google Earth and OSM polygon information shows strong resemblance. Figure 1 shows data for four selected estates that were listed in the LSMS-ISA survey as employer of smallholders or their relatives (see below). A comparison with harvested area from FAOSTAT indicates that the database covers all sugar estates and nearly all tea estates, the two crops for which production is dominated by large farms. For the other crops, which also involve smallholder production, we are unable to validate the coverage.

3.3. *Combining data on estates and smallholders*

The map in Figure 2 displays all estates and a 20 km buffer around them as well as all locations of smallholders available in the LSMS-ISA sample. It shows that many estates are clustered in certain areas which can be suspected to be especially suitable for agricultural production. Smallholders are distributed within and outside the proximity of estates and thus deliver enough possibilities for comparison of both groups. 36.8% of smallholders are within the 20 km range in the sample used for analysis.

With a mean size of 862 ha and the median at 385 ha, the estates mainly represent the businesses of large, partly owned by foreign companies. Additionally the sizes of the estates contribute to a homogeneous sample which is used to address the analysis of the interplay of estates with very large land holdings in the Malawian context to smallholders with small land holdings.

The distance thresholds up until which spillovers are considered is not a priori known. Most previous studies discussed in the literature section solve this by either setting a single threshold without further justification or varying the threshold and analysing the differences in the resulting estimates. We complement the variation approach by deriving a novel indicator for the

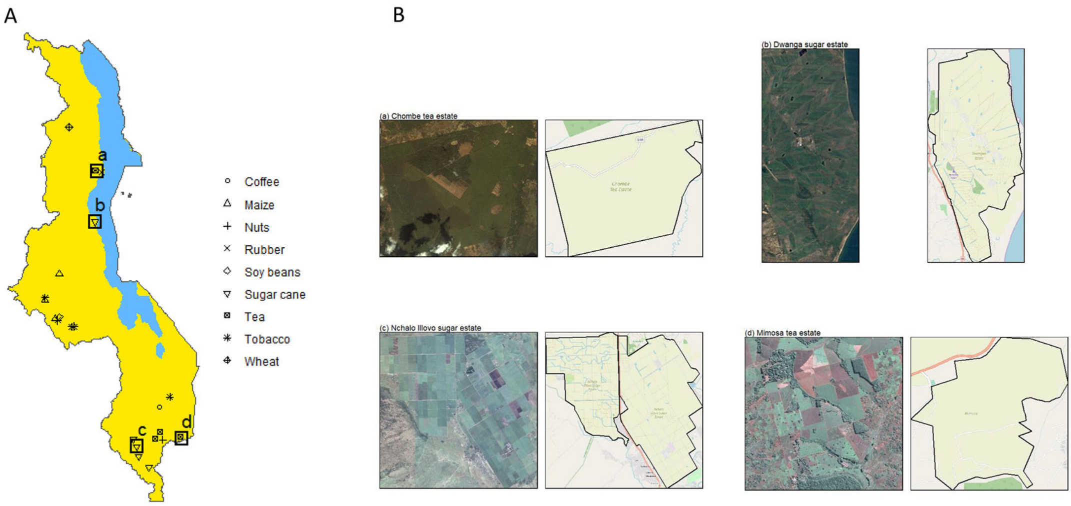

Figure 1. (A) location of large farms in Malawi and main crop based on OSM information and (B) comparison between Google Earth satellite images and OSM polygon for four selected estates that were listed in the LSMS-ISA survey as employer of smallholders or their relatives. Symbols to locate tea estates in (A) represent multiple clustered estates to improve visualisation. All Google Earth images are dated 28 October 2010 apart from Mimosa tea estate, which is dated 27 December 2009.

distance until which possible spillovers might occur from the distance of estate workers to the nearest estates. These are the group receiving wages from estates and possibly benefiting from their establishment through this wage channel. Workers on the estates can be identified in the LSMS-ISA data through the occupation and employer questions matched with the corresponding estate name in the OSM data. In cases where the estate name is not mentioned but workers are known to work on a tea or sugar estate they are matched to the nearest estate in our sample. Figure 3 shows the cumulative distribution of distances of the 167 estate workers in the sample. It shows that 90% of estate workers are within 20 km of the estate employing them. All cut-offs, which are considered in the empirical part, are shown as dashed lines in Figure 3. We use 5, 10 and 20 km thresholds to examine outcomes very proximate to estates and within the threshold found for estates workers. Additionally we analyse a 25-km threshold which to see how robust results from the 20 km cutoff are. The 50-km threshold is included as a comparison to previous studies (e.g. Ali et al., 2019) but is potentially too large in the Malawian context as evidenced by the distance of estate workers.

A limitation of the LSMS-ISA data is the offset in the geolocation of smallholders which is introduced to ensure confidentiality. To ensure that there is no overlap between the nearby and distant smallholders we use distance thresholds around estates to identify those nearby, exclude all farmers in between the distance threshold and 10 km from the threshold and only use those 10 km plus the distance threshold away from the estate as control group distant from the estate.

4. Econometric methods

Yields and individual economic outcomes were observed to vary strongly in-between nearby and distant groups for some crops but with varying directions in descriptive statistics. But all of these variables need not necessarily differ as a result of the proximity to estates. Another possible explanation could be the self-selection of estates into superior regions for example in terms of water availability or lower productivity of smallholders. Scholars have identified resource availability as the most important driver (Schoneveld, 2014). For example Kareem (2018) indicates positive effects of the availability of arable land and negative effects of the level of cereal

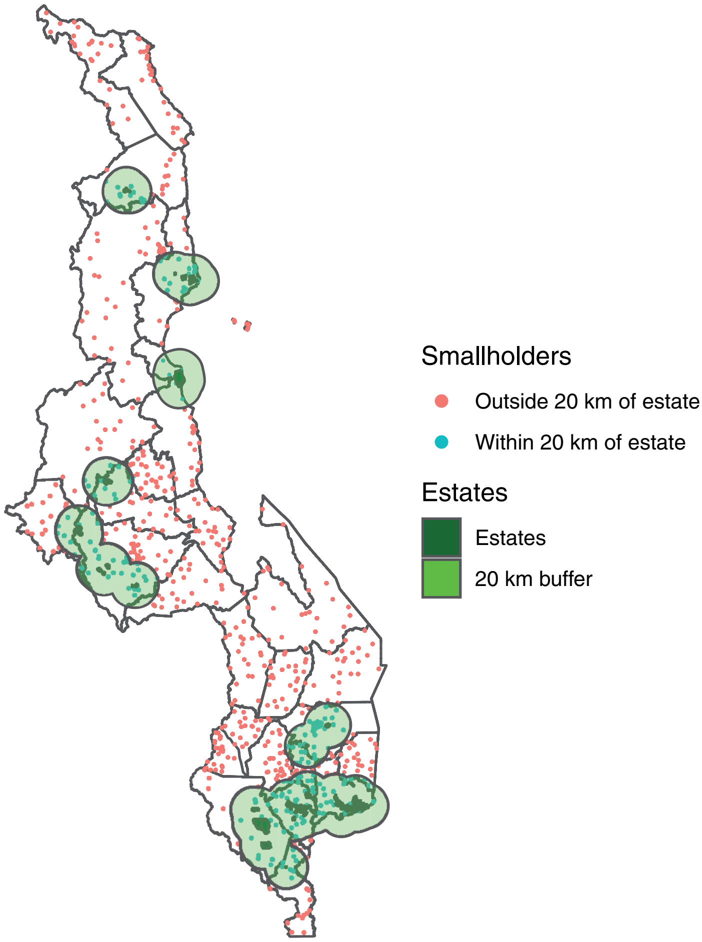

Figure 2. Locations of smallholder farmers and estates including 20 km buffers around estates.

