

Transport Infrastructure and Welfare

An Application to Nigeria

Rubaba Ali

Alvaro Federico Barra

Claudia N. Berg

Richard Damania

John Nash

Jason Russ



WORLD BANK GROUP

Agriculture Global Practice Group

May 2015

Abstract

Transport infrastructure is deemed to be central to development and consumes a large fraction of the development assistance envelope. Yet there is debate about the economic impact of road projects. This paper proposes an approach to assess the differential development impacts of alternative road construction and prioritize various proposals, using Nigeria as a case study. Recognizing that there is no perfect measure of economic well-being, a variety of outcome metrics are used, including crop revenue, livestock revenue, non-agricultural income, the probability of being multi-dimensionally poor, and local gross domestic product for Nigeria. Although the measure of transport is the most accurate possible, it is still endogenous because of the nonrandom placement of road infrastructure. This endogeneity is addressed using a seemingly novel instrumental variable termed the natural path: the time it would

take to walk along the most logical route connecting two points without taking into account other, bias-causing economic benefits. Further, the analysis considers the potential endogeneity from nonrandom placement of households and markets through carefully chosen control variables. It finds that reducing transportation costs in Nigeria will increase crop revenue, non-agricultural income, the wealth index, and local gross domestic product. Livestock sales increase as well, although this finding is less robust. The probability of being multi-dimensionally poor will decrease. The results also cast light on income diversification and structural changes that may arise. These findings are robust to relaxing the exclusion restriction. The paper also demonstrates how to prioritize alternative road programs by comparing the expected development impacts of alternative New Partnership for Africa's Development projects.

This paper is a product of the Agriculture Global Practice Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at rdamania@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Transport Infrastructure and Welfare: An Application to Nigeria¹

**Rubaba Ali, Alvaro Federico Barra, Claudia N. Berg,
Richard Damania, John Nash, and Jason Russ**

The World Bank

Key words: transport, roads, agriculture, development, Nigeria, Africa, multi-dimensional poverty index (MPI), NEPAD

JEL Codes: O1, I3, Q1, L9

¹ We are grateful Francisco Ferreira, Harris Selod, Somik Lall, Atsushi Iimi, Don Larson, Shahe Emran, Remi Jedwab, two anonymous reviewers, and the participants of the NEUDC conference for helpful comments and/or discussions.

Transport Infrastructure and Welfare: An Application to Nigeria

1. Introduction

Governments and donors in Sub-Saharan Africa have devoted considerable resources to the construction and rehabilitation of roads. An emphasis on transport infrastructure is evident in the lending pattern of the World Bank, which commits a larger share of resources to transport infrastructure than education, health and social services combined (World Bank 2007). Total transport commitments in fiscal year 2013 amounted to US\$5.9 billion and rural and inter-urban roads remained the largest sub-sector with 60 percent of lending in FY13 (US\$3.2 billion) (World Bank 2014). The rationale behind these investments is self-evident. Roads, while expensive, facilitate the creation of, and the participation in, markets and are deemed to be central to development. Africa has the lowest density of roads in the world, with 204 kilometers of road per 1,000 square km, nearly one-fifth the world average, and less than 30% of the next worst region, South Asia. Starting from such a low base, the potential for growth due to improvements in transportation infrastructure is presumed to be especially large in Africa.

However, the existing body of research about the impact of roads on economic well-being remains ambiguous, partially because it is hard to disentangle cause and effect. There is even less evidence on where investments might be the most transformative in creating new opportunities to link producers to markets. Given limited resources, there is a need for selectivity in deciding what investments should occur and where these should be located. This paper aims at tackling these issues by drawing on, and improving upon, the best data available, and by using a somewhat novel approach to overcome some of the technical challenges.

The two key challenges of estimating the impact of road networks on economic activity are well known. First is the difficulty of obtaining data which accurately reflect the conditions of the roads, and the cost of traveling along them. This is always a concern when dealing with road infrastructure—the quality of which is constantly in flux—but it is especially a challenge in Africa where infrastructure assessments are infrequent and rural roads are often unaccounted for. The second challenge is overcoming the potential

sources of endogeneity arising from the non-random placement of (i) roads, (ii) spatial sorting of households, and (iii) the geographic emergence of markets.² Such endogeneity could potentially bias econometric results.

Roads tend to be built so as to connect major economic activities, e.g. linking cities, markets, mines, or areas of high agricultural productivity. Hence estimates need to take account of reverse causality in looking at the impact of roads: on the one hand economic potential may determine where roads are built, on the other hand, roads may spur greater economic activity. In situations where natural experiments are not feasible, and panel data are unavailable, instrumental variables are the most commonly used technique to correct for these placement effects, which is the approach used in this analysis. While no instrument is perfect, this paper constructs a variable, termed the “natural path” (described in the data section), which we suggest greatly improves the efficiency of the estimates over, say, estimates produced using more common straight-line instruments.³

Spatial sorting by households could also potentially bias our estimates if, for example, a household moved to a particular location on the basis of a variable which we do not control for. Similarly, locations of the markets which we focus on in this paper (i.e. cities with populations greater than 100,000) may also be endogenous, as they emerged historically in locations of high economic potential. We address both of these potential biases by including carefully chosen control variables in our regression analysis (discussed later in the paper).

Another difficulty with estimating the impact of road networks on economic activity is generating a variable which properly captures the proximate benefits of both local and regional roads. We overcome this difficulty by calculating the actual cost of transporting goods to market along Nigeria’s existing road network. Taking into account the road classification, quality, type of paving and roughness of the terrain, the measure of transport cost to market that is calculated is perhaps the most accurate possible, given existing information.

² See Emran and Hou (2013) which discusses these three sources of bias in the context of rural China.

³ See also Faber (2014) which uses a similar instrument.

Nigeria has an extensive national road network of more 85,000 km of classified roads (Gwilliam 2011). Both paved and unpaved road network densities are more than twice as high as those for the peer group of resource-rich African countries, although still only half of the levels found in Africa's middle-income countries (Foster and Pushak 2011). According to the Africa Infrastructure Country Diagnostic benchmark study (Foster and Briceño-Garmendia 2008): if Nigeria wishes to meet its economic and social targets for transportation infrastructure it would need to invest \$1.2 billion annually for a 10 year period. However, it is important to evaluate the effect of investment on transportation infrastructure to justify large investment.

This paper attempts to provide a more complete picture of the extent to which household welfare and incomes are expected to improve with a given reduction in transport costs. We do so by considering several different outcome variables which we obtain from two household surveys, and a raster data set⁴ on local gross domestic product (GDP). The household surveys employed in this paper are the 2010 Living Standards Measurement Study - Integrated Survey on Agriculture (LSMS-ISA) for Nigeria, and the 2008 Nigeria Demographic Health Survey (NDHS). From these surveys, we are able to obtain several welfare indicators, including: revenue from crop production, revenue from livestock sales, non-agricultural income, employment, wealth, and a multi-dimensional poverty indicator. A key advantage of these household surveys is that the enumeration areas – the geo-locations – are recorded. The raster data set, which we obtain from Ghosh et al (2010), gives an estimate of local GDP at a very fine spatial level for the entire land area of Nigeria.

These measures also gives us the ability to study the effects of transport costs on both 'flow' measures of welfare, or 'stock' measures, which capture much longer term effects. Measures of income, such as crop revenue, livestock sales, and non-agricultural income are in the former category and will be impacted by idiosyncratic shocks or localized impacts, for instance, a bad harvest due to less than average rainfall, or a sudden illness of the household head. Meanwhile, although improving transportation infrastructure can lead to benefits in the short term, many of the benefits will not become apparent for many years after the improvement, at which point households and businesses

⁴ A raster is a matrix of cells where each cell contains spatial information, in our case local GDP.

have had time to adjust to the new equilibrium. To capture these benefits, we also study ‘stock’ measures of welfare, including a wealth index available in the NDHS, and a multi-dimensional poverty index (MPI) which we construct following Alkire and Santos (2010).⁵ In addition, we examine a plausible mechanism of income diversification and improvement in economic activity (all year employment of working age population) by which transport cost reduction may affect local GDP. By looking at these different indicators of welfare, we are able to disaggregate the benefits of a transport cost reduction and obtain insights into the causal pathways to poverty reduction.

The elasticities generated from the household survey and local GDP analysis are summarized in Table 1 below. We then use the elasticity on local GDP to forecast the economic impact of the improvement of the current network. Local GDP is used in these simulations because these data provides a baseline which is available throughout the entirety of the country, unlike household survey data which has only limited spatial coverage. Our forecasts allow for heterogeneous benefits depending on current levels of welfare, current transportation costs, and spatially varying transportation cost reductions from new road construction. This can enable decision-makers to maximize the efficiency with which they use scarce resources, by prioritizing construction of those roads which would have the biggest impact on economic growth and poverty reduction in the region.

Table 1: Estimated Elasticities

Welfare Indicator	Benefit (from a 10% reduction in transport costs)
Crop Revenue	6.4%
Livestock Revenue	3.4%
Non-Agricultural Income	3.3%
All year employment	0.4% (male), 0.3% (female)
Agricultural employment	-4.0% (male), -5.3% (female)
MPI Poverty Reduction	2.6%
Wealth Index	2.0%
Local GDP	5.0%

⁵ Poverty, when measured in a multi-dimensional fashion, which accounts for both physical and human capital, captures much more the ‘stock’, or cumulative welfare effects over time, and is thus a much better measure of welfare over the longer term.

Note: These benefits were estimated using the natural path IV.

2. Related Literature

This paper is related to a vast and rapidly growing literature on the effects of infrastructure on well-being and a continuing debate (among planners, policy makers, and academics) about the role of transport investments in economic growth. This debate has been fostered by limited evidence of a causal relationship and conflicting evidence provided by different studies on the relationship between the two (Gunasekara et al Lakshmanan 2007).

Approaches to addressing this issue have varied considerably and evolved over time. Researchers have examined the effects of road infrastructure and transport capital investments on aggregate productivity (usually measured by GDP or personal income), output elasticity and productivity in developed countries (Aschauer 1989, Lakshman and Anderson 2002, Lakshman and Anderson 2007, Chandra and Thompson 2000, Demetriades and Mamuneas 2000, Annala and Perez 2001, Foster and Araujo 2004, Ihuri and Kondo 2001, Lokshin and Yemtsov 2003, Nadiri and Mamuneas 1996, Munnell 1990, Shirley and Winston 2004, and Sturm 2001), and in developing countries (Deichmann et al 2002, Morrison-Paul et al 2001, Lokshin and Yemtsov 2003; Feltenstein and Ha 1995). The results however remain ambiguous with conflicting evidence of impacts in both developed and developing countries. To a large extent the contradictory evidence and the ensuing debates are a consequence of the identification and reverse causality problems.

A set of recent papers have used rigorous and compelling identification strategies to shed light on the impact of large transport infrastructure improvements (Michaels 2008, Donaldson 2012, Datta 2012, Faber 2012 and Banerjee et al 2012)⁶. One approach is to use panel data estimation methods (Dercon et al 2008, Khandker and Koolwal 2011). Regrettably, however, panel data on transportation costs with adequate observations over time are rare especially in a developing country context, and not available for our analysis involving Nigeria and other West African countries. Another approach is to use spatial panel data with natural experiments that exploit the historical context of transport infrastructure (Donaldson 2012, Jedweb and Moradi 2012, Banerjee et al 2012). Others

⁶ Elhance and Lakshmanan, 1988 and Ford and Poret, 1991 are examples of earlier papers that analyze the impact of aggregate transport investment in Mexico and highway improvement in Sri Lanka.

have used difference-in-difference (Datta 2012) or difference-in-difference with an IV (Faber 2014). In the absence of natural experiments and panel data, numerous studies have attempted to capture exogenous variations in transport cost by incorporating exogenous geographic features (Jacoby and Minten 2009, Shrestha 2012, Emran and Hou 2013).

Recent papers have looked at the mechanisms through which transportation costs impact wellbeing. One of these is that reducing transportation costs leads to greater access to markets, as well as a decrease in both trade costs and interregional price gaps (Donaldson 2012, Casaburi et al 2013). This further affects input and output prices of crops (Khandker et al 2006, Minten and Kyle 1999, Chamberlin et al 2007, Stifel and Minten 2008) as well as land value (Jacoby 2000, Shrestha 2010, Donaldson 2013, Gonzalez-Navarro and Quintana-Domeque 2010). Not surprisingly the literature also finds that access to good quality roads facilitates economic diversification (Gachassin et al 2010, Fan et al 2000, Mu and van de Walle 2007). Another mechanism is that reduced transport costs can insure farmers against negative shocks (Burgess and Donaldson 2012).

3. Data

This paper utilizes several different data sets to analyze the relationship between transportation costs that households incur to access the nearest market (defined as cities with population of at least 100,000) and several different measures of welfare.⁷ In order to do so, a very thorough road network was constructed for Nigeria, using several sources of data described in section 3.1 below. To correct for endogenous road placement, an IV approach is used and a instrument was generated for this paper which we refer to as the *Natural Path*, described in section 3.2. Finally, a multitude of welfare indicators are utilized in this paper, and these are described in section 3.3. (For summary statistics of the main variables used, see Appendix I.)

3.1 Travel costs to market

Throughout much of the literature on transport infrastructure, the variables of choice to measure infrastructure availability typically fall into three categories; local road

⁷ This paper analyzes the combined effect of both large transport infrastructure, such as highways, and rural roads, and thus differ from Michaels, 2008, Donaldson, 2012, Datta, 2012, Faber, 2012 and Banerjee, Duflo and Qian, 2012, which analyze the impact of large transport infrastructures, highways and railways. It also differs from Jacoby and Minten 2008, Dorosh et al 2010, Gibson and Rozelle (2003), Ali (2011), Khandker, Bakht and Koolwal (2011), Mu and van de Walle, 2007 which analyze the impact of smaller rural roads.

quality/density (e.g. Casaburi, et al 2013, Garcia-Lopez et al 2013, Gertler et al 2014), travel distance to destination (e.g. Martincus et al 2013, Banerjee, Duflo and Qian 2012, Stifel and Minten, 2008, Fafchamps and Moser, 2003; Jacoby, 2000; Minten and Kyle, 1999), and travel time to destination (e.g. Dorosh et al 2012). These are all attempts to proxy the true price of traveling or transporting goods throughout a road network. Proxies are necessary because obtaining true transport prices would require surveying every possible origin and destination to determine the local price of shipping goods, which would be very difficult, if not infeasible. These techniques, while certainly correlated with transport prices, are imperfect measures. The biggest shortcoming of these measures is the fact that they often fail to distinguish between roads of different qualities and across different terrains. It is certainly more costly to travel along an unpaved, tertiary road, or a road with a steep gradient, than it would be to travel down a flat, paved, national highway. However, simple measures of road infrastructure will not distinguish between these routes (although travel time to some extent can account for this by reducing speed on unpaved roads).

To better account for this, we use the Highway Development Management Model (HDM-4) and a mixture of GIS tools to estimate the actual cost of traveling along a road. This model penalizes routes which follow roads in less than perfect condition, that are unpaved, or that are not flat (See Appendix II for more details). While this is still not a perfect measure of transport prices, it is a significant improvement over the current state of the literature. The total travel cost to the cheapest market (defined as a city of at least 100,000 residents) is calculated using an iterative cost-minimizing process in which every possible travel path to every available market was calculated, and the least cost one chosen as the optimal route.⁸

3.2 The natural path instrument

As discussed above, it is well established in the literature that simple OLS regressions will often yield biased estimates of the effects of public investment due to the non-random placement of infrastructure. In cases where roads are built to connect regions of high economic potential, OLS will tend to overestimate the impacts of roads.

⁸This was done in ArcGIS using the Network Analysis Extension Closest Facility Tool.

If on the other hand, roads are built with poverty-reduction goals in mind, then OLS would instead underestimate the impacts.

In order to eliminate the bias from this endogeneity, we use an instrumental variable (IV) approach based on the topography of the land between the origin (households) and the destination (markets). Specifically, this variable, which we refer to as the “natural path”, is the time that it takes to walk along the time minimizing route from a given location to the nearest market, in the absence of roads. Faber (2014) made use of a similar instrument. Given that road construction costs are a function of segment length, and land topography, the natural path route is therefore highly correlated with the most cost effective place to construct a road network, if economic benefits were ignored. Moreover in the context of Africa it captures many of the historic trade and caravan routes where head-loading (walking) was the dominant pre-colonial mode of transportation.⁹ Therefore, it is plausible to suggest that any endogeneity in the road network from placement decisions (i.e. decisions to place roads in areas which would maximize economic benefits) is captured in the difference between the current road network, and the natural pathway. This instrument is strictly an improvement over “straight-line” instruments as the natural path more accurately represents what straight-line instruments are attempting to estimate, that is, the most cost effective route to connect two points, while excluding any other economic benefits. Details on the GIS algorithm and data used to construct the Natural Path are in Appendix III.

