INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

IFPRI

IFPRI Discussion Paper 02433

July 2026

Paving the Way

The Impact of Road Infrastructure on Agricultural Productivity in Isolated African Settings

Henry Kankwamba

Karl Pauw

Foresight and Policy Modeling Unit

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

AUTHORS

Henry Kankwamba (hkankwamba@gmail.com) is a Senior Lecturer in the Agricultural and Applied Economics Department of the Lilongwe University of Agriculture and Natural Resources, Lilongwe, Malawi.

Karl Pauw (k.pauw@cgiar.org) is a Senior Research Fellow in the Foresight and Policy Modeling (FPM) Unit at the International Food Policy Research Institute, Washington, DC.

Abstract

Road infrastructure is a key driver of agricultural growth and transformation. Using three large nationally representative, randomly sampled survey datasets and group-time difference-in-difference estimation strategies, this study establishes the relationship between paved road access and crop productivity in Malawi. Results reveal significant positive effects of access to paved roads on crop productivity, as measured by the average value product of land. They further reveal a growing impact over time, with estimated productivity increases ranging from 35% to 57% over a span of 10 years. This confirms that road infrastructure is a key enabler of sustained agricultural productivity growth, thus providing a strong justification for the allocation of more public resources toward investments in rural road networks that can improve market access, reduce transport costs and postharvest losses, and create incentives for the adoption of productivity-enhancing technologies.

Keywords: Agricultural productivity; paved roads; difference-in-difference, instrumental variables; DBSCAN; machine learning.

Acknowledgments

The modeling and data systems and analytical techniques were developed with financial support from the CGIAR Science Program on Policy Innovations. We would like to thank all funders who supported this research through their contributions to the CGIAR Trust Fund (www.cgiar.org/funders).

1 Introduction

Road connectivity is essential for smallholder agricultural productivity in sub-Saharan Africa. Critically constrained by factors such as limited access to markets, and high transaction costs, investments in rural infrastructure, especially paved roads, are essential for enhancing supply chain integration, economic growth, and poverty reduction in the region (Fuglie, 2018; Jayne et al., 2010). Notably, the average value productivity in sub-Saharan Africa is statistically lower than in Latin America and Asia (Djoumessi, 2021; OECD/FAO, 2016). While some authors have hypothesized that the productivity of land is limited by its abundance—owing to its high dependence on primary factors of production (Frisvold, 1995)—recent evidence suggests that farm sizes are shrinking and their isolation could be a major factor (Jayne & Sanchez, 2021). Rural areas—home to most of the farming population—are often geographically isolated, with poor road connectivity impeding farmers' ability to access markets, inputs, and services (Abate et al., 2020; Stifel & Minten, 2008). Addressing these infrastructural deficiencies has long been recognized as a priority for unlocking the region's agricultural potential. While other proximate constraints such as access to credit, climate risks, and institutional weaknesses may also impede agricultural productivity growth, the need for rigorous empirical evidence on the effect of road infrastructure on agricultural productivity remains a critical gap, particularly in contexts where high-quality data is scarce. In this study, we seek to fill this gap by evaluating the impact of road infrastructure on agricultural productivity in Malawi, a landlocked country in southeast Africa.

Malawi relies heavily on its agriculture sector, which generates approximately 30% of GDP and supports the livelihoods of more than 80% of the population (National Statistical Office, 2019). Agricultural productivity remains low in Malawi. A critical bottleneck to agricultural growth is poor infrastructure, particularly poor road networks (Lall et al., 2009). Most rural areas are only connected by unpaved and seasonally impassable roads, which hinder farmers' access to inputs, such as fertilizers and improved seeds, as well as to markets for their produce. Consequently, rural farmers face high transaction costs, reduced profitability, and limited incentives to invest in productivity-enhancing technologies (Zant, 2018). While the significance of road infrastructure in enhancing agricultural productivity, economic development, and agricultural transformation is widely acknowledged in the literature on development economics (Dorosh & Thurlow, 2018), the extent to which paved roads alleviate agricultural constraints and improve productivity remains empirically underexplored in Malawi and the broader African context. Further, rigorous evidence on its effects in isolated rural African settings remain limited. Existing research often relies on cross-sectional correlations, or cross-sectional variation, leaving open questions about magnitude, persistence, and mechanisms of these effects under real-world conditions (Dorosh et al., 2012; Lee et al., 2023). Such reliance may fail to isolate the treatment effect of road infrastructure from other confounding factors such as climate variability, institutional weaknesses, or pre-existing economic differences between well-connected and isolated areas. This paper contributes new empirical evidence by estimating the treatment effect of improved access to paved roads on smallholder farm productivity in Malawi using multiple identification strategies and three rounds of nationally representative household panel data.

Specifically, we seek to exploit spatial and temporal variation in the expansion of the paved road network and apply group-time difference-in-difference approaches with staggered adoption following Callaway and Sant'Anna (2021). Additionally, as a test of the robustness of our results, we employ an instrumental variable (IV) approach based on historical trade routes to address potential endogeneity in road placement. Our approach allows us to address selection bias, heterogeneous treatment effects, and potential violations of the parallel trends assumption. Existing studies often rely on administrative or aggregate data, which may obscure localized effects of infrastructure investments. By using georeferenced survey data and employing density-based spatial clustering of applications with noise (DBSCAN), an unsupervised machine learning technique, we create a unique panel dataset that links enumeration areas across surveys using 2 km^2 grid references. Our primary contribution is therefore methodological and context specific, as we provide credible long-term estimates of the average treatment effect of proximity to paved roads on farm productivity in a low-income agrarian African economy.

These methodological innovations allow us to analyze the effects of road construction at a granular level, capturing localized impacts that are often missed in aggregate studies. Drawing on spatial economic theories, such as Von Thünen's land-use model (Von Thünen & Hall, 1966) and central place theory (Christaller, 1933), we provide a comprehensive analysis of how road connectivity influences agricultural productivity. Our findings offer valuable insights for policymakers on the role of infrastructure in stimulating rural economies. This analysis is critical to Malawi's strategic development agenda, which prioritizes agricultural productivity growth and commercialization, agro-processing, and stronger agrifood value chain linkages between rural areas and secondary cities in particular.

The paper is structured as follows. Section 2 discusses the policy and legislative context for roads in Malawi. Section 3 reviews the literature on the link between infrastructure and agricultural productivity. Section 4 presents the methods employed, and Section 5 presents and discusses results.

2 Roads and Development in Malawi

Roads are important in Malawi's economic development. In 1973, almost a decade after Malawi became independent, Kamuzu Banda—the country's founding president—viewed paved roads as a sign of economic development in his essay in the first issue of *World Development* (Banda, 1973). Since then, roads have featured prominently in the country's development agenda. For instance, the Malawi Growth and Development Strategy (MGDS III, 2017–2022) emphasizes infrastructure development as a key priority for enhancing connectivity, reducing transport costs, and facilitating market access in support of agricultural development, diversification, regional trade, and economic growth (Government of Malawi, 2017b).

To that end, Malawi's government developed the Malawi National Transport Master Plan (MNTP), which articulates the strategic vision for the transport sector and plans to upgrade and expand the road network to meet current and future demand (Government of Malawi, 2017a). The country's long-term strategy document, Malawi 2063 (Government of Malawi, 2021), also identifies road infrastructure as a key priority. This is consistent with the aspirations laid out in the Agenda 2063 of the African Union (2015), which identifies roads as a catalyst for intra-African trade, investment, and tourism. In this regard, Malawi's partnerships with neighboring countries to establish regional transport networks or development corridors is notable, including the Mtwara Development Corridor (with Tanzania, Mozambique, and Zambia) and the SADC Sub Regional Transport and Trade Facilitation (with Mozambique) (African Development Bank, 2022).

Malawi's public road network comprises approximately 15,451 km. The network is categorized into five main classes: main, secondary, tertiary, urban, and district roads. The main, secondary, and tertiary roads, totaling about 10,603 km, form the country's primary road network (Government of Malawi, 2017a). Paved roads make up only 26% of the total road network (just over 4,000 km), while the remainder consists of gravel and earth roads. The quality of these roads is generally poor, with many unpaved roads becoming impassable during the rainy season. This reduces transportation efficiency and restricts year-round access to remote areas. Despite large-scale investment programs such as the Mtwala Development Corridor and the Road Sector Program (RSP), the paved road network in Malawi remained has shown no growth over the past decade (2015–2025) (Government of Malawi, 2017).

3 Literature review

The theoretical basis for the positive impact of infrastructure on agricultural productivity is well established in the fields of economic geography and spatial economics. Classical theories emphasize the central role of transport costs, mobility, and accessibility in determining patterns of production, land use, and economic activity. The Von Thünen model (Von Thünen, 1826; Von Thünen & Hall, 1966) identifies transport costs as a fundamental determinant of agricultural productivity, positing that farms closer to markets have stronger incentives to commercialize and invest in productivity-enhancing inputs (Walker, 2022). Similarly, central place theory (Christaller, 1933; Getis & Getis, 1966) underscores that improved road networks promote economic clustering, enhance rural market participation, and strengthen input-output linkages in local economies.

