

Better roads, better off? Evidence on improving roads in Tanzania

Christelle Dumas*

Ximena Játiva†

May 14, 2019

Preliminary and incomplete - please do not cite

Abstract

Spatial isolation is considered as one of the main determinants of poverty. Therefore, many transport investments are undertaken with a stated objective of poverty reduction. In our paper, we evaluate a Tanzanian program that rehabilitated 2500km of major roads between 2008 and 2013. We deal with endogenous placement issues with a household fixed-effect strategy combined with a propensity score matching to strengthen the common trend assumption. Contrary to most studies, we find damaging effects of the road on the rural population: we observe a reduction in life satisfaction, a reallocation of labor away from farm without many additional opportunities of work, that we trace back to a decrease in the price of the main product (rice). This is consistent with predictions obtained from trade models where rural households, with lower transaction costs, actually face competition from lower price goods.

*University of Fribourg, Bd de Perolles 90, 1700 Fribourg, Switzerland. Email: christelle.dumas@unifr.ch.

†University of Fribourg, Bd de Perolles 90, 1700 Fribourg, Switzerland. Email: ximena.jativa@unifr.ch

1 Introduction

Eighty percent of the extreme poor and seventy five percent of the moderate poor live in rural areas (Castaneda et al., 2016). Therefore, isolation is considered as one of the main contributors to poverty. According to the World Bank’s Rural Access Index, one billion rural dwellers do not have reliable transport access and only 34 percent of the rural Sub-Saharan African population live within two kilometers of an all-season road. Spatial isolation can impose serious constraints on agricultural production, access to health, education and work opportunities. As a consequence, infrastructure investment has been considered for a long time as a key aspect of development policy.¹ However, isolation also limits the competition from external producers and therefore might allow the subsistence of non-competitive rural farmers. The construction of transport infrastructures could well have damaging impacts on small-scale farm producers, at least in the short run on vulnerable populations. This paper identifies the adjustments made by households in an array of dimensions after a road rehabilitation program in Tanzania.

A growing literature seeks to identify the consequences of improved access to transportation infrastructure, be they railroads, highways or rural roads. Most of the literature concludes to positive effects of a better access. Better connectivity to other locations has been shown to reduce transport costs (Khandker et al., 2009; Donaldson, 2018), lower prices and increase the availability of non-local goods (Aggarwal, 2018; Khandker et al., 2009), increase the use of productivity-enhancing inputs such as fertilizers or hybrid seeds (Shamdasani, 2016; Shrestha, 2019), increase agricultural production (Khandker et al., 2009), reallocate labor from agriculture to urban activities (Asher and Novosad, 2018), facilitate agricultural specialization and market-orientation of farm activity (Qin and Zhang, 2016; Shamdasani, 2016), and increase income (Donaldson, 2018; Qin and Zhang, 2016). Only a few find no or very limited effects (Banerjee et al., 2012). These results are consistent with the combination of two previous strands of literature: the analysis of agricultural households in presence of market imperfections (Singh et al., 1986; de Janvry et al., 1991) and the quantification of transaction costs (Renkow et al., 2004; Jacoby and Minten, 2009; Minten and Kyle, 1999; Harrison et al., 2014). Additional effects identified by the infrastructure literature are the following: younger children seem to spend more time enrolled in school (Aggarwal, 2018; Adukia et al., 2017; Khandker et al., 2009), local markets develop (Mu and van de Walle, 2011), and in the case of major roads or highways, firms are more likely to export (Volpe Martincus et al., 2017; Storeygard, 2016) and manufacturing activity is stimulated (Ghani et al., 2016). In addition, the few papers interested in identifying heterogeneous impacts of roads identify stronger (positive) effects for the poor population (Khandker et al., 2009; Lokshin and Yemtsov, 2005).

Increased specialization and income arising from reduced transaction costs is also a result obtained in trade models, which address a similar question at a broader scale. However, while many economists agree on the value of trade in terms of growth in the long-run,² there is also an ample literature documenting the damaging effects of greater trade in the short-run. For instance, Goldberg and Pavcnik (2007) document the increase in inequality associated to the commercial opening of countries who liberalized in the 90’s. The arguments put forward to explain such

¹The World Bank Transport Business Strategy for 2008-2012 stated: “One of the best ways to promote rural development is to ensure good accessibility to growing and competitive urban markets.” (The World Bank, 2008, p. 3)

²Exceptions are Rodrik (2017); Williamson (2011).

damaging effects are the following. First, commercial opening changes all the vector of relative prices. Given that consumers differ in their consumption bundle and in their sector of activity, the whole population is not affected homogeneously (Porto, 2006). In particular, rural households may well suffer from increased price competition instead of benefiting from a greater market access. Second, at odds with the canonical trade models (Heckscher-Ohlin and Stolper-Samuelson), the factors are not perfectly mobile within the country. A host of evidence shows that, in developing countries, the sectoral reallocation is extremely limited.³ However, Haltiwanger et al. (2004) shows that within sector, labor might be reallocated towards more efficient plants.

The common ground between transport infrastructure investment and commercial liberalization is the reduction in transaction costs that they entail. It is thus surprising that the two literatures come to different conclusions. The theoretical analysis of the effects is of little guidance, unfortunately. As previously mentioned, the reduction in transaction costs changes the price of several goods, which then induce positive or negative changes in welfare. One approach consists in assessing the change by a first-order approximation based on the change in price and the net quantity consumed (Deaton, 1997). Households who are net consumers of the goods whose prices decline benefit from the change, while the net producers benefit from increases in the prices. This first-order approximation consists in neglecting the allocation changes made by households following the change in the price vector. However, Friedman and Levinsohn (2002) find that allowing for household responses roughly halves the welfare losses predicted by the first-order formulation. From a theoretical perspective, this suggests that a general equilibrium model may be needed to analyze the effect of a change in transaction costs, but many different predictions could be obtained depending on the assumptions made for the price elasticities of demand and supply.

In this paper, we evaluate the effects of a program that by 2013 upgraded 2'500km of major roads in Tanzania on: agricultural decisions, market participation for labor and products, prices, migration, consumption and individuals' life satisfaction. We combine a household panel data with a GIS dataset listing changes in road infrastructure in the country since 2008. This allows us to identify the effect of roads controlling for household fixed effects. In order to guarantee that the within household change would be similar among treated and non-treated households in absence of the treatment, we build a comparison group of the treated households by propensity score matching. We then assess in greater detail changes in employment using two rounds of the Labor Force Survey and a similar methodology.

Our paper is a valuable addition to the existing roads literature for the following reasons. First, the existing research is either on the impact of rural roads, i.e. roads that reach villages, or of railroads and has focused mainly on Asia or the United States. Little has been done on major roads, to the exception of Storeygard (2016); Ghani et al. (2016); Shrestha (2019) and few papers have focused on Sub-Saharan Africa (Casaburi et al., 2013; Burgess et al., 2015; Gachassin et al., 2010)⁴. There are few reasons to expect similar consequences from the improvement of rural roads and major roads. For instance, rural roads could well improve access to other villages and therefore improve the spatial integration of the labor market,

³See for instance, Revenga (1997); Harrison and Hanson (1999); Attanasio et al. (2004); Currie and Harrison (1997); Topalova (2004). Congruent with this conclusion is the evidence, in some countries, that spatial mobility is limited (Munshi and Rosenzweig, 2016).

⁴These papers address questions related to prices in rural markets, ethnic favoritism and poverty reduction.

while a major road would fail to have such consequences, simply because accessing the major road is too time-consuming. Covertly, main roads might trigger a larger reduction in the price of goods than rural roads. Insofar as the Tanzanian government decided to fight poverty using the instrument of major roads, it is important to assess whether this type of roads also fosters economic growth in rural areas. Second, we acknowledge the variety of adjustments households might implement in reaction to their new environment by using a wide range of outcomes rather than focusing on one dimension, compared to most papers in the road literature. This allows us to picture a situation where rural households, instead of reaching new markets, actually face new competitors and have to reduce their farm activity, increase off-farm work and see their life satisfaction decrease. This result is very different from the ones obtained previously in the literature and is in our view extremely important. This leads us to the fourth and main reason. If there are negative consequences of roads on poor rural households, even if they are transitory, then we need to be aware of them in order to design safety nets that would help these households to transition to their new economic environment.

The article is structured as follows. Section 2 describes the road improvements in Tanzania during the period 2008-2013 and section 3 presents the data. Section 4 describes the identification strategy and section 5 provides some preliminary results. Section 6 concludes the paper.

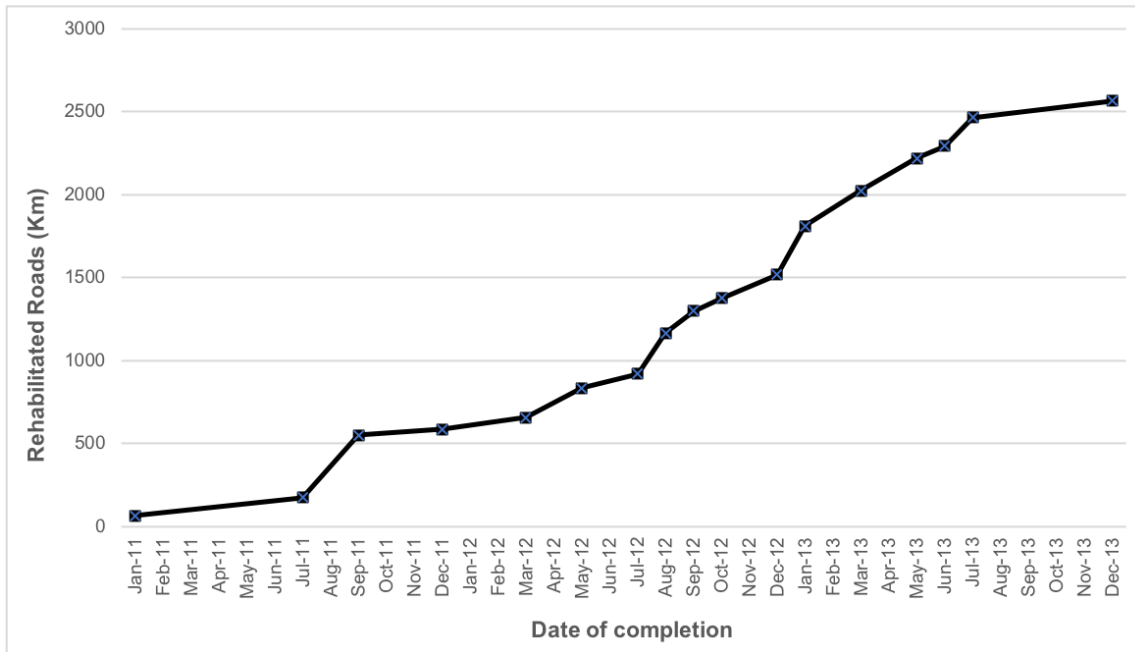
2 The road construction program

The road construction program takes place in a context where the Tanzanian infrastructure is considered as extremely poor: the country ranked 118th out of 134 economies in the infrastructure dimension of the World Economic Forum's Global Competitiveness Index in 2008-2009 (Schwab and Porter, 2008). This is due to both a low road density (96.5 meters per square kilometer compared to 296.95 in Kenya and Uganda) and to a poor condition of the road network: only 36.63% of road network was paved or sealed and was classified as in "good or fair condition" (Government of Tanzania, 2008).

The 2005 National Strategy for Growth and the Reduction of Poverty in Tanzania (MKUKUTA) identified the poor condition of the rural road network as one of the major impediments to growth and poverty reduction: insufficient and poor quality infrastructure hampers access to labor and input markets, hinders agricultural productivity, and generates significant losses for the rural poor. Therefore, the plan emphasized that investing in adequate road infrastructure had the potential to boost the rural economy, promote growth and bring down the level of rural poverty.

In this light, the Government of Tanzania adopted two important measures. First, the government launched the "10 Year Transport Sector Investment Programme (TSIP)", which identified nine road transport corridors (about 10'300 km of roads) that were key to enhance integration of regions within the country, as well as better connect the country to its neighbors. During the ten years of the program, all roads in the nine development corridors that were not in good condition were planned to be upgraded to paved standards. The program was launched with the goal of upgrading and rehabilitating approximately 6'000 km of roads in a period of 10 years. The Phase I of the TSIP was scheduled to be implemented between 2007/2008 and 2012/2013. Second, the "Roads Act" was launched in 2007, to establish the legal framework for the implementation of road investment (Government of Tanzania, 2007). This act gave the Tanzania National Roads Agency

Figure 1: Roads Rehabilitated (km) 2008-2013



(TANROADS) the responsibility to improve the Trunk and Regional Road Network, estimated to be 33'891 km of the 86'472 km total classified road network in Tanzania Mainland. The remaining network of Urban, District, and Feeder Roads is managed by the Local Government Authorities (LGA). The timing of these two measures made that the infrastructure works started after 2008. The second phase was scheduled to start in 2012-13 and end in 2016-17. However, a review of recent data on public works show that it has actually been postponed.⁵

The roads to be upgraded were selected to address three policy goals. First, promote economic growth. Second, improve economic wellbeing of the rural poor by enhancing production and market exchanges of both subsistence and cash crops. And third, take advantage of the geographic location of the country vis-à-vis of its landlocked neighboring countries.⁶

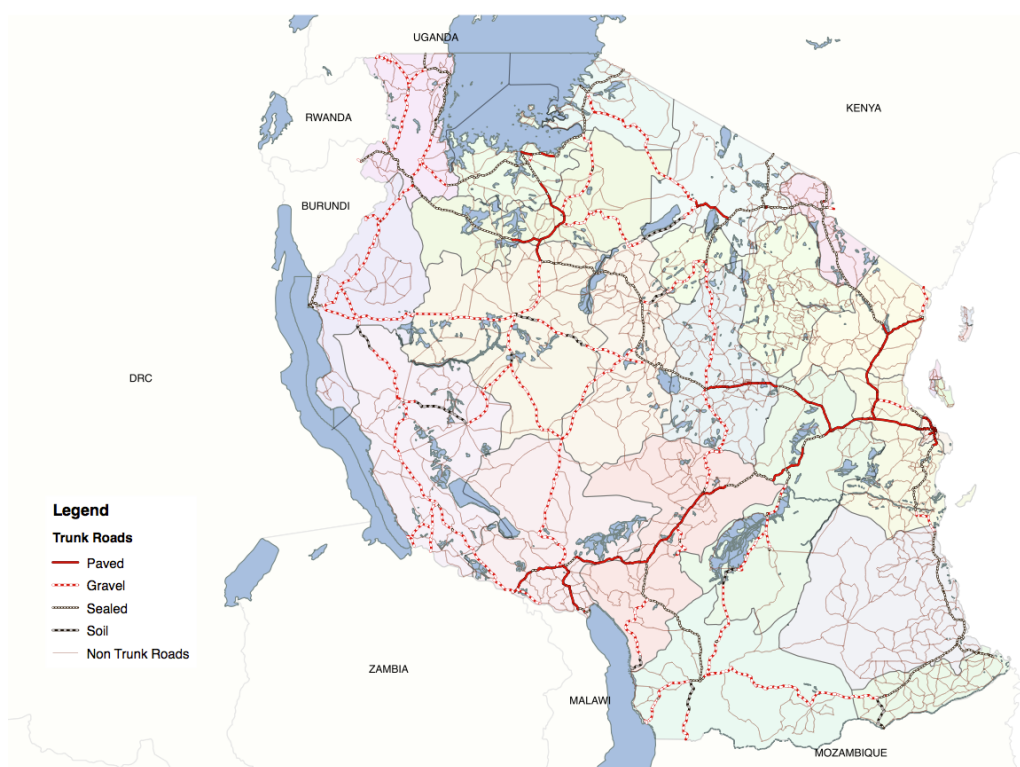
Importantly for this paper, the above-mentioned measures set up the guidelines to be considered to determine the allocation of funds and prioritization of road projects: fully funded projects related to maintenance of roads in bad condition were given first priority. Second, road segments that would enhance exports and key food crop production had to be given preference. Third, all regional centers should be linked with paved roads, and district headquarters should be connected with all-weather roads of at least gravel standards. Last, the projects should open up all major population areas to modern economy and trade. We will proxy these guidelines by a set of geo-referenced variables.