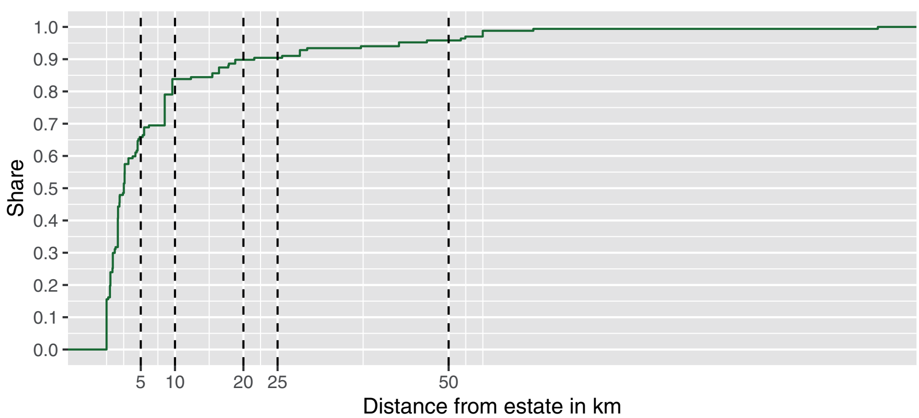

Figure 3. Empirical cumulative distribution function (CDF) of distance of estate workers from estates in km and distance thresholds at 5, 10, 20, 25 and 50 km.

yields on the size of investments per African country. Within countries estates are in areas with superior agroecological conditions (Julien et al., 2019). Availability of water has been found to be a major important driving factor in location choice for large-scale land investments in agriculture in SSA (Hirsch, Krisztin, & See, 2020; Lay & Nolte, 2018; Mazzocchi, Salvan, Orsi, &

Sali, 2018). At the same time lower variability in rainfall attracts migration of farmers from areas with high variability in rainfall and high probabilities of droughts (Lewin, Fisher, & Weber, 2012). Finally, variations in soil quality and elevation are important for decision making of estate owners when setting up their location and for smallholder farming simultaneously.

Therefore, suitable methods of analysis take this possibility of two-way causality into account. The use of counterfactuals is one suitable approach to do so. It identifies comparable smallholders within and outside a buffer around estates to estimate the outcomes a farmer close to an estate would have had if the estate had not been there. Within this framework the variables identified as drivers in the literature inform the selection for our matching and weighting. After matching covariates should be balanced across treatment and control groups.

In this study, we use a variant of propensity score matching (PSM) to identify the counterfactual (Rosenbaum & Rubin, 1983) as well as Inverse Probability Weighted (IPW) regressions (Imbens & Wooldridge, 2009). Similar matching techniques have been applied among others by Osabuohien et al. (2019) and Herrmann (2017). The IPW method is comparable to the methods used by Glover and Jones (2019). The comparison of the results of both techniques is useful to identify the dependence of estimated effects on the chosen method.

The propensity score can be used to match individual instances on several continuous variables and is commonly estimated using a logistic regression (Caliendo & Kopeinig, 2008). The propensity score then reflects the estimated propensity of an estate locating in a certain area given a control vector. Matching is conducted on the fitted values for the propensity p of proximity of an estate to a smallholder. I.e. calculated scores are used to match a smallholder close to an estate to the distant smallholder(s) which have the most similar propensity score. We use Nearest Neighbour Matching with a calliper. The introduction of a calliper of one standard deviation of the propensity score for matching farmers ensures that farmers who are too different are not matched to each other. This is ensured by dropping all instances where the matched nearest neighbors have a difference in propensity scores larger than one standard deviation. The calliper can alternatively be interpreted as a tool to establish Common Support of the propensity score. Where Common Support is not strictly met in this case, meaning that close by and distant farmers can have propensity scores at ends of the distribution which are not matched in the other group but distances between groups are not allowed to grow beyond the caliper threshold. Instances of close-to-estates locations households for whom this is not possible are dropped from the corresponding analysis.

The constructed matched dataset is then used to calculate the Average Treatment Effects on the Treated (ATT):

$$ATT = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] \tag{1}$$

where $Y_i(1)$ is the value of an outcome variable for a farmer if he is close to an estate, $Y_i(0)$ is the outcome for distant smallholders and D_i is the actually realised treatment (i.e. whether a smallholder is close to an estate). Hence in the equation first term on the right-hand side of the equation is the observed outcome of a nearby farmer and the second expectation is the value of that outcome he would have realised had he been far away from any estate. The last term is replaced by the counterfactual we create through matching; meaning a distant farming household in our sample which best resembles the estate-neighbouring smallholder. This term is additionally bias adjusted by replacing the outcome $Y_i(0)$ for distant smallholder i with the regression adjusted term

$$\widehat{Y}_i(0) = \alpha + \beta_c' X_i(0) + \epsilon_i \tag{2}$$

where the $X_i(0)$ are an additional set of control variables for the distant smallholder, β_c' are the regression coefficients. (Abadie & Imbens, 2006, 2011). So, the ATT formula using the fitted values $\widehat{Y}_i(0)$ from the above regression becomes:

$$ATT = E[Y_i(1)|D_i = 1] - E\left[\widehat{Y}_i(0)|D_i = 1\right] \tag{3}$$

We report the results for one-to-three matching, meaning that each smallholder in the threshold distance around an estate is paired with the three individuals distant from the estate. This is especially useful for the very large or very small distance thresholds (5 or 50 km) where group sizes are much smaller for one of the groups (Caliendo & Kopeinig, 2008; Smith, 1997). Matches were excluded from the analysis if there were less than 100 nearby smallholders growing a specific crop within the defined distance threshold. Finally, standard errors are calculated according to Abadie and Imbens (2006).

A second approach to estimating ATTs is Inverse Probability Weighting. Here the propensity score is used to generate weights for regressions. The regressions without weights have the form:

$$Y = \beta_0 + \beta_1 D + \beta_2 X + \epsilon \tag{4}$$

where Y are vectors of the respective outcome variables, β_0 is a constant term, D is a dummy vector for estate proximity with parameter β_1 and X is a matrix containing a set of vectors of control variables with parameters β_2. We then introduce weights for the regression to compute β_1 as an estimate for the ATT. The individual weights w_i are:

$$w_i = \begin{cases} 1 & \text{if } D_i = 1 \\ \dfrac{\widehat{p}_i}{1 - \widehat{p}_i} & \text{if } D_i = 0 \end{cases} \tag{5}$$

where the $\widehat{p}_i$ is the estimated propensity which is used to calculate the weights w_i and employ them for all individuals distant from estates i.e. with $D_i = 0$. I.e. the regression is weighted by a vector w with element i equal to w_i.