Empirical evidence across multiple contexts broadly supports these theoretical insights. In sub-Saharan Africa, Dorosh et al. (2012) demonstrate that improved road access is positively associated with agricultural production through greater access to inputs and markets. Their simulations show that reducing travel time through road investments raises production by as much as 70% for high-input farmers and 89% for low-input subsistence farmers. Likewise, in Madagascar, Stifel & Minten (2008) find that isolated communities experience higher poverty levels and lower agricultural productivity, illustrating the adverse welfare implications of geographic isolation. In Bangladesh, Khandker et al. (2010) report that newly connected villages experience substantial increases in output and revenue, while Ali et al. (2016) use panel data from Ethiopia to show that proximity to roads increases fertilizer adoption and farm income once unobserved heterogeneity is controlled for. Further, Abate et al. (2020) find that access to main markets stimulates both farm and nonfarm employment in Ethiopia, underscoring the broader transformative role of infrastructure in rural development.

Other studies have highlighted how road quality and market connectivity influence agricultural market integration, input access, and production outcomes. For example, Minten & Kyle (1999) show that variation in food prices in the former Zaire is largely explained by road quality through its impact on transport costs. Similarly, Shamdasani (2021) finds that better-connected farmers in India diversify their crops and adopt new technologies. Ouma et al. (2010), using a Heckman selection model, reveal that geographic factors influence the likelihood of participating in banana markets in

Rwanda and Burundi. Beyond production and market participation, accessibility also affects the delivery of agricultural extension and advisory services. In Ethiopia, Abate et al. (2020) find that extension workers spend less time interacting with farmers in remote areas. In Malawi, Ragasa and Mazunda (2018) show that farmers located farther from Agricultural Development and Marketing Corporation (ADMARC) markets are less likely to receive extension advice, while Hawonga and Misook (2018) highlight how poor infrastructure weakens the effectiveness of agricultural projects intended to improve farmer productivity.

Despite the preponderance of evidence suggesting a positive relationship between road infrastructure and agricultural performance, methodological limitations remain pervasive. Many of the existing studies rely on cross-sectional data, limiting the ability to infer causality. Some findings are further constrained by small or non-random samples (e.g., Hawonga & Misook 2018) or by potential reverse causality, where more productive regions attract road investments rather than roads driving productivity. Dorosh et al. (2012) acknowledge that their results identify correlations rather than treatment effects, despite employing IV approaches. These limitations underscore the need for longitudinal approaches that account for unobserved heterogeneity, mitigate omitted variable bias, and identify exogenous sources of variation in road placement.

This study contributes to and addresses several methodological weaknesses in the literature by exploiting empirical data within a natural experimental setting to more rigorously assess the treatment effect of road infrastructure on agricultural productivity. This approach builds on Hawonga and Masook (2018) by strengthening identification and inference using the representativeness of the data and rigor of methods. Complementing Ali et al. (2016), this study shows that when access to reliable infrastructure is staggered over time, the impacts could vary over time, by incorporating both spatial and temporal dimensions and using an improved version of panel data techniques. Further, the study complements Khandker et al. (2010) by bringing an African perspective on the impacts of road infrastructure on agricultural outcomes.

4 Methodology

4.1 Data sources and construction of the panel data

The econometric analysis is based on three rounds of Integrated Household Surveys (IHS) collected in 2010/11, 2016/17, and 2019/20 by Malawi's National Statistical Office (NSO). The IHS are nationally representative and data were collected using random stratified sampling, with sample sizes of 12,271, 12,447, and 11,434 households, respectively. Data collection for each round was scheduled over a 13-month period from April to April the following year, thus covering the full agricultural season using a detailed agricultural module. This makes these surveys suitable for tracking agricultural outcomes (National Statistical Office, 2020).

The IHS have unique identifiers for both districts and Traditional Authorities (TAs), which did not change during the three survey rounds. TAs are larger geographic units under the jurisdiction of a chief. While some TAs might be sufficiently homogeneous for the analysis, most tend to be quite heterogeneous due to their geographic size. The surveys also have unique enumeration area (EA) identifiers and household identifiers (National Statistical Office, 2020). Each of the 18,722 EAs include, on average, 50 households. Common physical features and socioeconomic characteristics of these communities, such as schools, hospitals, and roads, are reported separately. While EA identifiers are unique to a given wave of the IHS, they do not remain constant throughout the subsequent surveys unless the specific households are tagged to become panel households later. Importantly, the IHS surveys are georeferenced at the EA level (National Statistical Office, 2020). Given these unique characteristics, the IHS data is mostly used in its cross-section form or in a pooled fashion.

This study seeks to exploit a natural experiment where, over time, the presence of paved roads in EAs change. We identify four possible scenarios:

1. An EA had no paved roads in all the survey rounds. Technically, this would be the "never treated" group and therefore serves as a control.
2. An EA received a paved road in 2019 – a first treatment group.
3. An EA received a paved road in 2016 – a second treatment group
4. An EA had paved roads in all rounds – an "always treated" group (this group is dropped from the analysis due to lack of variation during the study period)

The four scenarios present a unique natural experiment where we can track agricultural productivity outcomes before and after a road infrastructure investment was made. Figure 1 shows a satellite image of a road that was unpaved in 2010 but was converted into a paved road by 2019. To exploit the full potential of the natural experiment, we require data that both identify all relevant scenarios and contain a sufficient number of observations to capture meaningful statistical variation.

The series of IHS surveys can potentially provide that variation but require some adjustments. The IHS data is georeferenced for each EA. The availability of geolocations allows us to identify centroids for each EA, defined as balance points that represent the average location of all known points within an EA. From that centroid, we draw a circle with a radius of 2 km. Next, we determine the existence of paved roads within that radius. The chosen radius reflects the Sustainable Development Goal indicator 9.1.1, which tracks the proportion of the rural population within a 2 km radius of an all-season road (United Nations, 2025).

Panel A: A cross section of the Lilongwe-Kasiya Road (S117) in 2010 before the paved road was constructed

Panel B: A cross section of the Lilongwe-Kasiya Road (S117) in 2019 after the paved road was constructed

Figure 1: Google Maps of the Lilongwe Kasiya (S117) road before (2010) and after (2019) a paved road construction.
Source: Google Earth (2025).

Notably, the IHS3 and IHS4 EAs were based on the 2008 Population and Housing Census (National Statistical Office, 2019), while the IHS5 was based on the 2018 census (National Statistical Office, 2020). Over time, the EAs were updated and some boundaries were redrawn (National Statistical Office, 2020). However, since the EAs are georeferenced, we can merge them over the surveys. Therefore, we merge EAs that are within the 2 km radius. To achieve that, we employ an unsupervised machine learning algorithm called density-based spatial clustering of applications with noise (DBSCAN) to construct a unique panel dataset linking observations across the surveys while accounting for spatial proximity with a 2 km radius.

The DBSCAN method uses latitude and longitude combinations to create unique identifiers that are based on closeness of the points in the surveys. Defining observations (or households) based on closeness across the IHS surveys would ensure consistency across the geographic regions. In practice, DBSCAN defines clusters as areas with high density of points separated by areas of lower density. Points that do not belong to any cluster because they are in a low-density region are classified as noise (*see* Figure 2). DBSCAN is non-parametric. Unlike other algorithms, DBSCAN does not require specifying the number of clusters in advance. For comparison, we use another level of clustering by matching TAs. The TAs merge one-to-one across the surveys but at the expense of considerable within heterogeneity due to the larger geographical size. That is, analysis at the TA level would consider a TA with one paved road as treated even though some households may live far away from the paved road. For this reason, we only show the distribution of treated and control groups and do not proceed with econometric assessment at the TA level (Table 1).

