

TI 2017-076/V
Tinbergen Institute Discussion Paper



The Impact of Road Development on Household Welfare in Rural Papua New Guinea

Martin Wiegand¹
Eric Koomen²
Menno (M.) Pradhan³
Christopher Edmonds⁴

1: VU Amsterdam

2: VU Amsterdam

3: VU Amsterdam; University of Amsterdam, the Netherlands; Tinbergen Institute, The Netherlands

4: Tokyo International University

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at the [Tinbergen Site](#)

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

The Impact of Road Development on Household Welfare in Rural Papua New Guinea

Martin Wiegand^{*}, Eric Koomen[†], Menno Pradhan[‡], and Christopher Edmonds[§]
July 2017

Abstract

In this paper we estimate the impact of road development on household welfare in rural Papua New Guinea over the period between 1996 and 2010, using two cross-sectional household surveys and corresponding road maps. To deal with endogenous placement of road infrastructure programs we employ a correlated random effects model that corrects for location-specific changes in road quality. We also use a newly developed quantile regression method to investigate whether road works are pro-poor. Estimates show that investments in sealing roads to nearest towns led to higher consumption levels and housing quality, and to less reliance on subsistence farming. Effects are stronger among poor, less educated, and female-led households.

^{*} Department of Economics of the Vrije Universiteit Amsterdam, the Netherlands.

[†] Department of Spatial Economics of the Vrije Universiteit Amsterdam, the Netherlands.

[‡] Department of Economics of the Vrije Universiteit Amsterdam and University of Amsterdam, the Netherlands.

[§] Institute for International Strategy, Tokyo International University, Kawagoe, Saitama Prefecture, Japan.

The authors recognize, with thanks, the advice and assistance of John Gibson in early discussions of the design and construction of the dataset compiled for this research. They also acknowledge financial support of the Asian Development Bank's financial support for this research through its technical assistance project 'Developing Impact Evaluation Methodologies, Approaches, and Capacities in Selected Developing Member Countries'.

1. Introduction

Transport access is one of the key elements necessary for economic development, especially in rural areas. For areas without access to waterways, roads generally offer the sole means for connecting people to markets and public services. Better roads have the potential to impact living standards through a number of mechanisms: They can reduce transport costs for agricultural goods and inputs, enable rural households to engage with the labor market, and permit larger truckloads and more frequent transport options. Improved market access can lead to a greater variety and lower prices of essential inputs and consumption goods, as well as higher prices and demand for local products (Gibson and Rozelle 2003). They may also attract financial service providers, facilitating agricultural investments and consumption smoothing (Binswanger et al. 1993). Living close to a road with higher traffic intensity can create demand for local businesses like roadside stalls. Better market access may also raise local productivity and wages, and facilitate the transformation from subsistence agriculture to growing cash crops or to non-agricultural activities, enabling diversification of household income sources (Mu and van de Walle 2011, Aggarwal 2017). Last of all, better roads may also enhance access to services like schools and hospitals, lower their cost, and improve their quality, e.g. because they are easier to reach for teachers and doctors as well as materials suppliers (Bell and van Dillen 2012).

All these factors suggest that better roads lead to higher average household consumption. Several existing studies confirm this to be the case (Knox et al. 2013). The distributional effects of better roads are less clear, and the empirical evidence on whether the poor benefit from roads in the same way as the non-poor remains inconclusive. Consumption gains from better roads could be relatively higher for the poor. For example, if the non-poor are able to compensate better for a lack of good roads because they have a better market position within the village, the poor would experience a relatively higher productivity gain from improved roads. But it may just as well be the case that the non-poor profit more from better roads, since they might be able to scale up agricultural production easier, or because the poor may be kept from road utilization due to transport costs.

In this paper we investigate the impact of road quality and access on household welfare in rural Papua New Guinea (PNG) between 1996 and 2010. Roads in rural areas remain scarce in PNG. In 2009, the country had a road density of 56 km per 1,000 square km, which is very low compared to PNG's neighboring countries (Indonesia had 250 km per 1,000 square km in the same year). In the same year,

only 13% of roads were sealed, while the majority of roads were gravel or dirt roads. Besides the setting of a country with limited rural road access, our study is also interesting because of the data we use. We combine two cross-sectional household surveys with GIS maps of the road system of around the time the household surveys were administered. The advantage of this setup is that we can employ a set of road variables that is directly linked to road infrastructure characteristics. Most other studies obtain their road access variables from the household survey and rely on travel times and distance to the nearest road. Our map data contains quality measures for each road section. In combination with the location of the household it allows us to calculate the distance to the nearest road, and the quality and length of the stretch of road that leads to the nearest town. The types of data we use in our analysis are available in many country settings, which makes the approach widely applicable and advantageous in terms of data collection costs and time.

Estimating the effects of road investments is complicated by the fact that government decisions about where to construct new roads or whether to rehabilitate or upgrade existing ones is endogenous with areas' growth and other development achievements. These decisions are often made based on unobserved factors like expected traffic volume, local productivity, investment cost, and political benefits of placing roads in particular areas—all factors that may also affect household welfare directly.¹ Existing research in road development impacts has used a variety of approaches to address this endogeneity problem. Instrumental variable estimation—which requires an exogenous variable that affects road development but has no direct effect on the outcome variable of interest—is one approach. For example, the straight line approach pioneered by Banerjee et al. (2012, 2004) uses the distance between the sampled household and the nearest straight line between major cities as an instrument for road access. The location of the major cities is assumed exogenous, as are the straight lines between them. Where panel data is available, an alternative approach to address potential endogeneity in road placement is to use time-invariant village or household fixed effects (Khandker et al. 2009, Khandker and Koolwal 2010) to assess the impact of road investments over the period covered by the panel. The fixed effects account for endogeneity caused by time-invariant characteristics of the location. The availability of multiple time periods further allows instrumentation using lagged outcomes (Dercon et al. 2009, Khandker and Koolwal 2011). Another technique often used when the road variable is binary (project road or not) is difference in differences (DID) estimation, often with propensity matching to allow the

¹ An example of political favoritism in road placement is provided by Burgess et al. (2015), who find that during periods of autocracy in Kenya, road construction in each district was governed strongly by whether the district had a large proportion of people from the same ethnic affiliation as the current president.

common trend assumption to hold conditional on covariates (Lokshin and Yemtsov 2005, Mu and van de Walle 2011). Estimation using regression discontinuity to study road impacts is rare, but one example is provided by Casaburi et al. (2013) in their study of road development impacts in Sierra Leone, where the roads selected for improvement were the ones that ranked highest on a priority ranking.

We use a village-fixed effects model to correct for endogenous placement of road infrastructure. In spite of the absence of panel data, we are able to estimate impact correcting for time-invariant location-specific factors, because we have census data on the road access variables for the time periods covered in different cross sections. Specifically, using the correlated random effects approach proposed by Chamberlain (1982), we correct for correlations of road quality with unobserved location-specific effects.² Access to road maps of the entire network from different years and knowledge of the location of surveyed households enables us to get results equivalent to the within estimator for panel data. We also apply a generalized quantile regression estimator developed by Powell (2016) to investigate how the effects of road infrastructure vary across the consumption distribution. This procedure allows to examine how different quantiles of consumption are affected while also accounting for covariates, making the estimates comparable to the results from our base specification. Our results generally point to beneficial effects of high-quality roads. The estimates suggest that upgrading one percent of the route leading to the nearest town from dirt to sealed road surface increases average household consumption by about 0.55%, raises the chance households live in a house with a high-quality roof by about 0.19%, and decreases the probability a household relies on subsistence farming by 0.14%. Furthermore, analyzing these impacts across different subgroups, we find that the effects on consumption are at least twice as high for households with less than 4 years of average education, an illiterate household head, or a female household head, when compared to their respective opposite subgroups. We find effects of similar magnitude for the upgrading of a gravel road to a sealed road. The differences between dirt and gravel roads are not significant for most outcomes. Generalized quantile regression estimates indicate that the effect of sealing gravel roads is higher for the poorest households, suggesting that road works can be considered anti-poverty measures in the case of rural PNG.

Our study contributes to the literature that rigorously assesses the impact of roads on rural development outcomes at the household level. Dercon et al. (2009) find that upgrading the nearest road to an all-weather road in rural Ethiopia led to a reduction in poverty of 6.9 percentage points.

² It should be noted that Khandker et al. (2009) and Khandker and Koolwal (2010) use the Chamberlain approach in their roads studies as well, however for quantile regressions and using panel data.

Furthermore, they find no evidence of heterogeneity of this effect with regard to household characteristics like size of landholdings, livestock holdings, or literacy of the household head. However, their estimates show the effect on consumption growth is larger for households with landholdings of at least a hectare and a literate household head. Dercon et al. (2012) obtain a complementary results, finding that remoteness from towns and poor roads are among the factors most associated with chronic poverty. Khandker et al. (2009) investigate how households in Bangladesh profited from road improvement projects. They predict that villages next to an improved road experience a reduction of poverty of 5 percentage points. The impact on household expenditure is higher for lower expenditure quintiles in this study, suggesting that road investments are pro-poor. However, using a larger dataset and controlling for other investment programs, Khandker and Koolwal (2010) find the opposite pattern. Mu and van de Walle (2011) find positive and significant average effects of rural road rehabilitation on local (commune level) market development in Vietnam using double difference and matching methods to address potential selection bias in identifying impacts. The authors note a tendency for poorer localities to have higher impacts due to lower levels of initial market development. A replication study by Nguyen (2016) confirms these results.

Research by Gibson and Rozelle (2003) examines the impact of road development in PNG, so comes closest to this paper. The study uses data from the 1996 PNG Household Survey to estimate the impact of travel time to the nearest road on household welfare. To correct for potential endogeneity, the authors use the year in which the national highway system of PNG first entered each district as an instrument for travel time to the nearest road. The rationale is that highway construction happened mainly according to geographical necessities—like having to start at the coast and proceeding inland—and was therefore independent of local characteristics like productivity and average income levels. Rural feeder road networks would follow highway construction, so that in districts where the highway entered early, a higher road density is expected. The estimates show that reducing the travel time to the nearest road to a maximum of two hours led to an overall reduction in poverty of between 5.8% and 11.8%. Our paper updates this earlier analysis, motivated by the availability of a new round of household survey and road maps covering the period in between both household surveys.