Using weights in a regression instead of matching on the propensity score can be interpreted as heavier punishing for control (distant) cases with low propensity scores. If both specifications produce statistically significant results, those are strong indicators for overall significant effects.

Both methods are applied to a range of dependent variables separately and estimated for each of five varying distance thresholds. First, we estimate all models for variables which need to be separated by crop which are yields, harvests and area the crop is grown on. Then income, assets and crop diversity variables which are aggregated at household level are used as outcome variables in the latter model specifications for both methods.

The number of treatment effects estimated becomes very large given the number of outcome variables and distance thresholds. To account for the possibility that significant effects show up by chance when testing the resulting hypothesis across specifications, the analysis is extended by adjusting p-values for false discovery rates (FDR). For FDR control we use the standard method of Benjamini and Hochberg (1995) which adjusts p-values for the number of tested hypothesis. This method orders and reweights p-values to adjust them for the expected proportion of hypothesis rejections which are type I errors (Anderson, 2008).

5. Empirical results

For matching methods it is important to assess the quality of the matches. Meaning the results of the matching need to show similar characteristics of the variables matched on in both control and treatment groups. These results are described below, followed by the presentation of results for both matching and inverse probability weighted regression estimates.

5.1. Propensity score estimation and balance

The estimation of the propensity score is based on a rich set of features available from the LSMS-ISA survey data. Established factors in literature that affect productivity of Malawian farms are Human Capital, quality of input markets, climatic and agroecological conditions as well as infrastructure (Julien et al., 2019). Recent research has confirmed these factors and extended them to other welfare outcomes such as income and assets. For example Bezu, Kassie, Shiferaw, and Ricker-Gilbert (2014) show the dependence of income and assets on deviations in rainfall, household size and farm size. These variables are captured in the LSMS-ISA sample and introduced into the econometric framework.

Controls include: average 12-months rainfall between 2001 and 2016, several average temperature indicators, precipitation averages for the wettest quarter and month, population density, elevation, several indicators for soil quality and distances to the next road, population centre, ADMARC (Agricultural Development and Marketing Corporation), auction, boma (district government office), border post and agricultural market. Table shows results of an example regression for setting the distance threshold at 20 km from which the fitted values of the propensity score are used for matching. The limiting assumption when using the propensity score for matching is the selection on observables which must be assumed to make causal inference. It has been argued that the most credible way to achieve this is to carefully choose the variables included according to theory as reflected here in the drivers of investments section of the literature review. Additionally, the consensus among researchers is to include as many variables as possible and needed to explain selection, in this case of investment locations (Caliendo & Kopeinig, 2008).

For assessing the quality of matches, the most important feature is the reduction in deviance of controls in treatment and control groups, meaning close-by and distant smallholders. Figure 4 shows the reduction of standardised mean differences of matching variables at different distance thresholds from estates. All included variables achieve significant reductions in standardised mean differences after matching. Matching variables are varied in the analysis to avoid identifying effects which occur only at certain predefined thresholds.

To control for variation in smallholder outcomes which do not influence the location decisions of estates, additional regression bias adjustments are introduced into the calculation of the ATT using the matched sample as described above. The variables included vary depending on the outcome variable under examination. The full set includes: number of children, old and adults in the household, gender, age and education of household head, days of hired labour, dummies for ploughing, inter-cropping and irrigation, quantities of organic and inorganic fertiliser as well as herbicide used, total farm size and number of coupons received from government or other sources for agricultural inputs. Likewise, the same sets are part of the corresponding IPW regressions as covariates.

5.2. Spillover effects from estates to smallholders

All hypotheses for potential spillover channels discussed in Section 2 are tested empirically. All household level outcomes are summarised as ATTs in percent for both estimation methods and at all distance thresholds in Figure 5. Treatment effects for outcome indicators which are further subdivided into three major crops are reported in Figure 6.

First, income is used as an outcome variable in the econometric model. Wage income is significantly higher in both PSM and IPW specifications but only in close proximity to the estate. At any distance threshold beyond 5 km the large effects of a magnitude of 67% (IPW) loose