3.3 Welfare indicators

In order to robustly estimate the welfare benefits of reducing transportation costs, we explore several different welfare indicators from three different sources of data, all of which are geolocated. The first data source described here is the 2010 Living Standards Measurement Study - Integrated Survey on Agriculture (LSMS-ISA) for Nigeria.¹⁰ The LSMS-ISA is a national survey on household welfare conducted by the Nigerian Bureau of Statistics and the World Bank’s Development Research Group (DEC). The panel is a 5,000-household subset of the 22,000-household nationally representative General

⁹ Animal traction was unavailable due to the high incidence of disease carried by the tsetse fly.

¹⁰ See Kuku-Shittu et al (2013) for a recent example using the Nigeria LSMS-ISA.

Household Survey (GHS).¹¹ LSMS-ISA provides information on total crop revenue, livestock revenue, and non-agricultural income over the past year, at the household level. Another household survey we exploit is the Nigeria Demographics and Health Survey (NDHS) from 2008, which contains an index on household wealth, and various measures of health and educational attainment of household members, which we use to construct a multi-dimensional poverty index. The NDHS is a nationally representative survey of nearly 50,000 Nigerians aged 15-59.¹² Finally, we use the Nigerian portion of the nighttime lights raster data set (of spatially disaggregated GDP for all of Nigeria) from the lights data set for the entire world developed by Ghosh et al (2010).

4. Empirical Framework

Our main identification strategy is to instrument for cost to market with the natural path variable (i.e. time it takes to walk to market along the natural terrain).

To illustrate the approach, consider the following model:

$$\ln Y_i^k = \beta_0 + \beta_1 \ln T_i + X_i' \gamma + \varepsilon_i \quad (1)$$

$$\ln T_i = \alpha_0 + \alpha_1 \ln N_i + X_i' \theta + \mu_i \quad (2)$$

where Y_i^k denotes the level of outcome k (agricultural revenue, livestock sales, non-agricultural income, multi-dimensional poverty index, wealth index, local GDP, income diversification, and all year employment) indicating welfare or employment of household i in case of the two household surveys, or location i in case of local GDP. T_i is the transport cost to market, X_i is a vector of household controls, and N_i is the natural path

¹¹ The LSMS-ISA is part of a \$19 million project of the Bill and Malinda Gates Foundation. In Nigeria, the LSMS-ISA data was collected twice over two seasons. The Post-Planting Survey was conducted August-October 2010. This was followed by the Post-Harvest Survey in February-April 2011. Each survey is made up of three integrated questionnaires: household, agriculture, and community. In addition, certain geo-variables are available (including information on agro-ecological zones). Each enumeration area is geo-located, allowing us to merge this data with spatial data from other sources. For the purposes of this analysis, we use the 2010 post-planting survey, mainly focusing on the agriculture questionnaire with a few variables (e.g. labor) from the household survey.

¹² The original purpose of the survey was to inform policy makers on a variety of issues mainly affecting women and children, including fertility preferences, infant and young children feeding practices, nutritional status, and early childhood and maternal mortality. For an explanation of the sampling procedure used in the NDHS, see Appendix IV.

variable. For local GDP analysis, household controls are replaced with geographic-level control variables.

The key parameter of interest is β_1 , the causal impact of the cost of traveling to the cheapest market, on household welfare. Three of the outcome variables analyzed represent sources of income (i.e. from crops, livestock, or nonagricultural sources) and are seemingly related. Thus, these three outcomes are also estimated using a seemingly unrelated regression (SUR) framework. Specifically, since transport costs are endogenous, we employ three-stage least squares (which combines SUR and IV methods). The remaining outcome variables are estimated using the customary two-stage approach. In all cases, the endogenous transport cost variable is instrumented using the natural path variable.

In addition to addressing the endogeneity of non-random road placement, we further consider the potential endogeneity stemming from non-random locations of households and markets. Failure to take these into account could yield potentially biased estimates. We address these sources of bias through carefully chosen control variables to be included in X_i .

Spatial sorting by households could potentially bias estimates if, for example, a household's location was determined by a variable that has not been controlled for. In the context of Nigerian farmers, this spatial sorting is arguably much less of a concern. Given the lack of a functional land market, it is unlikely that farming households would change locations, as moving the entire household would require abandoning one's land. In the Nigeria LSMS data, for example, 74% of land is inherited with less than 6% bought (the remaining land is either rented or used free of charge). Therefore, while individuals do migrate (usually to cities), it is rare for the household as a whole to relocate, especially to other rural areas. Even so, we do control for characteristics of the household head (i.e. age and literacy) in our regressions that may indicate whether a household will have the means or relocate.

The location of markets is largely determined by environmental and topographical factors. As discussed previously, cities tend to emerge historically in areas of high economic potential. To address the endogeneity of market locations we include fixed effects based on which city a household travels to according to the least cost path, which

we refer to as ‘marketshed fixed effects’. By doing so, we are accounting for any unobserved heterogeneity between market locations.¹³

As a robustness check for the validity of the instruments, we calculate a set of Conley Bounds (Conley et al (2012)), for the coefficients of interest. To illustrate this, let N_i be the IV and rewrite equation (1) as follows:

$$\ln Y_i = \beta_0 + \beta_1 \ln T_i + X_i' \gamma + \delta N_i + \varepsilon_i \quad (3)$$

The traditional IV strategy assumes that $\delta = 0$. Conley Bounds allow δ to be close but not actually equal to zero, in other words they allow the IVs to be only “plausibly exogenous”. By allowing the value of δ to vary, we can then test how sensitive the estimates are to different degrees of exogeneity.

5. Empirical Results

For each of our welfare measures, our identification strategy focuses on the use of our natural path instrumental variable. As a robustness check, we further report the results from using a Euclidean Distance IV,¹⁴ which yields very similar results. In the interest of brevity, our discussion focuses on the natural path results. To account for the fact that household observations are likely correlated within enumeration areas, we report results clustered at the enumeration area. Overall, the empirical results suggest that lowering the cost to market would yield significant benefits to rural households—though the impacts appear to depend on the source of income and location.

5.1 LSMS Measures of Household Welfare

We begin by presenting the effects of transport cost on income from different sources, such total revenue from crop sales, livestock sales, and non-agricultural income of the household over the past year, which are the flow measures of household welfare using the LSMS-ISA data for Nigeria.

¹³ As a robustness check, we instead controlled for regional fixed effects (e.g. Agro-ecological or geopolitical zones) in place of marketshed fixed effects. With the exception of livestock, which turned out insignificant, all outcome variable estimates remained robust.

¹⁴ Similar to the natural path variable, Euclidean distance measures the straight-line distance from the household (or cell, in the case of local GDP) to the least-cost market.

Crop Revenue

Our crop revenue regressions suggest that, on average, decreased transport costs lead to increased household revenue from crop sales. Preliminary SUR results are reported in column (1) of Table 2, which shows that that decreasing transport cost by ten percent would increase crop revenue by approximately 6.2 percent. While these preliminary SUR results are reassuring in that they conform to prior expectations, they must be treated with caution as they do not take account of the endogeneity of roads. Table 2, columns (2) and (3) report the three-stage least squares (3SLS) estimates where cost to market is instrumented by the Euclidean distance and natural path, respectively. These unbiased estimates of the effect of transport costs are slightly larger than the SUR coefficient (at -0.63 and -0.64 respectively). The natural path IV passes the Angrist-Pischke F Test of Weak Identification, with the F statistic far exceeding 10, the rule of thumb.

Livestock Sales

As with crop revenue, we report both our SUR and 3SLS estimates of livestock sales in Table 3. In both cases, we find that the estimated coefficient on cost to market is negative and significant. After considering the various robustness checks below, it would seem that overall the relationship between cost to market and the sale of livestock is not very robust. This might be in part due to the multiple roles of livestock as a store of value and capital good. Livestock sales in a given year are therefore driven by decisions on asset management (e.g., need for revenue to manage temporary household needs for cash such as weddings or natural disasters losses), much more than are crop sales

Non-Agricultural Income

Turning next to the relationship between access to markets and non-agricultural income, economic theory suggests that as transport costs decrease, more opportunities outside of the agricultural sector become available. To investigate this, we regress log non-agricultural income on the log of cost to market, holding constant household characteristics and marketshed fixed effects.

Table 4 reports the SUR estimates (column 1), and the 3SLS estimates (column 2) for non-agricultural income. The SUR estimates suggest that reducing transport costs by

10 percent increases non-agricultural income by 3.2 percent. After controlling for endogeneity, our 3SLS estimates find a higher increase in income: 3.3 percentage points.

Robustness Checks

As a robustness check, we estimate the impact of transport costs on the three sources of income separately. That is, we estimate three sets of two-stage least squares models and find very consistent estimates. Controlling for Agro-Ecological Zone fixed effects in place of the marketshed dummies yielded similar results for each of the three outcomes. Further, the estimated impact of transport costs on crop revenue and non-agricultural income was found to be robust to a number of alternative specifications. For example, they were robust to the inclusion and exclusion of land, labor, fertilizer, and irrigation control variables. Further, alternatives were tested such as credit, as well as an indicator of the presence or absence of a hospital within the community. These results are reported in Appendix VII.

To check the robustness of our estimates to the relaxation of the exclusion restriction on the natural path instrument, a set of Conley Bounds are calculated following Conley et al (2012), and reported in Table 10.¹⁵ For both the crop revenue and non-agricultural income, the 95% confidence interval suggests that the coefficient on the variable of interest remains consistently negative. Taken together with the Angrist-Pischke F statistic and first stage results, these findings suggest that our estimates are robust to relaxation of the exclusion restriction.¹⁶ In the case of livestock sales, the Conley Bound 95% confidence interval crosses the zero line. This, together with the abovementioned alternate specifications, suggests that these livestock estimates are not as robust as those of crop revenue and non-agricultural income.

5.2 NDHS Measures of Household Welfare

We now turn to the two measures of household welfare from the NDHS; the wealth index and the multidimensional poverty index.

¹⁵ Note that since the Conley Bound estimation is designed for two-stage least squares, and so for the purposes of this robustness check the three outcome variables are treated separately.

¹⁶ These first stage results are taken from separate two-stage least squares models.

Nigeria Demographic and Health Survey

The first indicator of household welfare is the “wealth index”, available in the DHS data.¹⁷ The second indicator, a multi-dimensional poverty measure, is generated specifically for this study. We follow Alkire and Santos (2010) to calculate the Multi-Dimensional Poverty Index (MPI) for each household. The MPI is a weighted sum of ten indicators of deprivations across three dimensions: education, health, and standard of living. We follow convention and use equal weights for each of the three dimensions and for indicators within dimensions. A household is considered to be multi-dimensionally poor if it is deprived in three of the ten weighted indicators. Table A4 in Appendix V gives more specific details on how this index was constructed.¹⁸

Household Wealth Index

Table 5 presents the results from regressing the wealth index on transportation costs (both in natural log form). Column (1) in Table 5 presents the coefficients from OLS estimation, and column (2) presents the coefficients from 2SLS estimation. The coefficient on the natural path instrument in the first stage is very highly statistically significant and positively related to transportation cost to the market, as expected. Our results indicate that a 10 percent reduction in transportation cost leads to a statistically significant 2.3 percent increase in the wealth index according to OLS estimation, and a 2.1 percent increase in the wealth index from our IV estimates.¹⁹ Again, our IV passes the Angrist-Pischke F Test of Weak Identification, with the F statistic far exceeding 10. The control variables tend to follow intuition. Households located in areas of higher agricultural potential, larger households, and households with more adults in the working

¹⁷ The wealth index is an estimate of a household’s long term standard of living. It is computed using data from the household’s ownership of consumer goods; dwelling characteristics; type of drinking water source; toilet facilities; and other characteristics that are related to a household’s socio-economic status (NPC 2009). To construct the index, each of these assets are assigned weights (factor scores) generated through principal component analysis, and the resulting asset scores are standardized in relation to a standard normal distribution with a mean of zero and standard deviation of one (Gwatkin et al. 2000). Each household is then assigned a score for each asset, and the scores are summed to arrive at a final number. More detailed information on the wealth index is generated can be obtained from Shea and Johnson (2004).

¹⁸ For robustness, we also calculate an additional MPI using data from the LSMS. Despite differences in the data, results from the LSMS generated MPI are quite similar to that from the NDHS. These results are presented in Appendix VI.

¹⁹ As with the regressions using LSMS data, we cluster our standard errors in the DHS regressions at the enumeration area.

age group (15-49 years for females and 15-59 for males) tend to have accumulated more wealth. Households which are agriculturally involved, rural households, and households with more young children have lower levels of wealth.

Multi-dimensional poverty

Turning next to the impact of transportation costs on the probability of a household being multi-dimensionally poor, we report two sets of results in Table 6: a standard linear probability (OLS and IV) and the marginal effects from maximum likelihood estimation (probit and IV probit). We present both results for robustness, but for space considerations, we only interpret probit models here since the two sets of estimates are broadly consistent.²⁰ Overall, decreasing a household's transport cost to market by 10 percent reduces a household's probability of being multi-dimensionally poor by 2.6 percent.

Our results also indicate that households that live in rural areas or are agriculturally involved are more likely to be multi-dimensionally poor, and households that live in areas with higher agricultural potential are less likely to be multi-dimensionally poor. Comparing the probit and IV probit marginal effects, we find that the IV probit estimate (0.26) is considerably larger than the probit estimate (0.08) indicating that the Probit model underestimates the effect of transportation costs on multi-dimensional poverty and that the IV estimation approach was important to obtain an unbiased, accurate measure.

The Conley Bounds in Table 10 show that with both the wealth index and the multi-dimensional poverty index, the coefficients on market cost are robust to a relaxing of the exclusion restriction. The coefficients remain within a small range, and of the same sign, when the exclusion restriction is relaxed.

5.3 Local GDP

Our next set of results looks at the impact of transport costs on local GDP. Our data on local GDP comes from Ghosh et al. (2010) which estimates a raster data set of local economic activity using nighttime light satellite imagery collected by the National Oceanic and Atmospheric Administration (NOAA). This data set spatially disaggregates

²⁰ Given that the outcome variable in this case is binary, probit may be more efficient but least squares may be more robust because it does not rely on distributional assumptions.

Nigeria's (among other countries) 2006 GDP into square pixels 30 arc seconds wide (approximately 1km²), using the fact that brighter lights at night are associated with higher levels of economic activity (see Ghosh et al. 2010 for additional details about how these data were generated). Given the granularity of our control data, we aggregated this data into square cells with sides measuring 5 arc minutes in length (approximately 10km). As control variables, we include total population within each cell,²¹ total population squared, the Euclidean distance to the nearest mining facility,²² as well as indicators measuring the agro-ecological potential yield of the land for four staple crops²³—cassava, maize, rice, and yams—and their squared terms (all variables are in natural logs). In addition to these control variables, marketshed fixed-effects are included in the regressions. This specification is tested both for all of Nigeria, and also for only rural areas.

Columns (1), (2), and (3) of Table 7 show OLS and IV results. The OLS estimate of the coefficient on transportation costs implies that a 10 percent reduction in transport costs increases local GDP by 5.4 percent. The IV estimates are slightly lower at approximately 5 percent. Both the coefficient on population, and its squared term are significant and positive, implying that there important agglomeration economies to local GDP. The negative coefficient on distance to mine implies that economic activity is, as we should expect, denser around mining facilities, although this relationship is not very strong. The coefficients on the agricultural potential of various crops are difficult to interpret because they are highly correlated with each other.

²¹ Population data is from Landsat and is available here: <http://web.ornl.gov/sci/landscan/>

²² Distance to the nearest mining facility is included because mines tend to be areas of great economic activity. In addition to the economic activity at the mine, mines can often generate economic spillovers for industries which service the mine and the mine's workers and their families. Data on mining facilities throughout Nigeria was obtained from the National Minerals Information Center of the USGS. The data set includes geo-referenced data on all mining facilities, active or closed, between 2006 and 2010. Because of the wide definition of what a mining facility actually is, we selected only a subset of mining facilities available to include in our data set. Facilities selected were those which involved the extraction of minerals or hydrocarbons from the ground (specifically coal, tin, iron, nitrogen and petroleum), or the processing of hydrocarbons. Mining facilities that were in the USGS data set but not included in this analysis include facilities like cement plants, or steel mills, which are likely concentrated in large cities or manufacturing areas. We also excluded plants that were labeled as being closed.

²³ Agro-ecological potential data is from GAEZ, a product of FAO. It considers climate and soil conditions to estimate the maximum potential yields in each region for a large number of crops. The data used in this model assumes climactic conditions similar to the 1961-1990 baseline level, and is calculated assuming low input systems.