The remainder of the paper is structured as follows: In section 2 we discuss the country context of the study. Section 3 provides detail on the data used. Section 4 outlines the estimation techniques we use. Section 5 presents and discusses estimation results. In Section 6, we offer some concluding points.

2. Context

With a population of roughly 8.1 million people in 2016, PNG is the largest and most populated country of the Pacific region (SPC 2017). The most recent estimate of the poverty incidence (2010) classifies roughly 40% of the population as poor, which is an increase of about 2 percentage points relative to the prior estimate from 1996 (Gibson 2012). A majority of PNG's population (roughly 85%) lives in rural areas and relies on sale of crops for cash income in their semi-subsistence livelihoods.

The country performs poorly on a number of indicators of social development.⁴ According to latest available estimates, only about 6 in 10 people in the country have access to safe drinking water (2012). Roughly the same proportion of the school-aged population is enrolled in primary education (2008). Health indicators are poor: the maternal mortality rate stood at 220 per 100,000 live births (2014), infant mortality was over 48 per 1,000 live births (2012), and the percentage of children below 5 years of age reportedly suffering malnutrition was over 18% (2005). These indicators are well below those of other countries with similar per capita incomes, and Papua New Guinea did not achieve any of the Millennium Development Goals by 2015. The country was ranked 155 out of 188 countries in the United Nations Development Programme human development index in 2010.

PNG is a small, open, export-oriented economy heavily reliant on extracted resource exports, particularly oil, liquid natural gas (LNG), and minerals/metal ores. Mining exports represent about 75% of total exports, while agriculture products represent about 20%, and forestry products the remaining 5%. Revised national accounts estimates for 2013 show that the three largest sectors of the economy (as a share of estimated gross domestic product) were: industry—including mining and quarrying—at about 38% of GDP; services (e.g., wholesale and retail trade) at around 32%; and agriculture, forestry, and fishing at about 30%. However, the agriculture sector plays a dominant role for a vast majority of the population, with estimates suggesting that roughly three quarters of the population depend on subsistence agriculture for their livelihoods.

During the years covered by the data used in this study (1996 to 2010), PNG's real per capita growth was moderate—averaging 2.6% per annum—and highly volatile, reflecting typical cycles of resource boom and bust associated with an export-dependent economy reliant on a small number of exports. High points during this period were in 1996 and 2010, when per capita growth reached 5.0%

⁴ The numbers reported in this section come from the Asian Development Bank's *Asian Development Outlook* and the World Bank's *World Development Indicators*.

and 5.3%, reflecting, respectively, an oil export boom and growth stimulus from expenditures to construct an LNG pipeline. However, the economy suffered a number of recessions brought on by commodity price falls or worsening conditions in key export markets. From 1997 to 2003, PNG's per capita GDP fell in every year, with an average decline of 3.5% in that period. Development and maintenance of PNG's road network suffered during the two decades following independence, when funding for road maintenance fell by half (Kwa et al. 2009). Existing roads generally fell into disrepair and there was very low investment in new roads. Government expenditure on infrastructure per capita reached its minimum in 2001, however large and sustained increases in funding only began in 2010 (Dornan 2016).

A number of other factors have made it difficult to construct new and maintain existing roads during this period: (i) limited road management capacity in the private sector due to unsteady provision of maintenance contracts, (ii) competition for construction equipment and skilled engineers between resource extraction enterprises and the Department of Works, and (iii) disputes with owners of land proximate to road work (Lucius 2010). Outright corruption has also adversely affected the quality of road expenditures (Dornan 2016). Lastly, PNG's geography and weather—with steep slopes and high seasonal rainfall in many regions (especially the densely populated agricultural heartland of the Highlands region)—increase road construction and maintenance costs.

3. Data

Two nationally representative cross-sectional household surveys conducted in PNG—the 1996 Papua New Guinea Household Survey (PNGHS 96) and the 2009–2010 Household Income and Expenditure Survey (HIES 09/10)—provide the primary source of data on household consumption and other indicators used in this study. Both surveys collect a broad range of variables, but we only use variables which were collected consistently in both waves. The 1996 household survey included a nationally representative sample of 830 rural households in 73 geographic clusters (census units). The 2009 survey collected information from 2,208 rural households in 125 clusters. We also make use of variables from the PNG Census 2000 Community Profile System (CPS 2000), which contains information on the location and population of all census units and towns from the census of 2000. This allows us to locate all census units from the two surveys as point locations on a map.⁵ The HIES 09/10 includes GPS coordinates of all

⁵ Since the PNGHS 96 used sampling based on census units from the census of 1990, on which data is unavailable, we first had to recode the 1990 census units. For this, we relied on the census unit names listed in

surveyed households allowing us to calculate fairly precise, household-specific distances to the nearest road. In addition, we make use of the Papua New Guinea Resource Information System (PNGRIS) of the PNG National Agricultural Research Institute. This spatial database contains information on elevation, climate, and other biophysical characteristics we include as control variables in our statistical analysis.

For data on status of road infrastructure over time, we rely on the road information data bank and geographical information system of the Road Asset Management System (RAMS). The RAMS project, initiated in 1998 by the PNG Department of Works (DoW), was intended to provide a road asset management database and analytical tools to inform policy makers about road maintenance needs and economic efficiency of investments in the road network (Jusi et al. 2003). We link initial RAMS data—based on road surveys conducted between 1999 and 2001—to the PNGHS 96 (we refer to the combined data as the 2000 map). Due to continued underinvestment in the transport sector, certain dimensions of the RAMS—particularly its traffic counts (vital to estimating a road’s value)—were not updated after 2001. However, the provincial works managers of the DoW were given financial support to update data on road conditions, and data collected by DoW provide the basis for our second dataset of road conditions in 2009. The road system was estimated to consist of roughly 26,000 km of roads in 2009. We link this dataset to the HIES 09/10 and we refer to this as the 2009 map.⁶

Both the 2000 and 2009 road maps include detailed information for each road segment. Most importantly, the surface type (sealed, gravel, dirt) and the condition of the segment (good, fair, poor) is characterized for all roads. However, the 2009 road map has much more extensive coverage of the road network—covering an additional 14,000 kilometers of roads not included in the 2000 map.⁷ Discussions

Gibson and Rozelle (1998) as well as the generous help of staff at the NSO. The HIES 09/10 was sampled from the census of 2000, which made the location of census units easy.

⁶ It should also be mentioned that at the time of writing this report (2016), the Papua New Guinea-Australia Transport Sector Support Program (TSSP) together with the DoW completed the Visual Road Condition Survey, the most detailed survey of the national road network to date. However, due to the sudden heavy rise in national road investments starting in 2011, we believe that the conditions in this survey do not adequately reflect the conditions around the time the HIES 09/10 was conducted.

⁷ Guidelines for the interpretation of the condition classifications are provided in the CAPE-PNG-9-Transport-Sector-Assessment. A road segment is labelled “good” if it is passable for a two-wheel vehicle in wet weather, “fair” if it requires periodic maintenance, and “poor” if it requires reconstruction or rehabilitation (<http://devpolicy.org/road-management-in-papua-new-guinea-part-1-the-maintenance-challenge-20140903/>). Road condition is itself a function (unknown to us) of a number of indicators also given in the dataset, including severe damage, roughness, edge damage, potholes, loss of gravel, and corrugation. For a large share of the provincial road stretches in the 2000 map, road condition is missing. However, the other road condition variables are usually available. We classify roads into good, fair, and poor where those categories are unknown, using a random forest algorithm with a tree size of 500. The model works very well: based on the provincial roads of 2000 with known condition, the out-of-bag classification error is 3.3%.

with key informants from DoW and other organizations involved in the construction of road data lead us to believe that the higher density of roads depicted in the 2009 map is a result of an improvement of the information contained in the map, rather than the construction of new roads. Between 2000 and 2010, the focus of road works in PNG was on maintenance and upgrading, but reportedly very few new roads were constructed during this period (World Food Programme and Logistics Cluster 2011).

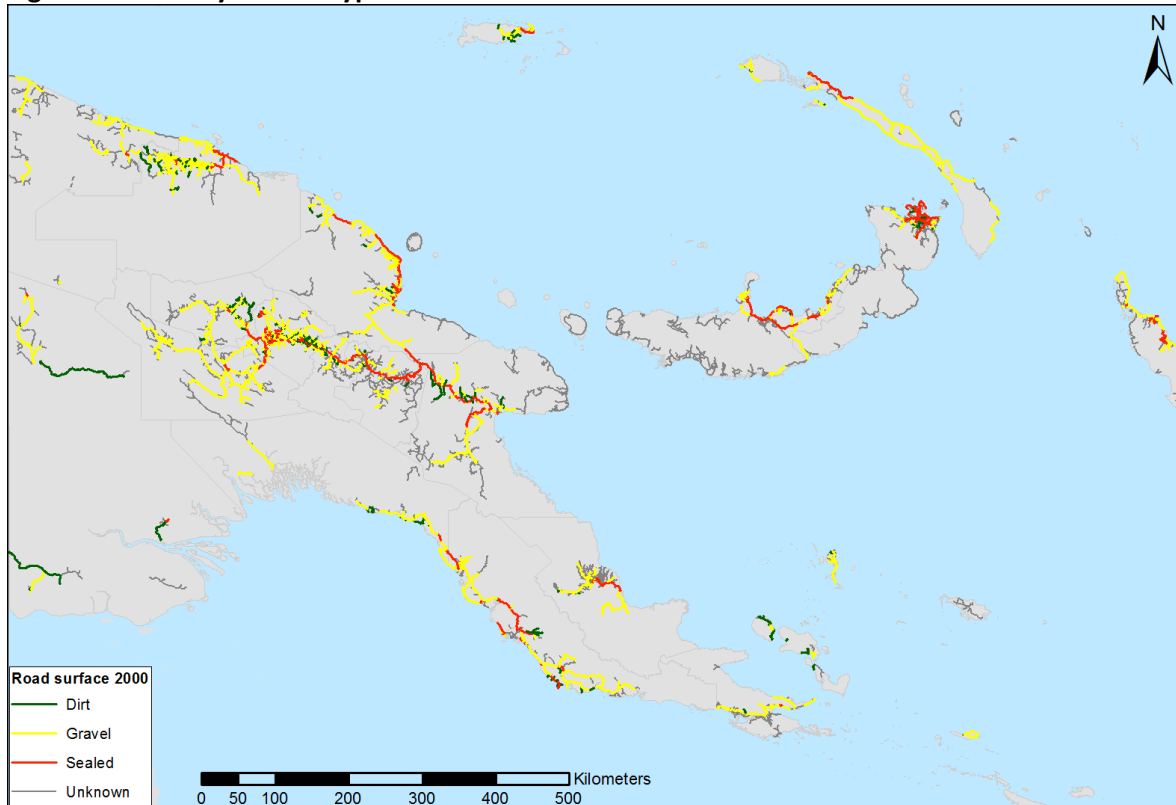
The two road datasets offer slightly different spatial representations of common road segments, with positional differences ranging up to several hundred meters. To ensure that our analysis is not influenced by differences in the coverage and spatial representation of the road network across the two years, we include the information on surface type and road condition of 2000 in the 2009 road map. The matching of roads was done based on road section IDs, and where those were lacking, on spatial proximity.