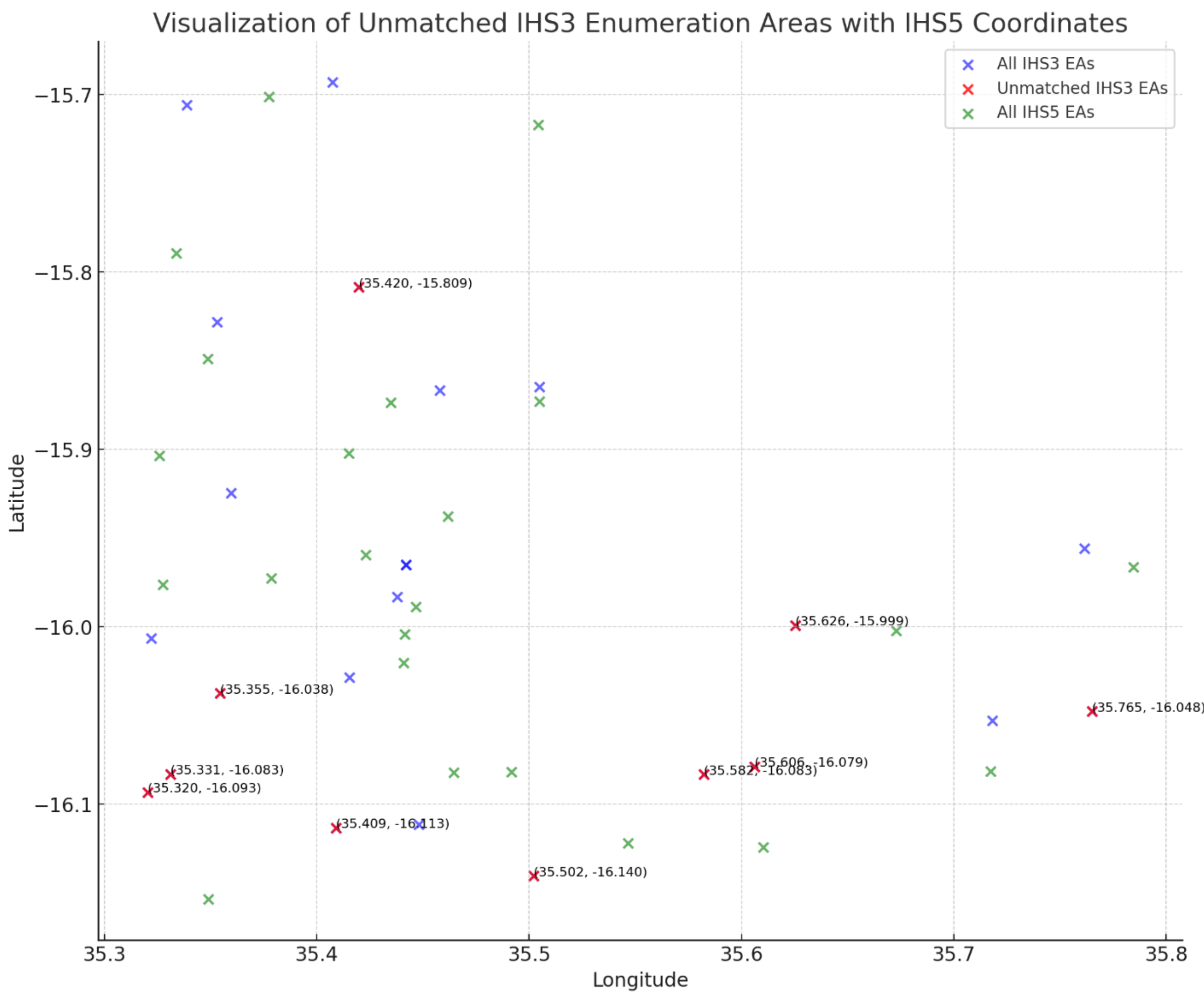

Figure 2: An example of DBSCAN output showing IHS3 EAs and IHS5 EAs and unmatched EAs

Table 1: Distribution of study communities at Traditional Authority and DBSCAN level

	(1) IHS3	(2) IHS4	(3) IHS5	(4) Total
(a) Traditional Authority level				
Control	227	245	261	733
Treated	38	35	58	131
Total	265	280	319	864
(b) DBSCAN 2 km radius clusters				
Control	648	699	624	1971
Treated	165	166	192	523
Total	813	865	816	2494

Note: The always-treated are those communities that had access to a paved road in IHS3 and the subsequent years. They corresponded to 2,813 clusters under DBSCAN.

4.2 Variable definitions and measurement

The main outcome variable of interest is farm-level productivity measured as the average value product (AVP) of land (Table 2). It is calculated as the price multiplied by quantity of the crop divided by a given cultivated plot area in a survey round (Chambers, 1989). The variable is presented in logarithmic scale because the variable is highly skewed to the right, driven by a small number of highly productive farmers and outliers. Taking natural logarithms compresses the skewness and produces a more symmetrical distribution—a suitable characteristic for further analysis using regression methods.

Landholding size (acres) is a key variable used in defining productivity. Improved road access lowers transport costs, thereby improving price and in turn encouraging farmers to expand cultivated land. This is consistent with the Von Thünen perspective (Parr, 2015). Notably, the landholding size variable is not used as an explanatory variable in the main model because it forms part of the outcome variable. However, it is used to balance covariates in the baseline and to demonstrate evolution of landholding sizes as paved roads are constructed overtime.

Adoption of soil and water conservation (SWC) and use of improved maize, inorganic fertilizer, manure, and pesticides serve to demonstrate that better road access reduces the cost and risk of adopting modern technologies (Mutenje et al., 2016). Higher rates of adoption in areas with paved roads could explain the shift toward high productivity. For example, increased adoption of SWC could improve soil health over time, thereby contributing to higher AVP. Similarly, inorganic fertilizer, pesticides, and organic manure could more easily reach farmers at a cheaper price, given reduced transport costs. Further, agricultural extension workers could more easily access farmers to introduce modern technology (Ragasa & Mazunda, 2018). Better road access could also encourage farmers to travel to demonstration sites or information dissemination shows, which could encourage diffusion of innovations and later improve productivity. These variables are constructed as dummy variables in the study.

Soil characteristics, as captured in the IHS data, are defined in terms of their main features, such as loam, sandy, and clay soils. While key characteristics remain stable across surveys, we acknowledge that adoption of sustainable agricultural practices could improve the structure of the soil and its relative acidity over time (Nyagumbo et al., 2022). However, the main physical features of the soil are used to improve precision and absorb baseline heterogeneity in productivity.

Table 2: Main variables and definitions

Variable	Definition
Farm level attributes	
Productivity	Average value product of land, i.e. price x quantity harvested/area (kg/acres)
Landholding size	Area cultivated in acres
Adopted improved SWC	Whether a plot uses soil and water conservation measures (Y/N)
Plots with loam soil	Whether plot has loam soil (Y/N)
Improved maize adopter	Whether a plot was planted with improved maize seed (Y/N)
Organic manure adopter	Whether a plot was using organic manure (Y/N)
Inorganic fertilizer adopter	Whether a plot was using inorganic manure (Y/N)
Pesticide adopter	Whether a plot was using pesticides (Y/N)
Access to extension	Whether a household had access to agricultural extension services (Y/N)
Household attributes	
Age of household head	Age of household head (Years)
Household is male-headed	Whether a household is male-headed (Y/N)
Household head is married	Whether the household head is married (Y/N)
Household size	Size of the household (count)

While household characteristics do not have strong impact in the relationship between paved roads and productivity, accounting for these baseline characteristics is important to ensure that treatment and control groups are well balanced. Nevertheless, age of the household head, though not directly linked to paved roads, could explain life-cycle effects on productivity and input use and help reduce heterogeneity in the model. Gender and marital status could explain gender differences in resource access and use and decision-making.

4.3 Data quality and treatment of noise in harvest, price data, and explanatory variables

It is well known that harvest data in IHS surveys contain substantial measurement error arising from recall bias, seasonality, aggregation costs across plots and seasons, and thin rural markets (Abay, 2020; Abay et al., 2023). We, therefore, implement several procedures to minimize the influence of noise while maintaining representativeness of the sampled data.

First, we construct the output value using household-reported sales prices whenever available, supplemented by EA-median prices for missing or outlier values for the same crop, and survey-round. This hybrid pricing strategy reduces measurement error common in Living Standards Measurement Study (LSMS) surveys. This strategy reduces both the item non-response and extreme outliers driven by one-off transactions or misreporting (Carletto et al., 2013). The measure also covers the rainy and dry seasons and is winsorized at the 1st to the 99th percentiles within each survey round. Winsorization is a statistical technique used to limit the influence of extreme values in a dataset by setting them equal to less extreme values at certain percentile thresholds (Carletto et al., 2013; Wollburg et al., 2024). To derive our productivity measure, the AVP, the total value of output was divided by total cultivated area. However, winsorization was applied after aggregating across crops, plots, and seasons but before dividing by the total cultivated area.

4.4 *Treatment variable and natural experiment design*

Our treatment variable is a dummy variable that captures the construction of paved roads in a cluster over the study period. Using the survey community modules and official records of road construction projects and triangulating them with Google Earth®, we identify areas that received new paved roads between 2010 and 2019. A community (EA) is treated if a paved road was constructed within a 2 km radius during the study period. The timing of treatment is assigned based on the year the road was completed. EAs that had access to roads throughout the study period (scenario 4) are defined as always treated and are dropped from the study. The communities that never had a road throughout the study period are retained and used as a control.

4.5 *Identification of treatment effects*

Road connectivity, as a measure of proximity to markets, is an important determinant of agricultural productivity, which is especially important in Malawi, where farming plays a crucial role in food security and poverty reduction. Investments in road connectivity can increase agricultural growth directly by raising the productivity of private inputs such as land, labor, and capital and by creating incentives for farmers to innovate and adopt modern technologies (see Romer, 1986). However, it is often difficult to establish the treatment effect of market proximity through roads on smallholder farm productivity due to endogeneity and omitted variable bias. Several sources of endogeneity in the context of Malawi are worth highlighting. First, since Malawi is a young democracy, polity plays a crucial role in the allocation of state resources and investments. As such, allocation and the types of road investments are primarily shaped by the endogenous decisions of elected public officers and their respective ethnicity and political ideology (Chinsinga, 2017).

In addition, Malawi has a very heterogeneous geography. The northern region is highly mountainous, making it difficult and expensive to develop infrastructure. The central region, in contrast, is mostly made up of the Lilongwe–Kasungu plain, which is more suited to infrastructure expansion. The southern region is also mountainous, but drops from 400 m above sea level over 180 km, to a low-lying area that is prone to flooding (Benson et al., 2005). These geographic variations also affect the allocation of road projects. Thus, the combination of these two major factors alone makes the attribution of road connectivity to markets to increased agricultural productivity problematic.

4.5.1 *Estimation by ordinary least squares regression*

Assuming that social planners make exogenous decisions devoid of partiality and political influence, and a flat plain earth, we can use pooled ordinary least squares (OLS) to run a regression of the association of access to a paved road and agricultural productivity (Amare et al., 2018; Savastano & Scandizzo, 2017). If we define Y_{pit} as the productivity of plot p belonging to community i, and a household j at time t, the regression model can be estimated as

$$\log(Y_{pijt}) = \alpha + \eta_j + \lambda_t + \beta_1 \log(D_{ijt}) + \mathbf{\Gamma}\bar{X}_{ijt} + \epsilon_{pijt} \tag{1}$$

where α is a regression constant, η_j is an area specific fixed effect, and D is the treatment variable capturing the association between road access and farm productivity. $\bar{X}_{pijt}$ summarizes a vector of explanatory variables. The setup adequately captures first difference effects but can be improved to capture difference-in-differences.