Columns (4), (5), and (6) of Table 7 show OLS and IV results when we only include rural areas of Nigeria. Examining the coefficient on transportation costs in the IV regression, we see that the effect of reducing transportation costs is slightly lower when urban areas are omitted; a 10 percent reduction in transportation costs implies a 4.5 percent increase in local GDP. However, a difference in means test shows that the coefficient on transportation costs for the full sample is statistically indistinguishable from that using only rural observations. Similarly, the coefficients on the control variables do not change significantly between columns (3) and (6), with the exception of the coefficient on population becoming insignificant (but the squared term remains significant, leading to similar interpretation). Again, the Conley bounds in Table 10 show that our coefficient on market cost remains negative and within a small range when the exclusion restriction is relaxed.

Given the spatial layout of the data used in the above regressions, there is a possibility that spatial autocorrelation could be biasing these results.. A Moran's I test confirms the presence of spatial autocorrelation in the residuals of the regressions in Table 7. In order to test for potential bias resulting from this, we employ a bootstrapping technique in which the data set is resampled 1,000 times in a way that ensures spatial independence of each of the samples.²⁴ The results from this bootstrap are shown in Table A11, Column 2, with the standard, non-bootstrapped results presented again in Column 1 for comparison. The coefficient of interest, on cost of market, while slightly lower in the bootstrapped model, is statistically indistinguishable between the two models. This implies that although spatial autocorrelation is present in the data, any bias it creates on our estimates is negligible. The result that the estimates from our non-spatial model are robust even in the presence of short-distance residual special autocorrelation (RSA) is not uncommon (Hawkins et al. 2007).

²⁴ Specifically, it is short-distance residual special autocorrelation (RSA) that has the potential to bias our estimates (Lichstein et al 2002). To ensure the spatial independence of each sample, a spatial correlogram of Moran's I based on the residuals from the regression in Table 7, column 3 was generated. The correlogram indicates that the extent of spatial autocorrelation is approximately 20kms. Therefore, samples were drawn in a way that ensured that each point was at least 30kms away from every other point in the sample (30km was chosen to be conservative). See Appendix A11 for more details on this process, and Dormann et al (2007) and Hawkins et al (2007) for other examples of this technique being used in practice.

5.4 Level and diversification of economic activity

To analyze whether lower transportation cost creates more, as well as diverse employment opportunities, we examine how transportation costs effect year round employment, and agricultural vs. non-agricultural employment for males and females. Our dependent variables consist of dummy variables indicating employment types of individuals within households- full time vs. less than full time/unemployed (Table 8), and agricultural vs. non-agricultural employment (Table 9). We also break up our analysis by gender, to allow for heterogeneous effects. Columns (3) and (6) in Table 8, indicate that a 10 percent reduction in transportation cost increases the probability of being employed all year round as opposed to being unemployed or seasonally employed by 4 percent for males and 3 percent for females. Columns (3) and (6) in Table 9 indicate that a 10 percent reduction in transportation cost will decrease the probability of being agriculturally employed for those who are employed (i.e. improve non-agricultural employment) by 4 percent among males and 5.3 percent among females. These results suggest that reducing transportation cost leads to an increase in both economic activity and diversification away from agricultural activity.

6. Economic Impact of Alternative Road Investments

In this final section, we use our estimate of the local GDP elasticity of transport costs to simulate the effect of several road infrastructure improvement projects which have been proposed by the World Bank, The African Development Bank (ADB) and the New Partnership for Africa's Development (NEPAD), a planning and coordinating technical body of the Africa Union. The approach is meant to be illustrative and could be applied to study any road improvement or new road construction project within Nigeria.

6.1 The projects

We present below an estimate of the impact of improving the portion of NEPAD's and ADB's Trans African Highway²⁵ project segments that runs through

²⁵ The system of Trans African Highways consists of 9 main corridors with a total length 59 100 km. The concept as originally formulated in the early 1970s, aims at the establishment of a network of all-weather roads of good quality, which would: a) provide as direct routes as possible between the capitals of the continent, b) contribute to the political, economic and social integration and cohesion of Africa and c) ensure road transport facilities between important areas of production and consumption.

Nigeria as shown in Figure 1. It is assumed in the simulations that each corridor would be improved from its current quality, to paved and good condition status. The baseline scenario is obtained from FERMA²⁶ and requires that 20 km need to be paved, while approximately 1,275 km need to be improved from poor to good and 815 km from fair to good condition.

6.2 Simulation Methodology

To calculate the change in transportation cost resulting from the improvement of each corridor, we follow the same procedure utilized in section 3.1 to estimate the travel cost to the cheapest market and compare these to current transport costs. The percentage change in transportation costs for each cell, if all three of the corridor improvement projects were completed, is shown in Figure 2.

New transport cost elasticities are calculated, one for each of the 6 geopolitical zones of Nigeria²⁷. This allows us to account for regional heterogeneity in the benefits of reducing transport costs. These elasticities are shown in Table 11 (the South East region dummy is omitted). Formally, local GDP increase calculation is given by:

$$B_j = \sum_i \eta_k * \tau_{ij} * y_i, \quad (4)$$

where B_j is the total increase to local GDP due to project j , η_k is the local GDP elasticity of transportation costs for region k (from Table 11), τ_{ij} is the percentage change in transportation costs in cell i due to project j , and y_i is the baseline GDP in cell i , from the local GDP data. This increase in GDP represents an increase in annual GDP over the baseline level. These benefits will accrue every year as long as the benefits from reducing transportation costs and baseline GDP levels, both remain constant.²⁸ For each project, we estimate the increase in local GDP in each grid cell separately, and then aggregate these benefits to arrive at an aggregate total benefit. The increase in local GDP is then summed up amongst all grid cells to arrive at an aggregate value.

²⁶ For more information about the FERMA road survey and GIS methodology see “Spatial Analysis and GIS Modeling to Promote Private Investment in Agricultural Processing Zones: Nigeria’s Staple Crop Processing Zones” presented at the Annual World Bank Conference on Land and Poverty 2013

²⁷ The 6 geopolitical zones of Nigeria are: South East, South South, South West, North Central, North East, and North West.

²⁸ This would be a dubious assumption over the long term, but might be a reliable approximation over a short, 3-5 year period.

The spatial approach also allows us to identify the number of beneficiaries and estimate the benefit per road kilometer improved. Given the inherent uncertainty involved in statistical analysis, we calculate total benefits given our preferred elasticities (point estimates given in Table 11), as well as a range of benefits representing the 95% confidence intervals around those point intervals.

6.3 Simulation Results

Benefit Point Estimates

We first present results from the point estimations of the elasticities in Table 11 for the benefits of the three NEPAD projects. These are shown in Table 12. The benefits from these projects are found to be quite large. The North-South Corridor, which is the longest road of the three projects, would result in estimated annual benefits of over \$1 billion. Annual benefits from the Northeastern and Southern corridors are significantly lower, at \$233 and \$529 million, respectively. Nevertheless, these roads are also shorter, potentially implying a lower cost of improvement. If all projects were completed, total estimated annual benefits would be \$1.8 billion. Note that the total benefits of all of the construction projects are not equal to the sum of the benefits of each of the projects individually, because there is some overlap between the project locations. Figure 3 shows where exactly the increase in local GDP would occur if all three of the corridor improvement projects were completed.

Turning to the third column of Table 12, we calculate the total benefit per KM of each corridor, which allows us to rank the projects according to their benefit-efficiency; i.e. assuming road improvement costs are uniform and equal across projects, which project gives us the most benefits per unit cost. We see that the North-South Corridor project would return annual benefits of approximately \$970,000 per km improved, significantly higher than that of the Northeastern Corridor, with benefits of \$250,000 per km improved, and moderately higher than the Southern Corridor, at \$730,000 per km improved.

Using Landscan population data, we also get an approximation of the number of people whose transportation costs to market would decline, as a result of each project. This is shown in column (4) Again, the North-South Corridor has the biggest impact here, benefiting 23.1 million people. The Northeastern and Southern Corridor projects would

benefit 14.9 and 9.2 million people, respectively. Figure 4 shows the total population affected if all three corridor improvement projects are completed. By dividing total benefits by the number of people affected, we arrive at estimated benefits per person affected (column 5). Despite only having the second largest overall impact, the Southern Corridor project has the largest benefit per capita, at \$57.7 per person benefited.

Range of Plausible Benefits

Given that statistical estimates are not precise, we also present benefit ranges based on the 95% confidence interval surrounding the estimated local GDP elasticity of transportation costs. These intervals are given in Table 13. The recalculated benefits for each of the NEPAD projects for these two elasticity bounds are given in Table 14. When all projects are completed, the estimated annual benefits range from \$1.2 billion to \$2.3 billion. Per capita and per km benefit could also easily be calculated for this range of benefits. Because the total number of people affected and the length of each road will not change, these values are not shown for brevity.

6.4 Road Prioritization

By analyzing small segments of each road separately, we can further prioritize these infrastructure projects not just by which overall project would have the largest impact, but within each project, which segments should be improved with greater urgency.

In order to analyze different segments of the road, we first divide roads into “marketsheds.” Recall that a marketshed is defined by the land area served by each city (with a population of at least 100,000 residents), when the transport costs of reaching the city are minimized. The size and shape of a city’s marketshed will therefore depend on both the road network around that city, and its proximity to other cities.

Table 15 shows a list of the marketsheds which contain a portion of one of the three NEPAD roads. Depending on the priorities of the policy maker, several alternative metrics may be employed to prioritize roads. If equitable and shared prosperity are the highest priorities, one might look to maximize the total population affected (column 6), or total population affected per KM of improved road (column 8). In either of these examples, improving roads around the Lagos marketshed would have the largest impact, affecting 8.6 million people, or 249,000 people per KM improved. Total population

affected per KM of improved road is displayed visually in Figure 5. If one is looking simply to generate the largest economic benefits, then total GDP increase (column 3) is the variable of interest. In that case, the portion of the road within the Lagos marketshed would again be the most beneficial to improve, as increasing GDP within that marketshed generates an estimated benefit of \$681 million. If the goal is to have the largest percentage increase in GDP, then local GDP increase as a share of total marketshed GDP (column 4) is the relevant metric, and the roads falling within the Abakaliki marketshed should take top priority. Finally, if economic efficiency is the most important, then GDP increase per km improved (column 7) should drive decision making. In this instance, again Lagos emerges as the top priority. Figure 6 displays visually the benefits per kilometer of each road segment.

7. Discussion and Conclusion

In summary, we find that reducing transportation costs in Nigeria would lead to a significant increase in several dimensions of welfare, an increase in economic activity, can increase all-year employment, and can allow for income diversification opportunities. It further suggests that income diversification and increase in income from different sectors lead to long-term wealth accumulation and poverty reduction. We also provide guidance on how the analysis conducted in the study can be used to prioritize road projects based on the objective of the planner.

The paper contributes to the literature on transport cost in a variety of ways. It creates a rich data set on transportation costs for Nigeria by combining data on quality and road networks from different sources. We deal with several potential sources of bias arising from the non-random placement of roads, spatial sorting of households and the geographic emergence of markets. We are thus able to disentangle cause and effect, using a novel instrumental variable, the time taken to reach markets using a natural pathway. Results from Conley Bounds demonstrate that our estimates are robust even in the case that the exclusion restriction assumption of our IV fails.

The above analysis, while by no means exact, represents a robust attempt at estimating the economic impact from several road improvement projects. Although we

believe we have used the best possible methods, and the best possible data, several shortcomings are acknowledged.

In calculating our estimated local GDP elasticity of transportation cost, we use data which itself is estimated. This adds an additional level of uncertainty to our estimates, but uncertainty which is unavoidable due to the fact that spatially disaggregated data on actual (non-estimated) GDP is not available. Additionally, even though our elasticity is based on estimated data, it falls within the range of other elasticities derived using survey data, suggesting that the estimate is plausible.

Another potential short-coming is the fact that we are using cross-sectional data, which can often make discerning causality very difficult. The instrumental variable technique employed is commonly used in the literature, and we believe our IV is a significant improvement over those used by other very well cited authors. Nevertheless, there is no such thing as a “perfect” instrument. For this reason, we have been careful to present our point estimates along with the respective Conley bounds, which give a range of estimates under the assumption that our instruments are not perfectly exogenous.

Finally, it is important to note that benefits from these road projects simulated in section 6 will not all occur immediately, nor all at once. They will likely cascade over time, as people begin learning of the new, lower transportation costs, and adjusting their behavior accordingly. Therefore, these estimates should be considered long-term annual benefits.

References

- Alkire, S. & Santos, M.E. (2010). Acute multidimensional poverty: A new index for developing countries. OPHI Working paper 38.
- Annala, C., & Perez, P. (2001). Convergence of public capital: Investment among the United States 1977–1996. *Public Finance and Management*, 1(2), 214–229.
- Aschauer, D. A. (1989). Is public expenditure productive. *Journal of Monetary Economics*, 23, 177–200. <http://www.worldbank.org/en/topic/transport/overview#2>.
- Banerjee, Abhijit, Esther Duflo, and Nancy Qian (2012) “On the Road: Access to Transportation Infrastructure and Economic Growth in China” NBER Working Paper 17897
- Burgess, Robin and Dave Donaldson, 2012, “Railroads and the Demise of Famine in Colonial India”, working paper
- Casaburi, Lorenzo, Glennerster, Rachel and Tavneet Suri “Rural Roads and Intermediated Trade: Regression Discontinuity Evidence from Sierra Leone”, MIT Working paper
- Chandra, A., & Thompson, E. (2000). Does public infrastructure affect economic activity? Evidence from the rural interstate highway system. *Regional Science and Urban Economics*, 30(4), 457–490.
- Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi (2012) “Plausibly Exogenous”, *The Review of Economics and Statistics*, 94(1): 260-272
- Datta, Saugato (2012), “The Impact of Improved Highways on Indian Firms”, *Journal of Development Economics*, 99(1): 46-57.
- Deichmann, U., Fay, M., Koo, J., & Lal, S. V. (2002). *Economic structure, productivity, and infrastructure quality in southern Mexico*. Washington, DC: World Bank.
- Demetriades, P., & Mamuneas, T. (2000). Inter-temporal output and employment effects of public infrastructure capital: Evidence from 12 OECD economies. *Economic Journal*, 110(465), 687–712.
- Dercon, Stefan, Daniel O. Gilligan, John Hoddinott, and Tassew Woldehanna (2008) “The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages” IFPRI Discussion Paper 00840
- Donaldson, David and Richard Hornbeck, 2013, “RAILROADS AND AMERICAN ECONOMIC GROWTH: A “MARKET ACCESS” APPROACH”, Working paper

- Donaldson, David. 2012. Railroads of the Raj: Estimating the impact of transportation infrastructure. Processed, mit.
- Dormann, Carsten, et al. "Methods to account for spatial autocorrelation in the analysis of species distributional data: a review." *Ecography* 30.5 (2007): 609-628.
- Dorosh, Paul, et al. "Road connectivity, population, and crop production in Sub-Saharan Africa." *Agricultural Economics* 43.1 (2012): 89-103.
- Emran, Shahe and Zhaoyang Hou (2013) "Access to Markets and Household Consumption: Evidence from Rural China" *Review of Economics and Statistics*, 95.2, pp. 682-697.
- Faber, Benjamin (2012), "Trade Integration, Market Size, and Industrialization: Evidence From China's National Truck Highway System", Working Paper, LSE.
- Fafchamps, M., Moser, C., 2003. Crime, isolation, and law enforcement. *J. African Economies* 12, 625–671.
- Fan, S., P. Hazell, and S. Thorat (2000). "Government Spending, Growth and Poverty in Rural India," *American Journal of Agricultural Economics* 82 (4), pp. 1038-1051.
- Foster, V., & Araujo, M. (2004). Does infrastructure reform work for the poor? A case study from Guatemala. Washington, DC: World Bank.
- Foster, Vivien, and Cecilia Briceño-Garmendia (2008) "Africa Infrastructure Country Diagnostic." *Overhauling the Engine of Growth: Infrastructure in Africa*. September. Washington, DC: World Bank.
- Foster, Vivien, and Nataliya Pushak. (2011) "Nigeria's infrastructure: a continental perspective." Policy Research Working Paper 5686. The World Bank. June, 2011.
- Feltenstein, A. and Ha, J. (1995). 'The role of infrastructure in Mexican economic reform.' *The World Bank Economic Review*, vol. 9, pp. 287-304.
- Gachassin, Marie, (2010) "Roads Impact on Poverty Reduction: A Case Study of Cameroon"
- Garcia-López, Miquel-Àngel, Adelheid Holl, and Elisabet Viladecans-Marsal. 2013. "Suburbanization and highways: When the Romans, the Bourbons and the first cars still shape Spanish cities." Working Paper
- Ghosh, Tilottama, et al. (2010) "Shedding light on the global distribution of economic activity." *The Open Geography Journal* 3.1: 148-161.