Figures 1 and 2 map the roads, distinguished by surface type, comprising identified stretches of the national network in 2000 and 2009. Table 1 tabulates the road lengths and conditions, as depicted on the maps. Comparison of the maps across the two periods makes it clear that most of the missing (unknown) road segments in 2000 are classified as dirt roads in 2009. This is in part due to the fact that the additional roads on the 2009 map are made up almost entirely of provincial roads, which were more likely than national roads to have a dirt surface. Tables 2 and 3 show the transitions in surface type and condition between the two years for those segments that are included in both maps. The tables reveal no consistent trend in development. Considering change in surface type, we observe that the length of roads upgraded (i.e., changes from gravel to sealed surface) was roughly offset by roads that deteriorated (i.e., gravel to dirt). The characterization of road condition captured in the 2000 and 2009 maps shows substantial improvement (most notably, road condition improving from poor to fair) alongside decline (mainly from good to fair).

Table 1. Extent and surface type/condition of the main PNG road network in 2000 and 2009

Surface	Condition	Roads in 2000 Length (km)	Share	Roads in 2009 Length (km)	Share
Sealed	Good	911	7.8%	1,799	7.0%
Sealed	Fair	914	7.8%	1,067	4.2%
Sealed	Poor	314	2.7%	371	1.5%
Gravel	Good	2,137	18.3%	1,096	4.3%
Gravel	Fair	1,649	14.1%	7,300	28.6%
Gravel	Poor	4,232	36.3%	5,726	22.4%
Dirt	Good	223	1.9%	166	0.7%
Dirt	Fair	63	0.5%	3,660	14.3%
Dirt	Poor	1,230	10.5%	4,332	17.0%
All	All	11,672	100%	25,517	100%

Figure 1. Roads by surface type in 2000



Using data from the road maps and household surveys, we construct variables indicating the length, surface type and condition of the road leading to the nearest town for households included in the sample. We consider the shortest route from the stretch of road that is closest to the household to the nearest town. We calculate the shares of this route by surface type (i.e., sealed, gravel, and dirt) and by the condition of the road (i.e., good, fair, and poor). For our analysis, we focus on households that are connected to a town by a road—excluding households that are located more than 15 km from any road. Further, we consider only towns that had more than 1,000 inhabitants according to the census 2011 and are within 5 km of the nearest road.⁸ Taken together, these restrictions mean that about 20% of the clusters of the two surveys are dropped in our analysis. The census units left out are mostly either on small islands (where roads may be of little or no importance anyway), or in the deep interior in the west of the Momase region or at the coast of Western province (both of which have a very low population density).

Figure 2. Roads by surface type in 2009

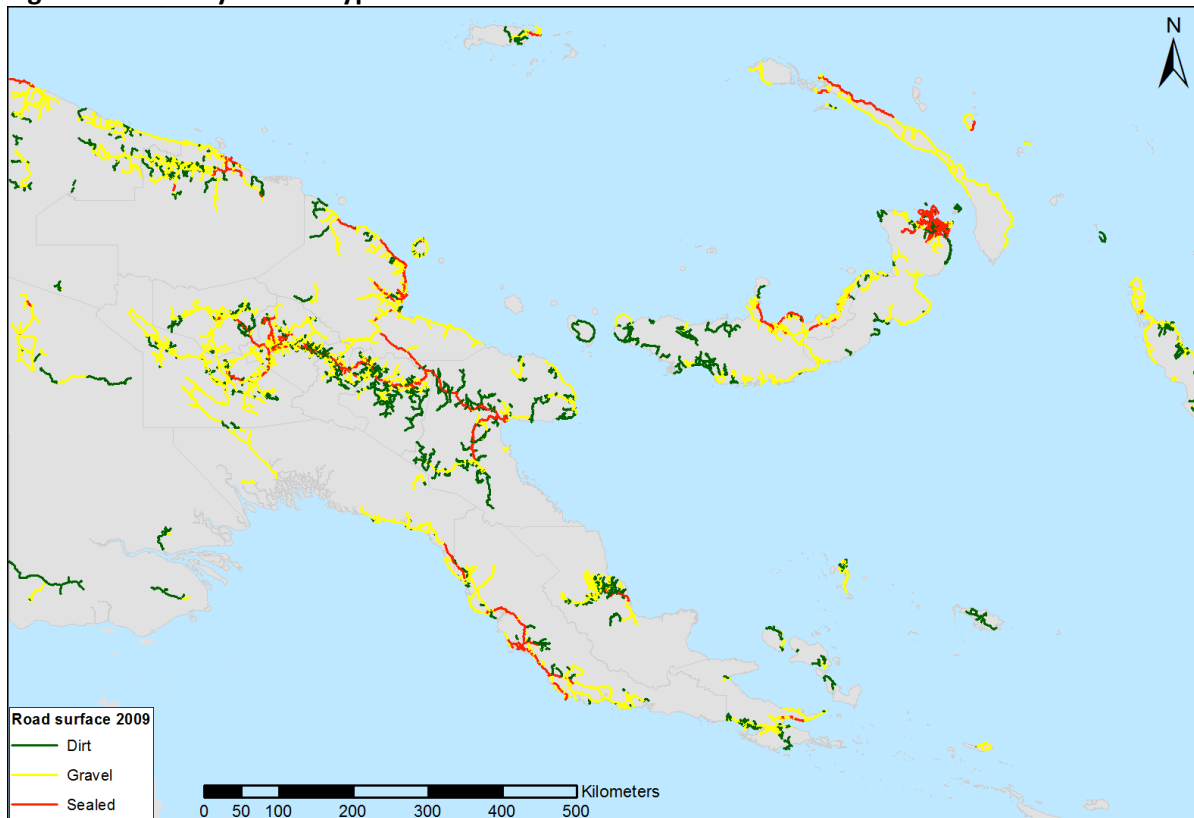


Table 2. Transition matrix comparing road segment surface types in 2000 and 2009

	Sealed '09	Gravel '09	Dirt '09	Total
Sealed '00	1,821	226	93	2,139
Gravel '00	683	6,502	832	8,017
Dirt '00	27	304	1,185	1,516
Total	2,531	7,031	2,110	11,672

Notes: Reported in kilometers, decreasing quality highlighted with *italics*.
 Statistics only listed for stretches where surface type is known in 2000.

Table 3. Transition matrix comparing road segment conditions in 2000 and 2009

	Good '09	Fair '09	Poor '09	Total
Good '00	1,077	1,531	662	3,270
Fair '00	925	970	731	2,626
Poor '00	457	2,994	2,326	5,776
Total	2,458	5,495	3,719	11,672

Notes: Reported in kilometers, decreasing quality highlighted with *italics*.
 Statistics only listed for stretches where road condition is known in 2000.

Table 4. Summary statistics for the analysis sample

	PNGHS 1996			HIES 2009/10		
	Mean	Std.Err.	N	Mean	Std.Err.	N
<i>Outcomes</i>						
Log real per adult-equivalent expenditure	0.309	0.050	680	0.157	0.050	1,477
Poverty	0.347	0.030	680	0.428	0.026	1,477
Home has a good roof	0.244	0.023	680	0.286	0.024	1,647
Engaged in subsistence farming	0.889	0.027	680	0.914	0.010	1,647
Someone in the household has wage job	0.264	0.032	680	0.155	0.022	1,636
Ratio of children age 7-17 going to school	0.445	0.029	484	0.590	0.027	1,056
<i>Location-specific control variables</i>						
Altitude (m)	1002	100.2	680	1183	50.52	1,647
Dummy, slope > 10 degrees	0.634	0.067	680	0.662	0.047	1,647
Dummy, land inundation occurs	0.236	0.058	680	0.242	0.043	1,647
Dummy, rainfall deficit is rare	0.290	0.064	680	0.321	0.049	1,647
Annual rainfall (m)	2.640	0.087	680	2.649	0.071	1,647
Log population of nearest town	8.615	0.172	680	8.802	0.127	1,647
<i>Household control variables</i>						
Household size	5.839	0.184	680	4.910	0.086	1,647
Ratio of household members under age 15	0.406	0.013	680	0.352	0.007	1,647
Ratio of household members above age 50	0.104	0.012	680	0.127	0.008	1,647
Age of household head	40.75	0.640	680	42.49	0.479	1,647
Household head is female	0.094	0.015	680	0.147	0.013	1,647
Household head is literate	0.525	0.038	680	0.474	0.021	1,646
Average years of schooling	3.392	0.198	680	4.373	0.188	1,633
Ratio of school children in primary school age	0.602	0.022	484	0.584	0.015	1,056
<i>Impact variables</i>						
Distance to nearest road (km)	0.609	0.196	680	0.906	0.177	1,647
Distance on road to nearest town (km)	49.99	6.526	680	36.50	3.407	1,647
Road to town: share of sealed road	0.351	0.048	680	0.403	0.038	1,647
Road to town: share of gravel road	0.538	0.047	680	0.475	0.042	1,647
Road to town: share of dirt road	0.111	0.026	680	0.122	0.022	1,647
Road to town: share of good sealed road	0.145	0.032	680	0.224	0.029	1,647
Road to town: share of fair sealed road	0.139	0.031	680	0.142	0.023	1,647
Road to town: share of poor sealed road	0.067	0.026	680	0.037	0.011	1,647
Road to town: share of good gravel road	0.131	0.027	680	0.088	0.023	1,647
Road to town: share of fair gravel road	0.184	0.030	680	0.293	0.037	1,647
Road to town: share of poor gravel road	0.222	0.039	680	0.094	0.018	1,647
Road to town: share of good dirt road	0.001	0.001	680	0	-	1,647
Road to town: share of fair dirt road	0.058	0.019	680	0.088	0.020	1,647
Road to town: share of poor dirt road	0.052	0.019	680	0.034	0.010	1,647

Means of log real per adult-equivalent expenditure and poverty are obtained using person sampling weights. All other means are obtained using household sampling weights.