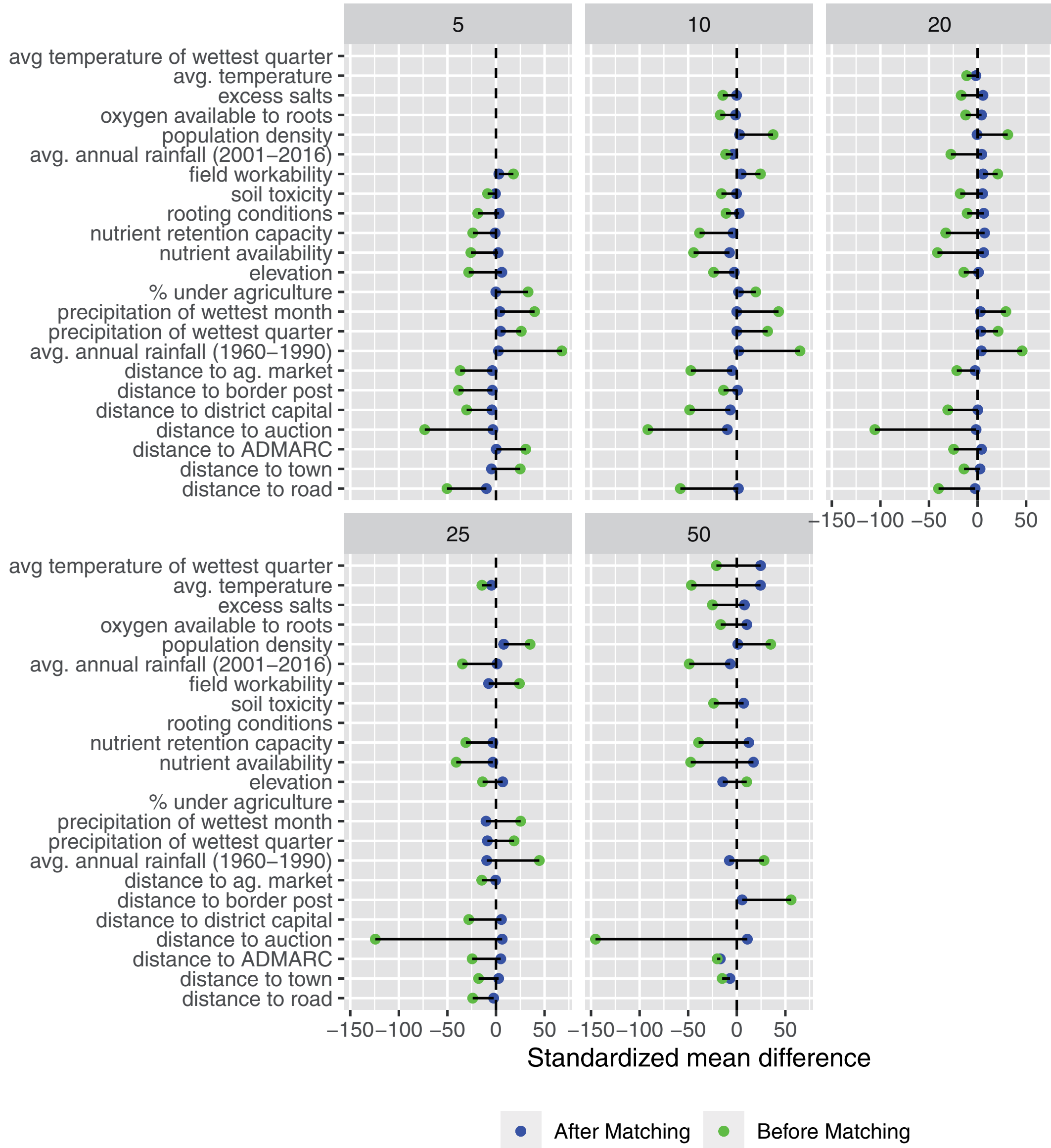

Figure 4. Standardised differences in means of control variables before and after matching at varying distance thresholds (5–50 km ranges around estates).

their significance. This also translates to a significant increase in total income but at a lower magnitude, which is feasible given that wage income is commonly only one component of total income besides farming and other household income sources. In the long run, income increases cannot be observed to also go along with higher assets. Except for one outlier, all specifications here show insignificant treatment effects for proximity to estates.

Splitting the sample of estates into those growing cash crops and those growing staple crops (maize) is predicted to increase wage incomes for smallholders nearby cash crop estates by Kleemann and Thiele (2015) in their theoretical model. Additional empirical evidence discussed in the literature review section suggest varying impacts for sugarcane and tea estates. In our sample there is very little common support of the propensity scores of the distant and nearby smallholder groups in the split sample. This is caused by substantial differences between the

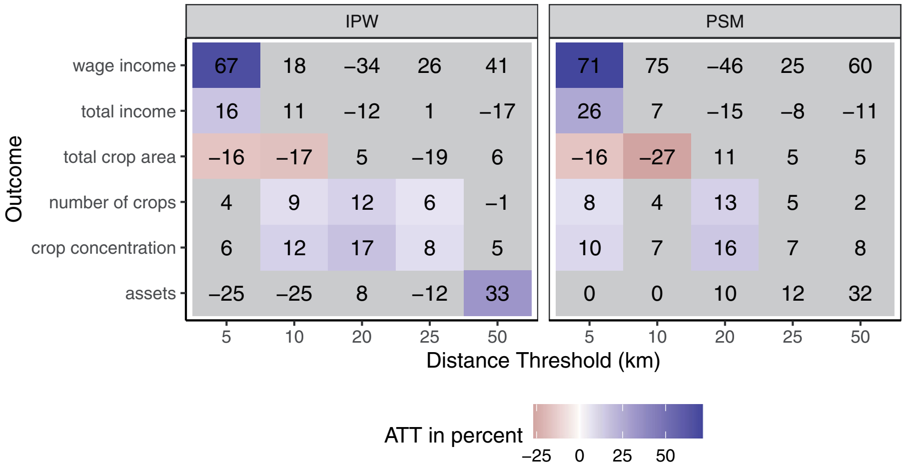

Figure 5. Heatmap of PSM and IPW ATTs of proximity to estates for dependent variables at household level in percent at varying distance thresholds.
Note. Statistically significant effects are coloured in blue or red, insignificant ($p > 0.05$) effects have grey background.

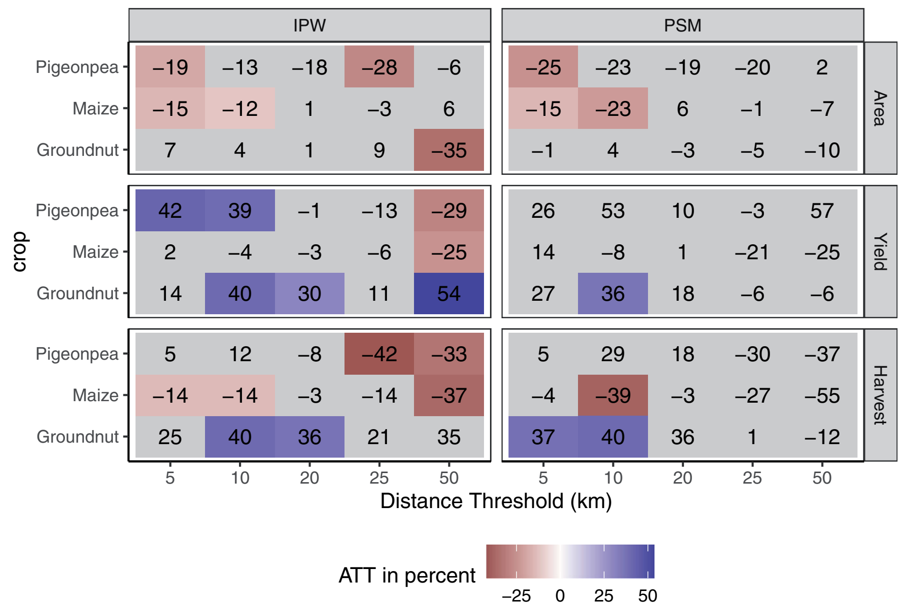

Figure 6. Heatmap of PSM and IPW ATTs of proximity to estate per crop in percent at varying distance thresholds for dependent variables field size, yields and harvests.
Note. Statistically significant effects are coloured in blue or red, insignificant ($p > 0.05$) effects have grey background.

groups in many of the variables used to construct the scores. Therefore it is not possible to match sufficiently high numbers of farmers in this setting which means we cannot obtain results to test the hypothesis of stronger income effects for certain estate crops.[3]

To examine spillovers in agricultural productivity yields for three major crops are used as outcomes. For maize with the exception of a single outlier, no evidence is found for any significant spillovers. For Pigeon pea and groundnut there is some indication for increased yields in proximity to estates, especially in a range of up to 10 km. These statistically significant increases range at magnitudes of 30–40%. For groundnut IPW results are confirmed by PSM while this is not the case for pigeon pea. Examination of the two components of yields, area and harvests shows which of the two drives the results for each crop. Areas are found to be significantly lower for pigeon pea up to 5 km from estates but not for groundnut. Combined with no significant increase in harvested quantities for pigeon pea, increased yields are a result of similar harvests but using less land. The contrary is true for groundnut. Here, yields increase because more is harvested on areas that are not significantly different. The area and harvest results show that while yields do not differ significantly, smallholders in ranges up to 10 km of estates harvest significantly less maize on significantly less land than their distant counterparts.

Knowledge spillovers could also occur if smallholders close to estates learn about or have access to a wider variety of crops. We therefore test whether the number of crops grown and crop concentration differ. Both variables exhibit small but significant increases of 8–13% in the number of crops grown and 8–17% in crop concentration. These results are persistent across several specifications of the distance threshold and estimation method.