- Gibson, John and Scott Rozelle (2003), “Poverty and Access to Roads in Papua New Guinea”, *Economic Development and Cultural Change*, 52(1): 159-185.
- Gonzalez-Navarro, Marco and Quintana-Domeque, Climent, *Roads to Development: Experimental Evidence from Urban Road Pavement* (February 23, 2010). Available at SSRN: <http://ssrn.com/abstract=1558631> or <http://dx.doi.org/10.2139/ssrn.1558631>
- Gunasekara, K., W.P. Anderson, and T. R. Lakshmanan (2008 November) “Highway Induced Development: Evidence from Sri Lanka”, *World Development*.
- Gwatkin, D.R., S. Rutstein, K. Johnson, R.P. Pande, and A. Wagstaff. 2000. Socio-economic differences in health, nutrition, and population. HNP/Poverty Thematic Group. Washington, D.C.: World Bank.
- Gwilliam, Ken (2011) “Africa’s Transport Infrastructure: Mainstreaming Maintenance and Management” World Bank.
- Hawkins, Bradford A., et al. "Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology." *Ecography* 30.3 (2007): 375-384.
- Ihori, T., & Kondo, H. (2001). The efficiency of disaggregate public capital provision in Japan. *Public Finance and Management*, 1(2), 161–182.
- Jacoby, H., 2000. Access to markets and the benefits of rural roads. *Econ. J.* 110, 713–737.
- Jacoby, Hanan G. and Bart Minten (2009) “On Measuring the Benefits of Lower Transportation Costs” *Journal of Development Economics* 89(1): 28-38
- Jedwab, Remi and Alexander Moradi (2012) “Colonial Investments and Long-Term Development in Africa: Evidence from Ghanaian Railroads” Working Paper
- Khandker, Shahidur and Gayatri Koolwal (2011), “Estimating the Long-Term Impacts of Rural Roads”, Policy Research Working Paper 5867, World Bank.
- Khandker, S. R., Z. Bakht, and G. B. Koolwal (2006, April). “The Poverty Impact of Rural Roads : Evidence from Bangladesh”. Policy Research Working Paper Series 3875, The World Bank.
- Kuku-Shittu, Oluyemisi, Astrid Mathiassen, Amit Wadhwa, Lucy Myles, Akeem Ajibola (2013) “Comprehensive Food Security and Vulnerability Analysis” IFPRI Discussion Paper 01275
- Lakshmanan, T.R. and W. Anderson, 2007, “Transport’s Role in Regional Integration Processes” *Market Access, Trade in Transport Services and Trade Facilitation*, Round Table 134. OECDECMT, Paris, pp. 45-71.

Lakshmanan, T.R. and W. Anderson, 2002. A White Paper on “Transportation Infrastructure, Freight Services Sector, and Economic Growth”, prepared for the U.S. Department of Transportation, Federal Highway Administration.

Levine, Ned. "CrimeStat III: a spatial statistics program for the analysis of crime incident locations (version 3.0)." *Houston (TX): Ned Levine & Associates/Washington, DC: National Institute of Justice* (2004).

Lichstein, Jeremy W., et al. "Spatial autocorrelation and autoregressive models in ecology." *Ecological monographs* 72.3 (2002): 445-463.

Lokshin, M., & Yemtsov, R. (2003). Evaluating the impact of infrastructure rehabilitation projects on household welfare in rural Georgia. Washington, DC: World Bank.

Martincus, Christian Volpe, Jerónimo Carballo, and Ana Cusolito. "Routes, exports, and employment in developing countries: Following the trace of the Inca roads." Inter-American Development Bank, mimeograph (2012).

Michaels, G. (2008): The Effect of Trade on the Demand for Skill | Evidence from the Interstate Highway System," *Review of Economics and Statistics*, 90(4).

Minten, B., Kyle, S., 1999. The effect of distance and road quality on food collection, marketing margin, and traders' wages: evidence from the former Zaire. *J. Devel. Econ.* 60, 467–495.

Morrison-Paul, C. J., Ball, E., Felthoven, R. G., & Nehring, R. (2001). Public infrastructure impacts on US agricultural production: A state level panel analysis of costs and netput composition. *Public Finance and Management*, 1(2), 183–213.

Mu, R. and D. van de Walle (2007, August). Rural Roads and Poor Area Development in Vietnam. Policy Research Working Paper Series 4340, The World Bank.

Munnell, A. (1990). How does public infrastructure affect regional economic performance. *New England Economic Review*, (September–October), 11–32.

Nadiri, I., & Mamuneas, T. (1996). Contributions of highway capital to output and productivity growth in the US economy and industries. Washington, DC: Administrative Office of Policy Development, Federal Highway Administration.

National Population Commission (NPC) [Nigeria] and ICF Macro. 2009. Nigeria Demographic and Health Survey 2008. Abuja, Nigeria: National Population Commission and ICF Macro.

Paul J. Gertler, Marco Gonzalez-Navarro, Tadeja Gracner, Alexander D. Rothenberg, “Road network: “The Role of Road Quality Investments on Economic Activity and Welfare: Evidence from Indonesia’s Highways”, working paper.

- Rutstein, Shea O. and Kiersten Johnson. 2004. The DHS Wealth Index. DHS Comparative Reports No. 6. Calverton, Maryland: ORC Macro.
- Shreshtha, Slesh A. (2012). Access to the North-South Roads and Farm Profits in Rural Nepal. Working Paper, National University of Singapore
- Shirley, W., & Winston, C. (2004). Firm inventory behavior and the returns from highway infrastructure investments. *Journal of Urban Economics*, 55, 398–415.
- Stifel, David, and Bart Minten. 2008. “Isolation and Agricultural Productivity.” *Agricultural Economics*, 39 (1): 1–15.
- Stifel, David, Bart Minten, and Bethlehem Koro (2012) “Economic Benefits and Returns to Rural Feeder Roads: Evidence from a Quasi-Experimental Setting in Ethiopia” IFRPI Policy Research Institute Working Paper
- Stifel, David, and Bart Minten. 2008. “Isolation and Agricultural Productivity.” *Agricultural Economics*, 39 (1): 1–15.
- Sturm, J. (2001). The impact of public infrastructure capital on the private sector of the Netherlands: An application of the symmetric generalized McFadden Cost Function. PFM. Vol. 1 No. 2
- Tobler, Waldo (1993) “Three presentations on geographical analysis and modeling non-isotropic geographic modeling speculations on the geometry of geography global spatial analysis”. NATIONAL CENTER FOR GEOGRAPHIC INFORMATION AND ANALYSIS. TECHNICAL REPORT 93-1. February 1993
- Uchida, H. and Nelson, A. (2009) “Agglomeration Index: Towards a New Measure of Urban Concentration.” Background paper for the World Bank’s World Development Report
- Warr, Peter, (2008), “How Road Improvement Reduces Poverty: The Case of Laos,” *Agricultural Economics*, 39, pp. 269-279.
- World Bank. 2007: Evaluation of World Bank Support to Transportation Infrastructure. Washington DC: World Bank Publications. World Bank. 2014. Sustainable transport for all : helping people to help themselves. IDA at work. Washington, DC : World Bank Group.

Table 2: Crop Revenue

Dependent Variable: ln(Crop Revenue)	(1) SUR	(2) 3SLS Euclidean Distance IV	(3) 3SLS Natural Path IV
ln(Cost to Market)	-0.625*** (-3.69)	-0.630*** (-3.34)	-0.643*** (-3.42)
Household agricultural labor	0.114 (1.29)	0.114 (1.29)	0.115 (1.30)
Household agricultural labor²	-0.014 (-1.47)	-0.014 (-1.47)	-0.014 (-1.48)
Land area	0.020*** (4.33)	0.020*** (4.33)	0.020*** (4.34)
Fertilizer	0.000 (-0.48)	0.000 (-0.48)	0.000 (-0.47)
Dummy=1 if irrigates land	0.611* (1.83)	0.611* (1.83)	0.611* (1.83)
Age of household head	0.020 (0.69)	0.020 (0.69)	0.020 (0.69)
Age of household head²	0.000 (-0.68)	0.000 (-0.68)	0.000 (-0.69)
Dummy=1 if household head is literate	0.386*** (2.53)	0.386*** (2.53)	0.386*** (2.53)
Constant	1.882* (1.65)	1.889* (1.65)	1.905* (1.66)
Marketshed fixed effects	Yes	Yes	Yes
Observations	910	910	910
First Stage Results			
IV: ln(Euclidean Distance)		0.589*** (34.44)	
IV: ln(Natural Path)			0.661*** (33.34)
Angrist-Pischke Test of Weak Identification		1186.22 P=0.0000	1111.64 P=0.0000

t-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1
Data: Nigeria LSMS-ISA 2010

Table 3: Livestock Sales

Dependent Variable: ln(Livestock Sales)	(1) SUR	(2) 3SLS Euclidean Distance IV	(3) 3SLS Natural Path IV
ln(Cost to Market)	-0.247* (-1.77)	-0.319** (-2.05)	-0.348** (-2.24)
ln(Cost of animals)	0.241*** (7.19)	0.241*** (7.19)	0.241*** (7.19)
Household agricultural labor	0.141* (1.95)	0.145** (2.00)	0.147** (2.02)
Household agricultural labor²	-0.012 (-1.56)	-0.012 (-1.60)	-0.013 (-1.61)
Land area	0.005 (1.20)	0.005 (1.27)	0.005 (1.29)
Fertilizer	-0.001 (-1.05)	-0.001 (-1.03)	-0.001 (-1.02)
Dummy=1 if irrigates land	0.134 (0.49)	0.136 (0.49)	0.136 (0.49)
Age of household head	0.036 (1.48)	0.036 (1.48)	0.036 (1.48)
Age of household head²	0.000 (-1.63)	0.000 (-1.64)	0.000 (-1.64)
Dummy=1 if household head is literate	-0.269 (-2.13)	-0.269 (-2.13)	-0.269 (-2.13)
Constant	-0.199 (-0.21)	-0.105 (-0.11)	-0.068 (-0.07)
Marketshed fixed effects	Yes	Yes	Yes
Observations	910	910	910
First Stage Results			
IV: ln(Euclidean Distance)		0.600*** (26.19)	
IV: ln(Natural Path)			0.664*** (25.42)
Angrist-Pischke Test of Weak Identification		685.88 P=0.0000	646.05 P=0.0000

t-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1
Data: Nigeria LSMS-ISA 2010

Table 4: Non-Agricultural Income

Dependent Variable: ln(Non-Agricultural Income)	(1) SUR	(2) 3SLS Euclidean Distance IV	(3) 3SLS Natural Path IV
ln(Cost to Market)	-0.317** (-2.12)	-0.312* (-1.87)	-0.326** (-1.96)
Household agricultural labor	0.263*** (3.25)	0.263 (3.25)***	0.263*** (3.25)
Household agricultural labor^2	-0.015** (-2.33)	-0.015 (-2.33)	-0.015** (-2.33)
Land area	0.001 (0.24)	0.001 (0.23)	0.001 (0.24)
Age of household head	-0.046* (-1.74)	-0.046* (-1.74)	-0.046* (-1.74)
Age of household head^2	0.000* (1.81)	0.000* (1.81)	0.000* (1.81)
Dummy=1 if household head is literate	0.449*** (3.30)	0.449*** (3.30)	0.449*** (3.30)
Total business expenses	0.000*** (5.48)	0.000*** (5.48)	0.000*** (5.48)
Constant	5.022*** (4.94)	5.016*** (4.91)	5.034*** (4.93)
Marketshed fixed effects	Yes	Yes	Yes
Observations	910	910	910
First Stage Results			
IV: ln(Euclidean Distance)		0.594*** (29.13)	
IV: ln(Natural Path)			0.672*** (27.93)
Angrist-Pischke Test of Weak Identification		848.71 P = 0.0000	779.96 P = 0.0000

t-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1
Data: Nigeria LSMS-ISA 2010

Table 5: Wealth Index

Dependent Variable: ln(Wealth Index)	(1) OLS	(2) IV Euclidean Distance	(3) IV Natural Path
ln(Cost to Market)	-0.235*** (-10.28)	-0.210*** (-8.654)	-0.204*** (-7.758)
Agricultural Potential	-0.0152 (-0.878)	-0.0159 (-0.922)	-0.0161 (-0.930)
Dummy=1 if household agriculturally involved	-0.235*** (-8.451)	-0.244*** (-8.746)	-0.246*** (-8.828)
Ln(age of household head)	-0.105*** (-3.233)	-0.103*** (-3.187)	-0.103*** (-3.176)
Female household head dummy	-0.0189 (-0.498)	-0.0207 (-0.544)	-0.0211 (-0.554)
ln(No. of household members)	0.0574* (1.831)	0.0563* (1.805)	0.0560* (1.796)
ln(No. of females aged 15 to 49 yrs)	0.0903*** (3.799)	0.0910*** (3.847)	0.0912*** (3.851)
ln(No. of males aged 15 to 59 yrs)	0.0521** (2.279)	0.0543** (2.385)	0.0548** (2.411)
ln(No. of children aged 0 to 5 yrs)	-0.0266 (-1.449)	-0.0258 (-1.412)	-0.0256 (-1.400)
Dummy=1 if rural	-0.441*** (-10.46)	-0.457*** (-10.92)	-0.461*** (-10.88)
Constant	12.97*** (109.8)	12.97*** (109.9)	12.97*** (109.8)
Observations	6,684	6,684	6,684
First Stage IV: ln(Euclidean Distance)		0.791*** (65.870)	
IV: ln(Natural Path)			0.852*** (33.030)
Angrist-Pischke Test of Weak Identification		4338.48 0.0000	1091.24 0.0000

Robust t-statistics clustered at the enumeration area in parentheses, *** p<0.01, ** p<0.05, * p<0.1
Data: Nigeria DHS 2008

Table 6: Multi-dimensional Poverty, NDHS

Dependent Variable: dummy=1 if poor	(1) OLS	(2) IV Euclidean Distance	(3) IV Natural Path	(4) Probit	(5) IV Probit Euclidean Distance	(6) IV Probit Natural Path
ln(Cost to Market)	0.086*** (6.940)	0.0769*** (5.843)	0.0669*** (4.823)	0.078*** (11.840)	0.304*** (9.440)	0.262*** (7.320)
Agricultural Potential	0.00934 (0.984)	0.00962 (1.008)	0.00991 (1.031)	0.007 (1.230)	0.036 (1.340)	0.036 (1.370)
Dummy=1 if household agriculturally involved	0.124*** (6.983)	0.128*** (7.158)	0.132*** (7.311)	0.0932*** (8.020)	0.421*** (8.080)	0.437*** (8.350)
Ln(Age of household head)	-0.0182 (-0.774)	-0.0190 (-0.804)	-0.0197 (-0.835)	-0.017 (0.770)	-0.077 (-0.78)	-0.079 (-0.8)
Female household head dummy	0.0627* (1.843)	0.0634* (1.854)	0.0641* (1.867)	0.036 (1.500)	0.16 (1.460)	0.167 (1.510)
ln(No. of household members)	0.0321 (1.443)	0.0325 (1.461)	0.0329 (1.479)	0.033 (1.570)	0.143 (1.560)	0.144 (1.570)
ln(No. of females aged 15 to 49 years)	0.114*** (6.656)	0.113*** (6.636)	0.113*** (6.613)	0.112*** (6.800)	0.492*** (6.740)	0.489*** (6.700)
Ln(No. of males aged 15 to 59 years)	0.0229 (1.253)	0.0221 (1.204)	0.0212 (1.154)	0.014 (0.820)	0.057 (0.770)	0.056 (0.760)
ln(No. of children aged 0 to 5 years)	0.00415 (0.330)	0.00384 (0.304)	0.00351 (0.277)	0.013 (0.990)	0.056 (0.960)	0.053 (0.910)
Dummy=1 if Rural	0.153*** (6.649)	0.160*** (6.902)	0.166*** (7.083)	0.139*** (9.290)	0.57*** (10.510)	0.597*** (10.760)
Constant	0.0668 (0.726)	0.0678 (0.735)	0.0689 (0.744)			
Observations	6,684	6,684	6,684	6,684	6,684	6,684
First Stage Results						
IV: ln(Euclidean distance)		0.791*** (65.870)			0.791*** (190.430)	
IV: ln(Natural Path)			0.852*** (33.030)			0.851*** (68.940)
Angrist-Pischke Test		4338.48	1091.24			
		0.0000	0.0000			

Robust t-statistics clustered at the enumeration area in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Data: Nigeria DHS 2008