However, due to the endogeneity presented, OLS estimates obtained will be biased and inconsistent. First, OLS may yield larger standard errors leading to inefficiency (Angrist & Pischke, 2009). Inefficiency in this respect may cause inference drawn from the estimates to be invalid (Wooldridge, 2013). Another more problematic consequence of the endogeneity caused may be that the estimates may confuse and exaggerate the effects on $\log\left(Y_{pij}\right)$, making interpretation of the coefficients difficult (Wooldridge, 2013). By implication, the results obtained may not be a true reflection of the actual effect of road connectivity on agricultural productivity.

4.5.2 Estimation by two-way fixed effects (TWFE) regression

In this study, we will use a difference-in-difference estimation method with multiple time periods (Callaway & Sant'Anna, 2021) to establish the direct effect of paved road infrastructure on agricultural productivity. In its canonical form, the DID method requires two discrete time periods: the time before a policy (pre-treatment), and the time after the policy implementation (post-treatment) (Nguyen, 2025). The DID further requires that there should be a group that is affected by the policy (the treated group) and a group that does not receive the policy (control group) (Baker et al., 2022a). Considering our dependent variable Y_{pijt}, we denote $Y_{pijt}(1)$ as the productivity outcome for a household that has access to a road in the community, and $Y_{pijt}(0)$ as the productivity outcome of a household without such access. The average treatment effect on the treated (ATT) is the difference in productivity outcomes $Y_{pijt}(1) - Y_{pijt}(0)$ averaged across individuals who had access to the road.

However, estimating this effect presents a fundamental missing data problem, since for any given household we can observe productivity either with access to the paved road or without such access, but never both states simultaneously. Specifically, one cannot observe the case of what could have happened to those who later had access to the road if they had not received access, and vice versa. DID designs overcome this problem by assuming that the pre-treatment trend would have been constant for both those who were treated and also for the group that was not treated. This is known as the parallel trend assumption (Baker et al., 2022b). Any change in this case would only be because of the deliberate policy change—in this case introducing road infrastructure to one group. Formally, if we let the $ATT = \delta$ and letting D be a dummy variable equal to 1 if the household has access to a paved road and 0, otherwise, then

$$
\begin{aligned}
\delta &\equiv \mathbb{E}\left[Y_{pij1}(1)\middle|D = 1\right] - \mathbb{E}\left[Y_{pij1}(0)\middle|D_i = 1\right] \\
&= \mathbb{E}\left[Y_{pij1}(1) - Y_{pij0}(1)\middle|D_i = 1\right] - \mathbb{E}\left[Y_{pij1}(0) - Y_{pij0}(0)\middle|D_i = 1\right] \\
&= \mathbb{E}\left[Y_{pij1}(1) - Y_{pij0}(1)\middle|D_i = 1\right] - \mathbb{E}\left[Y_{pij1}(0) - Y_{pij0}\ (0)\middle|D_i = 0\right] \qquad (2)
\end{aligned}
$$

The first block of equation (2) demonstrates the missing data problem and cannot be directly estimated using the data. The second part distributes the first equation by subtracting a component of a state of not being treated while being in the treatment category. This could be a case where a household could have all characteristics to have access to the road infrastructure but does not receive treatment and anticipates no such treatment in future. In regression form, we can rewrite equation (1) as

$$\log\left(Y_{pij}\right) = \alpha + \eta_j + \lambda_t + \beta_1 D_i + \delta(D_i \times T) + \mathbf{\Gamma}\bar{X}_{ij} + \epsilon_{it} \qquad (3)$$

so that the regression runs productivity on a dummy for the treated D, time $T = \lambda_t$, and location fixed effects η_j. This set up is known as the two-way fixed effects (TWFE) regression framework (Callaway & Sant'Anna, 2021; Nguyen, 2025). To identify treatment effects, the TWFE assumes parallel trends, treatment homogeneity, and constant effects (Callaway & Sant'Anna, 2021).

4.5.3 Estimation by group time average treatment effects

However, in our study, different communities received treatment at different times, and we cannot assume constant effects. We can improve this method by using a group-time fixed effects regression to account for scenarios (3) and (4)—when some households were only treated in 2016 and remained treated after and while others were yet to be treated in 2019. In this form we have three groups, and we define ATT for a treatment-timing-group g at a group point in time as the "group time average treatment effect" (Callaway & Sant'Anna, 2021), where the groups signify when they received treatment:

$$ATT(g,\tau) \equiv \mathbb{E}\left[Y_{i,\tau}(1) - Y_{i,\tau}(0) \middle| E_i = g\right] \quad (4)$$

Given this information, we can disaggregate the survey time τ and $g-1$ and calculate the group-time ATT as follows:

$$y_{it} = \alpha + \lambda_t + \eta_j + \sum_g \sum_t \delta_{g,t}(D_{it} \cdot Post_{it} \cdot \mathbb{I}_g) + \epsilon_{it} (5)$$

4.5.4 Estimation by instrumental variables: A robustness check

We assess the robustness of our DID results using an instrumental variable (IV) approach. In this approach, our identification strategy relies on the assumption that there is an exogenous source of variation in road connectivity that affects smallholder farm productivity. Several IVs have been used in the literature to identify effects of economic isolation, access to amenities, and infrastructure on agricultural outcomes. For example, Stifel and Minten (2008) used soil characteristics as an IV for access to roads in their assessment of effects of economic isolation on agricultural productivity. One critique of their IV is that soil characteristics could have a direct effect on agricultural productivity, thus violating the dictate of exogeneity.

To assess the effects of road connectivity on crop production, Dorosh et al. (2012) used terrain roughness as an IV for road placement. While it is reasonable to argue that terrain roughness explains the placement of road infrastructure, one can also—in the same vein—explain that farmers do not like to place their farms on rough terrains. Thus, in its given form, this instrument fails to satisfy the assumption that terrain roughness affects agricultural productivity only through its effects on road placement. Thus, if terrain roughness also affects agricultural productivity through farmers' preference to avoid rough terrain, the instrument would violate the exogeneity assumption. However, their method could be based on how they defined terrain roughness. The authors defined terrain roughness as the difference in elevation for a given georeferenced area. Given this context, it can be deduced that the instrument is merely a mathematical construction that could be related to roads, that is, roads are expensive to construct at varying and rough elevations, but farming is not necessarily dependent on elevation due to differences in crop portfolio (Dorosh et al., 2012).

We use the Euclidian distance to the historical slave trade routes as an IV for today's road connectivity because it not only preserves the historical account, but it also embodies a many of the institutional dimensions driving broader development. Distance to ancient slave trade routes is likely to be correlated with current proximity to major demand nodes. While slavery routes were a frightening site for locals, the caravans also moved significant volumes of other commodities, signalling future infrastructure investments. Further, colonial settlers in their battles against slave traders and carrying out missionary work were tracing the same slave trade routes and establishing forts and garrisons together with locals.

While it is very likely that the old slave trade routes may inform spatial patterns of road infrastructure over a century later, thereby positively affecting proximity to markets, it is highly unlikely that they explain smallholder farmers' productivity outcomes. Due to major technological advancements, population growth, and myriad other factors, it is also unlikely that the routes explain current smallholder farm productivity, except through their effect on current road connectivity. Our choice of this IV reflects the literature. A study by Donaldson (2018) used historical railroad expansion in colonial India as an instrument for today's transport infrastructure to estimate the impact of transport infrastructure, showing that regions with improved connectivity experienced large productivity gains. Our instrument is also similar in concept to Jedwab & Moradi (2016), who used Euclidian straight-line distances from defunct colonial railroads as an instrument for modern road infrastructure. They too argued that unused old railroads cannot influence today's productivity.

To implement our instrumental variables procedure, we proceed in two stages to get the reduced form

$$log\big(Y_{pij}\big) = \alpha + \eta_j + \varphi\, log(\widehat{D}_{ij}) + \boldsymbol{\Gamma}\bar{X}_{ij} + \epsilon \tag{6}$$

where $\log(\widehat{D}_{ij})$ is an instrumented road connectivity to potential markets, which is predicted from a first stage regression of the distance to the old slave trade routes as the key explanatory variable. The first stage regression is estimated as follows:

$$\log\big(D_{ij}\big) = k + \tau_j + \phi \log\big(R_{ij}\big) + \boldsymbol{\Phi}\bar{X}_{ij} + \xi \tag{7}$$

where k is the regression constant coefficient and τ_j are district level dummy coefficients. The parameter ϕ is the coefficient on the slave trade route $\log\big(R_{ij}\big)$. It is a first robustness check—an indicator of instrumental relevance. When it is statistically significant, it implies that our IV is relevant. That is, given the instrument, the first stage creates a natural experiment which we can use to exploit the variation in potential outcomes on productivity, that is, our instrument is as good as randomly assigned (Angrist & Pischke, 2009).

Figure 3 demonstrates the relationship graphically for a subsample, where distances ranged between 1 and 45 km. On the horizontal axis, the figure shows the distance between the household and the old slave trade route, while the vertical axis shows the distance to the paved road. Figure 3 clearly shows that there is a pattern between the instrument and the endogenous variable.