Finally, area under cultivation is an important measure for the availability of land for smallholders and is also directly linked to agricultural production. The estimates for ATTs of total crop area under cultivation show a reduction of 16–17% (IPW, higher for PSM) for smallholders residing within 10 km distance from an estate. This is in line with the results for a reduced area of the main crop for most farmers which is maize.

6. Discussion

Here, we first discuss empirical results and link them to previously outlined channels for spillovers. Second, we discuss the novel use of OSM data for spillover analysis in terms of its advantages and disadvantages for this study and future studies.

6.1. *Observed transmission channels for spillovers*

For income, the hypothesis of increases in wages and as a result total income is found to hold when households are within a range of 5 km to estates. This reiterates the findings from the analysis of distance of estate workers to the next estate. In our sample the median distance of worker location to the estate polygon is only 2.52 km and two-thirds are within a distance of 5.09 km. These results can help explain why previous literature such as Ali et al. (2019) or Glover and Jones (2019) did not find positive wage employment or wage income effects of agricultural investment sites. The occurrence of these spillovers is limited to close geographical proximity. Even for households close enough to estates wage employment only benefits a subset of smallholders who are employed on the farm. The much lower increase in total income than wage income is a result of the fact that total income is composed of several other components beyond wage income. Income generation, in line with previous studies therefore is only a benefit of estates for a small fraction of smallholders.

One reason income could potentially increase is because agricultural technology spills over from a large farm to surrounding small farms. This could increase yields and therefore harvests and income from agriculture. By using the yields of three major crops as outcome indicators we tested whether technology improvements materialise as increased productivity. Higher yields observed for groundnut are driven by larger harvests on not significantly larger fields. A possible explanation is offered by better access to extension services sometimes delivered privately (Tsusaka et al., 2016). This is because the lack of extension services and access to improved inputs are major limiting factors for productivity in groundnut production in Malawi (Owusu & Bravo-Ureta, 2022). If estates cooperate with input dealers this could also lead to higher availability of improved seeds which are more productive. This explanation is also in line with the arguments put forward by Kleemann and Thiele (2015) who theoretically derive a boost in agricultural production via reduced prices of higher inputs.

In the case of maize, in contrast to a recent analysis in Zambia (Lay et al., 2021), we do not observe significant treatment effects of proximity to estates on yields. However, both components area and harvests are somewhat lower. A reason for limited yield improvements could be the lack of a yield gap which Ali et al. (2019) argue to hinder productivity spillovers in the case of maize in Ethiopia. But given the focus of the analysed estates on cash crops this explanation only has limited explanatory power. It is more likely that this mismatch in estate and smallholder crops which was not the case in the sample of Lay et al. (2021) limits the potential for any spillovers. Given the differential effects across crops the hypothesis of technological spillovers can only be confirmed partly.

Moderate positive ATTs are significant for the number of crops grown and crop concentration. This also hints toward the explanations found for increases in groundnut and pigeon pea cultivation, namely better access to extension services and inputs. If these are more readily available farmers can diversify their crops. An important area for future research is to investigate the effect of higher crop diversity on income volatility and resilience in the context of spillovers from estates. While crop diversification commonly has a positive effect on income (Pellegrini & Tasciotti, 2014), the effects on its volatility are less researched, especially in the spillover context.

The hypothesis of reduced land holding sizes by smallholders is confirmed in our analysis. This could be the case because many of the estates are located in hilly or mountainous areas or next to lakes where expansion of smallholder land is not easily possible. Three other factors could be relevant: a shift of labour from agriculture to employment (on estates), out-migration from the treatment site or productivity spillovers leading to lower land requirements (Williams et al., 2021). While we cannot test for out-migration with the data used here, there is some evidence for both spillovers and more wage labour. Thus, both latter channels are likely contributors to the lower area for crop cultivation associated with proximity to estates.

In summary, household income is higher in close proximity to estates. Drivers for this result are productivity spillovers, shifts of labour and different crops grown. In closest proximity matched smallholders work significantly more for wages than distant counterfactual households.

An additional result of the empirical part is the dependence of many of the estimated effects and significance on the distance threshold. Therefore we argue that it is important to not arbitrarily set a single distance threshold but instead use other methods such as estimating treatment effects with a variety of cut-offs. FDR adjustment can help mitigate the problem of multiple testing in this case.

6.2. Potential and drawbacks of the OSM database

OSM data are available at the global level so it can potentially also be used to assess spillovers in other countries. Given the various unique advantages and its freely accessible nature outlined in the data section above, the research presented here demonstrates its previously unexploited potential for research in this field.

However, at the moment it is unclear if OSM data for other (African) countries contains relevant information on large farmers. To evaluate the use of OSM as a structural source of information on large commercial farms more research is needed. One way forward would be to conduct a systematic comparison between OSM-based database of large estates and the data in a nationally representative large farm surveys and national land registers.

A potential drawback of the OSM database is that it is difficult to assess its coverage of large farms in a country. If crop production is dominated by large farms (e.g. sugarcane and tea), a comparison with secondary sources, such as FAOSTAT, offers a solution (see additional OSM

Appendix). However, we were not able to evaluate the coverage for sectors that also feature a substantial share of smallholders. If our database does not include all, or the majority of large farms, this might bias the results, perhaps partly explaining why we find mixed or insignificant results.

Another limitation of the OSM data, which also holds for other large-scale farm data sources, is the lack of information on the capital inflow into each of the investments which can be an important proxy for the activity on a certain site. This would allow for further disaggregation of effects uncovered in this paper. The focus on land-size of investments has been criticised for example by Schoneveld (2014) in favor of capital used at each estate. The focus on large investments this study, however comes with two large advantages. First, the estates are not too heterogeneous as all of them are known to be of large size above 50 ha. Our data can secondly help mitigate some of the problems other datasets face in terms of limitation of the estates to those established in a certain time period. Time since investment in other studies, which use short panels of up to five years is often considered too short to unfold most of their potential effects (Glover & Jones, 2019). In contrast in the OSM data estates might not be limited backwards regarding their date of establishment – however, this remains speculative since dates are not captured in the OSM data.

7. Conclusion

In this paper we analyse the differences in economic outcomes of smallholder farmers closeby and distant from large estates using a data source (OSM) previously unexplored in this context. These are related to a key set of potential spillover channels. We find significant treatment effects for (wage) income in closest proximity, yields of groundnut and pigeon pea, crop diversity and cultivated area.

The OpenStreetMap data are validated and captures many of the major cash crop estates, especially tea and sugar estates in Malawi. While it is not a universal tool for the analysis of spillovers, it is a promising source for future studies requiring locations of estates and examining spillovers. The advantage of the dataset is that it provides polygons instead of only centroids and almost all polygons also list the crop grown on the estate. As in most datasets validation for example by remote sensing as proposed by Williams et al. (2021) is an important part of any research involving such data.

The channels found to be most important for long-run spillovers from estates to surrounding households are technology, wage employment and pressure on land. On the beneficial side, some individuals who live in closest proximity (<5 km) away from estates experience higher wages than distant counterfactual households. Differences associated with knowledge spillovers are observed within even larger distance thresholds. Here, yields of two cash crops promoted by some agricultural investors (groundnut and pigeon pea) are higher. The positive treatment effect in land allocated to groundnut cultivation and other crops also leads to higher crop diversity. These results point towards positive effects for at least a subset of smallholders and indicate at least some knowledge spillovers and wage generation from estates.