Table 4 provides summary statistics for the sample of rural households that is used in our estimations. This includes geographic and road access variables that were merged with the household survey data. Both household surveys include sections that allow calculation of per adult-equivalent household expenditure, either based on a closed recall method—as in the PNGHS 96—or consumption

diaries—as in the HIES 09/10, and regional poverty lines based on the cost of locally consumed foods.¹⁰ Across the two households surveys used in this investigation, average per adult-equivalent consumption decreased and poverty incidence increased slightly. School enrollment has increased, particular for secondary schools, and average years of schooling of adults has increased by almost one year. Households in the 2009/10 HIES were slightly older (higher average age of household members) and smaller (had nearly one person less, on average) compared to the households in the 1996 survey. It is also noteworthy that the routes taken by the sampled households to the nearest town are better than the average PNG road (see Table 1), signified by the higher shares of sealed and gravel roads. Particularly the share of dirt roads in our sample is fairly low with 11-12%.

4. Estimation

To estimate the causal relationship between the state of road infrastructure and material well-being, we propose a linear model of the form

$$y_{ijt} = \beta R_{jt} + \gamma D_j + \delta X_{ijt} + \mu_j + \tau_t + \varepsilon_{ijt}, \quad (1)$$

where y_{ijt} is a welfare measure of household i in census unit j at time t . R_{jt} and D_j are vectors of variables related to road infrastructure, X_{ijt} is a vector of exogenous control variables (at the household- or the census unit level, some varying over time and others time-invariant). μ_j denotes unobserved, time-invariant heterogeneity at the province- and census unit-level, respectively, τ_t is a time-fixed effect, and ε_{ijt} is an independent disturbance term.

Our first outcome variable y_{ijt} is the logarithm of real yearly consumption per adult-equivalent.¹¹ We divide yearly consumption by the respective regional rural poverty lines to calculate real consumption. A similar outcome variable is poverty status. We include it to examine specifically how the probability of being poor is affected by road infrastructure. Another indicator of well-being is housing quality.¹² For a lack of credible and intertemporally comparable housing value estimates, we chose to use a dummy defining whether or not the house has a good roof (i.e., a roof made of metal, tiles, or

¹⁰ We construct per capita expenditure as well as regional poverty lines as explained in Gibson and Rozelle (1998) and Gibson (2012). Particularly, we use the revised consumption figures, poverty lines, and sampling weights for the PNGHS 96 explained in Gibson (2012) to make expenditure and poverty comparable between the two surveys. For the HIES 09/10, Gibson (2012) suggests three different consumption figures. Due to evidence of diary fatigue, we use the figure based on the shortest time horizon (7 days). The poverty lines we use take the cost of a locally consumed food basket and add the non-food spending of households whose food expenditures exactly meet this cost (Lanjouw and Lanjouw 2001).

¹¹ Like Gibson and Rozelle (2003), we assign children aged between 0 and 6 years a weight of 0.5, while children older than 6 years as well as adults are assigned a weight of 1.

¹² Measures of housing quality were left out in the construction of the consumption figures (Gibson 2012).

cement). We expect that improved road access should increase income and reduce poverty, and with higher income and reduced transportation costs, improved access to roofing materials will lead people to improve their home.

We are also interested in the effects of infrastructure on rural employment and structural transformation. One common change observed among rural households as a result of improved access to markets is reduced dependence on subsistence farming. We create a dummy to indicate whether members of the household reported engaging in subsistence farming in the days prior to the survey date.¹³ Another important transformation may involve transitioning from informal employment to having a formal wage job. Accordingly, we use a dummy for whether a member of the household is formally employed as another outcome variable. We expect that better roads improve off-farm earning opportunities and therefore reduce the necessity for subsistence agriculture and increase likelihood of formal employment. Lastly, we examine the school-enrollment ratio of children in school age. Our hypothesis is that school enrollment increases with better infrastructure due to easier access to schools for both children and teachers.

We include two types of road infrastructure variables. D_j includes the logarithm of the Euclidean distance to the nearest road and the logarithm of the distance on that road leading to the closest town with a 2010 population above 1,000 people. Since we assume that no new roads were added between 1996 and 2010, the distance variables are time-invariant by construction. R_{jt} captures the shares of different types of road on the route between surveyed households and the nearest town. Since the shares add up to one for each route, the lowest quality category of road type is the left out variable in the estimation. This means that the coefficients β give the welfare impact of a relative change from the lowest quality road type to the other types. Road segments are upgraded, left to deteriorate, or remain the same type over time, so R_{jt} is time-varying.

Since there is no overlap in census units between the two surveys, we cannot difference out the term μ_j , and treating it as a random effect uncorrelated with all independent variables might lead to a biased estimate of β . Instead, we use the correlated random effects approach introduced by

¹³ The definitions vary slightly between survey rounds. For the PNGHS 96, the variable indicates that in the two weeks prior to the survey, at least one household member engaged in the production of sago, bananas, corn, sweet potato, cassava, taro, or other fresh fruits or vegetables without selling them. For the HIES 09/10, the variable indicates that in the week prior to the survey, at least one household member engaged in agricultural production for own consumption. The means of both variables are very close, as shown in Table 4.

Chamberlain (1982, 1984). Let Z_{jt} be the vector of potentially endogenous variables in equation (1). When this is available for both years, we can substitute μ_j with its linear projection on $Z_{j,00}$ and $Z_{j,09}$, and an independent random census unit-effect, ω_j :

$$\mu_j = \alpha_{00}Z_{j,00} + \alpha_{09}Z_{j,09} + \omega_j. \quad (2)$$

This way, we make Z_{jt} uncorrelated with the combined disturbance term, $\omega_j + \varepsilon_{ijt}$. The road type share variables R_{jt} are available for both years, so we include them in equation (2). Since distance variables D_j are time-invariant, they cannot be included in the model of μ_j and are assumed conditionally exogenous. Note that this assumption does not take account of migration in response to changes in road quality.¹⁴

We include a number of location-specific control variables in equation (1). Among these are the geoclimatic variables used in Gibson and Rozelle (2003)—namely, altitude (in meters), a dummy for whether the slope is above 10 degrees, a dummy indicating that the land is subject to flooding, a dummy indicating that rainfall deficits are rare, and annual rainfall (in meters). To control for the economic importance of the nearest town, we also include the logarithm of its population as measured in the census closest to the survey year.¹⁵ Population numbers could be endogenous, say, because more productive areas lead to more household welfare as well as faster population growth. To account for this potential source of endogeneity, we also include population figures, which are available for both years, in Z_{jt} in equation (2).

We also include some household-level controls. This demands caution, however, since changes in road access could alter household characteristics. For instance, the sector of work could change as a result of new opportunities created by changes in road access. This limits the choice of household variables we can consider exogenous in our model. We select a parsimonious set of variables that describes the composition and education level of adults in the household (see table 4). We report results with and without these households controls. We always include province dummies to deal with endogeneity of roads due to provincial differences in spatial patterns and the ability and political will to build, maintain, and upgrade roads, in all specifications. Last of all, we add a time-trend τ_t to account for overall shifts in the distribution of outputs.

¹⁴ For instance, if households responded to better roads by moving closer to them, a part of the effect of this change would be (falsely) attributed to a lower distance to the nearest road, and the estimates of road quality β would be biased downwards.

¹⁵ We approximate the population size for 1996 from the 2000 census, and the population size for 2009/10 with the 2010 census figure.

With regard to the shares of road types, we explore specifications with different levels of detail. A simple way to capture road quality is to consider only the surface type, i.e., whether a road is sealed, graveled, or a dirt track. A more detailed classification of road segments includes the surface type and the road condition, i.e., whether a road is in good, fair, or poor shape. The latter leaves us with nine categories, some of which are empty or have very few observations. To prevent share variables with almost no variation, we lump together road categories that accounted for less than 5% of total length. Accordingly, we combine fair and poor sealed roads into one category, and combine good, fair, and poor dirt roads into another category.

Our model specification rests on the assumption that all unobserved factors that are both contributing to the respective outcomes and correlated with road infrastructure are location-specific and fixed over the period between the two surveys. This may look like a daring presumption given that this period is 13 years long. Some areas may have gained or lost in population or economic importance in those years, potentially affecting infrastructure as well as household welfare. But we hope to capture these changes by including town population numbers from the two censuses in our model. In addition, rural economic output and poverty have stagnated over the study period, and there is no clear indication of structural change that might have significantly altered unobserved location-specific factors on a large scale. For these reasons we believe that assuming time-invariant cluster-level heterogeneity is reasonable in this case. Due to the fact that our road map of 2000 is less detailed than the 2009 map, we have missing information on surface type and condition for some of the road segments of 2000. For the routes used in the analysis, this information was lacking for 24% of the total distance. For that reason, we first drop from the estimation all observations where *all* road segments leading to the nearest town have unknown characteristics for 2000. This leaves a total of 20% of the total distance unknown for 2000. For the remaining observations, we simply assign the segments with unknown characteristics for 2000 the same characteristics as for 2009.¹⁶ Due to the lack of time-variation in the segments unknown for 2000 used in the correlated random effects correction, we only include the shares of those segments with characteristics known in both years, and the share of unknown segments as a control.

¹⁶ To check whether this decision is robust to changes in handling the missing data, we also did the analysis using imputations of the characteristics of the segments unknown in 2000 based on the changes in road types for the known parts. All results practically stay the same. The details and results of this exercise are available on request.