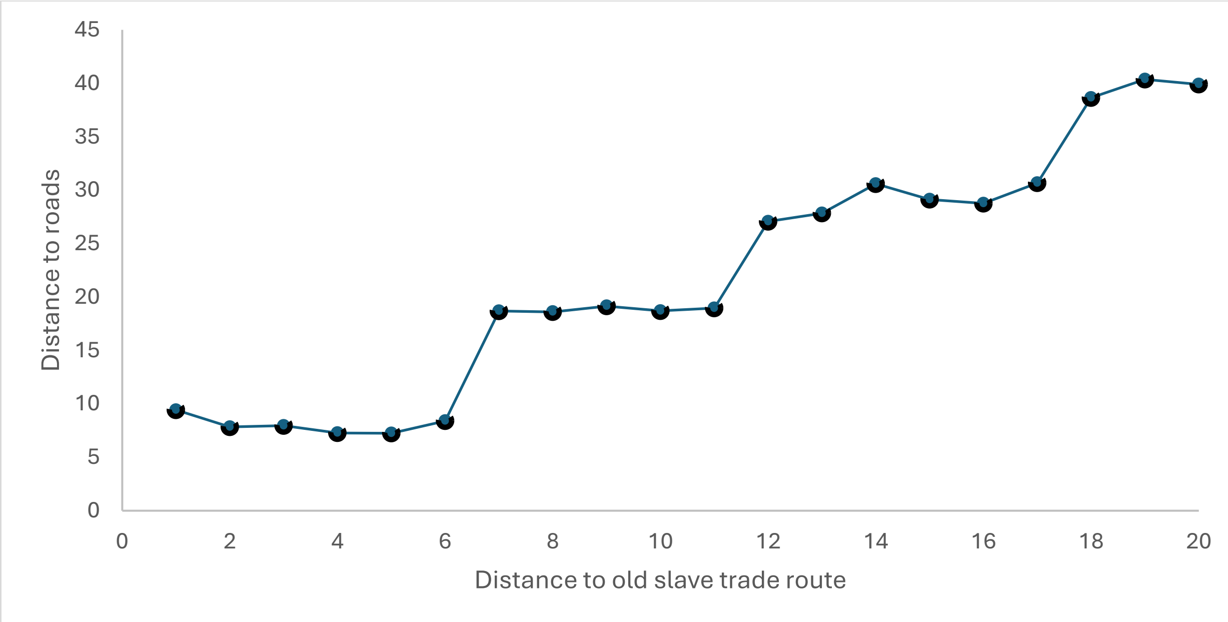

Figure 3: Average road proximity by distance to old slave trade routes (first stage IV regression).

Note: The graph shows the conditional expectation function (CEF) of distance to roads $E[D|Z = z, \boldsymbol{X}] = \int t\, f_D(t|Z = z, X)dt$ using ordinary least squares regression analysis with robust standard errors (not shown) from a subsample of distances ranging from 1 km to 45 km , where t are all possible values that D can take. No transformation has been made to the data yet. The graph shows a significant correlation between market potential and ancient slave trade routes.

Predicted values of distance to paved roads $\log(\widehat{D}_{ij})$ are then used in Equation 7. The first stage and the second stage complete what is known as the two-stage least squares estimation (2SLS) (Angrist & Pischke, 2009; Wooldridge, 2013). The strength of our instrument is formally tested using the F-statistic and the first-stage partial R-squared of the regression in Equation 8. A strong instrument should have a high F-statistic and a high first-stage partial R-squared. Third, we test for the presence of weak instrument bias using the Kleibergen-Paap F-statistic (Kleibergen & Paap, 2006).

Figure 4 demonstrates the relationship between the outcome variable and the IV. The results show the distance between a household and the old slave trade route for the same subsample that was used to construct Figure 3. With the IV on the horizontal axis and the outcome variable on the vertical axis, there is a clear pattern. Since the old trade routes no longer exist today, and technological change has occurred and people's settlements and productive activities have changed vastly, it is highly likely that the only way that the IV can affect the outcome variable is through the endogenous variable. Thus, the IV creates an ideal natural experimental variation when all the exogenous variables are accounted for.

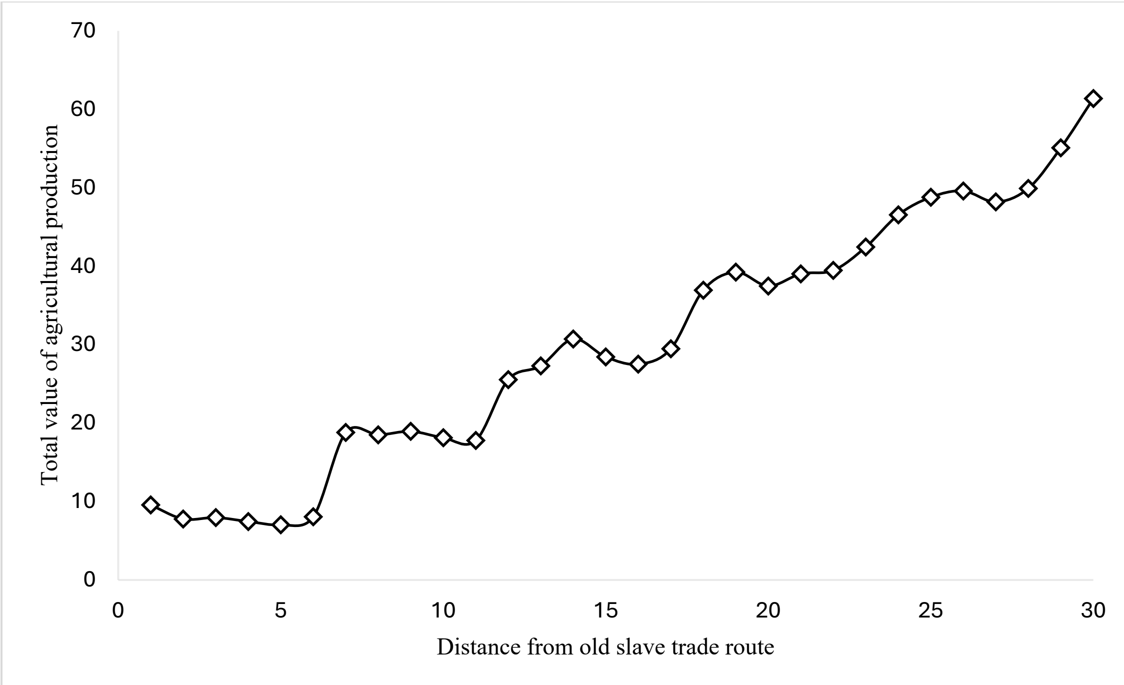

Figure 4: Average value of agricultural production per acre by distance to old slave routes (reduced form IV regression).

Note: The graph shows the conditional expectation of the dependent variable (y-axis) and the IV (x-axis). This is a reduced form part of the IV shown in equation 3 – i.e. $E\left[Y\middle|\widehat{D}=m,\boldsymbol{X}\right]$. This figure shows that areas closest to major slave trade routes have lower values of agricultural production. Considering the spikes and the troughs of this figure move together with the first stage (figure 3), this suggests that the two graphs are related. The old slave trade routes are long gone, but the pattern strongly suggests that the IV affects the dependent variable only through the endogenous variable, roads.

5 Results and discussion

5.1 Descriptive statistics of variables in the study

Table 3 presents descriptive statistics for key variables for the treated and control groups as measured in the IHS3 survey, our base year. Generally, the variables show no statistical differences with respect to proximity to the road during the first survey. Khandker et al. (2010) highlighted the importance of ensuring that both controls and treated groups have similar characteristics in the time before treatment. Results demonstrate that for most variables there are no significant differences between the groups. For example, the natural logarithm of productivity, a key outcome variable, shows no statistical difference between treated and control groups. Time invariant variables, such as soil characteristics, are not expected to change over time and are statistically not different in the base year. We also note that drivers of agricultural productivity, such as access to agricultural extension or adoption of inorganic or organic fertilizers, are not statistically different. An exception is the adoption of improved maize, which is significantly higher in areas closer to paved roads.

Household characteristics, however, showed some differences in the base year. While the average age of the household head is statistically not different between the groups, households closer to paved roads are statistically larger and their household heads are more likely married. The individual fixed effects components in the DID regression analysis seek to account for the initial conditions that exhibit difference in the baseline.

Table 3: Descriptive statistics of variables in the baseline (IHS3)

Variable	(1) Control	(2) Treated	(3) Difference	(4) t-stat	(5) p-value
Farm-level attributes					
Log productivity	6.571	6.606	0.035	0.461	0.645
	(0.031)	(0.073)	(0.077)		
Landholding size	1.136	0.887	-0.250	1.415	0.157
	(0.078)	(0.131)	(0.177)		
Adopted improved SWC	0.418	0.409	-0.009	0.856	0.392
	(0.004)	(0.009)	(0.010)		
Plots with loam soil	0.193	0.202	0.009	1.051	0.293
	(0.003)	(0.008)	(0.008)		
Improved maize adopter	0.278	0.317	0.039	4.303	0.000
	(0.004)	(0.009)	(0.009)		
Organic manure adopter	0.108	0.101	-0.007	0.948	0.343
	(0.003)	(0.006)	(0.006)		
Inorganic fertilizer adopter	0.603	0.590	-0.013	1.295	0.196
	(0.004)	(0.010)	(0.010)		
Pesticide adopter	0.231	0.029	-0.202	0.564	0.573
	(0.001)	(0.003)	(0.003)		
Access to extension	0.063	0.061	-0.002	0.365	0.715
	(0.002)	(0.004)	(0.005)		
Household-level attributes					
Age of household head	43.367	43.336	-0.031	0.097	0.928
	(0.131)	(0.289)	(0.321)		
Household is male-headed	0.783	0.753	-0.030	3.528	0.000
	(0.003)	(0.008)	(0.003)		
Household head is married	0.782	0.759	-0.023	2.814	0.005
	(0.003)	(0.008)	(0.008)		
Household size	4.908	5.139	0.231	5.127	0.000
	(0.018)	(0.044)	(0.045)		

Note: Log productivity is the natural logarithm of the average value product per acre of land in Malawi kwacha. SWC is soil and water conservation measures such as mulching, contour bunds, box-ridges, etc. Improved maize refers to hybrid and open pollinated maize varieties. Difference = Treated – Control. Standard errors in parentheses.

