The Effect of Road Access on Disease Incidence in Rural India

Yeshwant Chillakuru

Department of Economics, The George Washington University

2115 G St. NW Suite 340, Washington, DC 20052

1 May 2017

Abstract

Using data from the Indian Human Development Survey, the effect of road expansion on disease incidence in rural India was estimated at the village-level. Having a paved road, and proximity to a paved road for villages without a paved road had no impact on percent of villagers sick with any of the fourteen diseases surveyed. Roads had no significant relationship to disease incidence for categories of communicable and non-communicable diseases or for any individual disease. Results are robust with village and time-fixed effects. These results may be a byproduct of countervailing forces of increased access to medical treatment and increased detection and, thus, diagnosis of diseases.

I. Introduction

Improving transportation infrastructure is vital to India's long-term economic development. Paved roads are needed to facilitate trade between communities and improve access to markets, schools, and health facilities, especially in rural India, where most of the country's population lives. In 2000, India began the rural road expansion program known as Pradhan Mantri Gram Sadak Yojama (PMGSY), which aims to connect 170,000 rural villages with paved all-weather roads by 2020. As of 2010, 72,000 habitations have been connected by paved roads (World Bank, 2010). Rural villages are villages with a population of less than 1,000 people in the countryside. However, only a few studies, which will be discussed below, have examined the efficacy of road improvement on health outcomes in India. The Ministry of Rural Development in India assessed PMGSY in 2004 and claimed an overwhelmingly positive impact of roads on health outcomes, stating improved travel times, higher levels of patient satisfaction, and increased hospital visits in several districts. However, the assessment was limited by a small sample size, no control groups, and failure to control for confounding variables, such as increased disease communication due to improved travel conditions causing more hospital visits.

The effect of road expansion on health outcomes has recently been researched, but the results, thus far, offer limited understanding of roads' impact. Bell and van Dillen (2014) examined health outcomes after road expansion under PMGSY in rural villages in the upland region of Orissa, India, with a cross-sectional analysis of 2010 survey data. The theoretical framework underlying their analysis stated that roads would not alter the disease environment (e.g. sanitation, strains of local bacteria/viruses) of the villages in the short-term but would improve treatment times and outcomes as patients travel to health centers and as ambulances reach patients more easily. They found that roads had no effect on morbidity rates, mortality rates, or treatment times. Bell and van Dillen's findings refute the positive impact touted by the Ministry of Rural Development, providing evidence that Indian rural health services are so inadequate that improved access may not improve health outcomes of villagers. However, their data was limited to the Orissa uplands during one time period, which prevented the usage of a more robust, time and entity-fixed effects model.

The effect of road access can lead to unexpected negative health outcomes. Contrary to intuition, Djemai (2011) has shown that HIV morbidity increases with road access in six sub-Saharan African countries. Djemai proposes a framework of two opposing forces:

1. Increased road access allows individuals to access markets to purchase condoms and receive HIV-related education, thus reducing the risk of HIV infection.
2. Increased road access allows for more communicability through both high-risk mobile groups and

individuals having more sexual partners, thus increasing risk of HIV infection.

Djemai found that both effects independently occur and are significant, but the latter overpowers the former. By applying Djemai's framework with the work done by Bell and van Dillen, rural roads expansion can effect health outcomes in India through the following mechanisms:

1. By improving access to treatment facilities, roads will improve overall health outcomes. However, if rural health care facilities lack effective treatments or contain non-sterile pathogenic environments, health outcomes may worsen.
2. By accelerating economic development and thus personal hygiene and sanitation, roads will improve overall health outcomes.
3. By increasing the likelihood of communication from high-risk mobile groups, roads will worsen health outcomes related to communicable diseases (diseases that can be transmitted from person to person, such as tuberculosis), but will have no effect for non-communicable diseases (diseases that cannot be transmitted from person to person, such as cancer).

This paper uses panel data from the India Human Development Survey (IHDS) to demonstrate the relationship between rural road expansion and health outcomes, as defined by incidence of a major morbidity (i.e. long-term) diseases grouped into the following categories: any major morbidity disease incidence, communicable disease incidence, and non-communicable disease.Communicable disease incidence includes having at least one of the following: STD/AIDS, polio, or tuberculosis. Non-communicable disease incidence includes having at least one of the following: cataracts, high blood pressure, heart disease, diabetes, leprosy, cancer, asthma, paralysis, epilepsy, or mental illness. Any disease incidence includes having at least one communicable or non-communicable disease. Using a linear regression model revealed that road expansion, whether by decreasing distance to nearest paved road or building a paved road in the village, had no impact on the incidence of any disease in rural India.