Figure 1 shows that by the end of 2013, the road network showed significant improvements: 2'564 km of roads were rehabilitated to paved conditions. Figure 2 shows the state of the road network before (panel 2a) and after (panel 2b) the infrastructure works. Prior to the rehabilitation program the majority of the road network was either graveled (49.7%) or sealed (30.5%). Only 11% of the total main network was paved and in good condition, whereas at the end of 2013 this

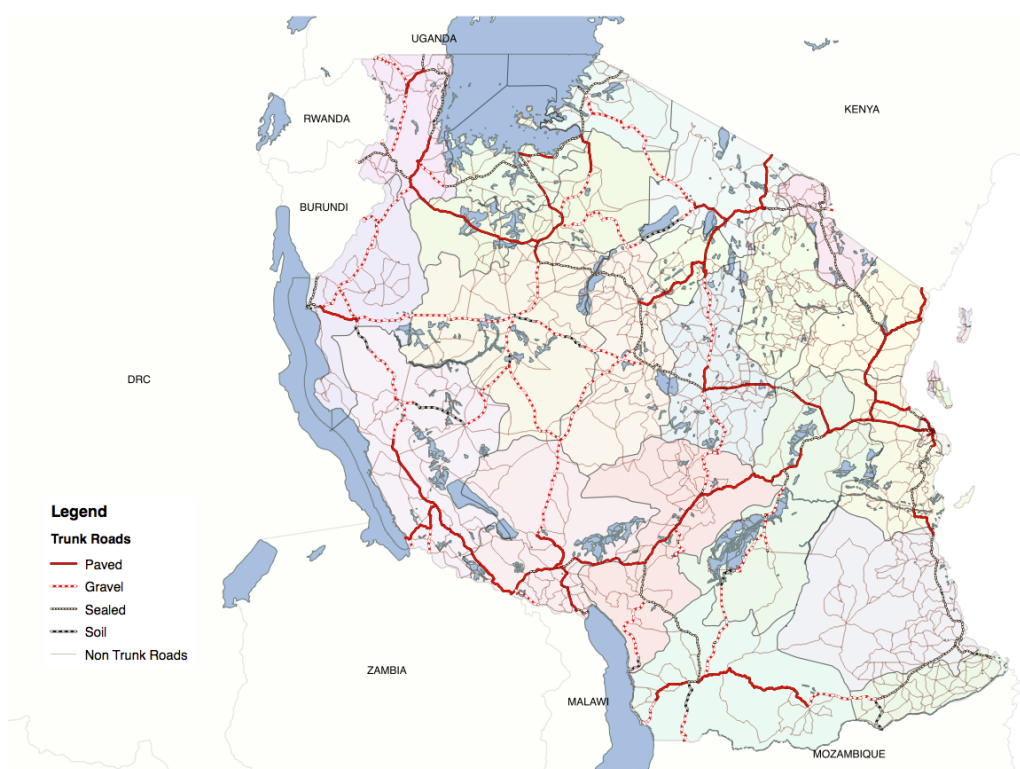
⁵Tanzania National Roads Agency, <http://tanroads.go.tz>

⁶Tanzania is an important transit gateway for six landlocked countries of Southern and Central Africa: Malawi, Zambia, DRC, Burundi, Rwanda, and Uganda.

Figure 2: Road Rehabilitation 2008-2013



(a) Road Network 2008



(b) Road Network 2013

percentage rose to 29.4%. As a result, Tanzania moved from the 118th to the 90th place of the infrastructure dimension of the Global Competitiveness Index in

2016-2017 (Schwab and Sala-i Martin, 2016).

The objective of this paper is to evaluate the consequences of this major roads' improvement program on the reduction of poverty and improvement of opportunities for Tanzanian rural households.

3 Data

We mostly use three sources of data: the detailed geolocated information on roads obtained from TANROADS, a household panel survey and two rounds of the Labor Force Survey.

3.1 TANROADS Data

The Tanzanian Road Agency has collected a geo-referenced dataset on road condition and rehabilitation, which we were able to access. It covers the entire road network and contain lines of primary and secondary roads, as well as information about their length, surface condition (paved, sealed, gravel and soil type), infrastructure projects, type of infrastructure project and date of completion. For the purpose of this study we are mostly interested in Trunk Roads for which rehabilitation works started and were completed between 2008-2013. For any point in the map, we can compute three variables: the geographic distance to the nearest road prior the intervention, the quality of this road (graveled, paved, sealed or soil), and the distance to the closest treated road.

3.2 LSMS-ISA Data

The LSMS-ISA (Living Standards Measurement Study - Integrated Surveys on Agriculture) data for Tanzania is a panel of three rounds of a nationally representative household sample, collected by the Tanzania's National Bureau of Statistics (NBS). The survey is made of 3'265 households in the first wave, clustered in 410 enumeration areas across mainland Tanzania. 258 of these enumeration areas are rural. The three first rounds were collected in 2008-2009, 2010-2011 and 2012-2013.⁷ The last round, collected in 2014-15⁸, draws 3'352 new households clustered in 419 enumeration areas, so as to ensure the representativity of the sample.⁹

These data are particularly suitable for our study. They have very detailed information on the household, agricultural and community dimensions. The wide variety of variables collected in the survey is important for the assessment of the consequences of the road rehabilitation since it may affect substantially the economic environment of the households. Most importantly, geo-location of villages is available, which allows us to know which households are treated and when, as well as match the data with a large set of variables obtained from other sources.

Table 1 summarizes our key variables, by reference period on which they are collected. Any information related to the agricultural activity of the household is collected over a specific long rainy season (roughly the period covering March to

⁷The first round was conducted from October 2008 to September 2009, the second from October 2010 to September 2011 and the third from October 2012 to November 2013.

⁸The collection took place from October 2014 to October 2015.

⁹Roughly a third of the original sample is re-interviewed but the data for this sub-sample is not yet released and in any case, would likely be too small to achieve identification. Several key variables are still missing from the release of the new dataset, which prevents us from extending our analysis to the fourth round, even as a new cross-section.

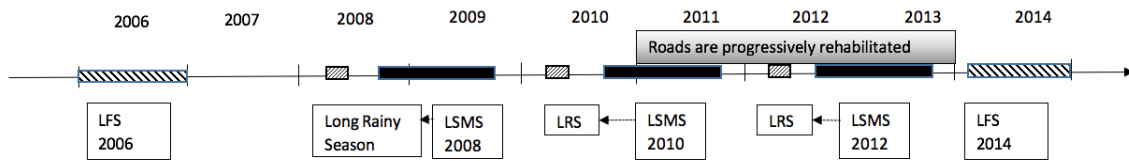
May), that is the same for all households in a given round. The survey collects any information related to inputs (market inputs, labor) and outputs (quantity, value) on each of the household plots. All information is aggregated at the household-year level. Information pertaining to market labor supply, household consumption, life satisfaction, and education choices are collected for a shorter reference period (varying from the last week for food consumption to the last year for large consumption items, for instance). Prices for food and standard goods are collected at the village level at the time of the survey. All variables used in the study are described in Appendix A.1. The difference in recall durations will impose constraints on the estimation. Figure 3 shows the timeline of the surveys.

Table 1: Outcome Variables

	Labor Market	Product Market	Welfare
Household	<i>Annual information</i> Off-farm Labor Hired-in Labor Household on-farm Labor <i>Last 7 days</i> Household wage labor	<i>Annual information</i> Harvest (kg, TSh)+ Sales (kg, TSh)+ <i>Last 7 days</i> Consumption(kg)+	<i>Annual information</i> PC real annual consumption Migration <i>Last 7 days</i> Subjective Welfare Durable assets Productive assets
Community		<i>Last 7 days</i> Prices <i>Annual information</i> Community projects	

+ By crop: rice, maize

Figure 3: Timeline of the surveys



Source: LSMS-ISA and LFS, author's computation.

3.3 LFS Data

We also use the Labor Force Survey (LFS) for the years 2006 and 2014 to further evaluate the effects of the program on labor market participation. These data are also of interest because they will allow to check the consistency of the results with the LSMS. The LFS is a nationally representative survey that gathered data on 16'445 households in 326 wards in 2006 and 11'473 in 363 wards in 2014. Each

village contains on average 28 households and approximately 52% of the sample is located in rural areas. We are able to geo-code these data to the ward level.¹⁰

3.4 Additional Data

We collected a set of geographic measures from multiple sources, in order to proxy for the main drivers of the policy. Table A1 presents the list of variables created as well as their corresponding sources. First, a list of major cities with more than 20'000 habitants according to the 2002 Tanzania Population Census, District Headquarters and Major Border Crossings were gathered from different sources in the form of a list and geo-referenced. Second, we obtained geo-referenced 2010 population density and IFPRI standardized Agroecological zones, both at 100mx100m resolution. Using village geographic coordinates from both the LSMS-ISA and the LFS surveys, we computed the euclidean distance to the nearest major city, district headquarter and border crossing.

We also build variables such as the labor market activity and education level in the LFS 2006 at the ward/district level that we then merge with the other datasets.¹¹ For example, to assess labor market activity, we compute the share of households in the ward who have at least one member providing wage work.

Finally, the data were also matched with rainfall data,¹² to control for productivity shocks. We check that the treatment does not correlate with shocks. Given that there is no such correlation, and that the estimates are the same with and without control for shocks, we simply provide the estimates without the additional control.

4 Estimation Strategy

We start describing the identification strategy for the LSMS panel and then explain how we adapt it to the LFS data.

4.1 Defining the treatment: timing and distance to roads

We exploit the fact that in 2008, no LSMS household was affected by the program. We therefore have a pre-treatment period. A small number of roads were rehabilitated before the second round of the panel and a much larger number of roads were rehabilitated before the third round of the panel.

However, the duration of the survey for each round exceeded one year and therefore the actual treatment depends on the exact date of interview of the household. More precisely, given that some outcomes are measured on a yearly basis, one needs to insure that the road was rehabilitated at least one year before the

¹⁰To do so, we use ward level boundary polygons from the 2002 and 2012 Tanzania Population Census obtained from the Tanzania Bureau of Statistics. We define centroids for all wards to obtain their geographic coordinates using Q-GIS.

¹¹We merge information at the ward level when the same ward can be found in the LFS 2006, if not we use the district-average information.

¹²Rainfall estimates are given by squares of roughly 10km x 10km. These data were obtained from http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.FEWS/.Africa/.DAILY/.ARC2/.daily/.est_prdp/datafiles.html. We compute the deviation from the mean in the area, divided by the standard deviation in the area using years 2001-2013. We construct two variables using rainfall from January to December of the current and previous year of the LSMS wave.

interview. While we could have allowed for lower-intensity treatment (not being treated for half of the year, and being treated for half of the year), we are also concerned of the potentially very disruptive effect of having works on the road. If the implementation of the program is disruptive during the works, then the no-treatment for half of the period is actually a negative treatment, with lower ability to use the transportation means, that would blur the estimated effect. Therefore, treatment is defined as such: for outcomes that are measured on a yearly basis, the road has to be finished at least one year before the interview; and for outcomes that are measured on lower-period basis (week, month), the road has to be finished at least a month before the interview.

The second difficult point in defining the treatment is to clarify who is treated and who is not when a road is rehabilitated. There is probably a continuous effect depending on the distance of the household to the road. However, households living further away from existing roads might substantially differ from the ones who are better connected. For this reason, we will control for household fixed-effects. Unfortunately, this does not guarantee that the evolution of these households would have been similar, in absence of the road rehabilitation. As a consequence, we want to implement a matching strategy to guarantee the comparability of treated and control villages. This requires to have a 0/1 treatment and to use a distance threshold for defining the treatment. However, we can perform the exercise for many thresholds and will do so for all consecutive kilometers between 20km and 50km. We cannot use a lower threshold because the estimation would rely on a too low number of treated households. Given that for yearly outcomes, the number of treated households is reduced, we face the same problem at an even lower threshold. For this reason, we only provide an assessment of the road program for treatments starting at 30km until 50km.

4.2 Estimation in the panel: Household fixed effect model with propensity score matching

A very nice feature of the panel is to allow controls for household fixed effects. We therefore compare the evolution of households who start benefiting from a rehabilitated road during the course of the panel to the evolution of other households. The strategy is therefore extremely close to a difference-in difference strategy except that we control for any unobserved household characteristic that is not time-varying. The results obtained from a standard DiD are similar to the ones presented in the paper.

However, a household fixed effect model assumes away different trends between treated and untreated households in absence of the treatment. This assumption is questionable, especially if treated and untreated households face different environment. We therefore use a propensity score matching (PSM) to build a control group that is similar to the treated group, which makes the assumption more likely to hold (Heckman and Navarro-Lozano, 2004; Smith and Todd, 2005; Caliendo and Kopeinig, 2008). Compared to a control strategy, PSM does not impose linear functional form restrictions in the estimation of the conditional expectation of the outcome variables. In addition, even if the PSM cannot take into account unobserved characteristics, this is a mild concern in our case given the use of household fixed effects.

4.2.1 Specification

Our specification is the following:

$$Y_{ivt} = \alpha + \delta_{1d}1(t = 2012)Road_{ivtd} + \delta_{0d}1(t = 2010)Road_{ivtd} + \theta Z_{ivt} + \phi_i + \beta_t + \epsilon_{ivt} \quad (1)$$

where Y_{ivt} is an outcome variable for household i in village v at time t . The outcome variables are listed in section 3. $Road_{ivtd}$ is a dummy indicating whether the household is located within a radius of d km of a treated road, d taking values from 20 to 50. We run regressions for all consecutive kilometers within that radius, i.e. distance thresholds, d . α is the constant term, β_t are the time fixed effects for each round of the survey, 2008, 2010 and 2012. Z_{ivt} is a vector of time varying household, farm and village characteristics. ϕ_i contains household fixed effects and ϵ_{ivt} is the error term (the standard errors are clustered at the village level). $1(it = 2012)$ takes the value one when the household is surveyed in 2012. As already pointed out, all households are pre-treatment in 2008, only a few of them are already treated in 2010 and the majority of treated households are observed in 2012. The parameter δ_{1d} is the effect of living less than d kilometers away from a road, once it has been rehabilitated and evaluated in 2012. We chose to discard the few households who are treated in 2010 (472 observations),¹³ so that the parameter δ_{0d} provides a test of the common trend assumption. Indeed, if treated and control households do not satisfy this criteria, then δ_{0d} will be significantly different from 0. As a result, our evaluation of the program mostly compares 2008 and 2012.¹⁴

4.2.2 Matching in the panel

We now turn to a discussion of the matching in the panel. First, given that we use a panel, we only need to match households based on their 2008 characteristics. As previously mentioned, there is little attrition in this panel. We construct a propensity score to be treated for each household since this allows us to use individual characteristics as predictors in the logit regression. For the estimation technique to be valid, we need to satisfy the unconfoundedness assumption, i.e. choose a set of covariates that explain both the treatment assignment and the outcome of interest (Heckman and Navarro-Lozano, 2004; Caliendo and Kopeinig, 2008).

First, we select households based on their “Distance to Nearest Major Road”. Indeed, we want to compare households who benefited from an improvement in their nearest road to those who did not, but had the same proximity to a road. We thus create “buffers” around roads that define which households might act as a control. By definition, treated households have to be in the buffer. For instance, if we assess the effect of living less than 30km away from a rehabilitated road, the households that are used in the control group also live less than 30km from a main road. As a result, we have as many buffers as treatments. Our procedure therefore prevents that we compare households close to a road to households living far away from a road. This set of potential controls being defined, we do a standard matching, based on observable characteristics. We also tried to simply use this distance to the nearest major road as a covariate in the matching but the common trend assumption was more often rejected and we decided to impose this constraint.

¹³More precisely, we exclude the households for which the short-period recall outcomes might be impacted by the road rehabilitation as early as 2010. No household has yearly outcomes impacted as early as 2010.

¹⁴We have also tried a different specification which keeps all the observations and where the treatment is being Post interacted with being close to rehabilitated road, and the results are similar.

For the matching, we use the main selection criteria for the improvement of roads in Tanzania. First, the state of the road prior 2008: only roads in bad condition were selected to be improved. This will be captured by three dummy variables on the quality of the closest road, i.e. gravel, paved and sealed.¹⁵ Second, the aim of the government of Tanzania was to improve trade and growth in the country by taking advantage of its geographic location vis-à-vis of its landlocked neighboring countries. We control for “Distance to the Nearest Border Crossing” to capture this effect. Third, the closest the household is to a major population center or a market increases the probability of having a road in good condition or, if that is not the case, to have the main road improved. Thus, we will control for the variables “Distance to the Nearest Population Center with 20’000 inhabitants”, “Distance to the Nearest District Headquarter” and “Distance to the Nearest Market”. The socio-economic importance of the region will be captured by a vector of village characteristics, including population density, agroecological zones and labor market availability. Finally, we include household and farm characteristics to find the best comparable households and to control for potential confounders influencing the outcome of interest. Table A2 and Table A4 in the Appendix present summary statistics of the variables used to predict the propensity score, as well as the propensity score estimation.

We use an Epanechnikov kernel matching procedure based on the estimated propensity scores. We choose this non-parametric matching estimation because it has the advantage of using more information to construct the counterfactual (Caliendo and Kopeinig, 2008). Compared to specification (1), it simply consists in running the household fixed effects regression with weights obtained from the logit, so that the control group is statistically comparable to the treatment group. We will report tests of common support, and balancing tests between treatment and control groups. We report two specifications for each outcome variable: the household fixed-effects with matching, and the household fixed-effects without matching but restrained on the common support. We do not report the pure household fixed-effects model (without any type of matching) between the common trend assumptions was more often rejected with this specification.

4.3 Estimation in the repeated cross-section: Difference-in-Difference with propensity score matching

The two rounds of the Labor Force Survey do not allow us to control for household fixed effects; however, we can still compare the evolution of treated and untreated households in a diff-in-diff strategy. We can control for constant characteristics at the group-level (the characteristics associated to living close to road that will be rehabilitated). The main difficulty compared to the panel consists in guaranteeing that the composition of the groups does not change too much between the two rounds. Using again a matching approach strengthens the validity of the identifying assumption. Here we need to perform two matches: one to create a control group for the pre-intervention wave in 2006, and another to create a control group for the post-intervention wave in 2014. The treatment is defined by distance to a rehabilitated road, as in the panel case.