Table 7: Local GDP

Dependent Variable: Local GDP	(1)	(2)	(3)	(4)	(5)	(6)
	OLS Full Sample	IV Euc Dist Full Sample	IV Nat Path Full Sample	OLS Rural Only	IV Euc Dist Rural Only	IV Nat Path Rural Only
ln(Cost to Market)	-0.543*** (-33.18)	-0.502*** (-26.57)	-0.496*** (-26.28)	-0.502*** (-29.03)	-0.450*** (-22.41)	-0.447*** (-22.30)
ln(Distance to Mine)	-0.0159 (-0.66)	-0.0246 (-1.03)	-0.0282 (-1.17)	-0.0295 (-1.20)	-0.0395 (-1.60)	-0.0424* (-1.71)
ln(Population)	0.128*** (3.08)	0.120*** (2.88)	0.101** (2.41)	0.00122 (0.02)	-0.00988 (-0.19)	-0.00650 (-0.13)
ln(Population)^2	0.0439*** (17.79)	0.0451*** (18.21)	0.0462*** (18.50)	0.0527*** (16.24)	0.0543*** (16.70)	0.0538*** (16.46)
ln(Cassava potential yield)	0.0191** (2.49)	0.0176** (2.30)	0.0160** (2.10)	0.0191** (2.46)	0.0172** (2.22)	0.0155** (2.01)
ln(Cassava potential yield)^2	0.00392** (2.09)	0.00371** (1.99)	0.00334* (1.80)	0.00352* (1.85)	0.00326* (1.72)	0.00285 (1.51)
ln(Yams potential yield)	-0.0190** (-2.07)	-0.0176* (-1.93)	-0.0175* (-1.92)	-0.0190** (-2.08)	-0.0172* (-1.88)	-0.0168* (-1.84)
ln(Yams potential yield)^2	-0.00360 (-1.60)	-0.00342 (-1.53)	-0.00339 (-1.52)	-0.00338 (-1.50)	-0.00313 (-1.39)	-0.00305 (-1.36)
ln(Maize potential yield)	0.064*** (4.49)	0.0649*** (4.58)	0.0603*** (4.02)	0.0749*** (5.30)	0.076*** (5.41)	0.069*** (4.69)
ln(Maize potential yield)^2	-0.0072*** (-3.45)	-0.0069*** (-3.34)	-0.0063*** (-2.94)	-0.0077*** (-3.73)	-0.0074*** (-3.59)	-0.0067*** (-3.12)
ln(Rice potential yield)	-0.00606 (-1.06)	-0.00620 (-1.09)	-0.00638 (-1.12)	-0.00661 (-1.14)	-0.00692 (-1.19)	-0.00700 (-1.20)
ln(Rice potential yield)^2	0.00022 (0.17)	0.00022 (0.16)	0.00009 (0.07)	0.00023 (0.17)	0.00019 (0.14)	0.00011 (0.08)
Constant	-0.811*** (-3.35)	-0.871*** (-3.60)	-0.748*** (-3.09)	0.397 (1.34)	0.334 (1.13)	0.397 (1.35)
Marketshed FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,728	10,728	10,607	9,899	9,899	9,797
First Stage Results						
ln(Natural Path)			0.7352*** (175.65)			0.7350*** (165.05)
ln(Euclidean Distance)		0.7330*** (176.29)			0.733*** (165.60)	
Angrist-Pischke Test		31076.43 P=0.0000	30852.25 P=0.0000		27424.55 P=0.0000	27242.72 P=0.0000

t-statistics in parentheses*** p<0.01, ** p<0.05, * p<0.1

Data: Various sources, see section 5.3

Table 8: All Year Employment

	Male			Female		
	(1) OLS	(2) IV Euclidean Distance	(3) IV Natural Path	(4) OLS	(5) IV Euclidean Distance	(6) IV Natural Path
Dependent Variable: Dummy=1 if employed all year round and 0 if not employed or seasonally employed						
ln(Cost to Market)	-0.0430*** (-2.971)	-0.0374** (-2.344)	-0.0407** (-2.281)	-0.0442*** (-5.427)	-0.0324*** (-3.517)	-0.0300*** (-2.979)
Agri. Potential	-0.0209* (-1.851)	-0.0209* (-1.861)	-0.0209* (-1.858)	-0.0101 (-1.634)	-0.00858 (-1.372)	-0.00861 (-1.376)
ln(Age)	0.490*** (17.78)	0.490*** (17.82)	0.490*** (17.81)	0.491*** (32.14)	0.491*** (32.24)	0.491*** (32.22)
Education level: Primary	0.156*** (4.699)	0.157*** (4.736)	0.157*** (4.720)	0.0718*** (5.461)	0.0714*** (5.589)	0.0719*** (5.636)
Education level: Secondary	0.0798** (2.339)	0.0815** (2.385)	0.0805** (2.353)	0.00451 (0.291)	0.00572 (0.386)	0.00659 (0.444)
Education level: Higher than secondary	0.0373 (0.921)	0.0394 (0.974)	0.0381 (0.939)	-0.00499 (-0.237)	-0.00245 (-0.117)	-0.00128 (-0.0616)
Dummy=1 if Rural	-0.0484* (-1.898)	-0.0521** (-2.018)	-0.0500* (-1.889)	-0.0151 (-1.078)	-0.0149 (-1.051)	-0.0165 (-1.132)
Marketshed FE	Yes	Yes	Yes	Yes	Yes	Yes
Religion FE	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,918	33,918	33,918	36,703	36,703	36,703
First Stage						
IV: ln(Euclidean Distance)		0.791*** (70.310)			0.79*** (70.360)	
IV: ln(Natural Path)			0.786*** (19.030)			0.793*** (19.620)
Angrist-Pischke Test of Weak Identification		4943.25 0.0000	362.16 0.000		4951.01 0.000	384.89 0.000

Robust t-statistics clustered at the enumeration area in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Data: Nigeria DHS 2008

Table 9: Income diversification

Dependent Variable: Agricultural employment among employed individuals	Male			Female		
	(1) Probit	(2) IV Probit Euclidean Distance	(3) IV Probit Natural Path	(4) Probit	(5) IV Probit Euclidean Distance	(6) IV Probit Natural Path
ln(Cost to Market)	0.095*** (7.140)	0.379*** (5.920)	0.403*** (5.910)	0.087*** (7.710)	0.425*** (7.040)	0.528*** (7.610)
Ln(Agri. Potential)	0.037*** (3.010)	0.163*** (3.080)	0.161*** (3.030)	0.012 (1.550)	0.065 (1.580)	0.062 (1.510)
ln(Age)	0.024 (1.000)	0.1 (0.980)	0.102 (1.000)	0.032*** (2.220)	0.167*** (2.240)	0.169*** (2.270)
Education level: Primary	-0.149*** (-5.19)	-0.571*** (-5.38)	-0.567*** (-5.33)	-0.09*** (-5.56)	-.402*** (-5.6)	-.390*** (-5.42)
Education level: Secondary	-0.262*** (-8.73)	-1.001*** (-9.1)	-0.996*** (-9.06)	-0.192*** (-11.55)	-0.915*** (-12.24)	-0.894*** (-11.83)
Education level: Higher than secondary	-0.401*** (-11.48)	-1.614*** (-10.27)	-1.61*** (-10.25)	-0.323*** (-18.57)	-1.95*** (-16.37)	-1.91*** (-15.79)
Dummy=1 if Rural	0.205*** (7.780)	0.852*** (7.950)	0.844*** (7.790)	0.135*** (7.280)	0.73*** (7.210)	0.679*** (6.400)
Market shed FE	Yes	Yes	Yes	Yes	Yes	Yes
Religion FE	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28,629	28,629	28,629	21,656	21,656	21,656
First Stage IV: ln(Euclidean Distance)		0.787*** (71.300)			0.786*** (63.030)	
IV: ln(Natural Path)			0.828*** (29.910)			0.781*** (17.080)

Robust t-statistics clustered at the enumeration area in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Data: Nigeria DHS 2008

Table 10: Conley Bounds

	Support for possible values of δ	95% Confidence Interval	
	IV: ln(Natural path)	Lower Bound	Upper Bound
ln(Crop revenue)	δ : [-0.0001, 0.0001]	-0.72	-0.123
	δ : [-0.001, 0.001]	-0.721	-0.122
	δ : [-0.01, 0.01]	-0.734	-0.108
ln(Livestock sales)	δ : [-0.0001, 0.0001]	-0.624	0.055
	δ : [-0.001, 0.001]	-0.625	0.057
	δ : [-0.01, 0.01]	-0.638	0.07
ln(Non-agri income)	δ : [-0.0001, 0.0001]	-0.65	-0.077
	δ : [-0.001, 0.001]	-0.651	-0.076
	δ : [-0.01, 0.01]	-0.662	-0.065
ln(Wealth Index)	δ : [-0.0001, 0.0001]	-0.238	-0.137
	δ : [-0.001, 0.001]	-0.239	-0.136
	δ : [-0.01, 0.01]	-0.251	-0.124
MPI	δ : [-0.0001, 0.0001]	0.039	0.089
	δ : [-0.001, 0.001]	0.038	0.09
	δ : [-0.01, 0.01]	0.026	0.101
ln(Local GDP) <i>full sample</i>	δ : [-0.0001, 0.0001]	-0.533	-0.459
	δ : [-0.001, 0.001]	-0.535	-0.458
	δ : [-0.01, 0.01]	-0.547	-0.445
ln(Local GDP) <i>rural only</i>	δ : [-0.0001, 0.0001]	-0.486	-0.407
	δ : [-0.001, 0.001]	-0.488	-0.406
	δ : [-0.01, 0.01]	-0.5	-0.394
All year employment (male)	δ : [-0.0001, 0.0001]	-0.081	-0.015
	δ : [-0.001, 0.001]	-0.082	-0.014
	δ : [-0.01, 0.01]	-0.093	-0.003
All year employment (female)	δ : [-0.0001, 0.0001]	-0.093	-0.05202454
	δ : [-0.001, 0.001]	-0.094	-0.05090694
	δ : [-0.01, 0.01]	-0.105	-0.04
Agricultural employment (male)	δ : [-0.0001, 0.0001]	0.0481	0.112
	δ : [-0.001, 0.001]	0.047	0.114
	δ : [-0.01, 0.01]	0.036	0.125
Agricultural employment (female)	δ : [-0.0001, 0.0001]	0.0404	0.0893
	δ : [-0.001, 0.001]	0.039	0.09
	δ : [-0.01, 0.01]	0.028	0.102

Table 11: Local GDP by Region

	(1) OLS Full Sample	(2) 2SLS- Nat Path Full Sample	(3) Market Cost Elasticities (Calculated from column 2)
ln(Market Cost)	-1.051*** (-12.68)	-1.095*** (-11.89)	-1.095*** (-11.89)
ln(Market Cost) * South West Region	0.475*** (5.13)	0.561*** (5.40)	-0.534*** (-3.84)
ln(Market Cost) * South South Region	0.417*** (4.59)	0.385*** (3.78)	-0.71*** (-5.17)
ln(Market Cost) * North East Region	0.731*** (8.40)	0.822*** (8.51)	-0.273** (-2.05)
ln(Market Cost) * North West Region	0.429*** (4.88)	0.533*** (5.44)	-0.562*** (-4.18)
ln(Market Cost) * North Central Region	0.436*** (5.06)	0.546*** (5.71)	-0.549*** (-4.14)
Other Variables in regression	ln(population), ln(population)^2, ln(Cassava potential yield), ln(Cassava potential yield)^2, ln(Yams potential yield), ln(Yams potential yield)^2, ln(Maize potential yield), ln(Maize potential yield)^2, ln(Rice potential yield), ln(Rice potential yield)^2, marketshed FE		
Observations	10728	10607	

t-statistics in parentheses*** p<0.01, ** p<0.05, * p<0.1

Data: Various sources, see section 5.3

Table 12: NEPAD Project Benefits

	(1) GDP Increase (Million USD, 2006 PPP)	(2) Road Length (Kms)	(3) GDP Increase Per KM (Million USD)	(4) Population Affected (Millions)	(5) Per Capita Benefit (USD per person affected)
All Projects	\$1,794	2,774	\$0.65	39.32	\$45.63
North-South Corridor	\$1,082	1,121	\$0.97	23.12	\$46.81
Northeastern Corridor	\$233	939	\$0.25	14.88	\$15.66
Southern Corridor	\$529	729	\$0.73	9.16	\$57.73

Table 13: Local GDP Elasticities 95% confidence interval

Region	(1) Point Estimate	(2) Standard error	(3) Upper Bound (95% confidence Interval)	(4) Lower Bound (95% confidence Interval)
South East	-1.095	0.0921	-1.2755	-0.9144
South West	-0.534	0.1039	-0.7377	-0.3304
South South	-0.71	0.1019	-0.9096	-0.5104
North East	-0.273	0.0966	-0.4623	-0.0837
North West	-0.562	0.0980	-0.7540	-0.3700
North Central	-0.549	0.0956	-0.7364	-0.3616

Table 14: NEPAD Project Benefits, 95% confidence interval

	(1) GDP Increase Lower Bound	(2) GDP Increase Point Estimate	(3) GDP Increase Upper Bound
All Projects	\$1,206	\$1,794	\$2,383
North-South Corridor	\$678	\$1,082	\$1,487
Northeastern Corridor	\$132	\$233	\$334
Southern Corridor	\$429	\$529	\$629

Table 15: Road Improvement Prioritization

Marketshed	(1) Length of Road Improved, Kms	(2) Total GDP, Millions USD PPP	(3) GDP Increase, Millions USD PPP	(4) Percentage increase in GDP (3/2)	(5) Total Population	(6) Population Effectuated	(7) GDP Increase per Km improved, Million USD/KM (3/1)	(8) Population Effectuated per KM of road (6/1)
Abakaliki	201.5	3,087	114.52	3.71%	3,592,510	3,587,821	0.568	17,806
Agbor	75.2	1,973	0.20	0.01%	773,825	33,465	0.003	445
Bauchi	52.1	2,808	0.21	0.01%	2,503,664	135,459	0.004	2,600
Benin	105	6,937	5.04	0.07%	1,606,343	158,835	0.048	1,513
Bida	28.4	1,607	0.91	0.06%	1,180,780	84,630	0.032	2,980
Enugu	89.1	7,556	133.43	1.77%	3,682,892	2,705,968	1.498	30,370
Ibadan	65.8	7,832	72.04	0.92%	2,587,179	1,219,364	1.095	18,531
Ijebu Ode	75.3	2,844	32.18	1.13%	788,070	576,516	0.427	7,656
Ilorin	190	4,682	59.80	1.28%	2,159,909	659,174	0.315	3,469
Kaduna	1737	9,512	55.59	0.58%	2,056,143	632,083	0.032	364
Kano	414	17,652	98.66	0.56%	13,158,975	7,488,524	0.238	18,088
Katsina	1265	5,410	80.88	1.49%	4,351,990	1,485,371	0.064	1,174
Lagos	34.6	41,860	681.38	1.63%	10,413,453	8,618,237	19.693	249,082
Maiduguri	252.9	4,787	37.12	0.78%	3,836,727	3,075,144	0.147	12,160
Minna	140	4,617	5.97	0.13%	1,780,793	187,871	0.043	1,342
Ogbomosho	56	887	21.43	2.42%	663,362	326,754	0.383	5,835
Okitipupa	73.6	1,201	1.37	0.11%	955,233	127,347	0.019	1,730
Onitsha	73.4	7,446	238.83	3.21%	2,778,550	1,935,824	3.254	26,374
Oyo	54.7	844	5.07	0.60%	528,668	203,382	0.093	3,718
Potiskum	267.6	3,807	24.49	0.64%	4,950,803	1,544,431	0.092	5,771
Shagamu	84.2	4,042	100.51	2.49%	653,758	653,758	1.194	7,764
Zaria	114.4	5,089	24.49	0.48%	4,496,220	3,056,001	0.214	26,713