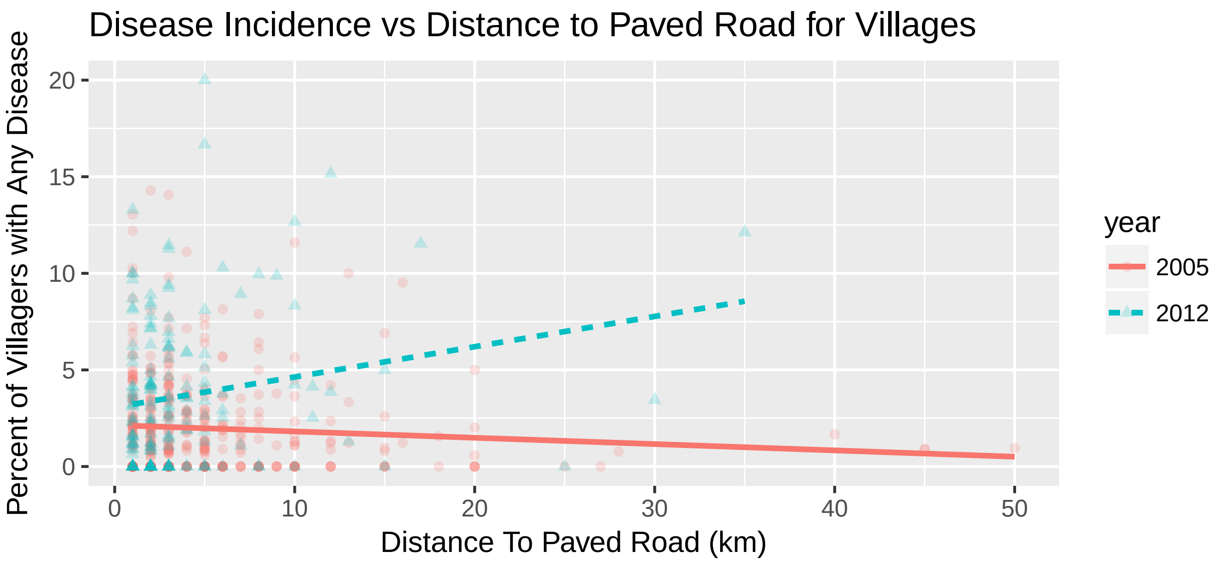
This paper is structured as follows. Section II will examine IHDS data, discussing its strengths and weaknesses. Section III will introduce the two main models that form the basis for the analysis. The Section IV discusses the main results from applying the models to the IHDS data. Section V examines the robustness of the findings. Section VI will summarize findings, and Sections VII, VIII, and IX contain references, tables, and the appendix, respectively.

II. Data

IHDS is a nationally representative survey of 41,554 households in 1,503 villages and 971 urban neighborhoods in India. IHDS-I was conducted from 2004-2005, and IHDS-II was conducted from 2011-2012. "Rural" as used in this analysis is determined by the guidelines used by IHDS when constructing their dataset. The data of interest for this study was divided into several levels. Health outcomes, such as incidence of diabetes or polio, and individual-level controls, such as education and alcohol consumption, were stored in the individual-level dataset. Household-level controls, such as income and access to electricity, were stored household-level dataset. Road data, such as distance to nearest paved road, and village-level controls were stored in the village-level dataset. The datasets were all encoded by IHDS to be mapped together. Each village had an I.D. number. Each household had an I.D. number that was mapped to the village I.D. number. Each individual had an I.D. number that was then mapped to the household and village I.D. numbers. Therefore, data was aggregated to the village level, which had the road data, and as a result, each individual village is the unit of observation. When aggregating data from the specific to the broad (i.e. individual/household to village), the mean was used if data was an interval or ratio variable. If the data was a categorical variable, dummy variables were created for each category, and the means of the dummy variables were aggregated to the village level. After data cleaning, the 2005 data (IHDS-I) contained 1411 observations (i.e. unique villages), and the 2012 dataset (IHDS-II) contained 1279 observations. 1217 villages were the same in both 2005 and 2012.

From the survey data, the impact of PMGSY on rural India can be seen. In 2005, 66.76% of villages in the IHDS survey had paved roads, but in 2012, 86.79% had paved roads. Additionally, for villages without paved roads, the closest paved road was 4.78 km away on average in 2005, but only 3.86 km away on average in 2012. Disease morbidity increased from 2005 to 2012 for all diseases in the dataset, except STD/AIDS, which remained approximately the same (Table 1 and Table 2). A plausible explanation for this increase from 2005 to 2012 may be improved diagnosis from medical campaigns. A simple analysis examining "Percentage of Village with Any Disease vs Distance to Road" reveals a negative relationship between distanceToPavedRoad and mbAnyDisease in 2005 but a positive relationship in 2012. The 2005 correlation opposes the prevailing notion that roads access would reduce disease incidence, but may be attributed to increased diagnosis.[1]

[1] Coefficients for the univariate OLS regressions lines in the graphs are -0.0394467 for 2005 and 0.1568583 for 2012.