The specification we implement in the LFS is the following:

$$Y_{ivt} = \alpha + \delta_d 1(t = 2014) Road_{ivtd} + \theta Z_{ivt} + \phi Road_{ivtd} + \beta_t + \epsilon_{ivt} \quad (2)$$

¹⁵Soil is the reference quality type variable.

where $Road_{ivtd}$ captures the average value of the outcome in the treated group. We use a similar set of controls Z_{ivt} as in the panel. Tables A3 and A5 present summary statistics of the variables used to predict the propensity score, as well as both waves propensity score estimations. As in the panel, we restrict the estimation to the region of common support and we perform balancing tests to ensure that treatment and control groups are not statistically different at their means.

5 Results

5.1 Matching results

We start by commenting the matching performed for the panel. We compute a propensity score for each household using a logit regression based on 45 baseline characteristics. Given the number of regressors in Table A4, the discussion here focuses on the estimated coefficient signs and significance rather than their magnitudes. The results suggest that geo-referenced controls, road and village characteristics are the best predictors for the treatment. Consistent with the program guidelines, households living closest to a graveled major road are more likely to have that road rehabilitated compared to soil roads, followed by paved and sealed roads. As stated in section 2, only roads in bad condition were selected to be improved and infrastructure works related to maintenance were given first priority. These results indicate that indeed roads that were in bad condition but didn't demand big infrastructure works were selected first¹⁶.

Furthermore, the closest a household is to a border crossing the higher the probability of having a road rehabilitated, in line with the government's objective of taking advantage of their location vis-à-vis of its bordering countries. Figure A1 shows that major roads connecting the country to its neighbors were prioritized. Likewise, households closer to a major population center are more likely to be treated. On the contrary, among measures of accessibility and market access, the closest a household is to a major road or to a market the less likely it is to be treated, conditional on the quality of the road. A few characteristics that may indicate higher living standards seem to have a negative effect on the probability of treatment, namely, durable and productive assets, a higher share of households with at least a member wage employed and the presence of a bank branch within 10km.

We evaluated the fulfillment of the common support condition separately for each treatment. Figure A1 displays the common support for the treatment at 20km¹⁷ and Figure A2 displays the spatial location of treated and control households on the common support, as well as the observations that are not used in the analysis for the same treatment. For the rest of the analysis, we limit the sample to the common support.

We carried out a balancing test of the difference in the means of the covariates between the control and the treated groups after the matching. Table A6 shows the quality of that matching. We achieve a perfect balancing on the baseline characteristics¹⁸ and we find no significant differences between treated and control groups

¹⁶Gravel roads are more likely than paved and sealed roads to be in bad condition due to weather and traffic volumes, but less than soil roads.

¹⁷We satisfied the common support condition for all treatments levels. Results are available from authors upon request.

¹⁸Results for all the other treatments for both the panel and the cross-section case are available from authors upon request.

based on observables. We will postpone the discussion on the balancing test results for the outcomes of interest.

For the repeated cross-section, we follow a similar methodology. Table A5 provides the results for the matching performed for guaranteeing the comparability of the households for the 2006 and 2014 waves. We use a total of 26 regressors, and the same set of geo-coded variables as for the panel case; but given that we can only geolocate households at the ward level, we lose some precision in the matching. However, we achieved an almost perfect balancing of the initial observed characteristics in both waves and a perfect matching for all outcome variables. The results are shown in Table A7. Finally, Figure A3 presents the results for the common support; this condition is achieved for all treatment levels in 2006 and 2014.

5.2 Panel results

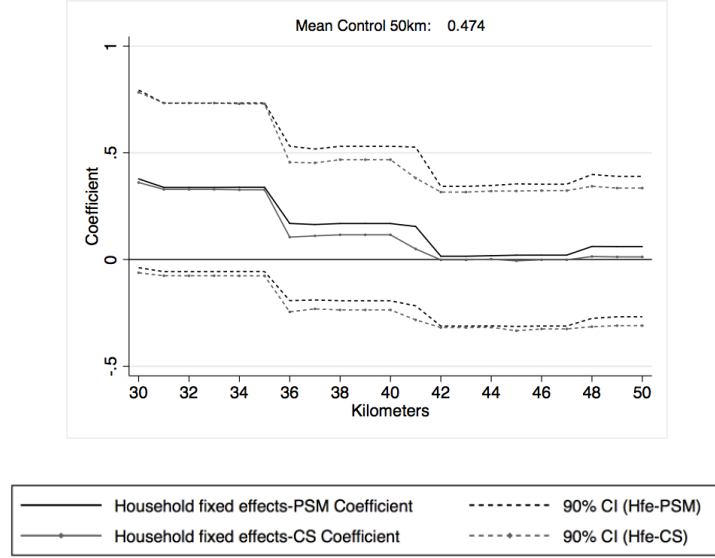
We start by assessing whether roads triggered more migration. This could be a potential threat to our identification strategy if better roads induce selective attrition in the panel. In addition, it is interesting to assess whether better roads induce changes where the subsistence means are obtained. However, traveling along a road of poor quality is unlikely to be a substantial share of the migration costs. Therefore, we expect that it has little effect except for temporary migration. The next figures display the estimates for each treatment cut-off (ranging from living less than 20km away from the road to living less than 50km away from the road). The provided estimates are the household fixed effects with propensity score matching and the household fixed effects without propensity score matching but on the common support. The two confidence intervals are at the 90% level. Figure 4 provides the effect of the treatment on the number of household members away at least one month in the previous year. The average of this variable in the control sample of households living less than 50 kilometers away from the road is 0.478. The figure shows that there is little evidence that better roads induce migration, and definitely not for households living more than 36 kilometers away from the road. The figure A5a provides the test of the common trend assumption (i.e. the δ_{0d} coefficients). It shows that the common trend assumption might be violated in the instance of the migration variable: households living less than 40 kilometers away from the road display a higher likelihood to migrate and this likelihood has increased as early as 2010. From now on and for the sake of brevity, we will only refer to the test of the common trend assumption when they seem to be rejected but the full set of results is available in the Appendix.

We now turn to a general and subjective welfare measure to assess whether households valued their access to a better road. The subjective welfare measure is based on a 7-point ladder question asked to household members. The exact wording of this question is “*How satisfied or dissatisfied would you say you are with your life?*”. The outcome variable used in this section corresponds to the average subjective welfare among all respondents in the household.¹⁹ Figure 5 shows that households who obtained a better road declare a *lower* life satisfaction. The decrease is actually quite substantial, since from an average of 3.7 it decreases by 0.5 (-13.5%) for households living less than 20 kilometers away from the road. The decrease is significantly different from zero for almost all treatments. However, the fact that it becomes non significantly different from zero for a treatment larger than 45 suggests that the households living further away from the road might

¹⁹As an alternative measure, we use the subjective welfare of the head of the household. Results are similar.

Figure 4: Migration

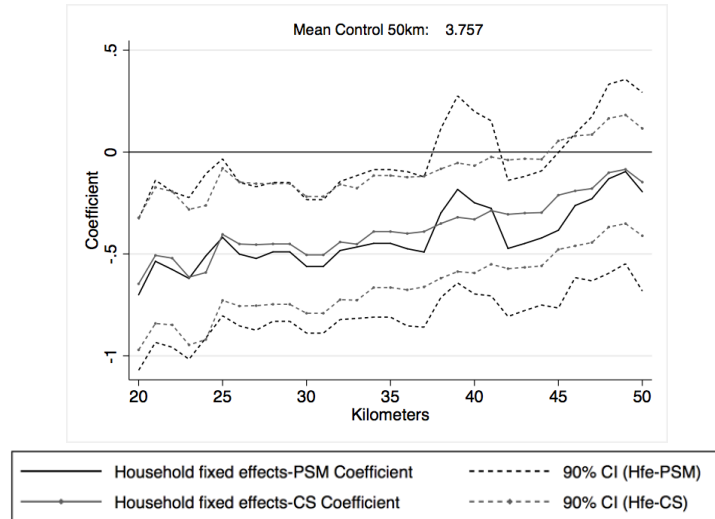
N. members away at least 1 month last 12m



actually have a no effect or a positive effect, which compensates for the loss for the households closer to the road. We will now try to understand where this reduction in life satisfaction might come from. One possibility is that roads trigger changes in the price system, which affects real income.

Figure 5: Subjective Welfare

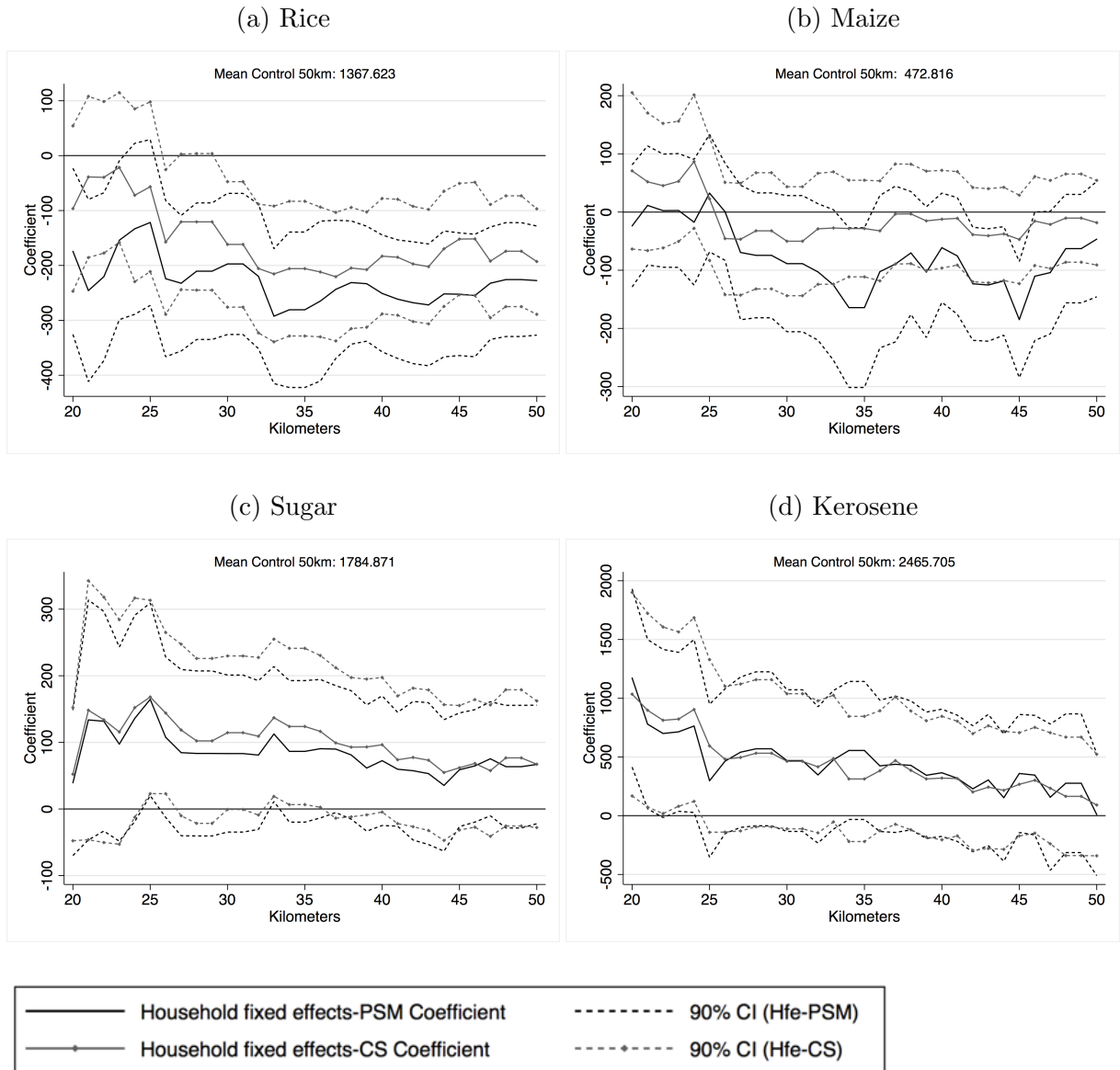
Life Satisfaction



We start by reviewing changes in the price of goods. We only have one observation per village for prices and only the price of available goods is reported. Therefore we need to focus on common goods. Rice and maize are by far the most traded products and they are the products for which we have enough observations to do the evaluation. Figure 6a displays a clear decline in the price of rice, for all treatment thresholds, and the decrease is very substantial compared to the mean. The point estimates for maize also tend to be negative (Figure 6b) but the coefficients are much less often significant. A decrease in prices would have a

negative impact on welfare for net producers and a positive impact on welfare for net consumers. Insofar as rice and maize are commonly produced by Tanzanian farmers, this decrease in rice price could explain the lower satisfaction. However, if transaction costs are lower, then other imported goods might also be cheaper (though it might not be automatically the case due to general equilibrium effects). We therefore check whether prices are generally lower, which presumably would limit the negative effect on welfare for producers. Figures 6c and 6d show that the price of sugar and of kerosene are not negatively impacted by road improvement. Therefore, the decrease in the price of produced goods might be a good explanation for the lower declared welfare.

Figure 6: Market Prices (TSh)



In order to confirm this interpretation, we check the effect of roads on harvest, sales and consumption for these two goods. Figure 7, first line, shows that the harvest for rice tends to be lower, but this is not significant at the 10% level, that the sales are significantly lower for all treatments and that the consumption does not change. All the variables are expressed in kilograms and not in value and

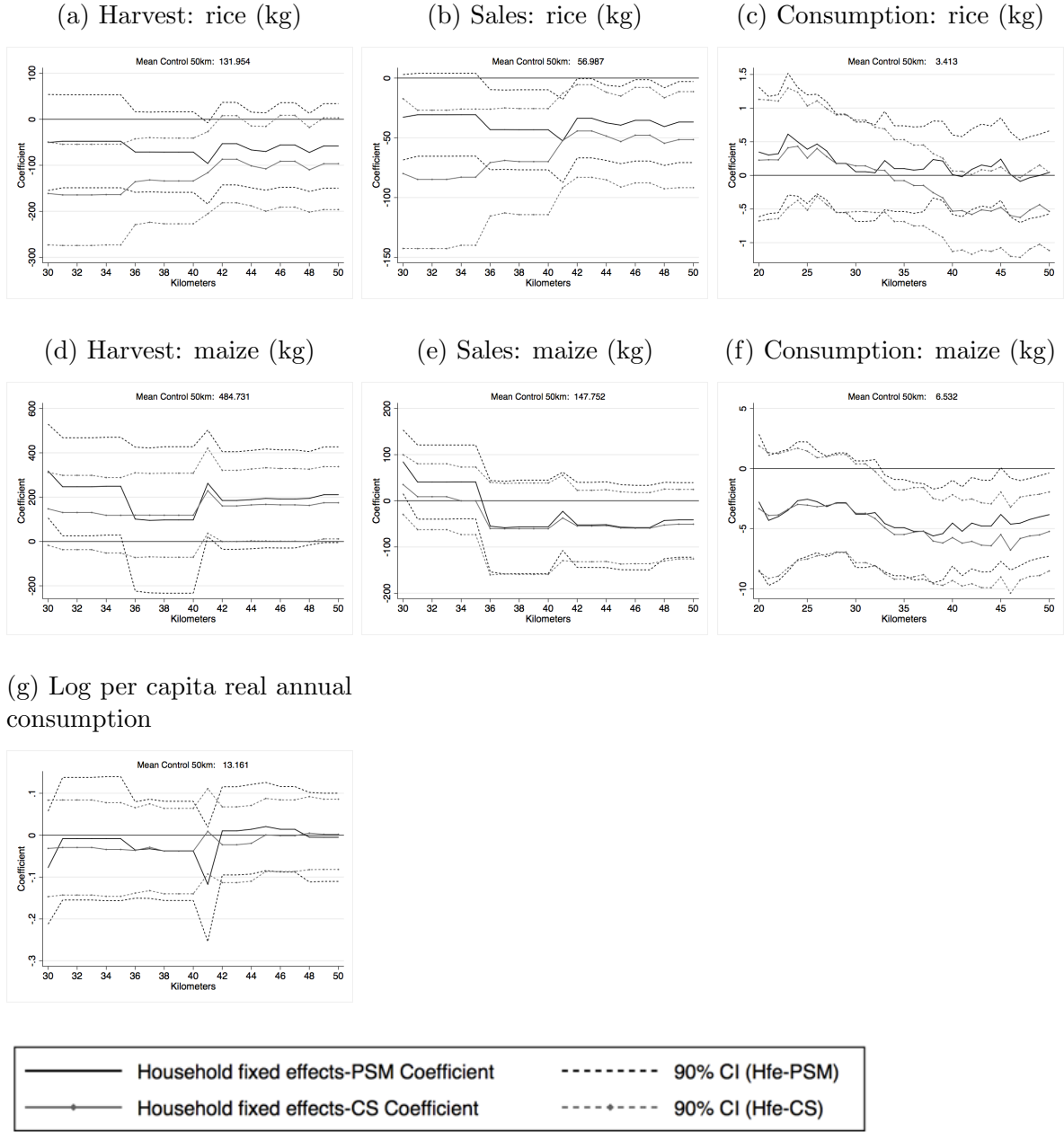
therefore are not mechanical effects of the reduction in price. Because the rice price is lower, households sell a lower share of their harvest, which might itself be somewhat lower. However, they fail to take advantage of rice price decrease because they do not consume more rice. We do not find any significant effect on the maize sector. In a way, it is surprising to find negative effects on the population that are associated to rice price decrease: indeed, if the sample is balanced between net producers and net consumers, the global effect could be close to zero. Actually, Table A6 shows that, despite the matching, there is actually an imbalance between the treatment and the control group in terms of crops: the treated households are clearly in a rice area, while the control households are in a maize area. Therefore we expect that the treated households are mostly rice net producers at baseline, which explains why the global effect is negative. We have tried to run separate matching regressions for net producers and net consumers but we do not manage to obtain a satisfying common support, due to the fact that treatment is correlated with crop production areas.