Figure 1: Map of NEPAD road projects

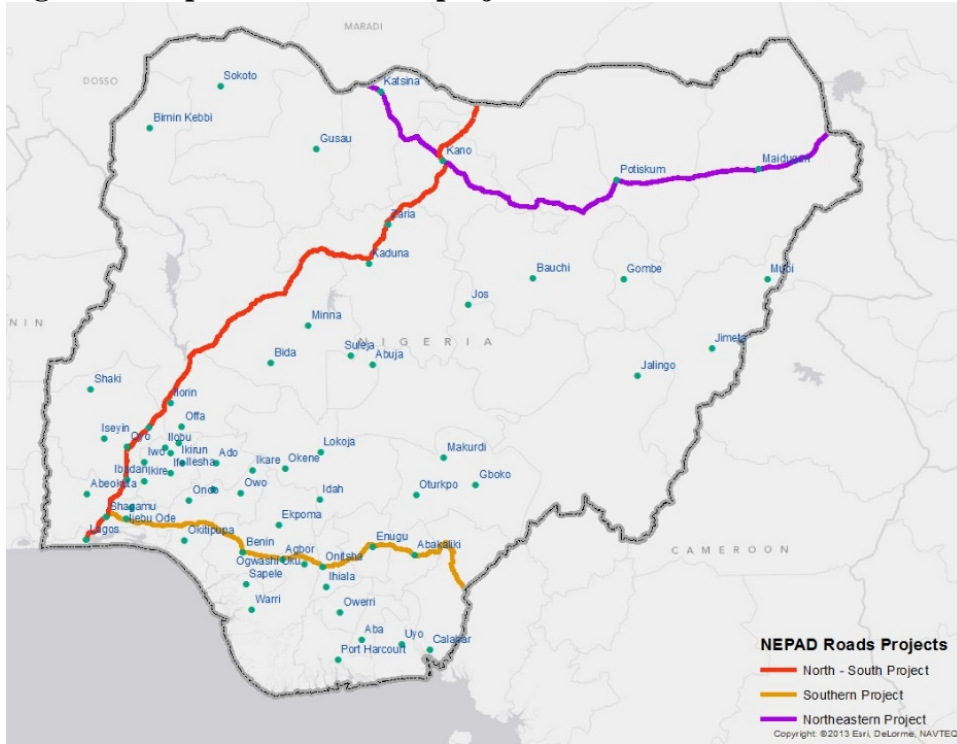


Figure 2: Percentage change in transportation costs

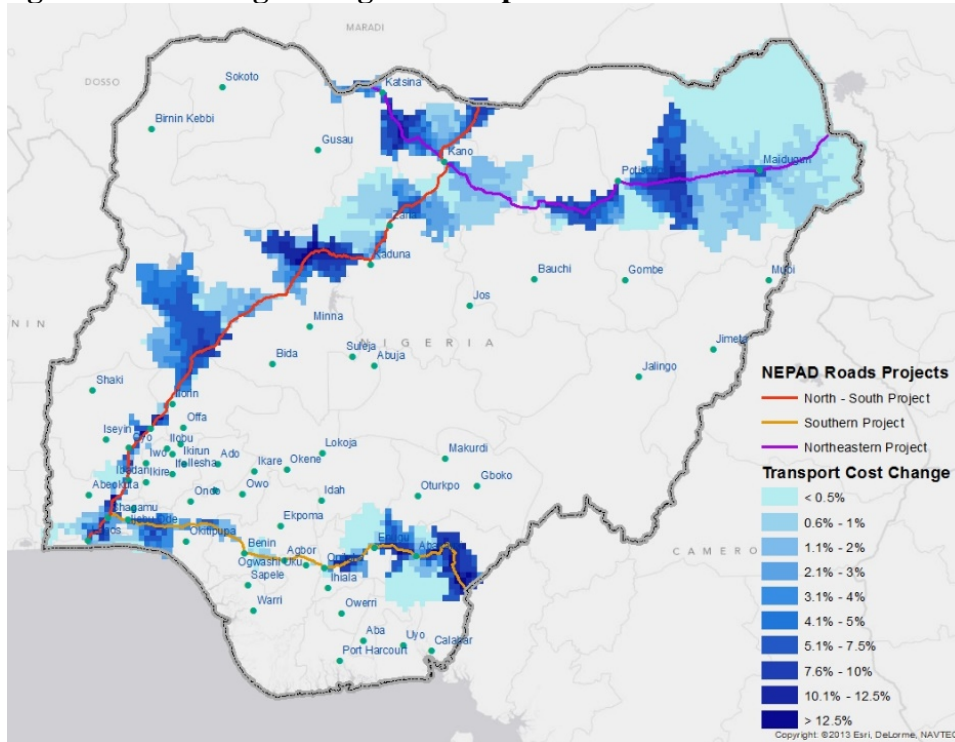


Figure 3: Increase in local GDP

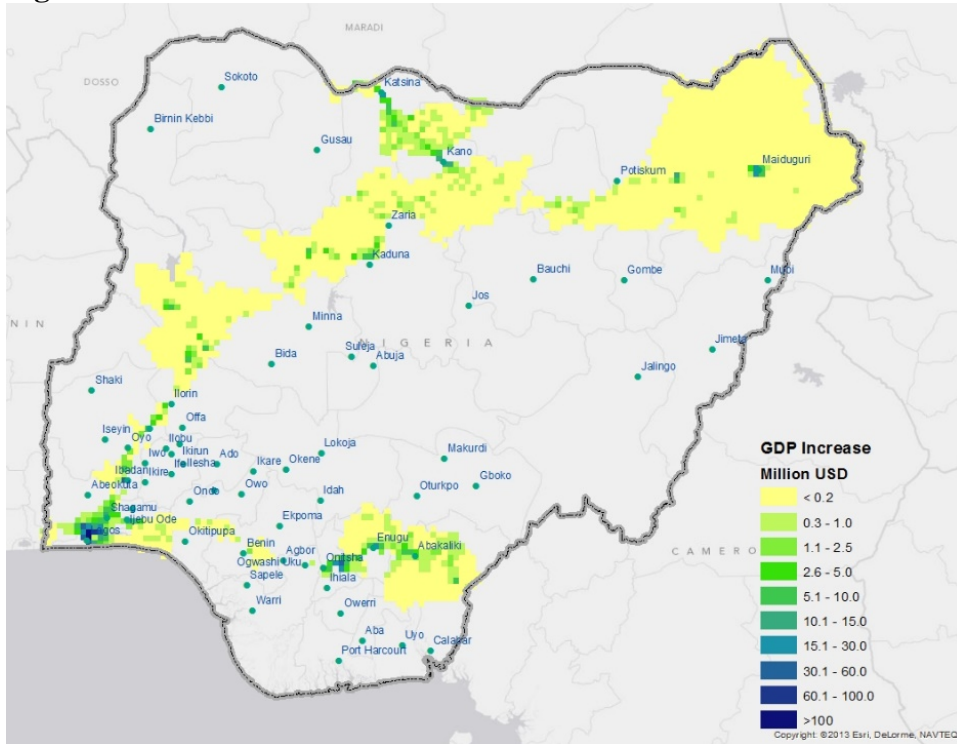


Figure 4: Population affected by Corridor improvement projects

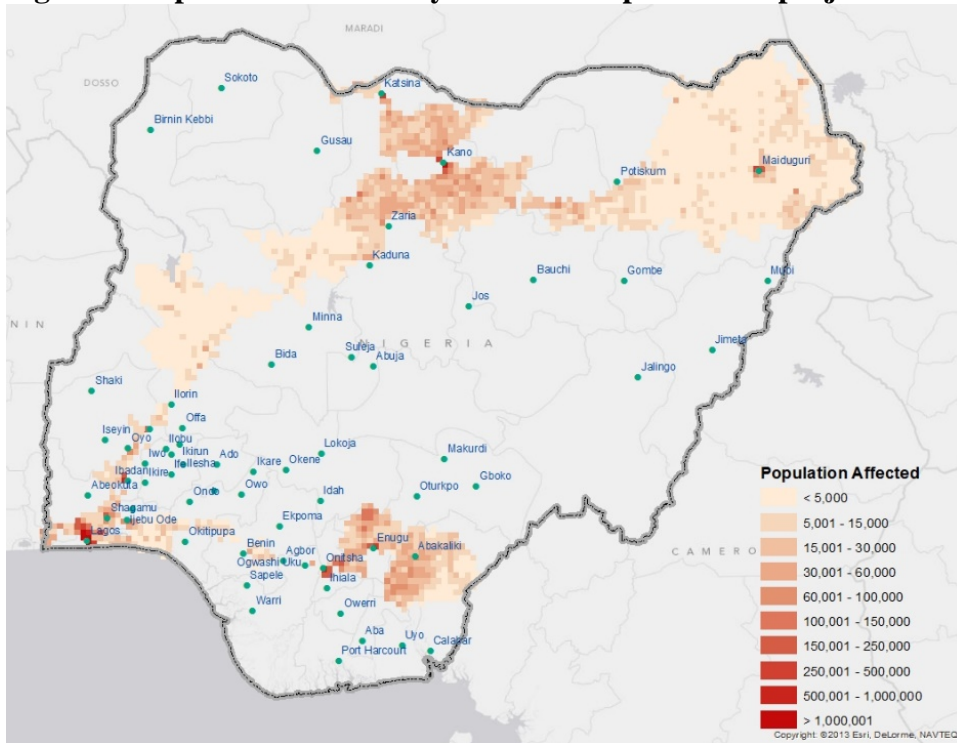


Figure 5: Population Benefited per KM Improved

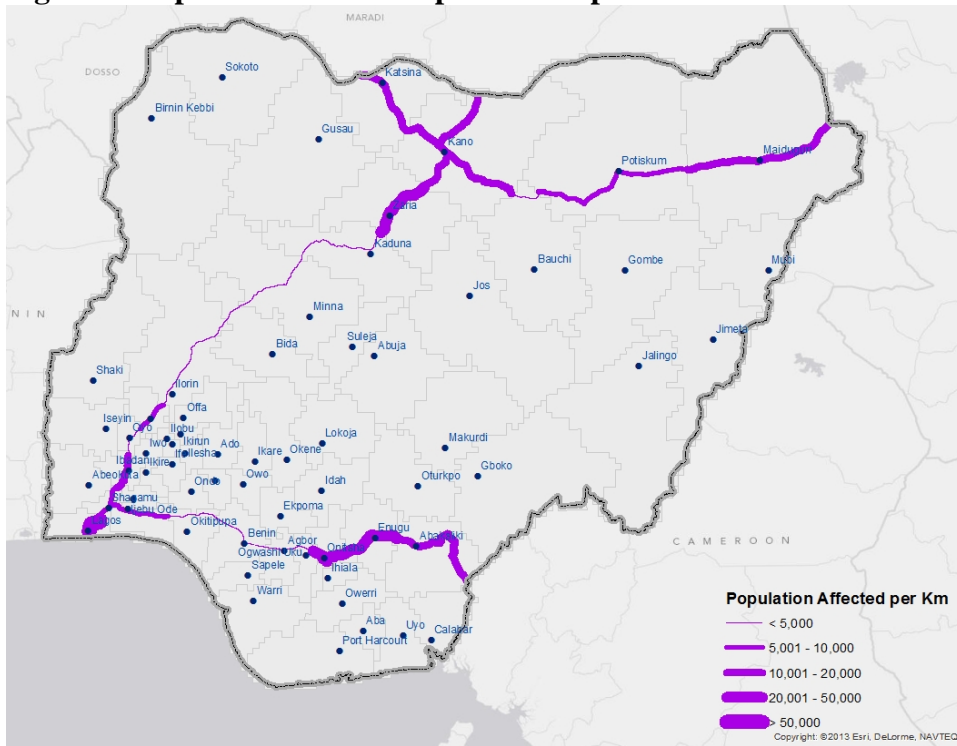
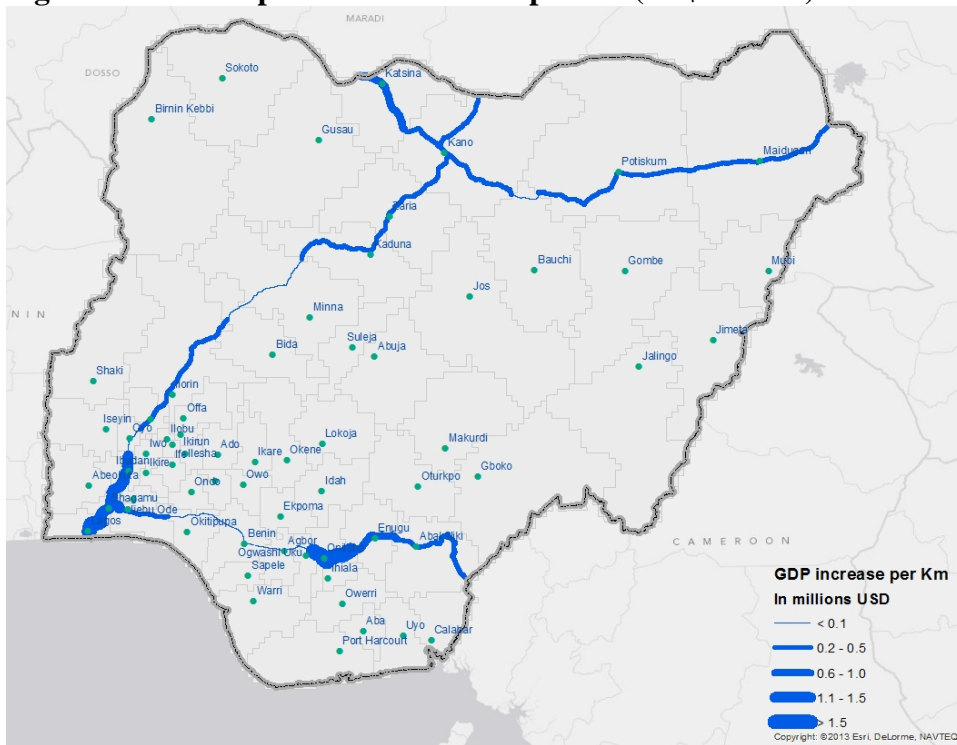


Figure 6: Benefits per KM of Road Improved (US\$ million)



Appendix I: Summary Statistics

Table A1: Summary Statistics, LSMS Nigeria

Variable	Obs	Mean	Std. Dev.	Min	Max	Label
Outcome						
crop	2,600	157.84	2,126.75	0.00	105,600.00	Crop revenue (USD)
income	2,600	1,281.07	35,930.45	0.00	1,320,013	Non-agriculture income (USD)
MPI	2,600	0.37	0.15	0.00	0.83	Multi-dimensional Poverty Index: Weighted sum of indicators
animalsale_usd	3,297	48.77	204.33	0.00	3,960.00	Total sales of livestock (USD)
Treatment						
total_cost	2,600	5.10	3.30	0.14	17.36	Cost to market (USD)
IV						
natpath_hrs	2,600	14.06	10.43	0.00	59.42	Natural path to market (hrs)
Controls						
age	2,600	51.40	15.10	15.00	110.00	Age of household head
dliterate	2,600	0.55	0.50	0.00	1.00	Dummy: hh head is literate
land	2,600	9.23	16.92	0.00	265.03	Land (km squared)
labor	2,600	2.92	2.14	0.00	18.00	Number of workers in the house
agrihome	2,600	2.02	2.03	0.00	17.00	Household members working on own plot
total_fertilizer	2,600	11.14	79.91	0.00	2,299.00	Total fertilizer used (kg)
dirrigate	2,600	0.95	0.21	0.00	1.00	Dummy: hh irrigates its plot
dwarmsemi-arid	2,600	0.31	0.46	0.00	1.00	Dummy: Tropical warm semi-arid
dwarmsubhumid	2,600	0.59	0.49	0.00	1.00	Dummy: tropical warm sub-humid
dwarmhumid	2,600	0.09	0.28	0.00	1.00	Dummy: tropical warm humid
dcoolsubhumid	2,600	0.01	0.11	0.00	1.00	Dummy: tropical cool sub-humid
_Izone_2	2,600	0.17	0.38	0.00	1.00	Dummy: north east
_Izone_3	2,600	0.19	0.40	0.00	1.00	Dummy: north west
_Izone_4	2,600	0.23	0.42	0.00	1.00	Dummy: south east
_Izone_5	2,600	0.14	0.35	0.00	1.00	Dummy: south south
_Izone_6	2,600	0.09	0.29	0.00	1.00	Dummy: south west
totalbiz_costs	2,600	613.69	25,888.87	0	1,320,000	Total business expenses (USD)
animalcosts_usd	3,297	58.90	313.26	0.00	10,312.50	Costs of livestock (USD)

Table A2. Summary Statistics, NDHS

	Mean	Std. dev	Min	Max
Outcomes:				
Wealth index	-12407	98657.2	-145026	305274
Multi-dimensionally poor (dummy)	0.70257	0.457	0	1
Variable of interest:				
Cost to market (USD)	5.652	4.097	0.290	30.301
Instruments:				
Time taken to reach market using natural path (hrs)	15.185	12.417	0	70.794
Controls:				
Agricultural potential (factor of ln agri. potential for cassava, maize and rice)	-0.043	0.966	-1.554	1.211
Dummy: Household agriculturally involved	0.732	0.443	0	1
Age of household head	40.031	11.268	17	99
Sex of household head	1.034	0.182	1	2
No. of household members	6.517	3.079	3	43
No. of female members in households aged 15-49 yrs	1.418	0.775	1	15
No. of female members in households aged 15-59 yrs	1.279	0.678	1	12
No. of children aged 0 to 5 yrs	1.706	0.860	1	9
Dummy=1 if type of residence is rural	0.713	0.452	0	1
Dummy=1 if north east	0.216	0.412	0	1
Dummy=1 if northwest	0.271	0.445	0	1
Dummy=1 if south east	0.080	0.271	0	1
Dummy=1 if south south	0.114	0.318	0	1
Dummy=1 if south west	0.130	0.336	0	1
No. of observations: 6684				

Table A3: Summary Statistics, Local GDP data sets

	Mean	Std. Dev.	Min	Max	Label
Local GDP	25.36888	145.3173	0	4469.6	Total income per cell, millions USD (2006 PPP)
Population	13.64	44.45	0	1,639.2	Population per cell (thousands), Landsat 2006
Cassava Potential Yield	833.5371	681.5004	0	2775	Yield Kg/ha, GAEZ FAO
Rice Potential Yield	494.7672	436.5076	0	1792	Yield Kg/ha, GAEZ FAO
Yams Potential Yield	609.166	383.2807	0	1747	Yield Kg/ha, GAEZ FAO
Maize Potential Yield	1209.439	625.5546	0	3556	Yield Kg/ha, GAEZ FAO
No. of Observations: 10015					

Appendix II: Transport Cost Calculation

To construct a measure of travel costs to the market in Nigeria we combine road survey data from the Federal Roads Maintenance Agency (FERMA) and World Bank's FADAMA²⁹ project, with GIS roads network data from Delorme³⁰. We used the Delorme data set for data on both the trunk roads and the rural network of Nigeria, and supplemented this data set by geocoding data on federal road attributes from FERMA as well as data on rural road attributes from the World Bank's FADAMA project.³¹ The Highway Development Management Model (HDM-4) programming tool, which accounts for the roughness of the terrain, quality and condition of the road, as well as country-level factors (such as the price of fuel, average quality of the fleet, the price of a used truck, and wages)³² was employed to compute travel costs from households (in the case of household level analysis) or cells (in the case of the local-GDP data analysis) to markets. The data used for the estimates in this paper were collected specifically for Nigeria, to best characterize the transportation conditions one would find there.