Our estimations of the model above are carried out using Ordinary Least Squares. Wooldridge (2010) recommends a generalized method of moments (GMM) approach to estimate correlated random effects models efficiently. However, since the set of variables included in the projection of μ_j is very small (containing only road type shares and the log population of the nearest town), GMM estimation leads to no improvements over OLS. All regressions are weighted using the sampling weights included in both surveys. Standard errors are adjusted for census unit-level clustering.

In addition, we are also interested in whether rural roads affect all households in the same manner. One open question, for example, is whether high education levels are complementary to road infrastructure, or whether it is mostly low-skilled labor that becomes more productive through better roads. Other sources of effect heterogeneity might be the gender and age of household members. For example, additional opportunities created by roads may help empower women and thereby have a larger effect on their welfare. Poor quality or lacking road infrastructure may trap older people and diminish their prospects more than those of young people due to physical constraints on walking long distances or transiting rough roads. To investigate whether this type of impact heterogeneity exists, we follow Dercon et al. (2009) and divide our sample in two subsamples to estimate the model separately by subsample. We define the subsamples on the basis of: (i) whether the average years of education for household members over age 20 is larger than 4 years, (ii) whether the household head is literate, (iii) whether the household head is female, and (iv) whether the ratio of household members above 50 exceeds 30%.

Last of all, we explore effect heterogeneity across the distribution of real consumption, which the impact estimates on poverty status do not capture. Estimates of road impacts on poverty show how roads affect the probability of being below or above the poverty line threshold, and ignore effects on people who stay poor or stay non-poor. Moreover, the effect of roads on poverty status depends on the shape of the initial consumption distribution, in particular the mass of people around the poverty line. However, we are also interested in how relatively poor households are affected by infrastructure compared to relatively rich households. To this end, we employ quantile regressions.

The principal idea of quantile regression is that, unlike in the linear regression framework, it is not the conditional expectation but a conditional quantile of the outcome that is a linear function of the covariates. We denote the τ^{th} quantile of consumption y_{ijt} conditional on road quality R_{jt} as $q(R_{jt}, \tau)$.

This is referred to as the structural quantile function. Each possible outcome can be related to this function:

$$y_{ijt} = q(R_{jt}, U_{ijt}^*), \quad U_{ijt}^* \sim \mathcal{U}(0,1), \quad (3)$$

where U_{ijt}^* is a non-separable disturbance term normalized to a standard uniform distribution. U_{ijt}^* determines the rank of the outcome within the conditional distribution, and is what causes heterogeneity in outcomes conditional on R_{jt} . U_{ijt}^* is also referred to as *prone*ness to score high in the conditional outcome distribution. If prone

ness is independent of road quality, this model can be identified and estimated via the moment conditions

$$E_y \left\{ R_{jt} \left[\tau - \mathbf{1}(y_{ijt} \leq q(R_{jt}, \tau)) \right] \right\} = 0. \quad (4)$$

Khandker et al. (2009) and Khandker and Koolwal (2010) use quantile regressions in their studies of road infrastructure. Similarly to our estimation above, they include correlated random effect terms as covariates in their quantile regressions to account for unobserved heterogeneity. However, a problem with their approach is that including control variables in a quantile regression model alters its interpretation. The τ^{th} consumption quantile for households with good roads is not the same as the τ^{th} quantile for households with good roads *and* low levels of education. Using a similar reasoning, including the correlated random effect terms—and thereby, implicitly, an approximation to the location-specific fixed effect—yields an interpretation different from the one associated with model (3). On the other hand, not conditioning on controls and the correlated random effect terms is likely to create biased estimates. It requires the assumption that U_{ijt}^* is independent of R_{jt} , which is quite strong, and it is likely to hold only conditionally on covariates.

To circumvent this problem, we apply a generalized quantile regression (GQR) model, as introduced by Powell (2016). Here, outcomes are modelled by the same structural quantile function as in (3). However, in GQR, the prone

ness term is itself dependent on covariates:

$$U_{ijt}^* = \lambda(C_{ijt}, U_{ijt}). \quad (5)$$

C_{ijt} is a vector of variables that are believed to influence the rank of y_{ijt} conditional on R_{jt} , and is called observed prone

ness. U_{ijt} is an error term. Identification now relies on an unconditional quantile restriction,

$$E_y \left[\tau - \mathbf{1}(y_{ijt} \leq q(R_{jt}, \tau)) \right] = 0, \quad (6)$$

and conditional quantile restrictions,

$$E_y \left\{ R_{jt} \left[\tau - \mathbf{1} \left(y_{ijt} \leq q \left(R_{jt}, \tau_{C_{ijt}} \right) \right) \right] \right\} = 0, \quad (7)$$

with the conditional quantiles $\tau_{C_{ijpt}} := P(y_{ijt} \leq q(R_{jt}, \tau) | C_{ijt})$.

While GQR leaves the structural quantile function (and therefore the interpretation of the conditional outcome distribution) unaltered, the restrictions (7) allow covariates C_{ijt} to update this distribution. This way, including covariates in the model does not alter the causal interpretation of the quantile impact estimates of road quality.

For our model, we focus on one of the shares of surface type in R_{jt} at a time, while the respective other shares as well as the time-invariant (log) distances to the nearest road and to the nearest town, D_j , are included in the set of observed proneness variables C_{ijt} . The main reason is that the more variables are conditioned on, the lower is the variance of the conditional outcome distributions, and the lower is the difference between quantile effects. So, if we considered the outcome distribution conditional on both R_{jt} and D_j , we would not expect to see much effect heterogeneity. As further proneness variables we include everything we controlled for in our most detailed specification above, which includes location- and household-specific control variables, correlated random effect terms, province dummies, and a time dummy. For estimation, we use the Stata routine *genqreg*, finding regression coefficients via a grid search. To obtain confidence intervals we apply a cluster bootstrap on the census unit-level, based on regional strata.¹⁷

5. Results

Our main estimation results are summarized in Tables 5 to 7. The columns show the results of a regression of different outcome variables on the logarithm of the distance to the nearest road, the logarithm of the distance to nearest town on that road, and the variables for the surface type of the route to the nearest town. For each outcome variable, we estimate three models. Model 1 is a linear regression of the model specified in equation 1, without the correlated random effect terms Z_{jt} and the household level controls. Specification 2 includes the correlated random effect terms. Specification 3 contains the correlated random effect terms as well as household-level control variables. As we choose

¹⁷ We compute bootstrap-*t* confidence intervals, as suggested by Efron (1979): For each quantile of interest, we make 999 draws of census units with replacement from each stratum. For each bootstrap sample, we reestimate the regression coefficient and draw another 25 bootstrap samples to get an estimate for the respective standard error. In the end we obtain the $(1 - \alpha)$ confidence interval $(\hat{\beta} - q_{1-\alpha/2} \hat{\sigma}, \hat{\beta} + q_{\alpha/2} \hat{\sigma})$, with coefficient estimate $\hat{\beta}$ and its standard error $\hat{\sigma}$, and where q_α denotes the α quantile of the bootstrap *t*-distribution.

a parsimonious set of household controls that leads to a considerable improvement in fit, this last specification is our preferred one. For the analyses by subgroups and the quantile regressions we also use the covariates of this specification.

Table 5. Impact of road type and distances on consumption and poverty status

	Log(real p.a.e. consumption)			Poverty Status		
	(1)	(2)	(3)	(1)	(2)	(3)
Dirt to Sealed	0.1935 (0.151)	0.4091** (0.188)	0.5516*** (0.201)	-0.0953 (0.093)	-0.0910 (0.122)	-0.1670 (0.131)
Dirt to Gravel	-0.0917 (0.175)	0.1028 (0.159)	0.1909 (0.161)	0.0777 (0.098)	0.0429 (0.109)	-0.0049 (0.112)
Gravel to Sealed†	0.2852** (0.115)	0.3063* (0.176)	0.3607** (0.165)	-0.1731*** (0.065)	-0.1340 (0.108)	-0.1621 (0.105)
Log total distance to nearest town	-0.1347*** (0.029)	-0.1482*** (0.030)	-0.1341*** (0.029)	0.0767*** (0.016)	0.0819*** (0.016)	0.0753*** (0.016)
Log distance to nearest road	0.0070 (0.017)	0.0090 (0.015)	0.0114 (0.015)	-0.0065 (0.010)	-0.0078 (0.009)	-0.0087 (0.009)
CRE terms included	No	Yes	Yes	No	Yes	Yes
Household Controls	No	No	Yes	No	No	Yes
R-squared	0.148	0.169	0.296	0.116	0.126	0.204
Census Units	155	155	155	155	155	155
Households	2153	2153	2148	2153	2153	2148
p-value CRE		0.001	0.000		0.039	0.018

Note: * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses, clustered at the census unit level. All regressions are weighted using person sampling weights from both surveys. Dirt road is excluded category for the share variables. Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications also include location-specific control variables (see Table 4) as well as province- and time-fixed effects. Columns labelled (2) in addition include the correlated random effect (CRE) terms, and columns labelled (3) include CRE terms and household-level control variables (see Table 4). The last row displays the p-value of a Wald test of the joint significance of the CRE terms.

†The coefficient for “Gravel to Sealed” is not actually part of the model and comes about by subtracting “Dirt to Gravel” from “Dirt to Sealed”.

Table 5 shows the impact of road condition variables on the logarithm of real per adult-equivalent consumption and the poverty status. The results indicate improved road conditions have a positive impact on consumption. An upgrade from dirt to sealed of one percent of the route to the nearest town leads to a 0.55% increase in consumption per household member. Upgrading one percent of gravel road to sealed road increases consumption by 0.36%. The difference between dirt and gravel roads is positive, but not significantly different from zero. An increase in the total length of the route by one percent is associated with a 0.13% decrease in consumption per household member. The estimates of the effect of quality roads on poverty in Table 5 are positive (i.e., lead to lower poverty) but not significantly different from zero.