On the downside however, the analysis confirms the pressure on land hypothesis leading to less cultivated area in general and less area cultivated with the main food crop maize. This is especially important since these effects are observable even beyond the threshold until which wage spillovers occur. Therefore, households who are either too far away or for other reasons do not benefit from income increases associated with estates could be left with less land to cultivate crops for own consumption. This is especially important because production of maize, which constitutes a large share of the diet of rural Malawian households is significantly lower in estate proximity. These results must be viewed in combination with the notion that outgrowers and employees on estates benefit from estates while other smallholder households do not (Herrmann & Grote, 2015).

In summary, the econometric analysis of OSM and survey data shows potential beneficial and adverse spillover effects from estates to surrounding small-scale farmers. Any policy and research focusing on this relationship should carefully evaluate which households are subject to which types of spillovers. This includes compensation or other types of support for households negatively affected by the estate. OSM data can be an important tool for identification of such estates.

Notes

1. A tabular overview including additional studies of a land reform in Malawi (Mendola & Simtowe, 2015) and a mixed methods study on effects of large-scale agricultural investments on smallholder in Kenya (Zaehringer et al., 2018) is provided in supplemental Table A1.
2. A standard definition of a Shannon index (S) is:

$$S = -\sum_{j=1}^{n}\left[\left(\frac{a_j}{\sum_{j'=1}^{n} a_{j'}}\right) * \ln\left(\frac{a_j}{\sum_{j'=1}^{n} a_{j'}}\right)\right]$$

   where a_j is the area used for production of crop j (with $a_j, a_{j'} = 1, \ldots, n$)).
3. Supplemental Tables S16 and S17 Show the divergence in explanatory variables for the propensity score construction and the logit model outcomes for its estimation. Supplemental Figure S2 depicts a histogram of estimated propensity scores separate by group, indicating that there is very little overlap i.e. common support between them.

Acknowledgements

Code for estimation of treatment effects, figures and tables in this article will be made available on the authors' GitHub pages.

ORCID

Julian Joseph http://orcid.org/0000-0002-3844-7807
Michiel van Dijk http://orcid.org/0000-0002-5207-7304
Tamás Krisztin http://orcid.org/0000-0002-9241-8628

References

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica: Journal of the Econometric Society*, *74*(1), 235–267. doi:10.1111/j.1468-0262.2006.00655.x

Abadie, A., & Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics*, *29*(1), 1–11. doi:10.1198/jbes.2009.07333

Ahlerup, P., & Tengstam, S. (2015). *Do the land-poor gain from agricultural investments? Empirical evidence from Zambia using panel data* (Working Papers in Economics No. 624). University of Gothenburg.

Ali, D. A., & Deininger, K. (2022). Institutional determinants of large land-based investments' performance in Zambia: Does title enhance productivity and structural transformation? *World Development*, *157*, 105932. doi:10.1016/j.worlddev.2022.105932

Ali, D. A., Deininger, K., & Harris, A. (2017). Using national statistics to increase transparency of large land acquisition: Evidence from Ethiopia. *World Development*, *93*, 62–74. doi:10.1016/j.worlddev.2016.12.027

Ali, D. A., Deininger, K., & Harris, A. (2019). Does large farm establishment create benefits for neighboring smallholders? Evidence from Ethiopia. *Land Economics*, *95*(1), 71–90. doi:10.3368/le.95.1.71

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry preschool, and early training projects. *Journal of the American Statistical Association*, *103*(484), 1481–1495. doi:10.1198/016214508000000841

Ben Slimane, M., Huchet-Bourdon, M., & Zitouna, H. (2016). The role of sectoral FDI in promoting agricultural production and improving food security. *International Economics*, *145*, 50–65. doi:10.1016/j.inteco.2015.06.001

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bezu, S., Kassie, G. T., Shiferaw, B., & Ricker-Gilbert, J. (2014). Impact of improved maize adoption on welfare of farm households in Malawi: A panel data analysis. *World Development*, *59*, 120–131. doi:10.1016/j.worlddev.2014.01.023

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72. doi:10.1111/j.1467-6419.2007.00527.x

Chinigò, D. (2016). Rural radicalism and the historical land conflict in the Malawian tea economy. *Journal of Southern African Studies*, *42*(2), 283–297. doi:10.1080/03057070.2016.1141514

Cotula, L., Oya, C., Codjoe, E. A., Eid, A., Kakraba-Ampeh, M., Keeley, J., … Rizzo, M. (2014). Testing claims about large land deals in Africa: Findings from a multi-country study. *The Journal of Development Studies*, *50*(7), 903–925. doi:10.1080/00220388.2014.901501

Darko, F. A., Palacios-Lopez, A., Kilic, T., & Ricker-Gilbert, J. (2018). Micro-level welfare impacts of agricultural productivity: Evidence from rural Malawi. *The Journal of Development Studies*, *54*(5), 915–932. doi:10.1080/00220388.2018.1430771

De Schutter, O. (2011). How not to think of land-grabbing: Three critiques of large-scale investments in farmland. *Journal of Peasant Studies*, *38*(2), 249–279. doi:10.1080/03066150.2011.559008

Deininger, K., & Byerlee, D. (2011). *Rising global interest in farmland: Can it yield sustainable and equitable benefits?* The World Bank.

Deininger, K., Savastano, S., & Xia, F. (2017). Smallholders' land access in Sub-Saharan Africa: A new landscape? *Food Policy*, *67*, 78–92. doi:10.1016/j.foodpol.2016.09.012

Deininger, K., & Xia, F. (2016). Quantifying spillover effects from large land-based investment: The case of Mozambique. *World Development*, *87*, 227–241. doi:10.1016/j.worlddev.2016.06.016

Deininger, K., & Xia, F. (2018). Assessing the long-term performance of large-scale land transfers: Challenges and opportunities in Malawi's estate sector. *World Development*, *104*, 281–296. doi:10.1016/j.worlddev.2017.11.025

Deininger, K., Xia, F., & Holden, S. (2019). Gendered incidence and impacts of tenure insecurity on agricultural performance in Malawi's customary tenure system. *The Journal of Development Studies*, *55*(4), 597–619. doi:10.1080/00220388.2018.1425395

Dhahri, S., & Omri, A. (2020). Does foreign capital really matter for the host country agricultural production? Evidence from developing countries. *Review of World Economics*, *156*(1), 153–181. doi:10.1007/s10290-019-00361-2

Glover, S., & Jones, S. (2019). Can commercial farming promote rural dynamism in sub-Saharan Africa? Evidence from Mozambique. *World Development*, *114*, 110–121. doi:10.1016/j.worlddev.2018.09.029

Herrmann, R. (2017). Large-scale agricultural investments and smallholder welfare: A comparison of wage labor and outgrower channels in Tanzania. *World Development*, *90*, 294–310. doi:10.1016/j.worlddev.2016.10.007