5.2 Testing the parallel trends assumption

The core identifying assumption of the DID approach is the parallel trends assumption. This assumption asserts that in the absence of treatment, the treated and control groups would have followed the same trend over time in their outcome variable. In this case, any observed difference in the outcome variables must be because of the paved road rather than the initial conditions or other confounding factors. The descriptive analysis has provided evidence that productivity levels and other observable characteristics are predominantly statistically similar ($p > 0.10$) in the initial conditions.

To assess the plausibility of the parallel trends assumption, we conducted a placebo test. First, we restricted the sample to EAs that never received paved roads during the entire study period. This is the "never treated" group. Within the never treated subsample, we randomly assigned a placebo treatment indicator to a subset of EAs. The assignment is designed to replicate the actual proportion and timing of treatment rollout observed in the real data. That is, we created artificial "treated" and "control" groups among units that were initially never exposed to the intervention of paved roads. This treatment is well known, by construction, to have no true treatment effect. We re-estimate the empirical specification on the placebo sample using the TWFE DID regression.

If our identifying assumptions are valid, the estimated coefficients on the placebo should be statistically indistinguishable from zero. Large or statistically significant placebo effects would indicate violations of the parallel trends' assumptions and the presence of time-varying confounders that could bias our main estimates.

Results of the placebo effects show that given the specification and estimator, the placebo effects are small in magnitude and are statistically not significant ($p > 0.10$) (Table 4). Such null findings provide strong supportive evidence that pre-existing differential trends or other confounders are not driving our main results. Consequently, the substantial and growing positive effects of paved road access on agricultural productivity can be interpreted as credible average treatment effects.

Table 4: Placebo difference in difference to test parallel trends assumption

	With No Controls	With Controls
	(1)	(2)
	Log Productivity	Log Productivity
Two-way fixed effects regression		
Placebo treatment (2010–2016)	0.295	0.298
	(0.216)	(0.216)
Placebo treatment (2010–2019)	0.305	0.308
	(0.209)	(0.210)

5.3 Difference-in-difference analysis of the impacts of paved road proximity on farm productivity

Table 5 and Figure 5 summarize the impacts of proximity to paved roads on farm productivity using sample means from the pooled cross-sections. The pooled cross-sections only rely on the comparisons of households within 2 km of a paved road and beyond as the sole distinguisher. The farm productivity effects of the paved road are averaged at $ATE = 32.97\% = (e^{0.285} - 1)$ in 2016 and 21% for the 2019 period. Notably, the lack of statistical significance raises concerns about the robustness of such findings. We call this approach to DID analysis naive, because it does not adequately consider (i) initial conditions, (ii) unobserved time invariant differences between treated and controlled groups, or (iii) dynamic effects over time.

Table 5: Naive difference-in-difference across the surveys using the full pooled sample

	(1) Treated	(2) Control	(3) Difference
(a) 2019 and 2010 difference-in-difference			
2019 (n=31,629)	9.549	9.328	0.221
	(0.048)	(0.025)	
2010 (n=17,935)	6.606	6.571	0.035
	(0.309)	(0.073)	
Difference-in-difference			0.187
			(0.322)
(b) 2016 and 2010 difference-in-difference			
2016 (n=25,177)	8.970	8.651	0.319
	(0.030)	(0.066)	
2010 (n=17,935)	6.606	6.571	0.035
	(0.309)	(0.073)	
Difference-in-difference			0.285
			(0.326)

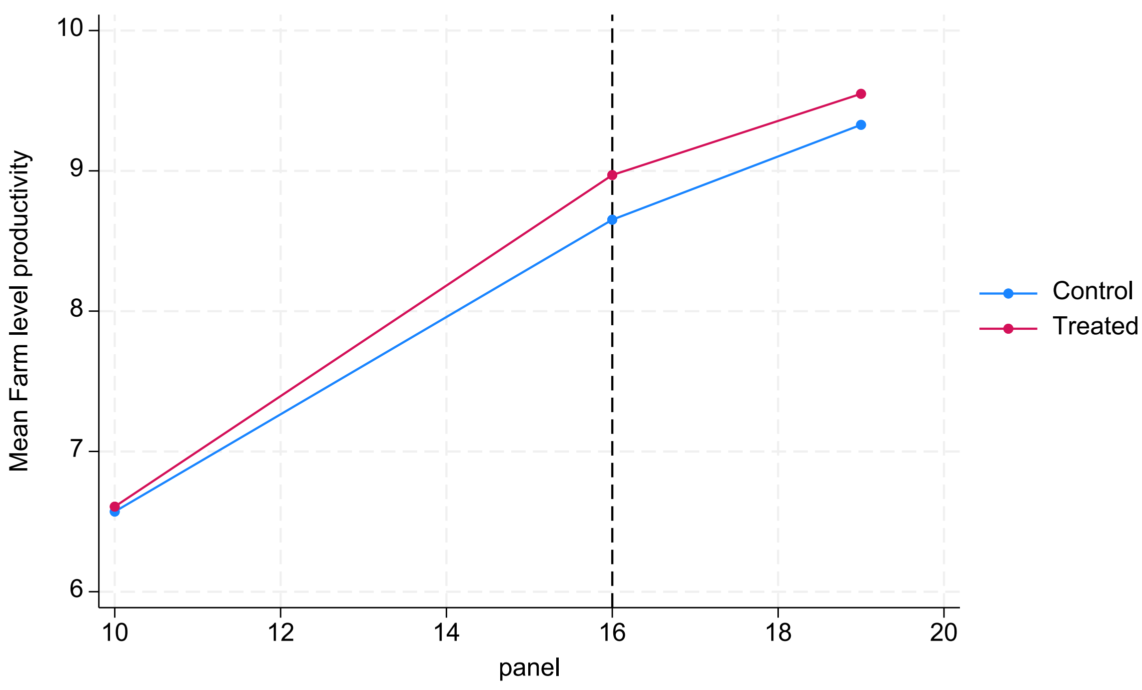

Figure 5: Impact of proximity to the road on agricultural productivity
Note: The indicator variable equals 1 if a community is less than 2 km away from a paved road. The figure is produced from the naive difference-in-difference model.

The study requires a proper TWFE DID design to address the concerns of the naive DID. Such a DID design would explicitly exclude the always treated group, which comprises households that were already close to a paved road in the baseline and, therefore, all periods. Adding this group of households to the study would dilute the treatment effect by conflating it with different baseline characteristics. Using the never treated group as the control ensures consistent counterfactual comparison.

Panel (a) in Table 6 summarizes TWFE regression results that mitigate biases associated with the naive DID approach. Taking the terminal as 2019, without any controls, the impact of a paved road is estimated to increase agricultural productivity by 35% (p<0.01) i.e. $TWFE = (e^{0.3} - 1) \times 100 = 35\%$. When all controls are accounted for the impact is 37% (p<0.01). This suggests that adding control variables strengthens the estimated treatment and that lack of control variables was biasing our estimates downward through omitted variable bias. Angrist and Pischke (2009) argue that including relevant variables to the regression model reduces residual variance and increases estimate precision. TWFE results are consistent with findings of Donaldson (2018), who observed that controlling for market-related variables increases benefits of infrastructure. Furthermore, our results are consistent with Dercon and Gollin (2014), who found that agricultural productivity might be influenced by other factors such as weather, the structure of labor supply, and government policies. Furthermore, in India, Asher and Novosad (2020) also found that road construction led to higher productivity.

Table 6: The impact of road infrastructure on agricultural productivity

	With No Controls	With Controls
	(1)	(2)
	Log Productivity	Log Productivity
(a) Two-way fixed-effects regression		
Treated	0.300***	0.313**
	(0.006)	(0.001)
(b) Callaway and Sant'Anna (2021): Group time average treatment effects on the treated		
Treated		
Survey (2016, IHS4)	0.272***	0.284***
	(0.002)	(0.001)
Survey (2019, HIS5)	0.454***	0.452***
	(0.001)	(0.002)
(c) Instrumental variable estimation		
Treated	0.944***	0.603***
	(0.050)	(0.049)

Note: The dependent variable is the natural logarithm of the average value product (AVP) per acre of land. For regressions with controls, household size, labor, access to agricultural extension, marital status, education level, age, were used as explanatory variables.

Panel (b) in Table 6 presents results of the group-time ATTs. Results indicate that the impact of proximity to paved roads on agricultural productivity was 33% (p<0.01) for those households that had access to newly constructed roads by 2016 (that is, places without roads in 2010/11), while those who were treated after 2016 saw their productivity increase by 57% (p<0.01). These results suggest a stronger effect over time. This is consistent with literature that indicates that agricultural yields increase when markets are nearer and inputs are easy to access (Ali et al., 2019; Asher & Novosad, 2020). Additionally, our results are consistent with Abate et al. (2020) who found that improving public infrastructure improves agricultural outcomes by increasing access to services such as extension. The group time effects suggest that the impact of paved roads on productivity is not immediate but rather strengthens over time, reinforcing the need for a dynamic perspective in evaluating infrastructure projects.