Table 6. Impact of road type and distances on subsistence farming and wage employment

	Household does subsistence farming			Someone in the household has wage job		
	(1)	(2)	(3)	(1)	(2)	(3)
Dirt to Sealed	0.0373 (0.070)	-0.0992 (0.094)	-0.1409 (0.091)	0.1105 (0.080)	0.1908 (0.145)	0.1653 (0.120)
Dirt to Gravel	0.0596 (0.065)	0.0419 (0.090)	0.0142 (0.082)	0.1093 (0.092)	0.0878 (0.128)	0.0952 (0.107)
Gravel to Sealed†	-0.0223 (0.037)	-0.1411** (0.054)	-0.1551*** (0.057)	0.0011 (0.079)	0.1030 (0.113)	0.0701 (0.097)
Log total distance to nearest town	0.0172* (0.010)	0.0138 (0.010)	0.0087 (0.009)	-0.0461** (0.020)	-0.0462** (0.022)	-0.0359* (0.020)
Log distance to nearest road	-0.0059 (0.005)	-0.0035 (0.005)	-0.0030 (0.004)	-0.0170* (0.009)	-0.0190** (0.009)	-0.0137 (0.008)
CRE terms included	No	Yes	Yes	No	Yes	Yes
Household Controls	No	No	Yes	No	No	Yes
R-squared	0.136	0.152	0.193	0.102	0.115	0.190
Census Units	155	155	155	155	155	155
Households	2323	2323	2312	2312	2312	2306
p-value CRE		0.008	0.002		0.129	0.154

Note: * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses, clustered at the census unit level. All regressions are weighted using household sampling weights from both surveys. Dirt road is excluded category for the share variables. Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications also include location-specific control variables (see Table 4) as well as province- and time-fixed effects. Columns labelled (2) in addition include the correlated random effect (CRE) terms, and columns labelled (3) include CRE terms and household-level control variables (see Table 4). The last row displays the p-value of a Wald test of the joint significance of the CRE terms.

†The coefficient for “Gravel to Sealed” is not actually part of the model and comes about by subtracting “Dirt to Gravel” from “Dirt to Sealed”.

Table 6 presents estimates of the impact of road type on the likelihood of engagement in subsistence farming and in wage employment. The point estimates of the effects on these two outcomes indicate that better quality roads facilitate the structural transformation from subsistence farming to economic activities that are more market integrated. A one percentage point increase in the share of sealed roads reduces the probability that a member of the household engages in subsistence farming by around 0.15 percentage points. For wage employment, the magnitude of the point estimate from dirt to gravel is similar, but not statistically significant.

Table 7 shows the effects on the likelihood of having a good roof and of school enrollment of children between 7 and 17. We find clear signs that improvements in roads lead to investment in housing. A one percentage point increase in sealed versus dirt roads increases the probability of having a good roof by 0.19 percentage points. Unsurprisingly—given the high transportation costs of tiles or corrugated sheet metal—the distance to the nearest road is significantly negatively correlated with good roofs.

Table 7. Impact of road type and distances on having a good roof and school enrollment

	Home has a good roof			Ratio of school children going to school		
	(1)	(2)	(3)	(1)	(2)	(3)
Dirt to Sealed	0.2024*** (0.069)	0.1822* (0.098)	0.1872** (0.092)	-0.0014 (0.080)	-0.0156 (0.127)	0.0103 (0.114)
Dirt to Gravel	0.1259* (0.067)	0.0613 (0.090)	0.0808 (0.082)	-0.1318 (0.083)	-0.2209** (0.099)	-0.2045** (0.090)
Gravel to Sealed†	0.0766 (0.055)	0.1208 (0.077)	0.1063 (0.074)	0.1304** (0.057)	0.2053* (0.112)	0.2149** (0.094)
Log total distance to nearest town	-0.0237 (0.017)	-0.0301* (0.018)	-0.0224 (0.018)	-0.0047 (0.017)	-0.0109 (0.017)	-0.0076 (0.016)
Log distance to nearest road	-0.0226*** (0.007)	-0.0195*** (0.007)	-0.0170*** (0.006)	-0.0241*** (0.009)	-0.0217** (0.009)	-0.0172** (0.008)
CRE terms included	No	Yes	Yes	No	Yes	Yes
Household Controls	No	No	Yes	No	No	Yes
R-squared	0.441	0.450	0.469	0.146	0.153	0.206
Census Units	155	155	155	155	155	155
Households	2323	2323	2312	1538	1538	1530
p-value CRE		0.018	0.023		0.582	0.564

Note: * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses, clustered at the census unit level. All regressions are weighted using household sampling weights from both surveys. Dirt road is excluded category for the share variables. Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications also include location-specific control variables (see Table 4) as well as province- and time-fixed effects. Columns labelled (2) in addition include the correlated random effect (CRE) terms, and columns labelled (3) include CRE terms and household-level control variables (see Table 4). The last row displays the p-value of a Wald test of the joint significance of the CRE terms.

†The coefficient for “Gravel to Sealed” is not actually part of the model and comes about by subtracting “Dirt to Gravel” from “Dirt to Sealed”.

The effects of road quality on school participation show a complex pattern, suggesting that upgrades from dirt to gravel roads substantially reduce the probability of school participation, while sealing gravel roads has the opposite effect. The finding that gravel roads relative to dirt roads reduce school enrollment is puzzling. A possible explanation may be that often the schools are not located in the nearest town, so the road quality variables do not accurately capture how road quality affects access to education services. A comparison of travel time variables from the HIES 09/10 to schools compared to other services that are usually only found in urban areas is consistent with this explanation. The distance variables show no correlation of distance with subsistence farming, a negative correlation of engagement in wage work with distance to the nearest town, and a negative correlation of school enrollment.

Tables 5 to 7 also contain the p-value of an F-test for the joint significance of the correlated random effect terms (only in specifications (2) and (3)). Except for the models using wage job and school

enrollment as outcomes, these p-values are below 0.05. This indicates that time-invariant, correlated random effects need to be accounted for.

We also explore a model that considers both surface type and road condition in the estimation. But even after reducing the number of categories to only six (as described above), the lower precision of estimates compared to having only three categories is high. The results are summarized in Appendix Tables 1 to 3. Conditional on surface type, we find no significant differences in effects by road condition. This can possibly be ascribed to the fact that road condition may have been assessed differently across provinces and time. Because the results are consistent with the simpler model and offer no new insights, we focus on the categorization by surface type only.

Next, we discuss the estimates disaggregated by subgroups of households to study heterogeneous effects of road quality. We only consider the models of consumption and poverty, since these are the key outcome variables of this paper. The results are reported in Tables 8 and 9, with the subsample defined in the top row. We find some striking differences when effects are estimated for households distinguished by education and gender. Households that have on average less than 4 years of education or where the household head is illiterate experience large and significant improvements in household consumption as a result of a greater share of sealed roads, and significantly more so than households with at least 4 years of education on average. The same effects are also observable in the estimates on poverty status. We also observe that female-headed households benefit more from road sealing in terms of consumption than male-headed households. In the regressions by households with more than 30% members above the age of 50, the differences in the effects of road quality are less pronounced. The only significant difference in effects is that the switch from dirt to gravel has a higher effect on lowering poverty for the households with many old people.

Last of all, we present the results of the generalized quantile regressions examining the effect of road quality on the log of consumption. Table 10 shows the results exemplary for the 10th, 30th, 50th, 70th, and 90th quantile. We report the confidence intervals from our cluster bootstrap. Figure 3 shows the marginal effects of upgrading from dirt to gravel and from gravel to sealed road by quantile. Note that all the control variables from the prior regression models are now included as proneness variables.

Table 8. Impact of road type and distances on log real per adult-equivalent consumption by subgroups

	Average years of schooling ≤ 4	Average years of schooling > 4	Household head illiterate	Household head literate	Household head male	Household head female	Household members above 50 ≤ 30%	Household members above 50 > 30%
Dirt to Sealed	0.7081*** (0.224)	0.0992 (0.214)	0.6732** (0.264)	0.3244* (0.192)	0.5004** (0.202)	0.9986** (0.395)	0.5371*** (0.203)	0.7231 (0.449)
Dirt to Gravel	0.1558 (0.163)	0.3296 (0.220)	0.0908 (0.184)	0.2563 (0.203)	0.2287 (0.168)	-0.4195 (0.318)	0.1880 (0.159)	0.5507 (0.411)
Gravel to Sealed†	0.5524*** (0.185)	-0.2304 (0.172)	0.5824** (0.246)	0.0681 (0.155)	0.2717* (0.154)	1.4182*** (0.381)	0.3491** (0.168)	0.1724 (0.424)
Log distance to nearest town	-0.1236*** (0.035)	-0.1812*** (0.031)	-0.0863** (0.036)	-0.1693*** (0.029)	-0.1393*** (0.028)	-0.0626 (0.051)	-0.1348*** (0.029)	-0.1185** (0.051)
Log distance to nearest road	0.0150 (0.018)	-0.0072 (0.016)	0.0152 (0.018)	-0.0031 (0.016)	0.0117 (0.015)	0.0035 (0.027)	0.0071 (0.015)	0.0422 (0.033)
R-squared	0.330	0.296	0.324	0.296	0.309	0.398	0.312	0.367
Census Units	146	144	148	151	154	106	154	123
Households	1249	899	1041	1107	1873	275	1836	312

Note: * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses, clustered at the census unit level. All regressions are weighted using person sampling weights from both surveys. Dirt road is excluded category for the share variables. Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications include location- and household-specific control variables (see Table 4), correlated random effect (CRE) terms, province- and time-fixed effects.

†The coefficient for “Gravel to Sealed” is not actually part of the model and comes about by subtracting “Dirt to Gravel” from “Dirt to Sealed”.