Characterization of network type and terrain

The road network of Nigeria includes three classes of roads: primary, secondary, and tertiary. Average vehicle speed and width of the main carriage road are used to characterize the differences among network types as follows:

Paved Road Speed (km/hr) by Network & Condition			
Road Condition	Primary 7m	Secondary 6m	Tertiary 5m
Flat	100	80	70
Rolling	80	70	60
Mountainous	60	50	40

²⁹ The FADAMA project is currently in its third stage. It was originally designed to improved utilization of irrigable land, implementing an innovative local development planning (LDP) tool, and building on the success of the community-driven development mechanisms. For more information about the survey and GIS methodology see "Spatial Analysis and GIS Modeling to Promote Private Investment in Agricultural Processing Zones: Nigeria's Staple Crop Processing Zones" presented at the Annual World Bank Conference on Land and Poverty 2013.

³⁰ Delorme is a company that specializes in mapping and GPS solutions and has the most comprehensive GIS data set on African roads.

³¹ Road segments missing from either of these data sets were deemed to be minor and categorized as tertiary class, unpaved, and in poor condition. Finally, all necessary adjustments were made through consultations with transportation experts familiar with Nigeria's road network in order to arrive at the final road network used in this study.

³² For more information on the HDM-4 model in general, see <http://go.worldbank.org/JGIHXVL460>. For more information on how the HDM-4 model was applied to Nigeria, see Appendix II.

Unpaved Road Speed (km/hr) by Network & Condition			
Road Condition	Primary 7m	Secondary 6m	Tertiary 5m
Flat	80	70	60
Rolling	60	50	40
Mountainous	40	30	20

where terrain type is defined using the following concepts and road geometry parameters:

- Flat. Mostly straight and gently undulating
- Rolling. Bendy and gently undulating
- Mountainous. Winding and gently undulating

TERRAIN TYPE	Number			Super-elevation (%)
	Rise & Fall (m/km)	Rise & Fall (#)	Horizontal Curvature (deg/km)	
	FLAT	10	2	
ROLLING	15	2	75	
MOUNTAINOUS	20	3	300	

Characterization of network type and condition

The International Roughness Index IRI (m/km) is used to define the differences in road condition by network as follows:

Paved Road IRI (m/km) by Network & Condition			
Road Condition	Primary 7m	Secondary 6m	Tertiary 5m
Good	2	3	4
Fair	5	6	7
Poor	8	9	10

Unpaved Road IRI (m/km) by Network & Condition			
Road Condition	Primary 7m	Secondary 6m	Tertiary 5m
Good	6	8	10
Fair	12	13	14
Poor	16	18	20

Characterization of vehicle type

A heavy truck was defined as the typical vehicle to model freight transport costs and it was assumed that it can transport an average weight of 15 tonnes (average net

weight). The following input data was gathered for the value of used vehicle, tire and fuel cost, maintenance labor cost and driver cost, among others.

FINANCIAL UNIT COSTS	HEAVY TRUCK
Used Vehicle Cost (US\$/vehicle)	70,000
New Tire Cost (US\$/tire)	800
Fuel Cost (US\$/liter)	0.77
Maintenance Labor Cost (US\$/hour)	4.73
Crew Cost (US\$/hour)	3.15

Using these parameters above, a final cost per ton-km for each road type is estimated (\$/ton/km):

Paved FLAT			
Road Condition	Primary	Secondary	Tertiary
Good	0.0526	0.0529	0.0533
Fair	0.0570	0.0583	0.0596
Poor	0.0617	0.0637	0.0986

Paved ROLLING			
Road Condition	Primary	Secondary	Tertiary
Good	0.0533	0.0531	0.0535
Fair	0.0577	0.0586	0.0599
Poor	0.0623	0.0643	0.0996

Paved MOUNTAINOUS			
Road Condition	Primary	Secondary	Tertiary
Good	0.0574	0.0562	0.0584
Fair	0.0620	0.0615	0.0644
Poor	0.0675	0.0676	0.1055

Unpaved FLAT			
Road Condition	Primary	Secondary	Tertiary
Good	0.0629	0.0673	0.0730
Fair	0.0795	0.0831	0.0867
Poor	0.0941	0.1017	0.1091

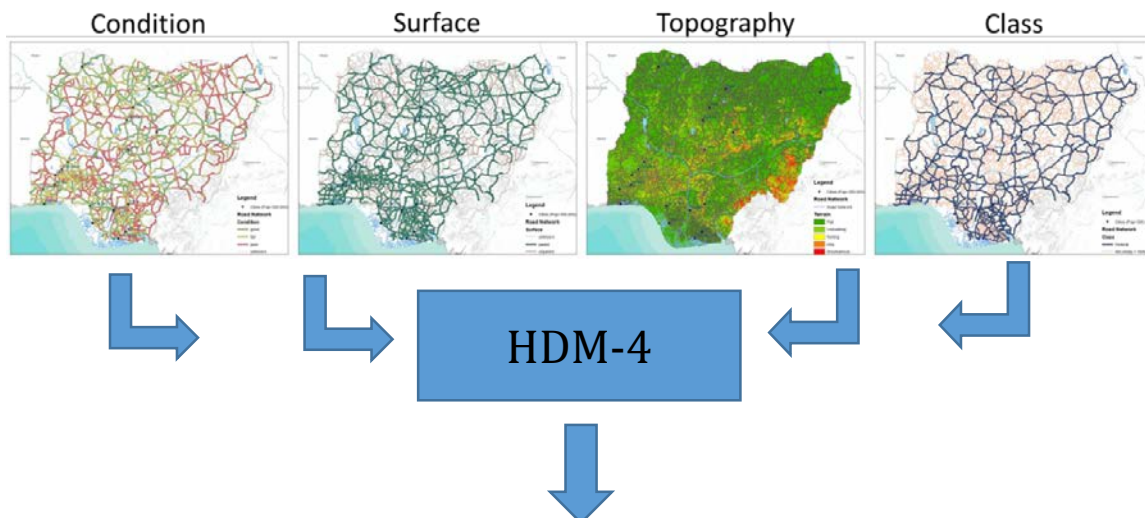
Unpaved ROLLING			
Road Condition	Primary	Secondary	Tertiary
Good	0.0618	0.0678	0.0752

Fair	0.0801	0.0837	0.0877
Poor	0.0945	0.1021	0.1095

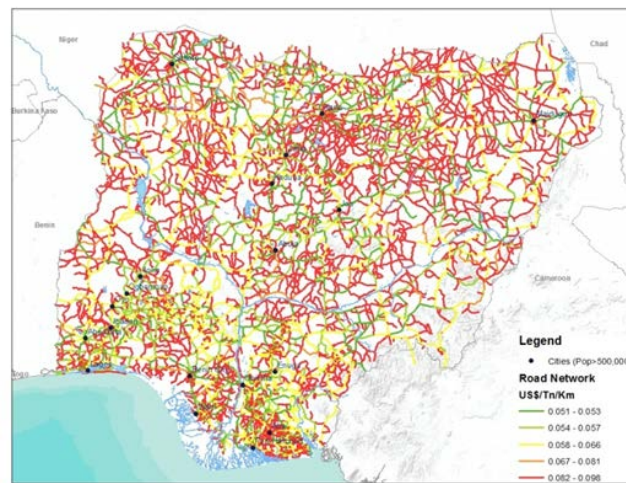
Unpaved MOUNTAINOUS			
Road Condition	Primary	Secondary	Tertiary
Good	0.0651	0.0748	0.0868
Fair	0.0820	0.0884	0.0974
Poor	0.0954	0.1038	0.1130

The last step is to add the individual transport cost for each combination of road types (54 in total) into the network data set segments and multiply them by the length of the roads (see figure 1). As a result a monetary road user cost can now be used as a measure of the amount of resistance required to traverse a path in a network, or to move from one element in the network to another. Higher impedance values indicate more resistance to movement, and a value of zero indicates no resistance. This way an optimum path in a network is the path of lowest impedance, also called the least-cost path. These costs per ton-km were used to calculate the cost it takes each georeferenced household or pixel centroid to transport one ton of goods to the nearest market.

Figure 7: Construction of road user cost network data set



Road User Costs Surface



Appendix III: Natural Path

To motivate the rationale for our proposed IV, it is instructive to briefly consider the source of the bias and the factors used to determine whether or not a public road investment is made. Benefits from new transportation infrastructure can come in many forms: (i) economic benefits, including the reduction of business costs and enhanced access to markets; (ii) social benefits, including improved social cohesion, faster diffusion of information, and better access to schools and hospitals; and (iii) political benefits, including the faster deployment of armies to quell unrest or conflict, and the pleasing of certain constituencies, among others.³³ Several factors are important in determining the costs of new road construction or improvement. The most important of these include the length of the road, and the topography of the land over which the road is being built.³⁴ Flatter land topography beneath the road is more desirable in that it is both cheaper to build the road upon, and to drive on once the road is built.

“Straight-line” or Euclidian distance IVs are commonly used in the literature because they are correlated with the costs of building the road, in particular distance but uncorrelated with the bias-causing economic benefits. The “straight-line” IVs therefore rest on the assumption that, regardless of topography and terrain, two points which are

³³ Indeed, many of these benefits spill over into all three categories.

³⁴ In addition to the standard costs, the opportunity cost of using public funds should also be considered. However, its inclusion in the decision-making process has no effect on the choice of instruments.

closer together (have a shorter straight line connecting them), have lower construction and travel costs. However, while useful, straight-line IVs ignore a major determinant of travel cost—the topography of the land—potentially making them weak instruments. Even if two points are very close together, they may not be easily accessible (e.g. if separated by a canyon or steep terrain), even if a road approximating that straight line existed. Thus, the natural path variable will be better correlated with our measure of transport, while still remaining exogenous, thus producing a more efficient estimate, especially in areas with steep or impassable terrain.

To construct the natural pathway instrument, we followed a similar approach that was used for the *Global Map of Accessibility* in the World Bank's World Development Report 2009 Reshaping Economic Geography (Uchida and Nelson 2009). An off-path friction-surface raster was calculated, which is a grid in which each pixel contains the estimated time required to cross that pixel on foot. We assume that all travel is foot based and walking speed is therefore determined by the terrain slope. The slope raster is taken from NASA's Shuttle Radar Topography Mission (SRTM) Digital Elevation Models (DEMs) which has a resolution of 90 meters. The typical velocity of a hiker when walking on uneven or unstable terrain is 1 hour for every 4 kilometers (4 km/hr) and diminishes on steeper terrain. We use a hiking velocity equation³⁵ (Tobler 1993) to reflect changes in travel speed as a function of trail slope:

$$W = 6e^{-3.5*|S+0.05|}$$

where W is the hiking velocity in km/hr and S is the slope or gradient of the terrain. Finally, we compute the time that it takes to travel from each point in Nigeria to each of our selected markets. The map of Nigeria is divided into a 'fishnet' grid of 10km² cells, with approximately 11,000 cells in total. Minimum travel times are calculated using the optimal walking path from the center of each of these 11,000 cells to each of the 65 markets. The algorithm utilizes a node/link cell representation system in which the center of each cell is considered a node and each node is connected to its adjacent nodes by multiple links. Every link has an impedance, which is derived from the time it takes to

³⁵ "Three presentations on geographical analysis and modeling non-isotropic geographic modeling speculations on the geometry of geography global spatial analysis". NATIONAL CENTER FOR GEOGRAPHIC INFORMATION AND ANALYSIS. TECHNICAL REPORT 93-1. February 1993

pass through the cell, according to the natural path friction cost surface, and takes into account the direction of movement through the cell. An ArcGIS/python script was written which creates an optimal path raster for each of the 65 selected cities/markets. This raster defines the optimal path (minimizing walking time), and then records the total time required in each cell. As a result we obtained an origin/destination travel time matrix of more than 11,000 rows (grid cells) and 65 columns (selected markets).

Appendix IV: Nigeria Demographic and Health Survey (NDHS)

Administratively, Nigeria is divided into 36 states and Abuja, the federal capital territory. Each state is subdivided into local government areas (LGAs), and each LGA is divided into localities. In addition to these administrative units, during the 2006 Population Census, each locality was subdivided into convenient areas called census enumeration areas (EAs). The sample frame for this survey was the list of EAs used in that census. The EAs were stratified separately by urban and rural areas. Rural areas are defined as a locality with a population of less than 20,000 constitutes.

The primary sampling unit (PSU), or cluster, for the 2008 NDHS was defined on the basis of enumeration area (EA) from the 2006 census frame. A minimum requirement of 80 households (400 population) for the cluster size was imposed in the design. If the selected EA has a population smaller than this minimum, a supplemental household listing was conducted in the neighboring EA. Although in Nigeria a majority of the population resides in rural areas, the urban areas in some states were over-sampled in order to provide reliable information for the total urban population at the national level. The target of the 2008 NDHS sample was to obtain 36,800 completed interviews. Based on the level of non-responses found in the 2003 Nigeria DHS, to achieve this target, approximately 36,800 households were selected, and all women aged 15-49 were interviewed. A requirement was to reach a minimum of 950 completed interviews per state. In each state, the number of households was distributed proportionately among its urban and rural areas. The selected households were then distributed in 888 clusters in Nigeria, 286 of which were urban area clusters, and 602 were rural area clusters. More details about the sample selection can be obtained in NPC (2009).

Appendix V: NDHS Multi-dimensional Poverty Index

Table A4: NDHS multi-dimensional poverty index components

Dimension	Indicator	Deprived if...	Relative Weight
Education	Highest degree earned	No household member has completed five years of education.	1.67
	Child School Attendance	Household has a school-aged child not attending school	1.67
Health	Child Mortality	Household has had at least one child aged 0-5 years die in the past 5 years.	1.67
	Nutrition	Household has a malnourished woman aged (15-49) or child aged (0-5).	1.67
Standard of Living	Electricity	The household has no electricity.	0.56
	Improved Sanitation	Household does not have improved sanitation.	0.56
	Safe Drinking Water	Household does not have access to improved water source.	0.56
	Flooring	The household has a dirt floor.	0.56
	Cooking fuel	The household uses dirty cooking fuel.	0.56
	Asset Ownership	The household does not own more than one bicycle, motorcycle, radio, fridge, TV, or phone and does not own a car.	0.56

World Health Organization (WHO) standards were followed in determining what to consider unimproved water sources, inadequate sanitation, and dirty cooking fuel.

A household is considered to be multi-dimensionally poor if its weighted sum of indicators was greater than 3. Note that the weights add up to about 10, the number of indicators (difference due to rounding).

Appendix VI: LSMS Multi-dimensional Poverty Index

To test the robustness of our multi-dimensional poverty index indicator, we constructed a second MPI using data from the LSMS. The index is constructed in a similar manner, with three main components each receiving equal weight: education, health and standard of living. Table A5 gives a detailed description of the components of this index.