Last, when aggregating consumption over all items in the last year, we do not find any significant effect. Unfortunately, we cannot build an income measure in the panel. At this stage, it is interesting to wonder what could explain the discrepancy between the results on life satisfaction and consumption. First, welfare might decrease even though consumption remains steady if households deplete their stock of assets to maintain their consumption level. Second, satisfaction is a measurement of the individuals utility (though imperfect) that takes into account how much the various consumption items are valued. If the household has to consume less appreciated goods in order to maintain its calorie intake, it will reduce his welfare but not necessarily its consumption level. Third, even though we fail to identify, at this stage, a clear effect on the production side (in particular, how much is produced), one has to recognize that a decrease in the price of one of the main products should bring changes in the strategies to secure livelihood. Also, the uncertainty associated to it might also reduce welfare without being associated to a reduction in overall consumption. We now try to delve deeper in this question, without aiming at definitely disentangling the various mechanisms.

Figure 8 assesses the impact of better roads on durable ownership and on productive assets ownership. Both variables are built with a principal component analysis run on a series of durable goods ownership (consumption goods for the first, productive assets for the second). The effect on durable assets is large and negative: treated households deplete assets. However, the common trend assumption seems to be rejected for this variable (Figure A5n), even though the margin of significance is lower than for the δ_{1d} coefficients. Of course, such a result would reconcile the conclusions on life satisfaction and overall consumption (first mechanism highlighted before) and point again to a reduction in household well-being. In the longer run, if households do not find new ways to secure livelihoods, their consumption would drop. The households do not seem to deplete their productive assets though. We will now try to understand better how households adjust their activity.

To do so, we turn to the information available in the panel on the labor market. One of the expected advantages of a better road is a better access to markets and therefore a better access to farm inputs and to labor demand. Figure 9a identifies that treatment above 40 kilometers leads to an increase in wage labor (at least one household member works for a wage). In the household fixed effect without PSM specification, we also find that these households tend to reduce their on-farm labor (Figure 9c). A better road does not increase the use of farm inputs (Figure 9b for

Figure 7: Harvest, Sales and Consumption

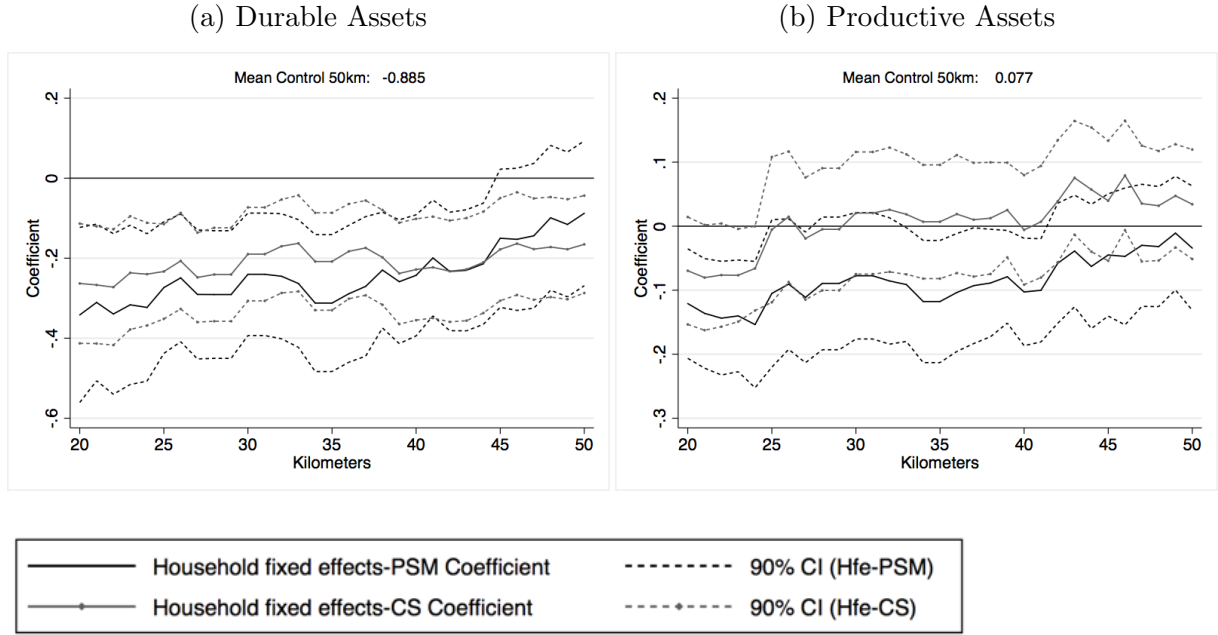


labor). However, given the product price decrease, farming becomes less attractive and this could be sufficient to counterbalance any improvement in market access. The analysis on the Labor Force Survey will help us be more specific on what households do as wage work for securing their livelihood.

5.3 LFS results

We first assess whether the results are consistent between the two surveys, despite the fact that the samples are different and that, in the LFS survey, we do not control for household fixed effects and implement a difference-in-difference strategy (with propensity score matching). Figure 10a shows that treated households living less than 40 kilometers away from the road reduce their economic activity: roughly 0.4 household members less have a current activity (against 2.4 on average in the

Figure 8: Assets (score)



control group).²⁰ Figure 10b confirms that treated households (but living far away from the road) increase wage work. Not only the result is the same than in the panel, but the cut-off is also the same. With the LFS, we also detect a small increase in number of households earning a wage for treatments below 30 kilometers. The reduction in the number of individuals with a current activity (-.4) is largely due to fewer individuals working on-farm (-.5) for households sufficiently close to the road and hardly compensated by self-employment (+.15 for the PSM specification) and wage work as previously mentioned. There are small effects on the likelihood to provide unpaid work but the prevalence of this type of work is small anyway. All in all, it seems that household reduce their economic activity, in particular on-farm, with only a very partial ability to compensate in self-employment and wage work. Actually, even income obtained from these types of activity (self-employment and wage work) does not seem to increase. Given the extent of reduction in the farming activity, the productivity in self-employment and wage work would have to be much larger than in farming to allow households to keep the income level they had before the reduction in prices.²¹ This does not seem to be the case and provides additional evidence that households living in rural areas are actually negatively impacted by the improvement of the road.

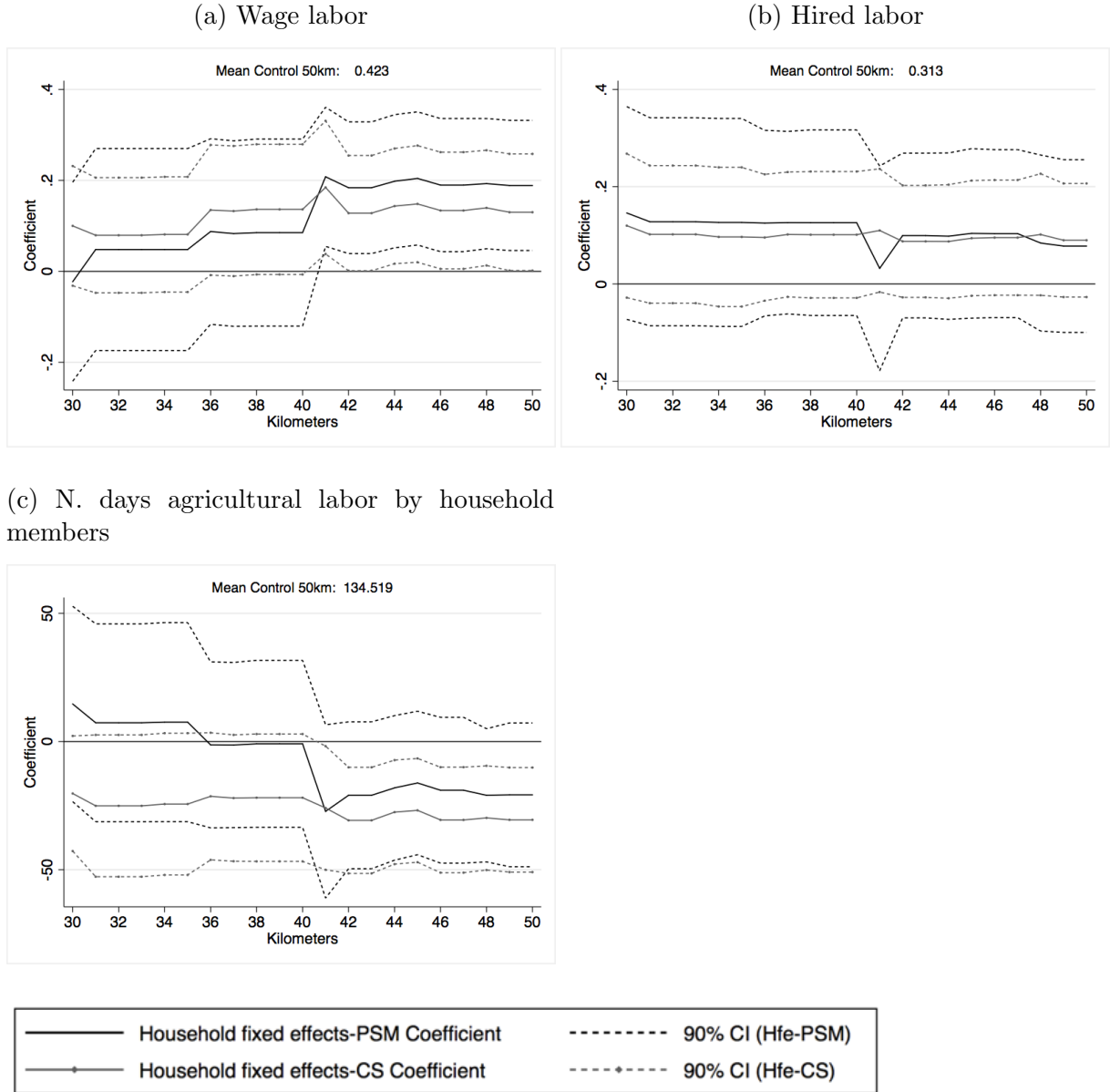
We finally provide information on the new sector of activity of individuals. Consistently, they are less employed in agriculture. The only sectors where we observe an increase are: hospitality industry (e.g. restaurants), construction and transport. This seems consistent with expected effects of a road improvement and that the area may benefit from more activity in these sectors. However, again, the new opportunities from these sectors for the households seem to be fairly low, given that they allow hiring between +.01 and +.04 household members more each.

Quite strikingly, households living further away from the road (more than 40 kilometers), might be the ones actually benefiting from it. Recall that the treat-

²⁰The number of household members is not impacted by the treatment. It is also a control variable in the matching.

²¹Income from farming activity is not collected in the LFS.

Figure 9: Labor Market

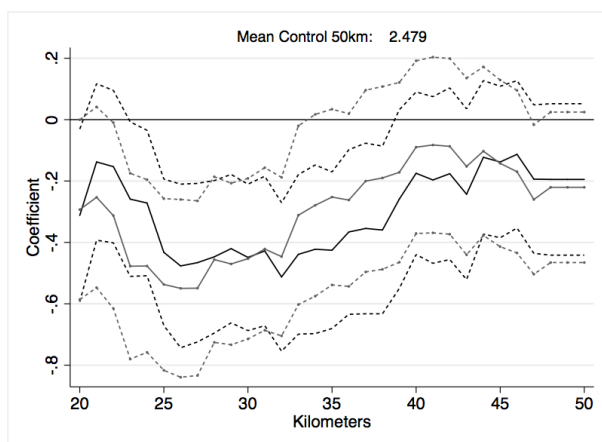


ment “below 40 kilometers” (for instance) encompasses households living very close to the road and some living much farther away, and therefore is an average between the ones who are “strongly” treated and the ones who are only “weakly” treated.²² When the size of the effect declines with the cut-off it indeed suggests that households who live closer to the road are more impacted than the ones living further away from the road. Generally, the treatment above 40 kilometers seem to be less negative than the treatment below 30 kilometers: they have a lower decrease in life satisfaction, a lower reduction in durable assets, a lower reduction in economic activity, but still an increase in wage work. This result based on the intensity in the treatment (even though we cannot quantify it) points to the same conclusion that improvement in roads was actually harmful for the households who lived close to them. However, it should be noted that households with a treatment above 40

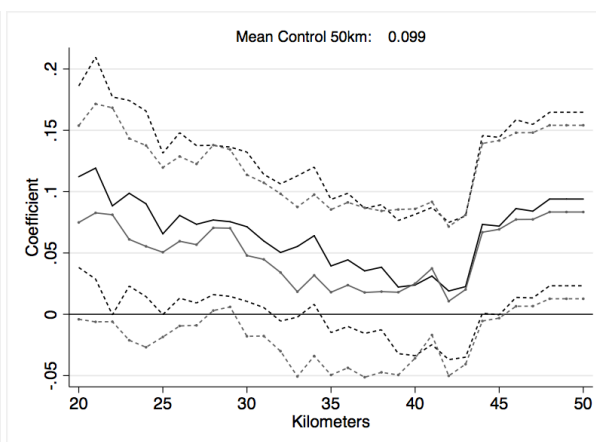
²²Ideally we would have liked to identify the treatment effect for each interval of distance, but this was not possible due to data limitations (number of observations and difficulty to identify a control group).

Figure 10: Labor Market
(N. household members)

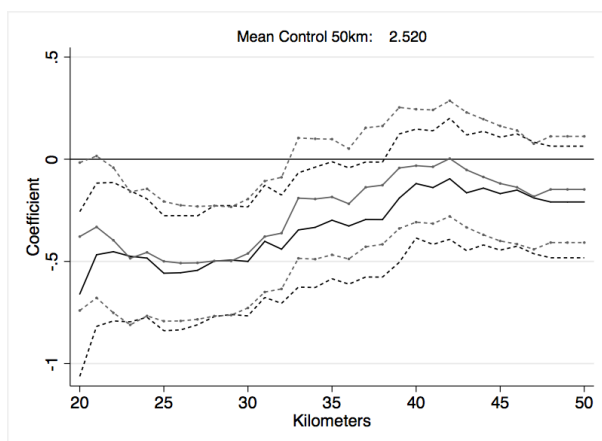
(a) With a current activity



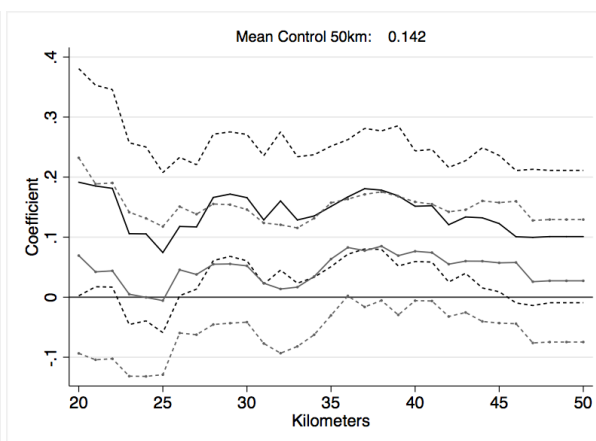
(b) Earn a wage



(c) Work on farm



(d) Self-employed



(e) Work unpaid

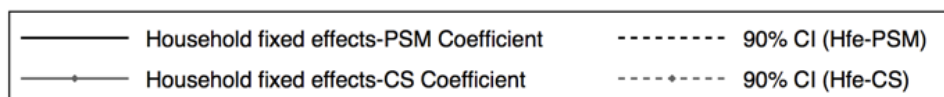
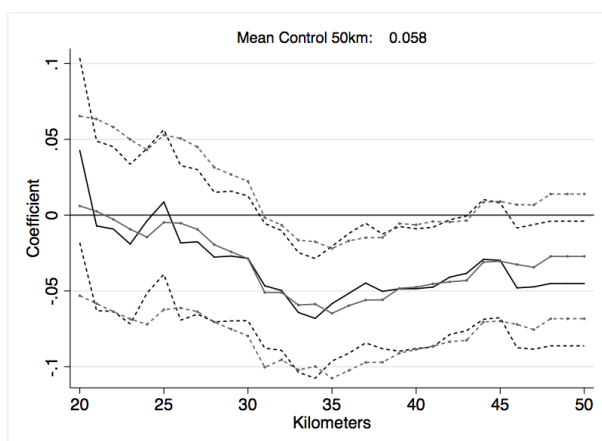
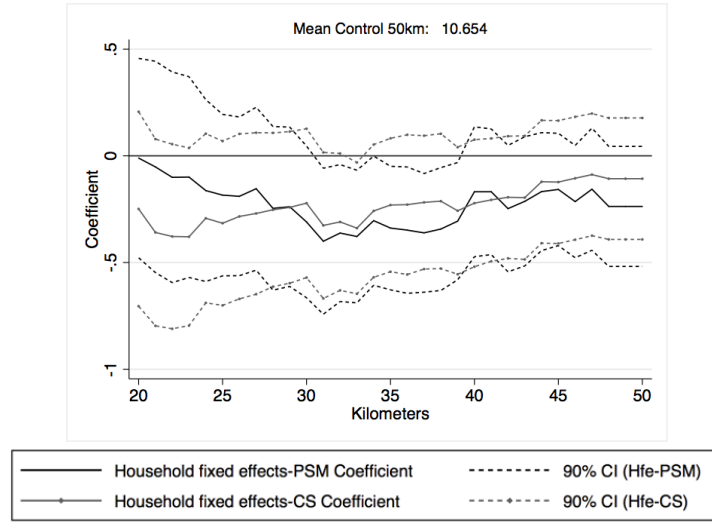


Figure 11: Income

Log household income paid and self employment
(2006 TSh)



kilometers were also hit by the reduction in rice price. This suggests that rice price decrease does not explain the whole effect. Given that we are not able to scrutinize each price independently, it is difficult to say if other prices were less affected in villages 40 kilometers away from the road. However, Figure 12a suggests that the agricultural sector over all was less impacted.