Herrmann, R., & Grote, U. (2015). Large-scale agro-industrial investments and rural poverty: Evidence from sugarcane in Malawi. *Journal of African Economies*, *24*(5), 645–676. doi:10.1093/jae/ejv015

Herrmann, R., Jumbe, C., Bruentrup, M., & Osabuohien, E. (2018). Competition between biofuel feedstock and food production: Empirical evidence from sugarcane outgrower settings in Malawi. *Biomass and Bioenergy*, *114*, 100–111. doi:10.1016/j.biombioe.2017.09.002

Hirsch, C., Krisztin, T., & See, L. (2020). Water resources as determinants for foreign direct investments in land - a gravity analysis of foreign land acquisitions. *Ecological Economics*, *170*, 106516. doi:10.1016/j.ecolecon.2019.106516

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86. doi:10.1257/jel.47.1.5

Julien, J. C., Bravo-Ureta, B. E., & Rada, N. E. (2019). Assessing farm performance by size in Malawi, Tanzania, and Uganda. *Food Policy.*, *84*, 153–164. doi:10.1016/j.foodpol.2018.03.016

Kareem, O. I. (2018). The determinants of large-scale land investments in Africa. *Land Use Policy*, *75*, 180–190. doi:10.1016/j.landusepol.2018.03.039

Kleemann, L., & Thiele, R. (2015). Rural welfare implications of large-scale land acquisitions in Africa: A theoretical framework. *Economic Modelling*, *51*, 269–279. doi:10.1016/j.econmod.2015.08.016

Lay, J., & Nolte, K. (2018). Determinants of foreign land acquisitions in low- and middle-income countries. *Journal of Economic Geography*, *18*(1), 59–86. doi:10.1093/jeg/lbx011

Lay, J., Nolte, K., & Sipangule, K. (2021). Large-scale farms in Zambia: Locational patterns and spillovers to smallholder agriculture. *World Development*, *140*, 105277. doi:10.1016/j.worlddev.2020.105277

Lewin, P. A., Fisher, M., & Weber, B. (2012). Do rainfall conditions push or pull rural migrants: Evidence from Malawi. *Agricultural Economics*, *43*(2), 191–204. doi:10.1111/j.1574-0862.2011.00576.x

Lovo, S. (2016). Tenure insecurity and investment in soil conservation. Evidence from Malawi. *World Development*, *78*, 219–229. doi:10.1016/j.worlddev.2015.10.023

Mazzocchi, C., Salvan, M., Orsi, L., & Sali, G. (2018). The determinants of large-scale land acquisitions (LSLAs) in Sub-Saharan Africa (SSA): A case study. *Agriculture*, *8*(12), 194. doi:10.3390/agriculture8120194

Mendola, M., & Simtowe, F. (2015). The welfare impact of land redistribution: Evidence from a quasi-experimental initiative in Malawi. *World Development*, *72*, 53–69. doi:10.1016/j.worlddev.2015.02.010

Nolte, K., & Ostermeier, M. (2017). Labour market effects of large-scale agricultural investment: Conceptual considerations and estimated employment effects. *World Development*, *98*, 430–446. doi:10.1016/j.worlddev.2017.05.012

Oberlack, C., Tejada, L., Messerli, P., Rist, S., & Giger, M. (2016). Sustainable livelihoods in the global land rush? Archetypes of livelihood vulnerability and sustainability potentials. *Global Environmental Change*, *41*, 153–171. doi:10.1016/j.gloenvcha.2016.10.001

Osabuohien, E. S., Efobi, U. R., Herrmann, R., & Gitau, C. M. W. (2019). Female labor outcomes and large-scale agricultural land investments: Macro-micro evidencefrom Tanzania. *Land Use Policy*, *82*, 716–728. doi:10.1016/j.landusepol.2019.01.005

Owusu, E. S., & Bravo-Ureta, B. E. (2022). Gender and productivity differentials in smallholder groundnut farming in Malawi: Accounting for technology differences. *The Journal of Development Studies*, *58*(5), 989–1013. doi:10.1080/00220388.2021.2008364

Pellegrini, L., & Tasciotti, L. (2014). Crop diversification, dietary diversity and agricultural income: Empirical evidence from eight developing countries. *Canadian Journal of Development. Studies/Revue Canadienne d'études Du Développement*, *35*(2), 211–227. doi:10.1080/02255189.2014.898580

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrica*, *70*, 41.

Sauer, C. M., Mason, N. M., Maredia, M. K., & Mofya-Mukuka, R. (2018). Does adopting legume-based cropping practices improve the food security of small-scale farm households? Panel survey evidence from Zambia. *Food Security*, *10*(6), 1463–1478. doi:10.1007/s12571-018-0859-3

Schoneveld, G. C. (2014). The geographic and sectoral patterns of large-scale farmland investments in sub-Saharan Africa. *Food Policy*, *48*, 34–50. doi:10.1016/j.foodpol.2014.03.007

Schuenemann, F., Thurlow, J., & Zeller, M. (2017). Leveling the field for biofuels: Comparing the economic and environmental impacts of biofuel and other export crops in Malawi. *Agricultural Economics*, *48*(3), 301–315. doi:10.1111/agec.12335

Smale, M., & Mason, N. (2014). Hybrid seed and the economic well-being of smallholder maize farmers in Zambia. *The Journal of Development Studies*, *50*(5), 680–695. doi:10.1080/00220388.2014.887690

Smith, H. L. (1997). Matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology*, *27*(1), 325–353. doi:10.1111/1467-9531.271030

Tsusaka, T. W., Msere, H. W., Siambi, M., Mazvimavi, K., & Okori, P. (2016). Evolution and impacts of groundnut research and development in Malawi: An ex-post analysis. *African Journal of Agricultural Research*, *11*(3), 139–158. doi:10.5897/AJAR2015.10167

West, J. J., & Haug, R. (2017). The vulnerability and resilience of smallholder-inclusive agricultural investments in Tanzania. *Journal of Eastern African Studies*, *11*(4), 670–691. doi:10.1080/17531055.2017.1367994

Williams, T. G., Trush, S. A., Sullivan, J. A., Liao, C., Chesterman, N., Agrawal, A., … Brown, D. G. (2021). Land-use changes associated with large-scale land transactions in Ethiopia. *Ecology and Society*, *26*(4), 34. doi:10.5751/ES-12825-260434

Zaehringer, J. G., Wambugu, G., Kiteme, B., & Eckert, S. (2018). How do large-scale agricultural investments affect land use and the environment on the western slopes of Mount Kenya? Empirical evidence based on small-scale farmers' perceptions and remote sensing. *Journal of Environmental Management*, *213*, 79–89. doi:10.1016/j.jenvman.2018.02.019

Zuka, S. P. (2019). Customary Land titling and inter-generational wealth transfer in Malawi: Will secondary Land rights holders maintain their Land rights? *Land Use Policy*, *81*, 680–688. doi:10.1016/j.landusepol.2018.11.039