As a robustness check to strengthen the interpretation of our results, we used an IV-approach leveraging proximity to 20th century slave trade routes as an exogenous instrument for paved road placement. Our IV estimate in panel (c) of Table 6 shows that proximity to paved roads could increase productivity by up to 82%. Using such an instrument is well aligned to studies by Michalopoulos and Papaioannou (2013), who assert that using historical instruments can help isolate exogenous variation in modern infrastructure placement. Further, Donaldson and Hornbeck (2016) used historical trade routes as an IV for railroads in the United States and observed larger effects on agricultural productivity. Khandker et al. (2010) used an IV to assess impacts of road placement in Bangladesh and showed that impacts were higher using IV than DID.

Our results confirming an increase in agricultural productivity due to road connectivity aligns with the classic Von Thünen agricultural land use model by reducing distance penalties (Horvath, 1967; Parr, 2015). Farmers near markets might grow high-value perishable crops as transport constraints relax, which is consistent with Von Thünen, who assumes that proximity to markets allows for intensified, high-value farming. Thus, infrastructural investments should prioritize connecting high potential agricultural areas to major markets. Our results are also consistent with Christaller's central place theory, which asserts that larger and better-connected centers serve as economic hubs and that improved transportation generally improves market potential (Getis & Getis, 1966).

5.4 Mechanisms

The staggered DID results indicate considerable heterogeneity in the impacts of paved road access on productivity. While late beneficiaries experienced larger productivity gains, all treated cohorts experienced productivity gains. The underlying mechanisms demonstrating changes in agricultural practices, input use, and general socioeconomic factors show a general increase over the three survey periods. Analysis of variance tests indicate that there are statistically significant differences in the means of the covariates over the three survey periods across all variables (Table 7).

Table 7: Evolution of key variables across the three survey periods

Variable	Pooled Mean	2010	2016	2019	F-stat
Farm characteristics					
Adopted SWC	0.474	0.475	0.452	0.491	15.28***
	(0.499)	(0.500)	(0.498)	(0.500)	
Plot with loam soil	0.181	0.163	0.148	0.219	8.73***
	(0.385)	(0.369)	(0.355)	(0.414)	
Improved maize adopter	0.236	0.361	0.251	0.157	137.61***
	(0.425)	(0.48)	(0.361)	(0.251)	
Organic manure adopter	0.214	0.084	0.219	0.280	156.25***
	(0.41)	(0.277)	(0.414)	(0.449)	
Inorganic fertilizer	0.646	0.671	0.661	0.619	21.99***
	(0.478)	(0.47)	(0.671)	(0.661)	
Access to agricultural extension	0.139	0.053	0.362	0.000	1564.89***
	(0.346)	(0.225)	(0.481)	(0.000)	
Pesticide adopter	0.033	0.015	0.010	0.062	54.35***
	(0.178)	(0.122)	(0.098)	(0.241)	
Market potential	77.197	67.814	70.214	89.899	48.22***
	(430.55)	(435.186)	(375.661)	(423.876)	
Demographic characteristics					
Household is male headed	1.315	1.241	1.296	1.372	12.82***
	(0.465)	(5847)	(1.241)	(1.296)	
Age of the household head	44.701	41.300	45.180	46.138	30.82***
	(16.072)	(15.692)	(16.302)	(15.821)	
Household head is married	0.711	0.747	0.713	0.689	6.67***
	(0.453)	(0.277)	(0.414)	(0.449)	
Household size	4.630	4.659	4.470	4.747	12.82***
	(2.025)	(2.13)	(1.959)	(2.013)	
Household head has no formal schooling	0.075	0.033	0.021	0.142	49.87***
	(0.263)	(0.179)	(0.145)	(0.349)	

Note: Standard deviations are in parentheses. *** means statistically significant at 99% confidence level.

The random variables in Table 7 were used in the analysis of the effects of access to paved roads on farm productivity using TWFE, staggered DID, and IV methods. The discussion follows the magnitudes of the full interaction terms in the model (Table 8) and the descriptive statistics (Table 7). The first main causal mechanism is that access to markets and complementary services and human capital drive agricultural productivity gains. Access to agricultural extension shows a positive and statistically significant effect in the 2019 cohort (0.203, $p<0.05$), implying that access to paved roads not only improves agricultural productivity but also enables delivery of productivity-enhancing services to farmers (Ragasa & Mazunda, 2018). This suggests that ease of access to roads facilitates farmers' access to technical husbandry advice (Abate et al, 2020). Further, when the household head has no formal schooling, productivity levels are significantly lower for the group that was treated in 2016 relative to the baseline in 2010 but increases steadily up through the 2019 survey (−0.032, $p<0.05$ in 2016 and 0.059, $p<0.1$ in 2016–2019, and 0.139, $p<0.1$ in 2019). This highlights the human capital contribution to agricultural productivity, such that at the beginning, there is a drop in productivity attributed to lack of information processing abilities relative to highly educated household heads. Thus, education amplifies paved road benefits.

The second channel through which paved road access drives agricultural productivity is effective input use (Table 8). Results indicate that while coverage of inorganic fertilizer use is above 60% across the three survey periods and varies significantly (Table 7), when incorporated in the regression framework, results are not statistically robust ($p>0.10$). However, improved maize seed adopters who were treated in 2016 saw significant declines in productivity (−0.028, $p<0.1$). Such declines could be attributed to cropland re-allocation to other high-value crops due to proximity to distant markets and reduced costs through the paved roads. Further, results show strong productivity gains among pesticide adopters (0.162, $p<0.1$). Such productivity gains imply that farmers treated in the 2019 survey period had access to farm inputs, reinforcing the reduced transport cost to input markets link.

Notably, market potential—measured as population density per kilometer of distance to the nearest urban center—captures how well connected a location is to large markets (Shrestha, 2020). In the model, the interaction between road treatment and market potential is consistently negative and statistically significant. The interaction showed declines in productivity, indicating that areas closer to major population centers marginally gain from access to paved roads compared to remote areas. This is consistent with the economic geography theory that infrastructure yields higher returns where market access was previously a binding constraint (Getis & Getis, 1966).

Table 8: Key impact channels influencing agricultural productivity

Covariate	2010 Cohort (2010)	2016 Cohort (2019)	2019 Cohort (2019)
Markets, services, and human capital			
Access to agricultural extension	0.008 (0.011)	0.042 (0.161)	**0.203**** (0.089)
Household head has no formal schooling	**−0.032**** (0.014)	**0.059*** (0.013)	**0.138*** (0.040)
Market potential	−0.002 (0.002)	**−0.009*** (0.002)	**−0.014**** (0.006)
Input use			
Pesticide adopter	−0.037 (0.023)	−0.003 (0.021)	**0.162*** (0.083)
Improved maize seed adopter	−0.003 (0.006)	**−0.028*** (0.011)	−0.021 (0.019)
Inorganic fertilizer adopter	−0.003 (0.007)	0.008 (0.007)	−0.014 (0.016)

Note: Coefficients represent differential effects relative to never-treated areas. Standard errors in parentheses. *** Statistically significant at 99% confidence level, ** Statistically significant at 95% level, * Statistically significant at 90% confidence level. Fully saturated model available on request.

6 Conclusion

This study provides empirical evidence on the impact of road infrastructure on agricultural productivity in Malawi. Using three large longitudinal nationally representative surveys spanning 10 years and multiple quasi-experimental approaches, the study finds that proximity to paved roads significantly increases the average value product of land, with estimated treatment effects ranging from 35% to 57% over a 10 year period. This suggests that the benefits of road infrastructure strengthen over time as market integration deepens. Robustness checks using instrumental variables suggest impacts that are strongly positive. Results align with Von Thünen's classic land use model and with central place theory, suggesting that improved connectivity fosters economic productivity. While the broad importance of roads is well established, this study advances the literature by (i) employing recent advances in difference-in-difference methodology that relax stricter identifying assumptions, (ii) documenting dynamic, long-term effects rather than short-run impacts, and (iii) providing detailed evidence from a data-scarce sub-Saharan African context using nationally representative surveys.

While the paper does not consider the costs of building new roads, which vary depending on location, terrain, and distance from other existing infrastructure, and which critically determine their cost-effectiveness, the results reinforce the general policy imperative of targeted road investments—especially in remote, high-potential agricultural regions—to facilitate commercialization, improve market access and efficiency, and ultimately reduce rural poverty. Other analytical tools such as computable general equilibrium (CGE) and spatial models could complement this strand of inquiry, shedding light on the economywide benefits and costs of investments once market feedback effects are considered or informing the spatial placement of new road infrastructure.

References

Abate, G. T., Dereje, M., Hirvonen, K., & Minten, B. (2020). Geography of public service delivery in rural Ethiopia. *World Development*, *136*, 105133. https://doi.org/10.1016/j.worlddev.2020.105133

Adamopoulos, T. (2011). Transport costs, Agricultural productivity, and cross-country income differences. *International Economic Review*, *52*(2), 489–521. https://doi.org/10.1111/j.1468-2354.2011.00636.x

African Development Bank. (2022). *Project appraisal report SADC sub-regional transport and trade facilitation project*.

Ali, D. A., Deininger, K., & Harris, A. (2016). Large farm establishment, smallholder productivity, labor market participation, and resilience: evidence from Ethiopia. *World Bank Policy Research Working Paper*, *7576*.