5.4 Robustness checks

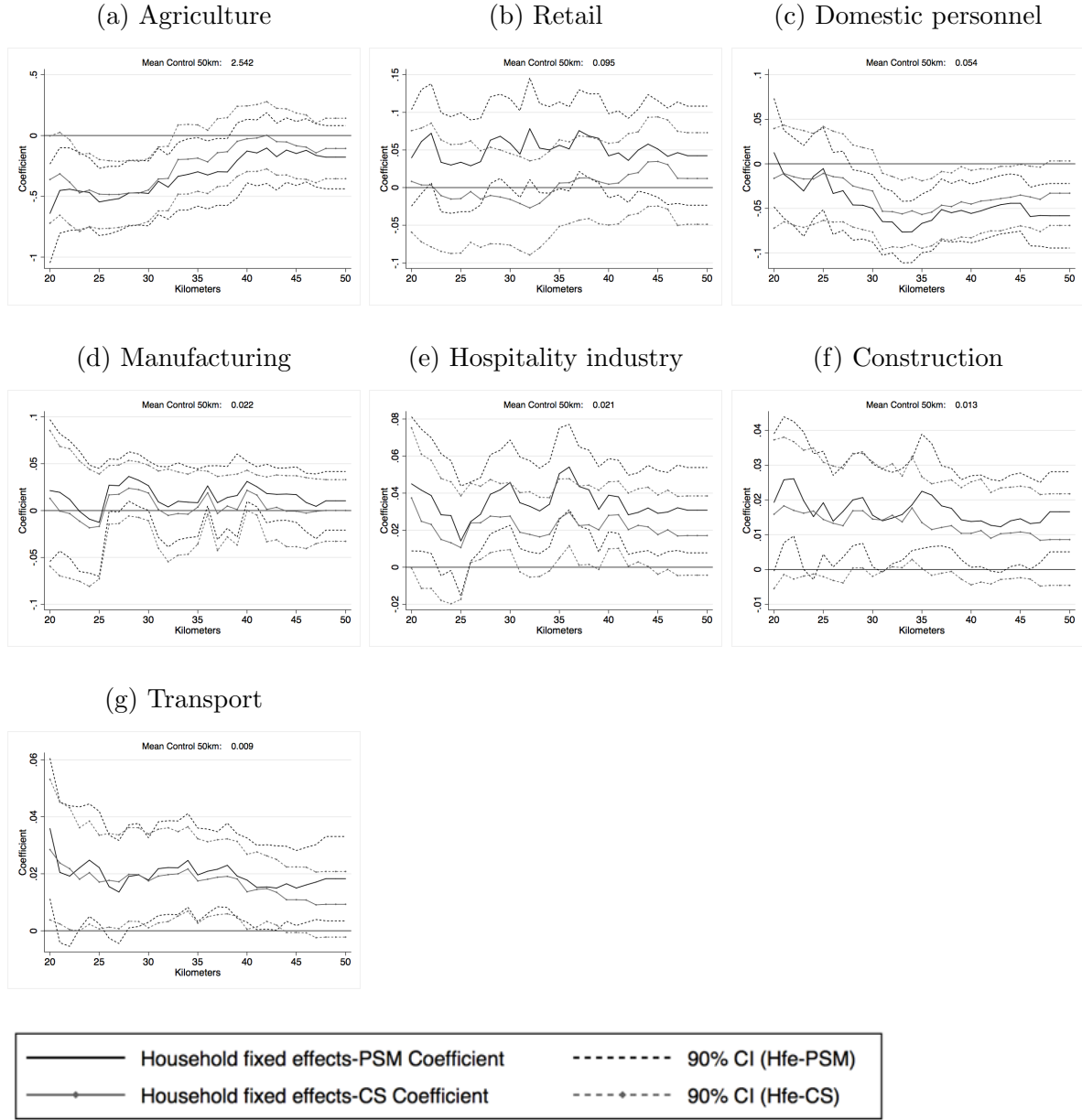
One possible threat to our identification could come from any other complementary policy that happened in response to the road infrastructure program. If, for instance, local governments decided to upgrade or construct roads connecting to the treated major roads, we would be evaluating the join effect of those policies. To address this concern, we test at the community level whether the treatment at the different thresholds had any impact on the implementation of community projects in the following areas: road construction, market construction, irrigation schemes, water supply and grain storage. We find no evidence that it is the case.

6 Conclusion

We evaluate the effects of road infrastructure improvements on market participation of labor and products, prices and welfare of rural households in Tanzania. Our empirical method consists on combining household fixed-effects and difference-in-difference with a propensity score matching. We provide tests of common support and balancing between the treatment and the control group that are satisfactory. The tests also do not reject the common trend assumption between the two groups in the vast majority of the outcome variables.

Contrary to most studies, we find damaging effects of the road improvements on the rural population: we observe a reduction in life satisfaction, a reallocation of labor away from farm without many additional opportunities of work, that we trace back to a decrease in the price of the main product (rice). This is consistent

Figure 12: Sector of activity (ISIC)
(N. household members)



with predictions obtained from trade models where rural households, with lower transaction costs, actually face competition from lower price goods. We do not identify effects on consumption and our negative effects on durable goods ownership are questionable but we show that there is evidence that household earnings decrease.

Our conclusion raises the question of the specificity of our results to the Tanzanian case. First, as recalled in the introduction, there are actually only few contributions which evaluate the impact of these type of roads. Often, rural roads (i.e. better connecting villages to neighboring cities) are evaluated. Articles on major roads (Storeygard, 2016; Ghani et al., 2016; Shrestha, 2019) use as outcomes exports and manufacturing and therefore say little on effects on the rural population. Second, the road rehabilitation program fulfilled several objectives: not only poverty reduction, but also improvements in the connectivity with the neighboring

countries. These countries are also producers of rice and maize and Tanzania is a natural export country for them. In our study, we are unable to tell where the competition comes from, because we do not have information on trade flows. One possibility is that the better connectivity with the neighboring countries triggered higher imports from them and therefore fiercer competition for the Tanzanian producers. We acknowledge that, had the choice of roads been different, the welfare effect on rural households could have been different as well. However, it is not innocuous that the actual placement of roads was as such. Under cover of poverty reduction, the governments may actually implement programs which simply aim at fostering economic growth. This paper does not address the question of the impact of roads on urban areas. However, preliminary estimations show that urban households (in the panel) actually benefit from the road improvement: they have a higher life satisfaction and a lower exposure to price shocks. It would be interesting as well to document the effect of roads in Tanzania on the farm households living in the neighboring countries. This is left for further research.

References

- Adukia, A., Asher, S., and Novosad, P. (2017). Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction. MPRA Paper 80194, University Library of Munich, Germany.
- Aggarwal, S. (2018). Do rural roads create pathways out of poverty? Evidence from India. *Journal of Development Economics*, 133(C):375–395.
- Asher, S. E. and Novosad, P. M. (2018). Rural roads and local economic development. Policy Research Working Paper Series 8466, The World Bank.
- Attanasio, O., Goldberg, P. K., and Pavcnik, N. (2004). Trade reforms and wage inequality in Colombia. *Journal of Development Economics*, 74(2):331–366.
- Banerjee, A., Duflo, E., and Qian, N. (2012). On the Road: Access to Transportation Infrastructure and Economic Growth in China. *NBER Working Paper*, (17897):1–51.
- Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., and Padró i Miquel, G. (2015). The Value of Democracy: Evidence from Road Building in Kenya. *American Economic Review*, 105(6):1817–1851.
- Caliendo, M. and Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*, 22(1):31–72.
- Casaburi, L., Glennerster, R., and Suri, T. (2013). Rural roads and intermediated trade: Regression discontinuity evidence from Sierra Leone.
- Castaneda, A., Doan, D., Newhouse, D., Nguyen, M. C., Uematsu, H., and Azevedo, J. P. (2016). Who are the poor in the developing world ? Policy Research Working Paper Series 7844, The World Bank.
- Currie, J. and Harrison, A. E. (1997). Sharing the Costs: The Impact of Trade Reform on Capital and Labor in Morocco. *Journal of Labor Economics*, 15(3):44–71.
- de Janvry, A., Fafchamps, M., and Sadoulet, E. (1991). Peasant household behaviour with missing markets: Some paradoxes explained. *Economic Journal*, 101(409):1400–1417.
- Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Press for Worldbank, Baktimore and London: John Hopkins U.
- Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, 108(4-5):899–934.
- Friedman, J. and Levinsohn, J. (2002). The Distributional Impacts of Indonesia’s Financial Crisis on Household Welfare: A “Rapid Response” Methodology. *World Bank Economic Review*, 16(3):397–423.
- Gachassin, M., Najman, B., and Raballand, G. (2010). The Impact Of Roads On Poverty Reduction : A Case Study Of Cameroon. *The World Bank Policy Research Working Paper*, (5209):1–39.

- Ghani, E., Goswami, A. G., and Kerr, W. R. (2016). Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing. *Economic Journal*, 126(591):317–357.
- Goldberg, P. K. and Pavcnik, N. (2007). Distributional Effects of Globalization in Developing Countries. *Journal of Economic Literature*, 45(1):39–82.
- Government of Tanzania (2007). The Roads Act, Act No.13 of 2007. pages 1–239.
- Government of Tanzania (2008). 10 Year Transport Sector Investment Programme (TSIP). pages 1–119.
- Haltiwanger, J., Kugler, A., Kugler, M., Micco, A., and Pages, C. (2004). Effects of tariffs and real exchange rates on job reallocation: evidence from Latin America. *Journal of Economic Policy Reform*, 7(4):191–208.
- Harrison, A. and Hanson, G. (1999). Who gains from trade reform? Some remaining puzzles. *Journal of Development Economics*, 59(1):125–154.
- Harrison, A. E., Lin, J. Y., and Xu, L. C. (2014). Explaining Africa’s (Dis)advantage. *World Development*, 63(C):59–77.
- Heckman, J. and Navarro-Lozano, S. (2004). Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models. *Review of Economics and Statistics*, 86(1):30–57.
- Jacoby, H. G. and Minten, B. (2009). On measuring the benefits of lower transport costs. *Journal of Development Economics*, 89(1):28–38.
- Khandker, S. R., Bakht, Z., and Koolwal, G. B. (2009). The Poverty Impact of Rural Roads: Evidence from Bangladesh. *Economic Development and Cultural Change*, 57(4):685–722.
- Lokshin, M. and Yemtsov, R. (2005). Has Rural Infrastructure Rehabilitation in Georgia Helped the Poor? *World Bank Economic Review*, 19(2):311–333.
- Minten, B. and 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.
- Mu, R. and van de Walle, D. (2011). Rural Roads and Local Market Development in Vietnam. *Journal of Development Studies*, 47(5):709–734.
- Munshi, K. and Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(1):46–98.
- Porto, G. G. (2006). Using survey data to assess the distributional effects of trade policy. *Journal of International Economics*, 70(1):140–160.
- Qin, Y. and Zhang, X. (2016). The Road to Specialization in Agricultural Production: Evidence from Rural China. *World Development*, 77(C):1–16.
- Renkow, M., Hallstrom, D. G., and Karanja, D. D. (2004). Rural infrastructure, transactions costs and market participation in Kenya. *Journal of Development Economics*, 73(1):349–367.

- Reventa, A. (1997). Employment and Wage Effects of Trade Liberalization: The Case of Mexican Manufacturing. *Journal of Labor Economics*, 15(3):20–43.
- Rodrik, D. (2017). *Straight Talk on Trade. Ideas for a Sane World Economy*. Princeton University Press, Princeton, New Jersey, USA.
- Schwab, K. and Porter, M. E., editors (2008). *The Global Competitiveness Report 2008–2009*. World Economic Forum, Geneva.
- Schwab, K. and Sala-i Martin, X., editors (2016). *The Global Competitiveness Report 2016–2017*. World Economic Forum, Geneva.
- Shamdasani, Y. (2016). Rural road infrastructure and agricultural production: Evidence from india.
- Shrestha, S. A. (2019). Roads, Participation in Markets, and Benefits to Agricultural Households: Evidence from the Topography-based Highway Network in Nepal. *Economic Development and Cultural Change*, forthcoming.
- Singh, I., Squire, L., and Strauss, J. (1986). *Agricultural Household Models: Extension, Application and Policy*. Johns Hopkins Univ. Press, Baltimore, MD.
- Smith, J. A. and Todd, P. E. (2005). Does matching overcome LaLonde’s critique of nonexperimental estimators? *Journal of Econometrics*, 125(1-2):305–353.
- Storeygard, A. (2016). Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa. *Review of Economic Studies*, 83(3):1263–1295.
- The World Bank (2008). Safe, clean, and affordable...transport for development : the World Bank Group’s transport business strategy for 2008-2012. Technical report.
- Topalova, P. (2004). Factor immobility and regional impacts of trade liberalization: Evidence on poverty and inequality from india.
- Volpe Martincus, C., Carballo, J., and Cusolito, A. (2017). Roads, exports and employment: Evidence from a developing country. *Journal of Development Economics*, 125(C):21–39.
- Williamson, J. G. (2011). *Trade and Poverty. When the Third World Fell Behind*. MIT Press, Cambridge, Massachusetts, USA.

A Appendix

A.1 Variable definition

Outcome variables in the LSMS:

Life satisfaction: Average self-reported life satisfaction among all respondents in the household on a scale from 1 to 7, where 1 corresponds to very dissatisfied and 7 to very satisfied.

Consumption rice (kg) last 7d: Household consumption of rice (kg) over the last week.

Consumption maize (kg) last 7d: Household consumption of maize (kg) over the last week.

Prices (TSh): Village level prices (TSh) of rice husked, maize grain, sugar and kerosene.

Wage labor: Dummy informing if at least one member of the household engaged in wage labor over the last year.

Hire labor: Dummy informing if the household hired labor over the last year.

N. days agri. labor by hhm: Number of farming days in the last rainy season performed by household members.

Harvest rice (kg): Harvest of rice (kg) over the last rainy season.

Sales rice (kg): Sales of rice (kg) over the last rainy season.

Harvest maize (kg): Harvest of maize (kg) over the last rainy season.

Sales maize (kg): Sales of maize (kg) over the last rainy season.

Log PC real annual consumption: Logarithm of the per capita annual household consumption. The price deflator is computed by the World Bank (see the data documentation).

N. members away at least 1m last 12m: Number of household members that lived away for at least one month over the last year.

Outcome variables in the LFS:

N. household members have current act.: Number of household members that had an economic activity over the last week.

N. household members earned wage: Number of household members that earned a wage over the last year.

N. household members self employed: Number of household members self-employed over the last year.

N. household members work unpaid: Number of household members that worked unpaid over the last year.

N. household members work on farm: Number of household members that worked on farm over the last year.

Isic: work in agriculture: Number of household members that worked in agriculture over the last year.

Isic: work in manufacturing: Number of household members that worked in manufacturing over the last year.

Isic: work in construction: Number of household members that worked in construction over the last year.

Isic: work in retail: Number of household members that worked in retail over the last year.

Isic: work in transport: Number of household members that worked in transport over the last year.

Isic: work in hospitality industry: Number of household members that worked

in hotels and restaurants over the last year.

Isic: work as domestic personnel: Number of household members that worked as domestic personnel the last year.

Log Hh income paid and self employment (2006 TSh): Logarithm of the total household income paid and self employment (TSh) adjusted to 2006 values.

Control variables:

Head Male: Dummy variable equal to 1 if the household's head is male.

Head Age: Age in years of the household head.