Table 9. Impact of road type and distances on poverty status by subgroups

	Average years of schooling ≤ 4	Average years of schooling > 4	Household head illiterate	Household head literate	Household head male	Household head female	Household members above 50 ≤ 30%	Household members above 50 > 30%
Dirt to Sealed	-0.2415 (0.161)	0.1219 (0.131)	-0.1222 (0.211)	-0.0893 (0.124)	-0.1914 (0.124)	-0.1346 (0.240)	-0.1084 (0.139)	-0.4431* (0.257)
Dirt to Gravel	0.0430 (0.123)	-0.0532 (0.138)	0.0725 (0.144)	-0.0094 (0.136)	-0.0526 (0.109)	0.3771* (0.206)	0.0326 (0.117)	-0.4613** (0.210)
Gravel to Sealed†	-0.2846** (0.136)	0.1752 (0.128)	-0.1946 (0.210)	-0.0799 (0.100)	-0.1388 (0.094)	-0.5117** (0.196)	-0.1410 (0.111)	0.0182 (0.246)
Log distance to nearest town	0.0812*** (0.025)	0.0805*** (0.018)	0.0371 (0.024)	0.0907*** (0.019)	0.0753*** (0.015)	0.0475 (0.031)	0.0765*** (0.017)	0.0512 (0.033)
Log distance to nearest road	-0.0122 (0.012)	0.0079 (0.010)	-0.0161 (0.011)	0.0063 (0.010)	-0.0106 (0.009)	0.0203 (0.017)	-0.0054 (0.009)	-0.0376* (0.019)
R-squared	0.232	0.212	0.240	0.206	0.216	0.369	0.211	0.392
Census Units	146	144	148	151	154	106	154	123
Households	1249	899	1041	1107	1873	275	1836	312

Note: * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses, clustered at the census unit level. All regressions are weighted using person sampling weights from both surveys. Dirt road is excluded category for the share variables. Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications include location- and household-specific control variables (see Table 4), correlated random effect (CRE) terms, province- and time-fixed effects.

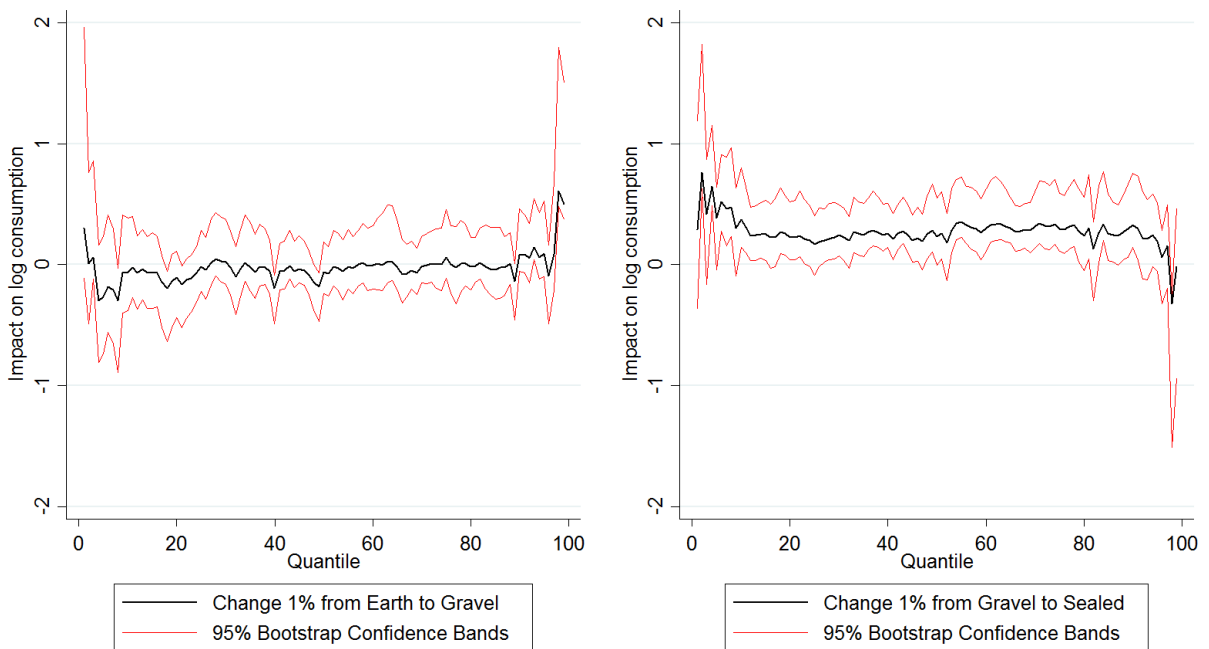
†The coefficient for “Gravel to Sealed” is not actually part of the model and comes about by subtracting “Dirt to Gravel” from “Dirt to Sealed”.

Table 10. Generalized quantile regressions of consumption on road type

Quantile	Log(real per adult equivalent consumption)				
	10%	30%	50%	70%	90%
Dirt to Gravel	-0.060 [-0.378, 0.387]	0.020 [-0.165, 0.371]	-0.061 [-0.241, 0.187]	-0.015 [-0.150, 0.234]	0.080 [-0.053, 0.462]
Gravel to Sealed	0.373*** [0.143, 0.803]	.244** [0.076, 0.497]	0.234* [-0.004, 0.548]	0.318*** [0.138, 0.602]	0.330*** [0.145, 0.754]
CUs	155	155	155	155	155
Households	2153	2153	2153	2153	2153

Note: * p<0.10, ** p<0.05, *** p<0.01. 95% bootstrap-*t* confidence intervals in parentheses, obtained using a cluster bootstrap at the census unit level with 999 replications. All regressions are weighted using person sampling weights from both surveys. All regressions use log distance to nearest town, log distance to nearest road, location- and household-specific variables (see Table 4), correlated random effect (CRE) terms, province dummies, and time dummies as proneness variables.

Figure 3. Road type coefficients by consumption quantile



Results suggest that the change from gravel to sealed road has an effect on consumption that is highest for the poorest 10%, for whom average impacts range between 0.3 and 0.76. Beyond that there is no clear indication for heterogeneous impacts along the distribution of consumption. For upgrades from a dirt road, there is no visible pattern of effect heterogeneity, with coefficients being small and insignificant for most quantiles. We interpret the results from the GQR model to indicate that sealing roads disproportionately benefits the poorest households in PNG, so if anything can be regarded as a pro-poor policy measure.

6. Conclusion

In this paper, we examined the impact of changes in the quality of PNG's road network on rural household welfare over a 13-year period. We find evidence that sealing roads leads to improvements in household welfare, and that the effects are modestly higher for disadvantaged households. Poorer households and households with lower education levels benefit relatively more from better roads—which suggests investment in upgrading of rural roads is a pro-poor policy. This complements the argument by Gibson and Rozelle (2003) that due to the sparse road network and the remoteness of many poor households in PNG, infrastructure spending may be one of the few feasible targeted antipoverty measures. One explanation for the disproportionate impact is that connecting rural households to local markets benefits smallholder farmers in particular. This is consistent with our results suggesting that sealing roads supports the structural transformation of households away from subsistence farming to more market-oriented activities. We do not find clear evidence for effects of road quality on access to education services.

The impact estimates rely on administrative road inventory data in combination with repeated household cross section data which include the geographical coordinates of surveyed households. With this data, it is possible to estimate a model that is akin to a model with village-fixed effects. The data we have at our disposal is available in many other countries at relatively low cost (compared with the costs of collecting new data), and thus the method lends itself for replication elsewhere, and again in PNG when another household survey is conducted.

References

- Aggarwal, S. 2017. "Do Rural Roads Create Pathways out of Poverty? Evidence from India." Working Paper. Available online at: http://www.isb.edu/faculty/shilpa_aggarwal/files/shilpa-aggarwal-rural-roads.pdf (last accessed on July 17, 2017).
- Banerjee, A., E. Duflo, and N. Qianx. 2012. "On the Road: Access to Transportation Infrastructure and Economic Growth in China," NBER Working Paper No. 17897 (March 2012).
- _____. 2004. "The Railroad to Success: The Effect of Access to Transportation Infrastructure on Economic Growth in China," Massachusetts Institute of Technology (MIT) Working Papers, MIT.
- Bell, C., and S. Dillen. 2012. "How does India's Rural Roads Program Affect the Grassroots? Findings from a Survey in Orissa." Policy Research Working Paper Series 6167, The World Bank.
- Binswanger, H. P., S. R. Khandker, and M. R. Rosenzweig. 1993. "How Infrastructure and Financial Institutions Affect Agricultural Output and Investment in India." *Journal of Development Economics*, 41(2), 337–66.
- Burgess, R., R. Jedwab, E. Miguel, A. Morjaria, and G.P.I. Miquel. 2015. "The Value of Democracy: Evidence from Road Building in Kenya." *American Economic Review*, 105(6), 1817–51.
- Casaburi, L., R. Glennerster, and T. Suriy. 2013. "Rural Roads and Intermediated Trade: Regression Discontinuity Evidence from Sierra Leone," Harvard University Department of Economics Working Paper.
- Chamberlain, G. 1982. "Multivariate Regression-Models for Panel Data." *Journal of Econometrics*, 18(1), 5–46.
- Chamberlain, G. 1984. "Panel data", in *Handbook of Econometrics*, Volume 2, eds. Griliches, Z. and Intriligator, M. D. Elsevier. 1247-1318.
- Dercon, S., D.O. Gilligan, J. Hoddinott, and T. Woldehanna. 2009. "The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages." *American Journal of Agricultural Economics*, 91(4), 1007–21.
- Dercon, S., J. Hoddinott, and D. Woldehanna. 2012. "Growth and Chronic Poverty: Evidence from Rural Communities in Ethiopia." *The Journal of Development Studies*, 48:2, 238-253.
- Dornan, M. 2016. "The Political Economy of Road Management Reform: Papua New Guinea's National Road Fund." *Asia and the Pacific Policy Studies*, vol. 3, no. 3, 443–457.
- Gibson, J., and S. Rozelle. 1998 (May). *Results of the Household Survey Component of the 1996 Poverty Assessment for Papua New Guinea*. Tech. rept.
- Gibson, J., and S. Rozelle. 2003. "Poverty and Access to Roads in Papua New Guinea." *Economic Development and Cultural Change*, 52(1), 159–185.
- Gibson, J. 2012. *Papua New Guinea Poverty Profile. Based on the Household Income and Expenditure Survey*. Tech. rept. Available online at: <http://www.planning.gov.pg/images/dnpm/pdf/PNG-Poverty-Profile-2012.pdf> (last accessed on December 21, 2016).
- Government of Papua New Guinea (2013). *National Budget 2013: Volume 1, Economic and Development Policies*. Department of Treasury, Port Moresby.
- Jusi, P., M. Roy, S. Jarvenpaa, B. Neausemale, and E. Sangrador. 2003. "Road Asset Management System Implementation in Pacific Region: Papua New Guinea." *Transportation Research Record: Journal of the Transportation Research Board*, 1819: 323–332.