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

Amare, M., Jensen, N. D., Shiferaw, B., & Cissé, J. D. (2018). Rainfall shocks and agricultural productivity: Implication for rural household consumption. *Agricultural Systems*, *166*, 79–89. https://doi.org/10.1016/j.agsy.2018.07.014

Angrist, Joshua. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Asher, S., & Novosad, P. (2020). Rural roads and local economic development. *American Economic Review*, *110*(3), 797–823. https://doi.org/10.1257/aer.20180268

Baker, A. C., Larcker, D. F., & Wang, C. C. Y. (2022a). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, *144*(2), 370–395. https://doi.org/10.1016/j.jfineco.2022.01.004

Baker, A. C., Larcker, D. F., & Wang, C. C. Y. (2022b). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, *144*(2), 370–395. https://doi.org/10.1016/j.jfineco.2022.01.004

Banda, H. K. (1973). 'Our position now is even better….' *World Development*, *1*(3–4), 13–28. https://doi.org/10.1016/0305-750X(73)90058-2

Benson, T., Chamberlin, J., & Rhinehart, I. (2005). An investigation of the spatial determinants of the local prevalence of poverty in rural Malawi. *Food Policy*, *30*(5–6), 532–550. https://doi.org/10.1016/j.foodpol.2005.09.004

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Chinsinga, B. (2017). The Green Belt Initiative, politics and sugar production in Malawi. *Journal of Southern African Studies*, *43*(3), 501–515. https://doi.org/10.1080/03057070.2016.1211401

Christaller, W. (1933). *Die zentralen Orte in Suddeutschland*. Fischer.

Dercon, S., & Gollin, D. (2014). Agriculture in African development: Theories and strategies. *Annual Review of Resource Economics*, *6*(1), 471–492. https://doi.org/10.1146/annurev-resource-100913-012706

Djoumessi, Y. F. (2021). What innovations impact agricultural productivity in sub-Saharan Africa? *Journal of Agriculture and Food Research*, *6*, 100228. https://doi.org/10.1016/j.jafr.2021.100228

Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, *108*(4–5), 899–934. https://doi.org/10.1257/aer.20101199

Donaldson, D., & Hornbeck, R. (2016). Railroads and American economic growth: A "market access" approach. *The Quarterly Journal of Economics*, *131*(2), 799–858. https://doi.org/10.1093/qje/qjw002

Dorosh, P., & Thurlow, J. (2018). Beyond agriculture versus non-agriculture: Decomposing sectoral growth–poverty linkages in five African countries. *World Development*, *109*, 440–451. https://doi.org/10.1016/j.worlddev.2016.08.014

Dorosh, P., Wang, H. G., You, L., & Schmidt, E. (2012). Road connectivity, population, and crop production in sub-Saharan Africa. *Agricultural Economics*, *43*(1), 89–103. https://doi.org/10.1111/j.1574-0862.2011.00567.x

Frisvold, G. (1995). Sources of agricultural productivity growth and stagnation in sub-Saharan Africa. *Agricultural Economics*, *13*(1), 51–61. https://doi.org/10.1016/0169-5150(95)01143-9

Fuglie, K. (2018). R&D capital, R&D spillovers, and productivity growth in world agriculture. *Applied Economic Perspectives and Policy*, *40*(3), 421–444. https://doi.org/10.1093/aepp/ppx045

Getis, A., & Getis, J. (1966). Christaller's central place theory. *Journal of Geography*, *65*(5), 220–226. https://doi.org/10.1080/00221346608982415

Government of Malawi. (2017a). *Malawi National Transport Master Plan*.

Government of Malawi. (2017b). *The Malawi Growth and Development Strategy (MGDS III) (2017–2022)*.

Government of Malawi. (2021). *Malawi's vision: An inclusively wealthy and self reliant nation | Malawi 2063*.

Hawonga, T., & Misook, L. (2018). Factors affecting the performance of agricultural project from the perspectives of agricultural extension workers: A case study of Malawi. *Journal of Agricultural Extension & Community Development, 25*(2), 111–120. http://doi.org/10.12653/jecd.2018.25.2.0111

Horvath, R. J. (1967). Von Thünen and urban sprawl. *Annals of the Association of American Geographers*, *57*(4), 811–812. https://doi.org/10.1111/j.1467-8306.1967.tb00638.x

Jayne, T. S., Mather, D., & Mghenyi, E. (2010). Principal challenges confronting smallholder agriculture in sub-Saharan Africa. *World Development*, *38*(10), 1384–1398. https://doi.org/10.1016/j.worlddev.2010.06.002

Jayne, T. S., & Sanchez, P. A. (2021). Agricultural productivity must improve in sub-Saharan Africa. *Science*, *372*(6546), 1045–1047. https://doi.org/10.1126/science.abf5413

Jedwab, R., & Moradi, A. (2016). The permanent effects of transportation revolutions in poor countries: Evidence from Africa. *Review of Economics and Statistics*, *98*(2), 268–284. https://doi.org/10.1162/REST_a_00540

Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2010). *Handbook on Impact Evaluation*. The World Bank Group.

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*(1), 97–126. https://doi.org/10.1016/j.jeconom.2005.02.011

Lall, S. V., Wang, H., & Munthali, T. (2009). *Explaining high transport costs within Malawi: Bad roads or lack of trucking competition?* (5133) World Bank Policy Research Working Papers.

Lee, H. B., McNamara, P. E., & Ho, H. (2023). Road accessibility and agricultural extension services in Malawi. *Agriculture & Food Security*, *12*(1), 3. https://doi.org/10.1186/s40066-023-00410-y

Mccracken, J. (1983). Planters, peasants and the colonial state: The impact of the native tobacco board in the central province of Malawi. *Journal of Southern African Studies*, *9*(2), 172–192.

Michalopoulos, S., & Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary African development. *Econometrica*, *81*(1), 113–152. https://doi.org/10.3982/ECTA9613

Minten, B., & Kyle, S. (1999). The effect of distance and road quality on food collection, marketing margins, and traders' wages: evidence from the former Zaire. *Journal of Development Economics*, *60*(2), 467–495. https://doi.org/10.1016/S0304-3878(99)00049-8

National Statistical Office. (2019). *2018 Malawi Population and Housing Census Main Report*.

National Statistical Office. (2020). *Malawi Fifth Integrated Household Survey (IHS5): Basic Information Document*.

Nguyen, M. (2025). *Foundations of Data Analysis*. Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-01858-8

OECD/FAO. (2016). *OECD-FAO Agricultural Outlook 2016-2025*. OECD. https://doi.org/10.1787/agr_outlook-2016-en

Ouma, E., Jagwe, J., Obare, G. A., & Abele, S. (2010). Determinants of smallholder farmers' participation in banana markets in Central Africa: The role of transaction costs. *Agricultural Economics*, *41*(2), 111–122. https://doi.org/10.1111/j.1574-0862.2009.00429.x

Parr, J. B. (2015). Overlooked aspects of the Von Thünen system. *Spatial Economic Analysis*, *10*(4), 471–487. https://doi.org/10.1080/17421772.2015.1076577

Ragasa, C., & Mazunda, J. (2018). The impact of agricultural extension services in the context of a heavily subsidized input system: The case of Malawi. *World Development*, *105*, 25–47. https://doi.org/10.1016/j.worlddev.2017.12.004

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, *94*(5), 1002–1037. https://doi.org/10.1086/261420

Savastano, S., & Scandizzo, P. L. (2017). *Farm Size and Productivity A "Direct-Inverse-Direct" Relationship*. http://econ.worldbank.org.

Shamdasani, Y. (2021). Rural road infrastructure and agricultural production: Evidence from India. *Journal of Development Economics*, *152*, 102686. https://doi.org/10.1016/j.jdeveco.2021.102686

Stifel, D., & Minten, B. (2008). Isolation and agricultural productivity. *Agricultural Economics*, *39*(1), 1–15. https://doi.org/10.1111/j.1574-0862.2008.00310.x

United Nations. (2025). *Do you know all 17 SDGs?* Https://Sdgs.Un.Org/Goals.

Vail, L. (1975). The making of an imperial slum: Nyasaland and its railways, 1895-1935. In *Source: The Journal of African History* (Vol. 16, Issue 1). https://www.jstor.org/stable/181100

Von Thünen, J. H. (1826). Der isolirte staat in Beziehung auf landwirshaft un Nationaloekonomie. *Wirschaft Und Finanz*.

Von Thünen, J. H., & Hall, P. (1966). *Von Thünen's (The) Isolated State (in its Relation to Agriculture and National Economy)*. Pergamon Press.

Walker, R. T. (2022). Geography, Von Thünen, and Tobler's first law: Tracing the evolution of a concept. *Geographical Review*, *112*(4), 591–607. https://doi.org/10.1080/00167428.2021.1906670

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). South-Western CENGAGE Learning.

Zant, W. (2018). Trains, trade, and transaction costs: How does domestic trade by rail affect market prices of Malawi agricultural commodities? *The World Bank Economic Review*, *32*(2), 334–356. https://doi.org/10.1093/wber/lhx011

ALL IFPRI DISCUSSION PAPERS

All discussion papers are available **here**

They can be downloaded free of charge

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
www.ifpri.org

IFPRI HEADQUARTERS
1201 Eye Street, NW
Washington, DC 20005 USA
Tel.: +1-202-862-5600
Fax: +1-202-862-5606
Email: ifpri@cgiar.org