Fraction hh members 0-6 years old: Share of household members aged 0 to 6 years old.

Fraction hh members 7-12 years old : Share of household members aged 7 to 12 years old.

Fraction hh members 13-18 years old : Share of household members aged 13 to 18 years old.

Fraction male hh members >18 years old : Share of household members who are male and older than 18 years old.

Fraction female hh members >18 years old: Share of household members who are female and older than 18 years old.

Household size: Number of people living in the household.

More than primary Head: Dummy equal to 1 if the household head has attained at least primary education.

N. adult members with > primary: Number of adults aged 18 and older in the household that attained at least primary education.

Durable Assets (score): Principal component analysis (PCA) asset index constructed on the basis of 30 household durable assets, i.e. radio, telephone, refrigerator, beds, etc.

Productive Assets (score): Principal component analysis (PCA) asset index constructed on the basis of 20 household productive assets, i.e. carts, tractor, hoes, etc.

Land title: Dummy variables equal to 1 if the household holds a land title for at least one of their plots.

Land rights: Dummy variable equal to 1 if the household has the right to sell or to use as collateral at least one of their plots.

Used org fertilizer: Dummy variable equal to 1 if the household uses organic fertilizer in at least one plot.

Used inorg fertilizer: Dummy variable equal to 1 if the household uses inorganic fertilizer in at least one plot.

Used pesticides: Dummy variable equal to 1 if the household uses pesticides in at least one plot.

Erosion: Dummy variable equal to 1 if the household has had an erosion issue for at least one plot.

Irrigation: Dummy variable equal to 1 if the household has at least one irrigated plot.

Tropic-warm/humid agroecological zone: Dummy variable equal to 1 if the climate is Tropical-warm and humid and 0 if the climate is Tropical-cool and humid.

Quality of land: Total surface area of land (in acres) of good, average and bad quality.

2008 Distance (Km) to Nearest:

Major Road: Distance (km) to the nearest major road.

Pop Center with >20'000: Distance (km) to the nearest population center with more than 20,000 inhabitants.

Border Crossing: Distance (km) to the nearest border crossing.

District Headquarters: Distance (km) to the nearest district headquarters.

Quality of closest road: Dummy variables equal to 1 if the closest major road is graveled, paved, sealed or in soil condition.

Average hh size in the ward: Average household size in the ward.

Population Density: 2010 population density (people per km^2) in the ward.

Share hhs with a member wage employed: Share of household members in the ward with at least a member engaged in wage employment based on the labor force survey sample.

Share illiterate people >15 years old : Share of illiterate people 15 years old and older in the ward based on the labor force survey sample.

N. farmer cooperative groups: Number of farmer cooperative groups/association in the village.

Services available (less 10km): Dummy variables indicating if there is a bank branch, a savings and credit cooperative organization (SACCOs), a daily market and a weekly market available within a radius of 10km.

Services available (less 2km): Dummy variables informing if there is a primary school, a clinic, a shop and a market available within a radius of 2km.

Table A1: Geovariables Description

Variable	Source	Data type	Resolution	Description
2008 Distance (km) to Nearest				
Major Road	TANROADS	Shapefile	N.A.	Euclidean distance computed using QGIS
Population Center with $> 20'000$	City population	Text List	N.A.	Georeferenced using QGIS
Border Crossing	Tracks4Africa	Text List	N.A.	Georeferenced using QGIS
District Headquarters	Statoids	Text List	N.A.	Georeferenced using QGIS
Quality of the closest Road				
	TANROADS	Shapefile	N.A.	Euclidean distance computed using QGIS
2010 Population Density (people per km^2)				
	WorldPop	Raster	0.008333 dd	2010 Population Density Range (people per km^2), with national total adjusted to match UN population division estimates
Agroecological Zone				
	IFPRI	Raster	0.008333 dd	Agro-ecological zones created using IFPRI standardized AEZ data based on elevation, climatology

Table A2: LSMS-ISA Summary Statistics

	Mean	(SD)	Mean	(SD)
<i>Treatment</i>				
Road rehabilitation 20km	0.117	(0.004)		
Road Rehabilitation 30km	0.189	(0.005)		
Road Rehabilitation 50km	0.344	(0.006)		
<i>Household Characteristics</i>				
Head Age	48.184	(0.195)		
Head Male	0.760	(0.005)		
Fraction hh members 0-6 years old	0.205	(0.002)		
Fraction hh members 7-12 years old	0.147	(0.002)		
Fraction hh members 13-18 years old	0.140	(0.002)		
Fraction male hh members >18 years old	0.234	(0.002)		
Household size	5.872	(0.042)		
<i>Education</i>				
More than primary Head	0.474	(0.006)		
Education's head missing	0.015	(0.002)		
N. adult members with > primary	1.290	(0.015)		
<i>Asset ownership (score)</i>				
Durable Assets	-1.057	(0.020)		
Productive Assets	0.095	(0.012)		
<i>Farm's Characteristics</i>				
Land title	0.104	(0.004)		
Land rights	0.719	(0.006)		
Land rights missing	0.061	(0.003)		
Used org fertilizer	0.196	(0.005)		
Used inorg fertilizer	0.111	(0.004)		
Used pesticides	0.155	(0.004)		
Inputs missing	0.008	(0.001)		
Erosion	0.222	(0.005)		
Irrigation	0.028	(0.002)		
Erosion/irrigation missing	0.068	(0.003)		
Tropic-warm/humid agroecological zone	0.549	(0.006)		
<i>Land Quality</i>				
Good quality*area	0.117	(0.004)	3.476	(0.238)
Avg quality*area	0.189	(0.005)	2.313	(0.074)
Bad quality*area	0.344	(0.006)	0.230	(0.020)
Soil quality missing			0.067	(0.003)
<i>2008 Distance (KMs) to Nearest</i>				
Major Road (GIS)	0.760	(0.005)	19.182	(0.273)
Pop Center with 20'000	0.205	(0.002)	56.997	(0.495)
Market	0.147	(0.002)	79.352	(0.644)
Border Crossing	0.140	(0.002)	163.421	(1.336)
District Headquarters	0.234	(0.002)	34.408	(0.345)
<i>Quality of closest road</i>				
Gravel	5.872	(0.042)	0.281	(0.006)
Paved	0.474	(0.006)	0.207	(0.005)
Sealed	0.015	(0.002)	0.482	(0.006)
Soil	1.290	(0.015)	0.030	(0.002)
<i>Village Characteristics</i>				
Average hh size in the ward	-1.057	(0.020)	6.073	(0.025)
Population Density	0.095	(0.012)	3.595	(0.032)
Population density missing			0.011	(0.001)
Share hhs with a member wage employed	0.104	(0.004)	0.141	(0.001)
Share illiterate people >15 years old	0.719	(0.006)	0.302	(0.002)
<i>Services available(less 10km)</i>				
Bank	0.061	(0.003)	0.222	(0.005)
SACCOs	0.196	(0.005)	0.512	(0.006)
Market (daily)	0.111	(0.004)	0.385	(0.006)
Market (weekly)	0.155	(0.004)	0.468	(0.006)
	0.008	(0.001)		
	0.222	(0.005)		
	0.028	(0.002)		
	0.068	(0.003)		
	0.549	(0.006)		
Values at baseline				

Table A3: LFS Summary Statistics

	2006		2014	
	Mean	(SD)	Mean	(SD)
<i>Treatment</i>				
Road rehabilitation 20km	0.147	(0.035)	0.100	(0.030)
Road Rehabilitation 30km	0.216	(0.041)	0.271	(0.047)
Road Rehabilitation 50km	0.390	(0.049)	0.456	(0.055)
<i>Household Characteristics</i>				
Head Age	46.219	(0.385)	45.877	(0.510)
Head Male	0.740	(0.009)	0.744	(0.016)
Fraction hh members 0-6 years old	0.210	(0.005)	0.197	(0.006)
Fraction hh members 7-12 years old	0.135	(0.003)	0.141	(0.005)
Fraction hh members 13-18 years old	0.107	(0.002)	0.105	(0.004)
Fraction male hh members >18 years old	0.255	(0.004)	0.270	(0.008)
Fraction female hh members >18 years old	0.293	(0.004)	0.287	(0.007)
Household size	4.716	(0.101)	4.715	(0.122)
<i>Education</i>				
More than primary Head	0.493	(0.012)	0.548	(0.020)
Education's head missing	0.000	(omitted)	0.008	(0.002)
N. adult members with > primary	1.329	(0.037)	1.629	(0.071)
<i>2008 Distance (KMs) to Nearest+</i>				
Major Road (GIS)	20.570	(1.944)	22.049	(1.599)
Pop Center with >20'000	52.894	(3.266)	54.472	(3.583)
Border Crossing	194.920	(13.802)	180.682	(13.745)
District Headquarters	27.709	(1.718)	32.735	(1.704)
<i>Quality of closest road+</i>				
Gravel	0.275	(0.044)	0.354	(0.052)
Paved	0.177	(0.038)	0.153	(0.037)
Seated	0.509	(0.050)	0.493	(0.055)
Soil	0.039	(0.019)	0.000	(omitted)
<i>Village Characteristics+</i>				
Average hh size in the ward	5.946	(0.140)	5.956	(0.160)
Population Density	1.116	(0.180)	0.777	(0.071)
Share hhs with a member wage employed	0.147	(0.010)	0.146	(0.008)
Share illiterate people >15 years old	0.273	(0.012)	0.298	(0.014)
Tropic-warm/humid agroecological zone	0.481	(0.050)	0.419	(0.053)
<i>Services available(less 2km) +</i>				
Primary School	0.171	(0.005)	0.947	(0.011)
Clinic	0.114	(0.007)	0.664	(0.035)
Shop	0.178	(0.006)	0.936	(0.013)
Market	0.108	(0.008)	0.582	(0.036)

Means and standard deviations computed using sample design weights. +2006 value

Table A4: LSMS-ISA Propensity Score Matching: Logit 20km - Buffer Any Road 20km

<i>Household Characteristics</i>		<i>Land Quality</i>	
Head Age	0.0157	(0.036)	Good quality*area
Head Age ²	-0.0002	(0.000)	Avg quality*area
Head Male	0.6300**	(0.299)	Bad quality*area
Fraction hh members 0-6 years old	0.5453	(0.972)	Soil quality missing
Fraction hh members 7-12 years old	0.5078	(0.943)	<i>2008 Distance (KMs) to Nearest</i>
Fraction hh members 13-18 years old	1.4847*	(0.866)	Major Road
Fraction male hh members >18 years old	-0.2300	(0.733)	Pop Center with >20'000
Log. Household size	-0.2234	(0.319)	Market
<i>Education</i>			Border Crossing
More than primary Head	0.1289	(0.302)	District Headquarters
Education's head missing	-1.6443**	(0.830)	<i>Quality of closest road</i>
N. adult members with > primary	-0.085	(0.163)	Gravel
<i>Asset ownership (score)</i>			Paved
Durable Assets	-0.1623*	(0.086)	Sealed
Productive Assets	-0.3861*	(0.213)	<i>Village Characteristics</i>
<i>Farm's Characteristics</i>			Average hh size in the ward
Land title	0.4395	(0.350)	Population Density
Land rights	0.2453	(0.261)	Population density missing
Land rights missing	-0.0351	(0.410)	Share hhs with a member wage employed
Used org fertilizer	0.3318	(0.265)	Share illiterate people >15 years old
Used inorg fertilizer	0.3977	(0.277)	<i>Services available(less 10km)</i>
Used pesticides	-0.1366	(0.286)	Bank
Erosion	-0.3751*	(0.222)	SACCOs
Irrigation	-0.9154*	(0.490)	Market (daily)
Tropic-warm/humid agroecological zone	0.4480**	(0.214)	Market (weekly)
Constant	-15.8821****	(1.254)	
Pseudo R2	0.238		
Observations	1022		

Heteroscedasticity corrected and clustered standard errors are presented in parentheses.

* p< 0.10, ** p<0.05, *** p<0.01, **** p<0.001

Treatment: Roads Rehabilitated between 2011-2013. Estimation at baseline.

Table A5: LFS - Propensity Score Matching: Logit 20km - Buffer Any Road 20km

	2006		2014	
<i>Household Characteristics</i>				
Head Age	-0.0142	(0.014)	0.029	(0.031)
Head Age ²	0.0001	(0.000)	-0.0004	(0.000)
Head Male	-0.1131	(0.205)	0.1866	(0.342)
Fraction hh members 0-6 years old	0.2365	(0.859)	-0.0533	(0.461)
Fraction hh members 7-12 years old	0.2062	(0.702)	0.0266	(0.544)
Fraction hh members 13-18 years old	0.3251	(0.546)	-0.0877	(0.623)
Fraction male hh members >18 years old	0.4995	(0.387)	0.0302	(0.585)
Log. Household size	0.1629	(0.243)	0.0974	(0.167)
<i>Education</i>				
More than primary Head	0.0315	(0.110)	-0.2707	(0.217)
Education's head missing	0	(.)	-0.2814	(0.416)
N. adult members with > primary	-0.1228	(0.074)	-0.082	(0.101)
<i>2008 Distance (KM)s to Nearest+</i>				
Major Road	1.39E-06	(0.000)	-0.0002*	(0.000)
Pop Center with >20'000	-0.0293	(0.041)	0.0107	(0.038)
Border Crossing	-0.0036	(0.005)	0.0028	(0.008)
District Headquarters	-0.0253	(0.047)	-0.0759	(0.054)
<i>Quality of closest road+</i>				
Gravel	1.2289	(1.139)	3.7725*	(2.068)
Paved	-0.9823	(2.291)	2.7459*	(1.420)
<i>Village Characteristics+</i>				
Average hh size in the ward	-0.5713	(0.876)	-0.8228	(0.530)
Population Density	-1.6585	(1.857)	-1.0118	(0.743)
Share hhs with a member wage employed	-4.8053	(8.376)	10.0594**	(4.314)
Share illiterate people >15 years old	2.4564	(5.921)	-8.0934*	(4.081)
Tropic-warm/humid agroecological zone	2.0353**	(0.881)	0.6393	(1.141)
<i>Services available(less 2km)+</i>				
Primary School	-5.4388	(10.683)	17.1844	(26.026)
Clinic	-13.3242	(13.601)	0.8406	(1.902)
Shop	25.4096	(21.245)	9.5332	(15.896)
Market	-11.7854*	(6.435)	-1.2597	(2.629)
Constant	4.0781	(11.784)	-21.2738	(17.697)
Observations	4573		1415	