Table A5: LSMS multi-dimensional poverty index components

Dimension	Indicator	Deprived if...	Relative Weight
Education	Highest degree earned	No household member has completed six years of education, i.e. earned at least the First School Leaving Certificate (FSLC)	1/6
	Child School Attendance	Any school-aged child is not attending school (children 6-16)	1/6
Health	Child Mortality	Any child has died in the family	1/6
	Nutrition	Any household member has gone to sleep hungry during the past week	1/6
Standard of Living	Electricity	The household has no electricity	1/18
	Improved Sanitation	The household does not have a toilet that flushes or a ventilated improved pit, or must share one with other households	1/18
	Safe Drinking Water	Household does not have access or must walk more than 30-minutes round trip to get safe water. (safe water includes: pipe borne water, bore hole/hand pump, well/spring protected, rainwater)	1/18
	Flooring	The household has a straw, dirt, sand, or mud floor	1/18
	Cooking fuel	The household cooks with firewood, coal, grass, or kerosene (as opposed to electricity or gas)	1/18
	Asset Ownership	The household does not own more than one radio, TV, bike, motorbike, or fridge, and does not own any landline, car or other vehicle	1/18

For consistency, similar controls were used in regressions on the LSMS MPI as were used for other LSMS regressions. As with the NDHS MPI, we estimate two linear probability models (OLS and IV), and two maximum likelihood models (probit and IV probit) Table A6, columns (1) report the OLS and IV estimates, respectively, of the effect that log cost to market has on the households' probability of being multi-

dimensionally poor. We find that the coefficient on market cost is positive and significant at the one percent level in both regressions. Consistent with a positive bias, the IV estimates is smaller in magnitude, between at 0.073, compared with 0.104 for OLS.

Table A6, column (3) reports the probit marginal effects which are nearly identical to the OLS estimate in column (1). Column (4), which reports the IV probit estimate shows a much larger estimated impact, more than doubled. A ten percent decrease in transport costs decreases the probability of being multidimensional poverty by roughly 2 percent.

As a robustness check of the exogeneity of our IVs, we report the Conley Bounds in Table A7. From the 95% confidence intervals reported, we see as we increase the correlation between the IV and the outcome variable the range of estimated values widens, but remains positive. Taken together with the first stage test results, these test statistics suggest that our IVs have power in explaining the variation in cost to market across the households.

In general, the coefficient on market cost for the MPI constructed using the NDHS is very similar to that for the MPI constructed using the LSMS. The NDHS coefficients on market cost for IV and IV probit are 0.065 and 0.243, respectively. For the LSMS MPI, those coefficients are 0.073 and 0.208. In fact, a modified t-test confirms that there is no statistical difference between these estimates.

Table A6: Multi-dimensionally poor, LSMS

Dependent Variable: dummy=1 if MPI poor	(1) OLS	(2) IV Natural Path	(3) Probit	(4) IV Probit Natural Path
ln(Cost to Market)	0.104*** (4.68)	0.073*** (3.02)	0.114*** (4.47)	0.208*** (2.77)
hh agricultural labor	0.061*** (4.40)	0.063*** (4.59)	0.065*** (4.43)	0.184*** (4.64)
hh agricultural labor squared	-0.006*** (-3.64)	-0.006*** (-3.72)	-0.006*** (-3.71)	-0.018*** (-3.81)
land	-0.001 (-1.41)	-0.001 (-1.15)	-0.001 (-1.57)	-0.002 (-1.29)
fertilizer	0.000 (0.71)	0.000 (0.74)	0.000 (0.68)	0.000 (0.67)
dummy=1 if irrigates land	-0.048 (-1.02)	-0.048 (-1.07)	-0.096 (-1.38)	-0.280 (-1.32)
dummy=1 if tropical warm subhumid	-0.140*** (-5.10)	-0.148*** (-5.37)	-0.153*** (-4.93)	-0.449*** (-5.04)
dummy=1 if tropical warm humid	-0.170*** (-4.14)	-0.181*** (-4.42)	-0.193*** (-4.07)	-0.535*** (-4.44)
dummy=1 if tropical cool humid	-0.339*** (-4.80)	-0.350*** (-5.16)	-0.370*** (-4.93)	-0.999*** (-4.77)
age of hh head	-0.003 (-0.88)	-0.003 (-0.99)	-0.003 (-0.78)	-0.010 (-0.90)
age squared	0.000 (1.18)	0.000 (1.25)	0.000 (1.03)	0.000 (1.11)
dummy=1 if hh head is literate	-0.258*** (-13.76)	-0.261*** (-13.97)	-0.271*** (-14.24)	-0.775*** (-13.34)
Constant	0.740*** (6.46)	0.808*** (7.02)		
Observations	2,600	2,600	2,600	2,600
First Stage Results				
IV: ln(Natural Path)		0.631*** (19.28)		0.631*** (19.32)
Angrist-Pischke Test		371.71 P=0.0000		

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

Data: Nigeria LSMS-ISA 2010

Table A7: Conley Bounds, MPI LSMS

	Support for possible values of δ	95% Confidence Interval	
	IV: ln(Natural path)	Lower Bound	Upper Bound
ln(MPI LSMS)	$\delta: [-0.0001, 0.0001]$	0.025	0.120
	$\delta: [-0.001, 0.001]$	0.024	0.121
	$\delta: [-0.01, 0.01]$	0.009	0.135

Calculated using the code from Conley et al (2012).

Appendix VII. Robustness Checks for LSMS results

To test the robustness of the estimated impact of transport costs on different sources of income, we estimate several alternative specifications. These include (1) estimating the impact on each source of income separately in a two-stage least squares model, (2) controlling for agro-ecological zone fixed effects in place of the marketshed dummies, (3) excluding land, labor, fertilizer, and irrigation controls, and (4) introducing additional controls including access to credit and the presence of a hospital in the community. These results are reported below.

Table A8

Dependent Variable: ln(Crop Revenue)	(1) 2SLS with marketshed FE	(2) 2SLS with AEZ FE	(3) 2SLS with Fewer Controls	(4) 2SLS with More Controls
ln(Cost to Market)	-0.377** (-2.54)	-0.970*** (-4.83)	-0.288** (-1.98)	-0.487*** (-2.64)
hh agricultural labor	0.234*** (3.66)	0.170** (2.56)		0.236*** (3.39)
hh agricultural labor squared	-0.026*** (-4.16)	-0.021*** (-3.43)		-0.026*** (-3.92)
land	0.012*** (3.51)	0.018*** (4.19)		0.011*** (2.91)
fertilizer	0.001 (1.04)	0.001* (1.67)		0.001 (0.99)
dummy=1 if irrigates land	0.780** (2.47)	0.497 (1.20)		0.638** (2.08)
age of hh head	0.022 (1.47)	0.013 (0.77)	0.025* (1.69)	0.021 (1.29)
age squared	-0.000 (-1.52)	-0.000 (-0.65)	-0.000* (-1.66)	-0.000 (-1.32)
dummy=1 if hh head is literate	0.189** (1.97)	0.298*** (2.65)	0.173* (1.76)	0.221** (2.09)
dummy=1 if hospital in village				-0.001 (-0.19)
dummy=1 if hh head has access to credit				0.189* (1.84)
Constant	1.777* (1.84)	2.090*** (3.31)	1.818* (1.94)	2.003* (1.95)

Marketshed fixed effect	Yes	No	Yes	Yes
Agro Ecological Zone fixed effect	No	Yes		
First Stage Results				
IV: ln(Natural Path)	0.661*** (33.34)	0.631*** (19.28)	0.676*** (33.22)	0.653*** (28.03)
Angrist-Pischke Test	1111.64 P=0.0000	371.66 P=0.0000	1103.70 P=0.0000	785.54 P=0.0000
Observations	2,600	2,600	2,600	2,228

Data: Nigeria LSMS-ISA 2010

Table A9

Dependent Variable: ln(Livestock Sales)	(1)	(2)	(3)	(4)
	2SLS with marketshed FE	2SLS with AEZ FE	2SLS with Fewer Controls	2SLS with More Controls
ln(Cost to Market)	-0.173 (-1.30)	-0.071 (-0.51)	-0.132 (-1.03)	-0.149 (-1.06)
ln(cost of animals)	0.205*** (5.25)	0.183*** (4.46)	0.205*** (5.30)	0.192*** (4.92)
hh agricultural labor	0.113 (1.63)	0.113* (1.68)		0.122 (1.63)
hh agricultural labor squared	-0.006 (-0.81)	-0.006 (-0.72)		-0.007 (-0.80)
land	0.005 (1.56)	0.006* (1.72)		0.003 (0.73)
fertilizer	-0.000 (-0.77)	-0.000 (-0.88)		-0.000 (-0.87)
dummy=1 if irrigates land	0.143 (0.56)	0.231 (0.82)		0.119 (0.47)
age of hh head	0.030** (2.04)	0.024 (1.61)	0.039*** (2.61)	0.030* (1.87)
age squared	-0.000** (-2.19)	-0.000* (-1.89)	-0.000*** (-2.70)	-0.000* (-1.86)
dummy=1 if hh head is literate	-0.186* (-1.95)	-0.199** (-2.04)	-0.204** (-2.11)	
dummy=1 if hospital in village				0.001 (0.21)
dummy=1 if hh head has access to credit				-0.064 (-0.60)
Constant	0.048 (0.07)	0.062 (0.12)	-0.043 (-0.07)	-.140614 -0.21
Marketshed fixed effects	Yes	No	Yes	yes
Agro-ecological zone fixed effects	No	Yes	No	No
First Stage Results IV: ln(Natural Path)	0.668*** (28.76)	0.641*** (22.82)	0.671*** (28.43)	-0.057 (-0.40)
Angrist-Pischke Test of Weak Identification	827.19 P=0.0000	520.90 P=0.0000	808.31 P=0.0000	617.09 P=0.0000
Observations	1,742	1,742	1,742	1,530

Data: Nigeria LSMS-ISA 2010

Table A10

Dependent Variable: ln(Non-Agricultural Income)	(1)	(2)	(3)	(4)
	2SLS with marketshed FE	2SLS with AEZ FE	2SLS with Fewer Controls	2SLS with More Controls
ln(Cost to Market)	-0.386** (-2.45)	-0.446*** (-3.48)	-0.360** (-2.34)	-0.385** (-2.35)
age of hh head	-0.041** (-2.09)	-0.049** (-2.52)	-0.022 (-1.13)	-0.044* (-1.89)
age squared	0.000** (1.97)	0.000** (2.34)	0.000 (1.18)	0.000* (1.76)
dummy=1 if hh head is literate	0.385*** (3.52)	0.385*** (3.50)	0.390*** (3.49)	0.312***
land	0.002 (0.60)	0.002 (0.82)		-0.001 (-0.37)
hh labor	0.253*** (2.78)	0.295*** (3.11)		0.251*** (2.61)
hh labor squared	-0.014 (-1.60)	-0.016* (-1.68)		-0.014 (-1.53)
Total business expenses	0.000*** (23.52)	0.000*** (29.27)	0.000*** (24.41)	0.000*** (25.44)
dummy=1 if hospital in village				0.006* (1.76)
dummy=1 if hh head has access to credit				0.323*** (2.67)
Constant	1.777* (1.84)	4.780*** (8.40)		5.004*** (7.51)
Marketshed fixed effect	Yes	No		Yes
Agro Ecological Zone fixed effect	No	Yes		No
First Stage Results IV: ln(Natural Path)	0.672*** (27.93)	0.641*** (22.73)	0.673*** (28.29)	.6550552 24.68
Angrist-Pischke Test	779.96 P = 0.0000	516.54 P=0.0000	800.57 P=0.0000	609.00 P=0.0000
Observations	1,355	1,355	1,355	1,088

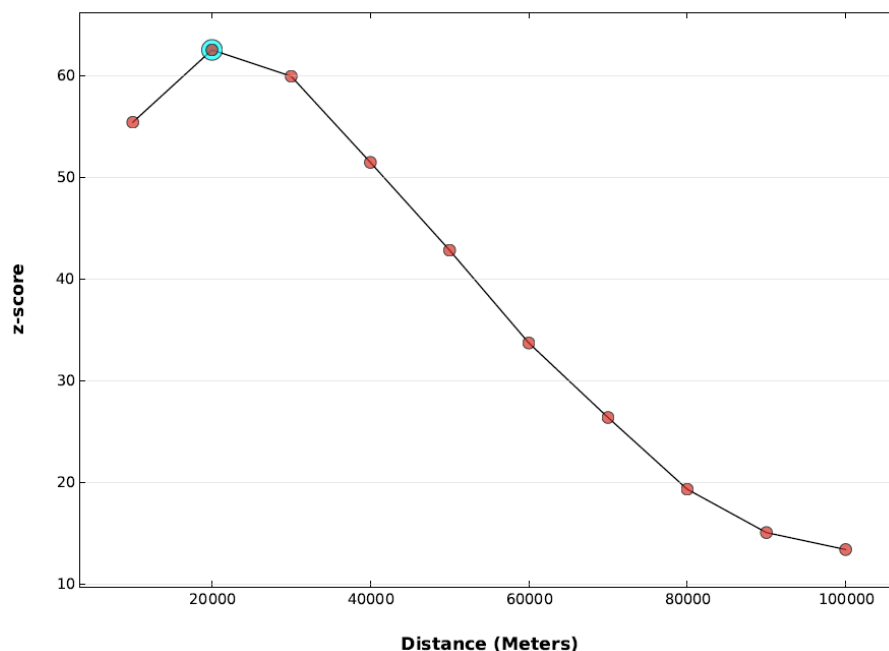
Data: Nigeria LSMS-ISA 2010

Appendix VIII: Spatial Autocorrelation Bootstrapping

Given the gridded form of the data in the local GDP regression, regression estimates produced using it are susceptible to being biased due to short-distance residual spatial autocorrelation (RSA). Positive spatial autocorrelation means that geographically nearby values of a variable tend to be correlated. This can cause statistical dependence of the residuals of a regression, which would lead to a biased linear estimate. The coefficients would be biased because areas with a higher concentration of events will have a greater impact on the model estimate and precision will be overestimated because concentrated events tend to have fewer independent observations than are being assumed. (Levine 2004). Nevertheless there are authors that have found non-spatial model to be robust and unbiased for several data sets see Hawkins et al. (2007)

We use a spatial bootstrapping technique to test for a possible bias in our estimates. The first step of implementing this technique involves determining the geographic extent of the spatial autocorrelation. The standard regression is run, in this case, the model shown in Table 7, column 3, and the residuals are used to calculate a correlogram. The correlogram graphs the correlation of the residuals with each other versus their distance from each other, by using Moran's I. The correlogram calculated from our model is shown in Figure 8. Peaks reflect distances where the spatial processes promoting clustering are most pronounced and are an indicator of the extent of the short-distance RSA in the model. In this case, it occurs at 20 kms, implying that all points more than 20 kms apart are spatially independent.

Figure 8: Spatial Autocorrelation by Distance



The correlogram shows the Moran's I z-score versus distance. The peak z-score gives the maximum extent of the spatial autocorrelation.

The next step is to resample the data set. This is done in a similar way to the standard bootstrapping methodology, with two important distinctions. The first is that the data set is sampled without replacement. And second, it is done in a way that ensures that all points are spatially independent from each other. To be conservative, we used a 30 km buffer, 10 kms larger than that suggested by the correlogram. This buffer, along with the geography of Nigeria, allows for approximately 500 points to be selected, and to be spatially independent of each other. 1,000 samples were generated and each sample was regressed separately, with the mean of each respective coefficient calculated to arrive at the bootstrapped coefficient, and the standard deviation of the respective coefficients used to calculate t-statistics.

The results shown in Table A11 show that there is no significant difference between the coefficient on cost to market in the standard model and the bootstrapped estimate, implying that spatial autocorrelation has no statistically detectable influence on our estimate.

Table A11: Local GDP, Spatial Bootstrap

Dependent Variable: Local GDP	(1) IV Nat Path Full Sample	(2) IV Nat Path Bootstrapped
ln(Cost to Market)	-0.496*** (-26.28)	-0.4624*** (-4.39)
ln(Distance to Mine)	-0.0282 (-1.17)	-0.0254 (0.23)
ln(Population)	0.101** (2.41)	0.0226 (-0.06)
ln(Population)^2	0.0462*** (18.50)	0.0523*** (-2.61)
ln(Cassava potential yield)	0.0160** (2.10)	0.0155 (-0.39)
ln(Cassava potential yield)^2	0.00334* (1.80)	0.0033 (-0.33)
ln(Yams potential yield)	-0.0175* (-1.92)	-0.0168 (0.42)
ln(Yams potential yield)^2	-0.00339 (-1.52)	-0.0035 (0.35)
ln(Maize potential yield)	0.0603*** (4.02)	0.1043 (-0.31)
ln(Maize potential yield)^2	-0.0063*** (-2.94)	-0.0104 (0.35)
ln(Rice potential yield)	-0.00638 (-1.12)	-0.0056 (0.19)
ln(Rice potential yield)^2	0.00009 (0.07)	0.0001 (-0.01)
Constant	-0.748*** (-3.09)	-0.8026 (0.38)
Marketshed Fixed Effects	Yes	Yes
Observations	10,607	Obs/sample: 500 Samples: 1,000

t-statistics in parentheses*** p<0.01, ** p<0.05, * p<0.1

Data: Various sources, see section 5.3

Coefficients given in column 2 are the mean coefficients of each variable from the 1,000 regressions of each sample. T-statistics are calculated by replacing the standard errors with the standard deviation of the sample estimates.