- Khandker, S.R., Z. Bakht, and G. B. Koolwal. 2009. "The Poverty Impact of Rural Roads: Evidence from Bangladesh." *Economic Development and Cultural Change*, 57(4), 685–722.
- Khandker, S.R. and G. B. Koolwal. 2010. "How Infrastructure and Financial Institutions Affect Rural Income and Poverty: Evidence from Bangladesh." *Journal of Development Studies*, Taylor & Francis Journals, vol. 46(6), pages 1109-1137.
- Khandker, S.R. and G. B. Koolwal. 2011. "Estimating the Long-Term Impacts of Rural Roads: A Dynamic Panel Approach," *Policy Research Working Paper*. Washington, D.C. : The World Bank.
- Knox, J., Daccache A., and T. Hess. 2013. "What is the impact of infrastructural investments in roads, electricity and irrigation on agricultural productivity? (systematic review)," Collaboration for Environmental Evidence (CEE) Syntheses, *CEE review* 11-007. Available online at <http://www.environmentalevidence.org/wp-content/uploads/2014/05/CEE11-007.pdf> (last accessed December 6, 2016).
- Kwa, E., S. Howes, and S. Lin. 2010. *Review of the PNG–Australia Development Cooperation Treaty (1999)*. OECD. Available online at <https://www.oecd.org/countries/papuanewguinea/45827611.pdf> (last accessed November 25, 2016).
- Lanjouw, J. and P. Lanjouw. 2001. "How to compare apples and oranges: Poverty measurement based on different definitions of consumption." *Review of Income and Wealth* 47(1): 25-42.
- Lokshin, M. and R. Yemtsov. 2005. "Has Rural Infrastructure Rehabilitation in Georgia Helped the Poor?" *World Bank Economic Review*, 19(2), 311–33.
- Lucius, D. 2010. Civil Works Capacity Constraints (unpublished report). Asian Development Bank, Port Moresby.
- Mu, R. and D. van de Walle. 2011. "Rural Roads and Local Market Development in Vietnam." *Journal of Development Studies*, 47(5), 709–34.
- Nguyen, C. 2016. "Impacts of Rural Road on Household Welfare in Vietnam: Evidence from a Replication Study." Economics Discussion Papers, No 2016-40, Kiel Institute for the World Economy.
- PNG National Statistical Office. 2013. *PNG HIES Statistical Summary Tables* (web version). Washington, D.C. : World Bank Group. Available online at <http://documents.worldbank.org/curated/en/329781468289823696/PNG-HIES-Statistical-Summary-Tables-Web> (last accessed November 25, 2016).
- Powell, D. "Quantile Treatment Effects in the Presence of Covariates. 2016. Available online at http://works.bepress.com/david_powell/4/ (last accessed December 21, 2016).
- Secretariat of the Pacific Community. 2017. Prism online statistical database. Available online at: <https://prism.spc.int/regional-data-and-tools/population-statistics> (last accessed July 26, 2017).
- Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press Books, The MIT Press, edition 2.
- World Food Programme and Logistics Cluster. 2011 (February–March). *Papua New Guinea Emergency Preparedness: Operational Logistics Contingency Plan Part 2 - Existing Response Capacity and Overview of Logistics Situation*. Available online at <http://reliefweb.int/sites/reliefweb.int/files/resources/PNG%20Logistics%20CP%20-%20Part%202%20-%20Existing%20Response%20Capacity%20and%20Logistics%20Overview.pdf> (last accessed December 21, 2016).

Appendix

Appendix Table 1. Impact of road type on consumption and poverty status using detailed road type variables

	Log(real p.a.e. consumption)			Poverty Status		
	(1)	(2)	(3)	(1)	(2)	(3)
Sealed and good	0.1109 (0.161)	0.3729 (0.228)	0.4260* (0.230)	-0.0620 (0.102)	-0.0710 (0.145)	-0.1002 (0.151)
Sealed and not good	0.2908 (0.180)	0.4802** (0.217)	0.6203*** (0.213)	-0.1342 (0.102)	-0.1559 (0.128)	-0.2254* (0.130)
Gravel and good	-0.0390 (0.273)	0.0779 (0.232)	-0.0191 (0.215)	0.0093 (0.135)	0.0294 (0.138)	0.0816 (0.132)
Gravel and fair	-0.1017 (0.179)	0.1025 (0.166)	0.2390 (0.168)	0.0896 (0.105)	0.0333 (0.111)	-0.0349 (0.116)
Gravel and poor	-0.0958 (0.172)	0.1893 (0.188)	0.2735 (0.194)	0.0986 (0.099)	0.0297 (0.113)	-0.0231 (0.116)
Log total distance to nearest town	-0.1324*** (0.028)	-0.1278*** (0.031)	-0.1177*** (0.028)	0.0756*** (0.016)	0.0718*** (0.016)	0.0677*** (0.016)
Log distance to nearest road	0.0071 (0.017)	0.0025 (0.016)	0.0043 (0.015)	-0.0071 (0.009)	-0.0052 (0.010)	-0.0057 (0.009)
CRE terms included	No	Yes	Yes	No	Yes	Yes
Household Controls	No	No	Yes	No	No	Yes
R-squared	0.150	0.176	0.303	0.117	0.134	0.211
Census Units	155	155	155	155	155	155
Households	2153	2153	2148	2153	2153	2148
p-value CRE		0.016	0.000		0.041	0.019

Note: * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses, clustered at the census unit level. All regressions are weighted using person sampling weights from both surveys. Dirt road is excluded category for the share variables. Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications also include location-specific control variables (see Table 4) as well as province- and time-fixed effects. Columns labelled (2) in addition include the correlated random effect (CRE) terms, and columns labelled (3) include CRE terms and household-level control variables (see Table 4). The last row displays the p-value of a Wald test of the joint significance of the CRE terms.

Appendix Table 2. Impact of road type on subsistence farming and wage employment using detailed road type variables

	Household does subsistence farming			Someone in the household has wage job		
	(1)	(2)	(3)	(1)	(2)	(3)
Sealed and good	0.0540 (0.077)	-0.0670 (0.103)	-0.0823 (0.099)	0.1082 (0.084)	0.2255 (0.154)	0.1850 (0.132)
Sealed and not good	0.0152 (0.070)	-0.0735 (0.110)	-0.1272 (0.103)	0.1254 (0.080)	0.1062 (0.139)	0.0992 (0.119)
Gravel and good	0.1313* (0.073)	0.0488 (0.102)	0.0415 (0.093)	0.1339 (0.101)	0.2157 (0.158)	0.2180 (0.140)
Gravel and fair	0.0306 (0.064)	0.0789 (0.089)	0.0401 (0.080)	0.0088 (0.089)	-0.0577 (0.127)	-0.0400 (0.106)
Gravel and poor	0.0457 (0.076)	-0.0045 (0.099)	-0.0247 (0.091)	0.2497* (0.127)	0.2628* (0.141)	0.2502** (0.124)
Log total distance to nearest town	0.0171* (0.010)	0.0176 (0.011)	0.0125 (0.010)	-0.0483** (0.019)	-0.0486** (0.020)	-0.0388** (0.018)
Log distance to nearest road	-0.0046 (0.005)	-0.0026 (0.005)	-0.0019 (0.005)	-0.0175* (0.009)	-0.0190** (0.009)	-0.0139* (0.008)
CRE terms included	No	Yes	Yes	No	Yes	Yes
Household Controls	No	No	Yes	No	No	Yes
R-squared	0.140	0.161	0.203	0.113	0.136	0.206
Census Units	155	155	155	155	155	155
Households	2323	2323	2312	2312	2312	2306
p-value CRE		0.002	0.001		0.079	0.089

Note: * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses, clustered at the census unit level. All regressions are weighted using household sampling weights from both surveys. Dirt road is excluded category for the share variables. Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications also include location-specific control variables (see Table 4) as well as province- and time-fixed effects. Columns labelled (2) in addition include the correlated random effect (CRE) terms, and columns labelled (3) include CRE terms and household-level control variables (see Table 4). The last row displays the p-value of a Wald test of the joint significance of the CRE terms.

Appendix Table 3. Impact of road type on having a good roof and school enrollment using detailed road type variables

	Home has a good roof			Ratio of school children going to school		
	(1)	(2)	(3)	(1)	(2)	(3)
Sealed and good	0.2015** (0.081)	0.2416** (0.108)	0.2369** (0.101)	-0.0143 (0.090)	0.0087 (0.129)	0.0155 (0.120)
Sealed and not good	0.2038*** (0.076)	0.2120** (0.105)	0.2209** (0.097)	0.0194 (0.082)	-0.0790 (0.116)	-0.0361 (0.107)
Gravel and good	0.2048** (0.082)	0.1375 (0.120)	0.1379 (0.109)	-0.0441 (0.102)	-0.1836 (0.132)	-0.1743 (0.118)
Gravel and fair	0.0710 (0.072)	0.0306 (0.104)	0.0589 (0.094)	-0.2170** (0.086)	-0.2224** (0.098)	-0.1923** (0.089)
Gravel and poor	0.1489 (0.090)	0.1137 (0.096)	0.1238 (0.093)	-0.0645 (0.102)	-0.2619* (0.133)	-0.2810** (0.125)
Log total distance to nearest town	-0.0239 (0.017)	-0.0222 (0.018)	-0.0138 (0.018)	-0.0042 (0.016)	-0.0121 (0.017)	-0.0089 (0.015)
Log distance to nearest road	-0.0215*** (0.007)	-0.0216*** (0.007)	-0.0195*** (0.006)	-0.0235*** (0.009)	-0.0171** (0.008)	-0.0129* (0.007)
CRE terms included	No	Yes	Yes	No	Yes	Yes
Household Controls	No	No	Yes	No	No	Yes
R-squared	0.443	0.457	0.475	0.152	0.171	0.221
Census Units	155	155	155	155	155	155
Households	2322	2322	2312	1538	1538	1530
p-value CRE		0.020	0.014		0.261	0.224

Note: * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses, clustered at the census unit level. All regressions are weighted using household sampling weights from both surveys. Dirt road is excluded category for the share variables. Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications also include location-specific control variables (see Table 4) as well as province- and time-fixed effects. Columns labelled (2) in addition include the correlated random effect (CRE) terms, and columns labelled (3) include CRE terms and household-level control variables (see Table 4). The last row displays the p-value of a Wald test of the joint significance of the CRE terms.