Sample design adjusted standard errors are presented in parentheses

* p< 0.10, ** p<0.05, *** p<0.01, **** p<0.001

Treatment: Roads Rehabilitated between 2011-2013. + 2006 value

Figure A1: LSMS-ISA Common Support Treatment 20km

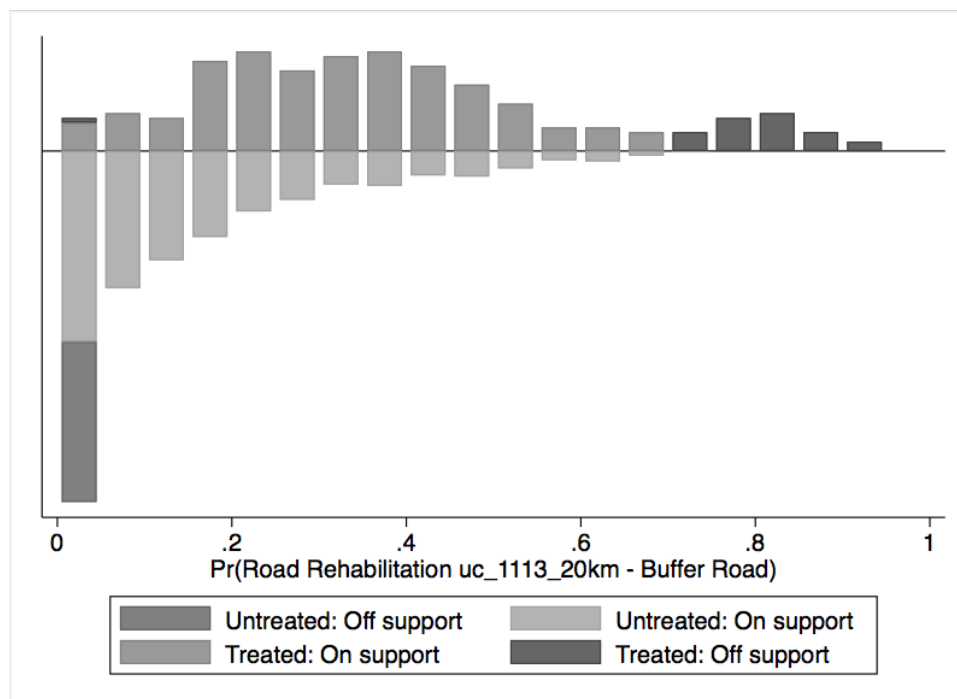
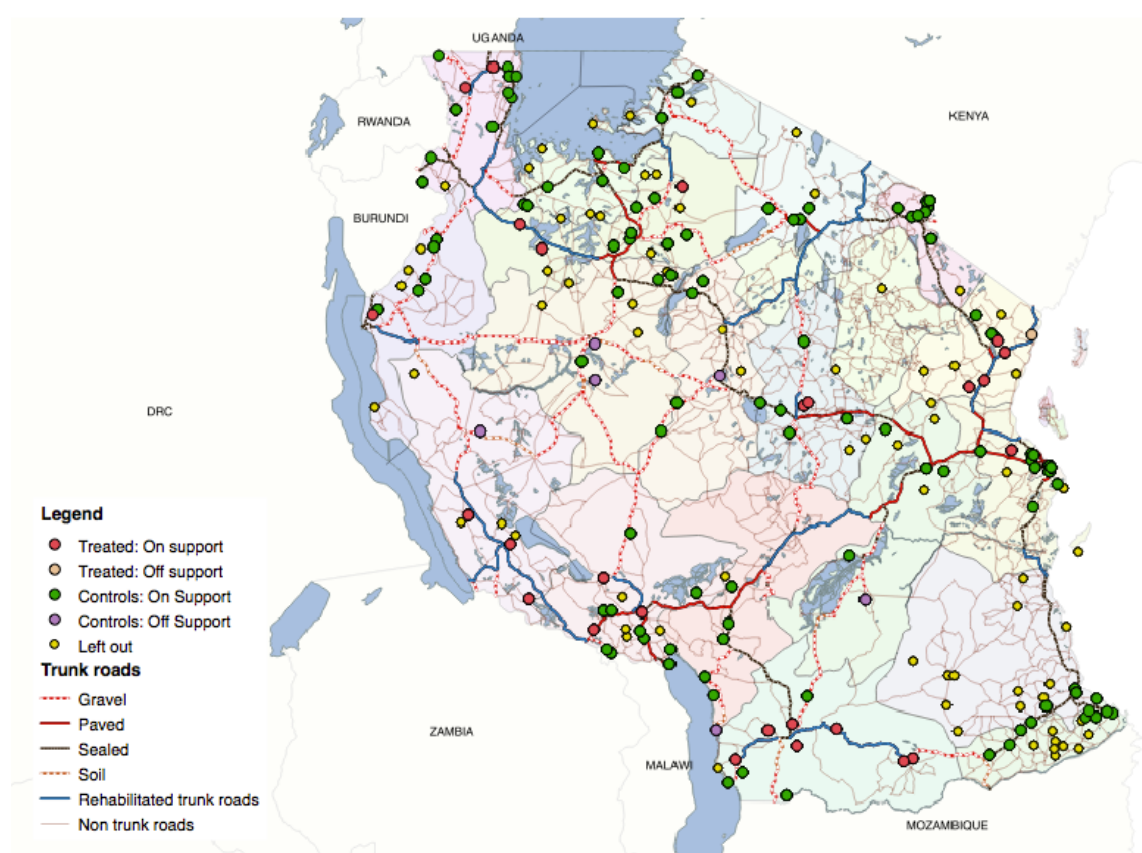


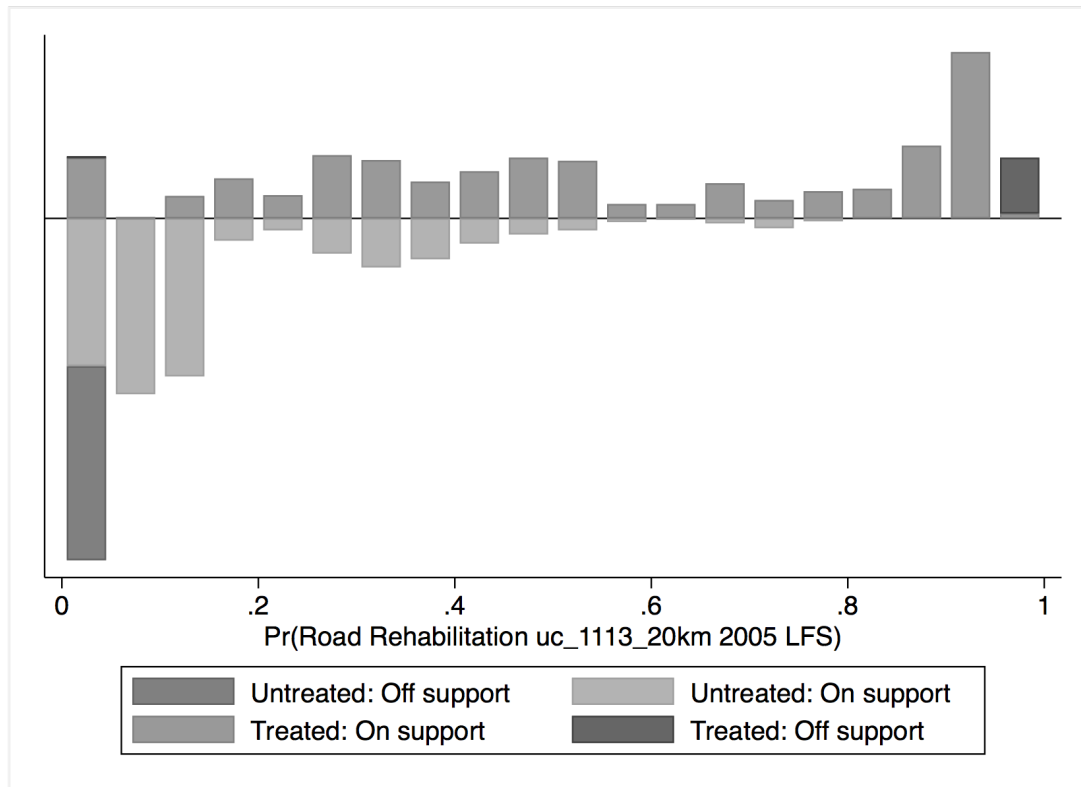
Figure A2: LSMS-ISA Common Support Treatment 20km (Buffer road)



Source: LSMS-ISA and TANROADS GIS data, author's computation.

Figure A3: LFS Common Support Treatment 20km (Buffer road)

(a) Common Support 2006



(b) Common Support 2014

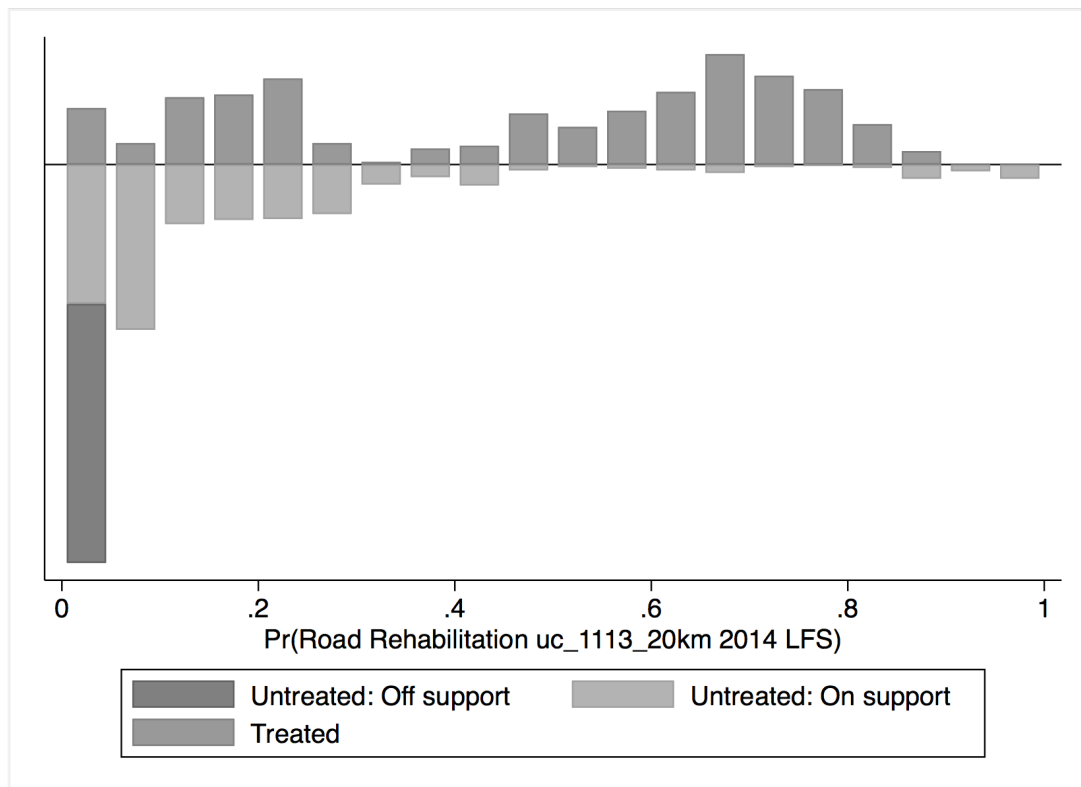
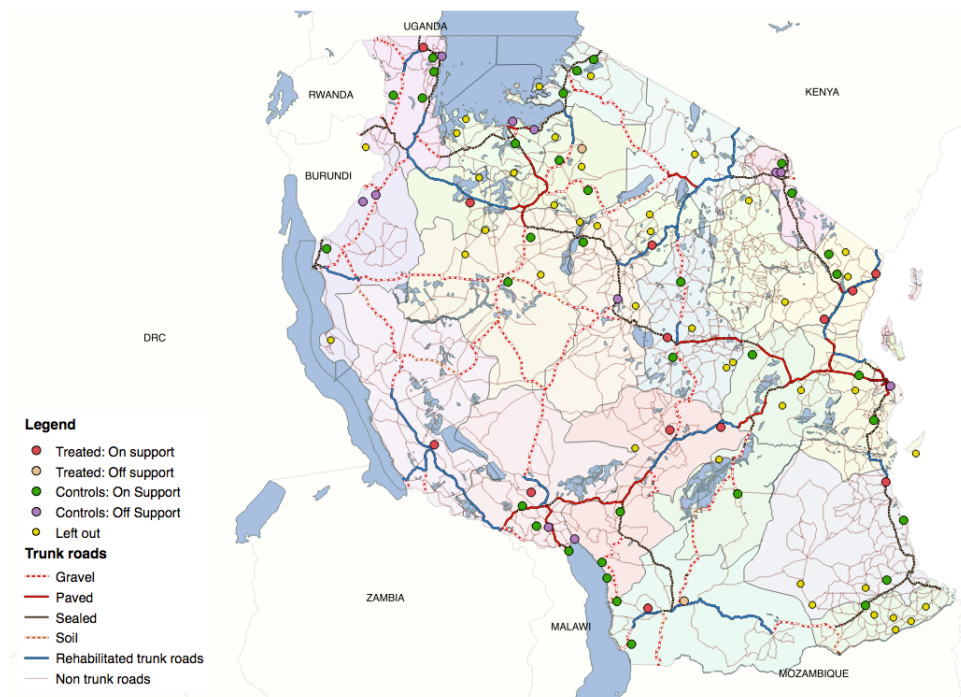
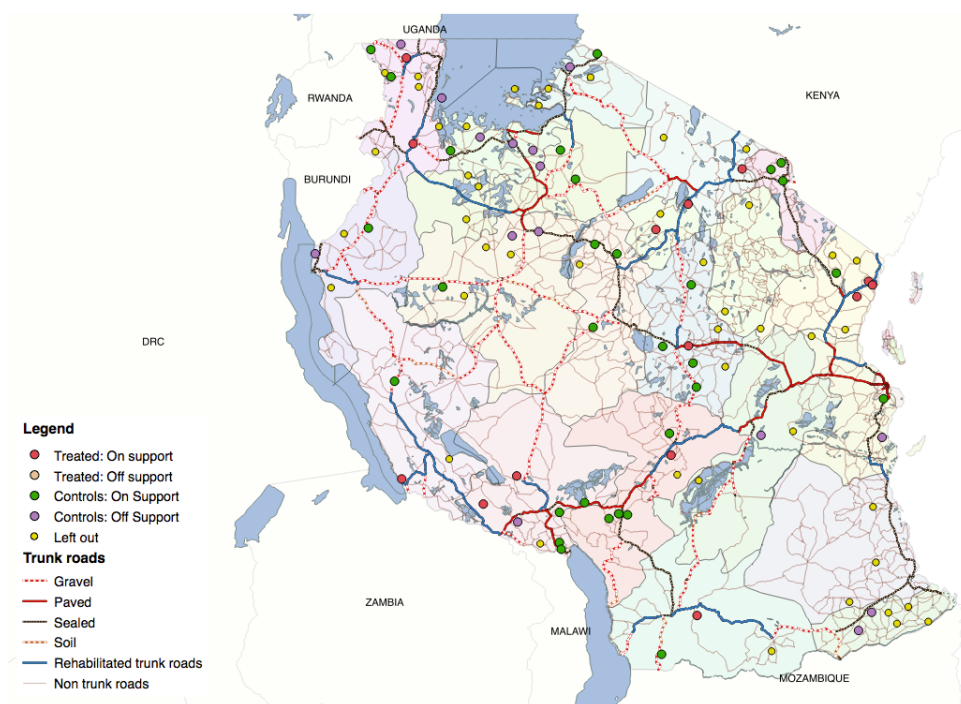


Figure A4: LFS Common Support Treatment 20km (Buffer road)

(a) Common Support 2006



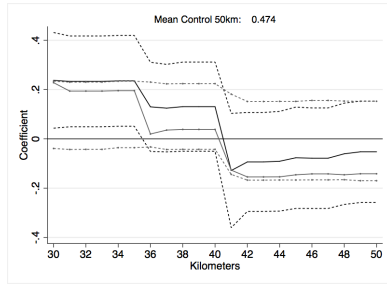
(b) Common Support 2014



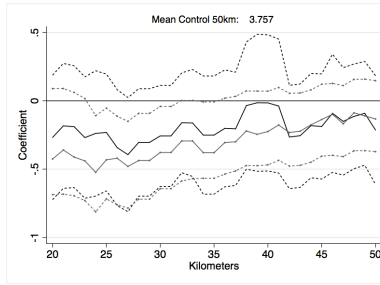
Source: LSMS-ISA and TANROADS GIS data, author's computation.

Figure A5: Common Trend

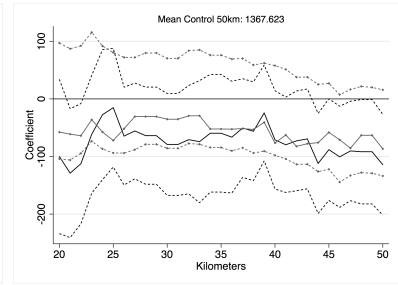
(a) N. members away at least 1 month last 12m



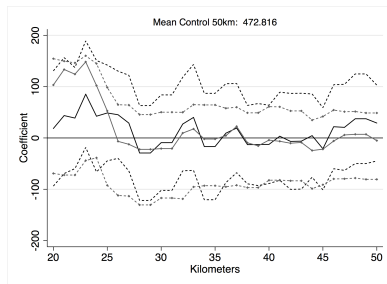
(b) Life Satisfaction



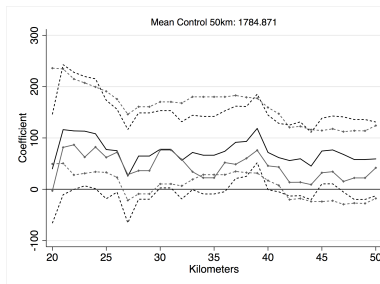
(c) Price rice husked (TSh)



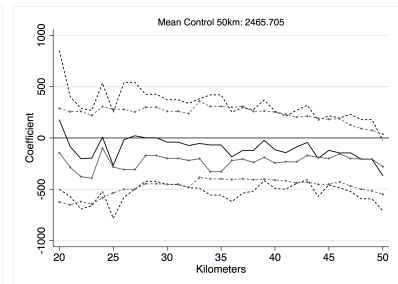
(d) Price maize grain (TSh)



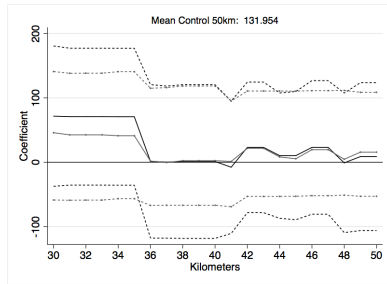
(e) Price sugar (TSh)



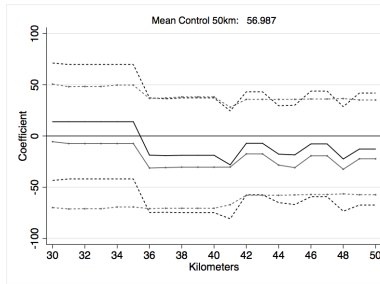
(f) Price kerosene (TSh)



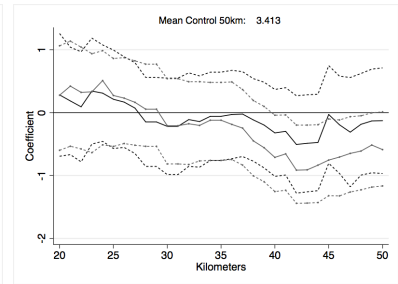
(g) Harvest rice (kg)