Appendix

Table A1. Summary of relevant empirical studies

Study	Level	Dependent variable	Independent variable	Direction of effect	Method
Ben Slimane et al. (2016)	55 developing countries	Agricultural production	FDI in agriculture	insig.	3SLS
		Food security	FDI in agriculture	+	
Dhahri & Omri (2019)	50 developing countries	Agricultural production	FDI in agriculture	+	Panel fixed effects
Herrmann (2017)	Tanzanian smallholders	Income	Outgrower	+	Propensity score matching
		Income	Estate worker	+	
Deininger & Xia (2016)	Mozambican smallholders	Yields	Proximity to estate	insig.	Diff-in-diff
		Access to inputs	Proximity to estate	+	
		Labour demand	Proximity to estate	+	
Glover & Jones (2019)	Mozambican smallholders	Access to inputs	Proximity to estate	–	Inverse probability weighting
Zaehringer, Wambugu, Kiteme, & Eckert (2018)	Kenyan smallholders	Occurence of pests	Proximity to estate	+	Interviews
Ali et al. (2019)	Ethiopian smallholders	Fertilizer use	Proximity to estate	+/-	Fixed effects
		Yields	Proximity to estate	(+)	
Ali et al. (2017)	Ethiopian smallholders	Labour demand	Proximity to estate	(+)	Descriptive statistics
Osabuohien et al. (2019)	Tanzanian smallholders	Income	Proximity to estate	–	Propensity score matching
	Female Tanzanian smallholders	Income	Proximity to estate	+	
Ahlerup and Tengstam (2015)	Zambian smallholders	Income	Proximity to estate	+	Fixed effects
Deininger and Xia (2018)	Malawian smallholders	Agricultural production	Proximity to estate	–	Fixed effects
		Yields	Proximity to estate	insig.	
Mendola and Simtowe (2015)	Malawian smallholders	Agricultural production	Land transfer	+	Propensity score matching
		Yields	Land transfer	+	
Herrmann and Grote (2015)	Malawian smallholders	Income	Proximity to estate	+	Propensity score matching
		Poverty	Proximity to estate	–	
Herrmann et al. (2018)	Malawian outgrowers	Food crop production	Proximity to estate	+	Propensity score matching/ endogenous switching regression
		Food crop yield	Proximity to estate	insig.	
		Food crop land	Proximity to estate	insig.	
Lay et al. (2020)	Zambian smallholders	Maize yields	Proximity to estate	+	Diff-in-diff

Note: Effects: + positive, - negative, insig. Insignificant.

Table A2. Descriptive statistics of household level outcome indicator variables for overall, nearby and distant to estate samples and *t*-tests for equal means

Characteristic	Overall $N = 9463$	Nearby ($\leq$20 km) $N = 6758$	Distant (>20 km) $N = 2705$	Difference[a]	95% CI	*p*-value
Wage income (MWK)	52,689 (194,234)	44,648 (182,167)	72,779 (220,283)	−28,131	−37,503, −18,760	<0.001
Total income (MWK)	215,411 (285,011)	207,850 (264,821)	234,251 (329,288)	−26,401	−40,369, −12,434	<0.001
Total area (ha)	0.80 (0.80)	0.82 (0.82)	0.74 (0.72)	0.08	0.05, 0.11	<0.001
Assets (TLU)	0.29 (1.30)	0.32 (1.41)	0.23 (0.99)	0.08	0.03, 0.13	<0.001
Number of crops grown				−0.17	−0.22, −0.13	
1	2101 (22%)	1596 (24%)	505 (19%)			
2	3868 (41%)	2797 (41%)	1071 (40%)			
3	2378 (25%)	1651 (24%)	727 (27%)			
4	812 (8.6%)	527 (7.8%)	285 (11%)			
5	241 (2.5%)	149 (2.2%)	92 (3.4%)			
6	55 (0.6%)	32 (0.5%)	23 (0.9%)			
7	8 (<0.1%)	6 (<0.1%)	2 (<0.1%)			
Crop concentration (Shannon index)	0.65 (0.43)	0.62 (0.42)	0.71 (0.43)	−0.08	−0.10, −0.07	<0.001

Note: CI: Confidence Interval. Values are presented as Mean (SD); *n* (%).
[a]Welch Two Sample *t*-test; Standardized Mean Difference

Table A3. Descriptive statistics of crop level outcome indicator variables for overall, nearby and distant to estate samples and t-tests for equal means

Group	Stat. per crop	Overall	Nearby (≤20 km)	Distant (>20 km)	Difference[a]	95% CI	p-value
Beans	Area (ha)	0.24 (0.22) $n = 1254$	0.26 (0.23) $n = 940$	0.18 (0.18) $n = 314$	0.08	0.05, 0.10	<0.001
	Harvest (kg)	26 (58) $n = 1254$	27 (58) $n = 940$	23 (58) $n = 314$	3.6	−3.8, 11	0.3
	Yield (kg/ha)	170 (490) $n = 1254$	169 (500) $n = 940$	175 (460) $n = 314$	−6.4	−67, 54	0.8
Groundnut	Area (ha)	0.29 (0.24) $n = 1352$	0.30 (0.24) $n = 1{,}002$	0.26 (0.24) $n = 350$	0.04	0.01, 0.07	0.007
	Harvest (kg)	230 (301) $n = 1352$	228 (304) $n = 1002$	234 (295) $n = 350$	−5.8	−42, 31	0.8
	Yield (kg/ha)	1,031 (1,183) $n = 1352$	1,005 (1,186) $n = 1002$	1,105 (1,175) $n = 350$	−101	−245, 44	0.2
Maize	Area (ha)	0.42 (0.29) $n = 8934$	0.44 (0.29) $n = 6441$	0.36 (0.28) $n = 2493$	0.07	0.06, 0.09	<0.001
	Harvest (kg)	313 (350) $n = 8934$	316 (353) $n = 6441$	304 (340) $n = 2493$	12	−4.3, 28	0.15
	Yield (kg/ha)	1,003 (1,094) $n = 8934$	962 (1,077) $n = 6441$	1,109 (1,129) $n = 2493$	−146	−198, -94	<0.001
Nkhwani	Area (ha)	0.20 (0.19) $n = 3292$	0.20 (0.19) $n = 2287$	0.19 (0.20) $n = 1005$	0.01	−0.01, 0.02	0.3
	Harvest (kg)	68 (135) $n = 3292$	64 (135) $n = 2287$	76 (134) $n = 1005$	−12	−22, -2.0	0.019
	Yield (kg/ha)	546 (1,094) $n = 3292$	496 (1,072) $n = 2287$	658 (1,133) $n = 1005$	−162	−246, -78	<0.001
Pigeonpea	Area (ha)	0.25 (0.22) $n = 2545$	0.28 (0.23) $n = 1497$	0.22 (0.19) $n = 1048$	0.06	0.05, 0.08	<0.001
	Harvest (kg)	60 (99) $n = 2545$	63 (103) $n = 1497$	55 (94) $n = 1048$	8.3	0.56, 16	0.036
	Yield (kg/ha)	330 (535) $n = 2545$	324 (542) $n = 1497$	339 (525) $n = 1048$	−15	−57, 27	0.5

Note: Values are presented as Mean (SD); n: No. obs.; CI: Confidence Interval.
[a]Welch Two Sample t-test.