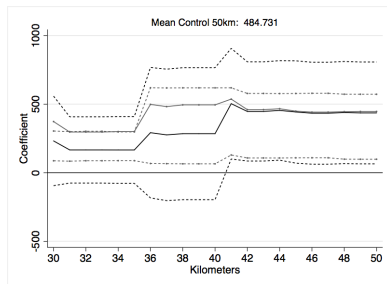
(h) Sales rice (kg)



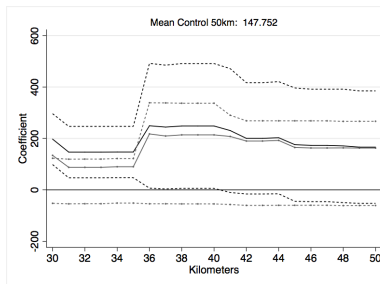
(i) Consumption rice (kg)



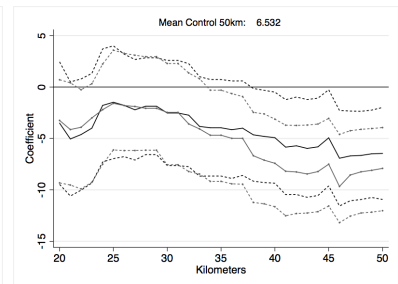
(j) Harvest maize (kg)



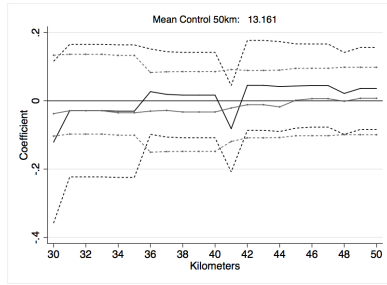
(k) Sales maize (kg)



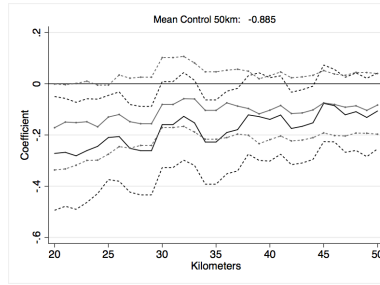
(l) Consumption maize (kg)



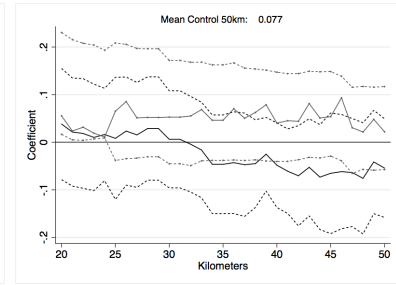
(m) Log PC real annual consumption



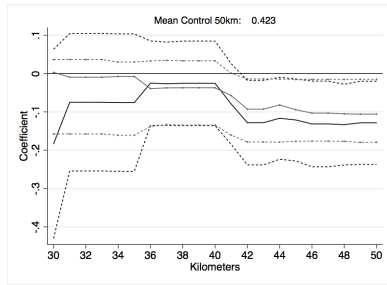
(n) Durable Assets (score)



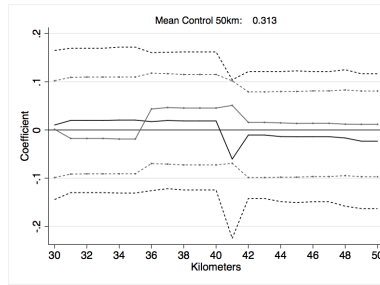
(o) Productive Assets (score)



(p) Wage labor



(q) Hired labor



(r) N. days agricultural labor by household members

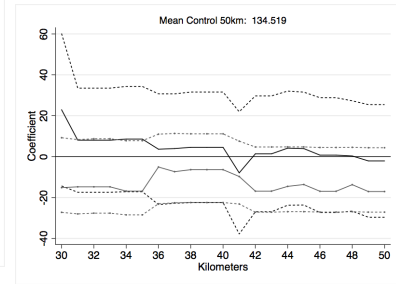


Table A6: LSM-ISA Balancing Test

Variable	Mean Treated	Mean Control	Diff.	
<i>Panel A</i>				
<i>Household Characteristics</i>				
Head Male	0.789	0.781	0.008	(0.028)
Head Age	47.760	48.005	-0.245	(1.109)
Head Age ²	2536.926	2574.646	-37.721	(116.940)
Fraction hh members 0-6 years old	0.204	0.206	-0.002	(0.013)
Fraction hh members 7-12 years old	0.154	0.152	0.002	(0.010)
Fraction hh members 13-18 years old	0.139	0.135	0.004	(0.011)
Fraction male hh members >18 years old	0.238	0.238	0.000	(0.014)
Household size	5.286	5.203	0.083	(0.180)
<i>Education</i>				
More than primary Head	0.497	0.513	-0.016	(0.034)
Education's head missing	0.006	0.004	0.002	(0.005)
N. adult members with > primary	1.143	1.206	-0.063	(0.073)
<i>Asset ownership (score)</i>				
Durable Assets	-1.305	-1.260	-0.045	(0.083)
Productive Assets	-0.040	-0.038	-0.001	(0.029)
<i>Farm Characteristics</i>				
Land title	0.103	0.104	-0.001	(0.021)
Land rights	0.743	0.740	0.003	(0.030)
Land rights missing	0.046	0.041	0.005	(0.014)
Used org fertilizer	0.229	0.231	-0.003	(0.029)
Used inorg fertilizer	0.211	0.188	0.024	(0.027)
Used pesticides	0.149	0.143	0.005	(0.024)
Erosion	0.246	0.250	-0.004	(0.030)
Irrigation	0.029	0.017	0.012	(0.010)
Erosion/Irrigation missing	0.063	0.065	-0.002	(0.017)
Tropic-warm/humid agroecological zone	0.469	0.462	0.006	(0.034)
<i>Land Quality</i>				
Good quality*area	3.292	3.350	-0.058	(1.464)
Avg quality*area	2.377	2.504	-0.127	(0.352)
Bad quality*area	0.245	0.256	-0.011	(0.130)
Soil quality missing	0.063	0.065	-0.002	(0.017)
<i>2008 Distance (Km) to Nearest</i>				
Major Road	8.558	8.543	0.015	(0.433)
Pop Center with >20'000	45.574	43.667	1.908	(2.042)
Market	75.599	75.371	0.228	(3.185)
Border Crossing	134.439	134.037	0.402	(6.639)
District Headquarters	32.211	30.172	2.039	(1.844)
<i>Quality of closest road</i>				
Gravel	0.406	0.431	-0.026	(0.034)
Paved	0.269	0.208	0.0609**	(0.029)
Sealed	0.326	0.361	-0.035	(0.032)

Table A6: LSM-ISA Balancing Test

Variable	Mean Treated	Mean Control	Diff.	
<i>Village Characteristics</i>				
Average hh size in the ward	5.653	5.634	0.019	(0.098)
Population Density	3.171	3.215	-0.044	(0.153)
Population density missing	0.023	0.022	0.001	(0.010)
Sh. hhs with a member wage employed	0.110	0.110	0.000	(0.006)
Sh. illiterate people >15 years old	0.299	0.294	0.005	(0.009)
<i>Services available (less 10km)</i>				
Bank	0.189	0.212	-0.023	(0.027)
SACCOs	0.583	0.565	0.018	(0.034)
Market (daily)	0.291	0.285	0.006	(0.031)
Market (weekly)	0.514	0.548	-0.034	(0.034)
<i>Panel B</i>				
<i>Outcome Variables</i>				
Subjective Welfare: Life	3.895	3.688	0.206	(0.127)
Durable Assets (score)	-1.305	-1.260	-0.045	(0.083)
Productive Assets (score)	-0.040	-0.038	-0.001	(0.029)
Consumption rice (kg) 7d	1.617	2.970	-1.3535****	(0.328)
Consumption maize (kg) 7d	9.967	8.234	1.734	(1.343)
<i>Outcome Variables last year*</i>				
Wage labor	0.348	0.295	0.053	(0.047)
Hire labor	0.379	0.395	-0.016	(0.049)
N. days agri. labor by hhm	102.136	110.590	-8.453	(9.907)
Harvest rice (kg)	192.576	9.663	182.9132****	(43.691)
Sales rice (kg)	137.258	4.476	132.7813***	(41.382)
Harvest maize (kg)	293.485	402.741	-109.2561**	(46.718)
Sales maize (kg)	34.015	114.464	-80.4484****	(22.264)
Log PC real annual consumption	13.050	13.027	0.023	(0.057)
N. members away at least 1m last 12m	0.167	0.260	-0.093	(0.057)

* p< 0.10, ** p<0.05, *** p<0.01, **** p<0.001

Treatment: Roads Rehabilitated between 2011-2013. Test at baseline

*Test at 30km

Table A7: LFS Balancing Test (2006)

Variable	Mean Treated	Mean Control	Diff.	
<i>Panel A</i>				
<i>Household Characteristics</i>				
Head Male	0.743	0.723	0.020	(0.031)
Head Age	46.211	44.786	1.426	(1.284)
Head Age ²	2406.234	2252.429	153.805	(127.237)
Fraction hh members 0-6 years old	0.208	0.243	-0.035*	(0.019)
Fraction hh members 7-12 years old	0.132	0.137	-0.005	(0.013)
Fraction hh members 13-18 years old	0.108	0.102	0.006	(0.008)
Fraction male hh members >18 years old	0.267	0.238	0.028*	(0.017)
Fraction female hh members >18 years old	0.285	0.280	0.006	(0.012)
Household size	4.500	4.836	-0.336	(0.303)
<i>Education</i>				
More than primary Head	0.460	0.437	0.023	(0.044)
Education's head missing	0.000	0.000	0.000	(.)
N. adult members with > primary	1.145	1.132	0.013	(0.125)
<i>2008 Distance (Km) to Nearest+</i>				
Major Road	7.307	7.477	-0.170	(1.884)
Pop Center with >20'000	38.272	54.189	-15.918	(11.048)
Border Crossing	173.804	184.530	-10.726	(42.717)
District Headquarters	23.860	24.753	-0.893	(4.650)
<i>Quality of closest road+</i>				
Gravel	0.419	0.493	-0.074	(0.177)
Paved	0.166	-0.015	0.181	(0.116)
Sealed	0.415	0.500	-0.085	(0.195)
Soil	-0.001	0.021	-0.022	(0.019)
<i>Village Characteristics+</i>				
Average hh size in the ward	5.638	5.938	-0.300	(0.373)
Population Density	0.638	0.622	0.016	(0.138)
Share hhs with a member wage employed	0.149	0.119	0.031	(0.028)
Share illiterate people >15 years old	0.270	0.317	-0.048	(0.045)
Tropic-warm/humid agroecological zone	0.600	0.324	0.275	(0.169)
<i>Services available (less 2km)+</i>				
Primary school	0.175	0.157	0.018	(0.015)
Clinic	0.105	0.091	0.014	(0.023)
Shop	0.196	0.160	0.0357**	(0.017)
Market	0.093	0.081	0.012	(0.019)

Table A7: LFS Balancing Test (2014)

Variable	Mean Treated	Mean Control	Diff.	
<i>Panel B</i>				
<i>Household Characteristics</i>				
Head Male	0.798	0.720	0.078	(0.086)
Head Age	42.749	48.277	-5.528*	(2.981)
Head Age ²	2023.860	2625.183	-601.323*	(326.210)
Fraction hh members 0-6 years old	0.204	0.165	0.039**	(0.019)
Fraction hh members 7-12 years old	0.147	0.152	-0.005	(0.015)
Fraction hh members 13-18 years old	0.103	0.128	-0.025	(0.017)
Fraction male hh members >18 years old	0.282	0.272	0.011	(0.038)
Fraction female hh members >18 years old	0.263	0.283	-0.020	(0.033)
Household size	4.558	4.299	0.259	(0.250)
<i>Education</i>				
More than primary Head	0.582	0.529	0.053	(0.070)
Education's head missing	0.012	0.006	0.007	(0.008)
N. adult members with > primary	1.603	1.529	0.074	(0.176)
<i>2008 Distance (Km) to Nearest+</i>				
Major Road	9.037	10.660	-1.624	(2.157)
Pop Center with >20.000	45.422	45.410	0.013	(11.506)
Border Crossing	137.271	167.211	-29.940	(49.786)
District Headquarters	24.435	24.449	-0.014	(4.217)
<i>Quality of closest road+</i>				
Gravel	0.560	0.621	-0.060	(0.205)
Paved	0.270	0.178	0.091	(0.149)
Sealed	0.170	0.201	-0.031	(0.150)
Soil	0.000	0.000	0.000	(.)
<i>Village Characteristics+</i>				
Average hh size in the ward	5.661	5.274	0.387	(0.301)
Population Density	0.936	0.726	0.209	(0.273)
Share hhs with a member wage employed	0.198	0.180	0.018	(0.035)
Share illiterate people >15 years old	0.255	0.255	0.001	(0.039)
Tropic-warm/humid agroecological zone	0.410	0.273	0.137	(0.159)
<i>Services available (less 2km)+</i>				
Primary school	0.976	0.983	-0.007	(0.012)
Clinic	0.687	0.790	-0.103	(0.114)
Shop	0.971	0.978	-0.007	(0.021)
Market	0.565	0.741	-0.175*	(0.101)

Table A7: LFS Balancing Test (2006)

Variable	Mean Treated	Mean Control	Diff.	
Panel C				
<i>Outcome Variables</i>				
<i>N. household members:</i>				
Have current act.	2.476	2.341	0.134	(0.213)
Earned wage	0.072	0.077	-0.006	(0.023)
Self employed	0.168	0.102	0.066	(0.076)
Work unpaid	0.100	0.104	-0.005	(0.037)
Work on farm	2.417	2.411	0.006	(0.201)
Isic: work in agriculture	2.450	2.440	0.010	(0.196)
Isic: work in manufacturing	0.078	0.037	0.041	(0.044)
Isic: work in construction	0.006	0.009	-0.003	(0.005)
Isic: work in retail	0.053	0.056	-0.003	(0.021)
Isic: work in transport	0.003	0.007	-0.005	(0.004)
Isic: work in hospitality industry	0.030	0.018	0.012	(0.016)
Isic: work as domestic personnel	0.096	0.105	-0.010	(0.037)
Log Hh income paid and self employment (2006 TSh)	10.288	10.141	0.147	(0.234)

Sample design adjusted standard errors are presented in parentheses

* p< 0.10, ** p<0.05, *** p<0.01, **** p<0.001

Treatment: Roads Rehabilitated between 2011-2013. + 2006 value

Figure A6: Treatment Distribution by distance to a treated road

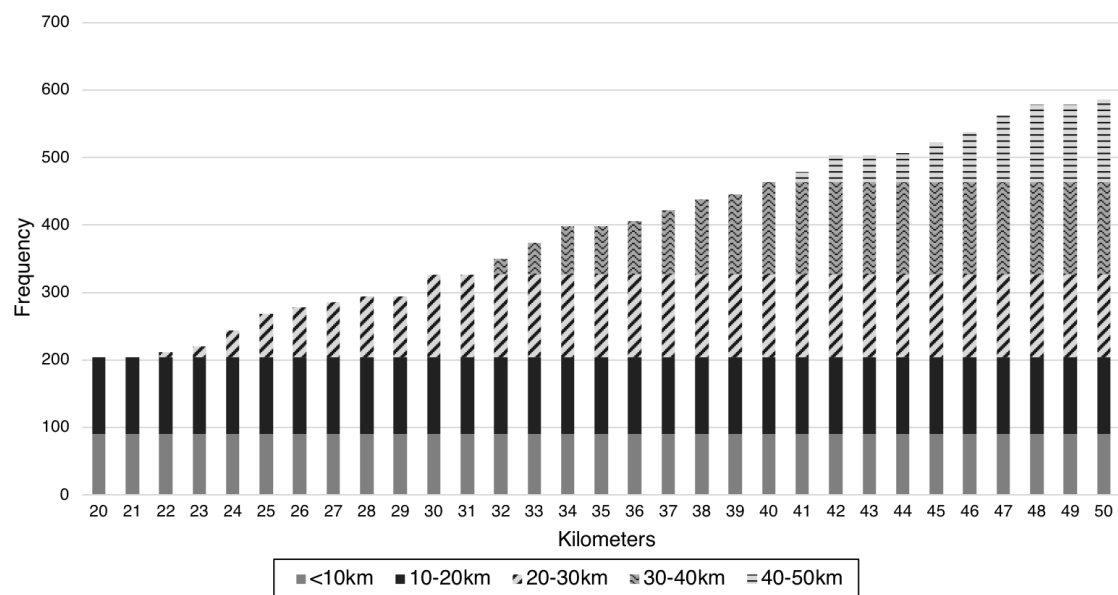


Table A8: LSMS-ISA Observations (3 rounds pooled)

	Household fixed effects-CS	
	Treated	Control
Outcome 7d		
20km treatment	676	2003
50km treatment	2223	2352
Outcome last year		
30km treatment	370	834
50km treatment	1237	1754
	Village fixed effects-CS	
	Treated	Control
Prices		
20km treatment	58	148
50km treatment	182	236

Table A9: LFS Observations (by round)

	Diff-in-diff-CS	
	Treated	Control
2006		
20km treatment	1055	2651
50km treatment	2967	3088
2014		
20km treatment	327	737
50km treatment	1047	996