III. Model

Two OLS regressions models with time and village-fixed effects will be used to analyze the impact of road expansion on incidence of disease:

$$diseaseIncidence_{it} = \beta RoadPaved_{it} + \delta_1 Household_{it} + \delta_2 Village_{it} + \gamma_{it} + \varepsilon_{it} \tag{1}$$

$$diseaseIncidence_{it} = \beta ln(DistToPavedRoad)_{it} + \delta_1 Household_{it} + \delta_2 Village_{it} + \gamma_{it} + \varepsilon_{it} \tag{2}$$

$diseaseIncidence_{it}$ is the percent of villagers in village i during year t (t is 2005 or 2012) that fall into one of three disease categories:

- Sick with at least one of the 14 diseases collected in the IHDS ($diseaseIncidence_{it}$ is set to $mbAnyDisease_{it}$)
- Sick with at least one of three communicable diseases, STD/AIDS, Polio, or Tuberculosis ($diseaseIncidence_{it}$ is set to $mbComDisease_{it}$)
- Sick with at least one of the remaining non-communicable diseases ($diseaseIncidence_{it}$ is set to $mbNonComDisease_{it}$)

$Household_{it}$ represents village-level controls of individual and household characteristics and focus on capturing changes in economic growth, literacy, and improved household sanitary conditions. This includes mean household income in each village, percent of villagers literate in each village, percent of villagers that smoke tobacco in each village, percent of households that own a toilet in their home in each village, percent of households that have electricity in each village, percent of households in each village that have seen a doctor in the past five years, and the percent of households in each village that have a piped water system (e.g. plumbing).[2] $Village_{it}$ represents controls for characteristics of the village and focus on capturing variations in village health resources from 2005 to 2012. These include the number of health subcenters, primary health centers, community health centers, private clinics, and private hospitals.[3] γ_{it} represents village-fixed effects, and ε_{it} is the error term. The only difference between Models 1 and 2 is the primary independent variable of interest: $RoadPaved_{it}$ vs $ln(DistToPavedRoad)_{it}$.[4] Equation (1) allows us to understand the impact of how introducing a paved road affects $diseaseIncidence_{it}$, while Equation (2) allows us to understand how becoming closer to a road for villages with no paved road can improve $diseaseIncidedence_{it}$. $roadPaved_{it}$ indicates if there is or is not a paved road in the village. $ln(DistToPavedRoad)_{it}$ indicates the natural log distance to the nearest paved for villages without a paved road. Equation (2) is only applied to villages without a paved road in order to separate the effects of a road being present or not versus improvement in distance to the road.

As demonstrated by Djemai (2011), the introduction of paved roads may increase disease communicability as mobile groups at high-risk for diseases like HIV may travel more through an area, increasing the risk of communicability. To isolate the impact of increased communicability on disease incidence, I performed separate regressions for communicable vs non-communicable diseases as noted above. For example, if road access is bringing high-risk groups, who partake in risky sexual behavior, STD/HIV incidence would increase, while non-communicable diseases, like mental illness would not be effected.

IV. Results

Table 3 details the results of a simple univariate OLS regressions for Equation (1). A positive correlation, significant at the 1% level, exists for both general disease incidence and non-communicable disease incidence with presense of a paved road, but not for communicable disease or STD/AIDS incidence. Table 4 details the results of a simple univariate OLS regression for Equation (2) and demonstrates a negative relationship,

[2] *Household* controls encoded as `income` (in units of a hundred thousand Indian Rupees), `literate`, `smokeTobacco`, `ownToilet`, `electricity`, `seenDoctor`, `PctPipedWater`.

[3] *Village* controls encoded as `ImmuniCamps`, `healthSubCntr`, `primaryHealthCntr`, `commHealthCntr`, `pvtClinic`, `pvtHospital`. Addititionlly, if preceded by `dist`, then variable refers to distance to the health facility in kilometers.

[4] The natural log of distance from village to a paved road was used in order to account for the likely non-linear relationship of distance to paved road on village health outcomes. For example, a reduction in distance from 5km to 1km would likely have a much greater impact on road access and, theoritically, health outcomes than a reduction in distance from 100km to 96km, even though the change in distance is -4km both times. The natural log transformation internalizes the non-linear relationship to the model.

significant at the 1% level, for only non-communicable disease incidence and distance to paved road. In the following sections, the $Household$ and $Village$ controls, especially "smokeTobacco" eliminate the significant relationship between roads and health outcomes.

General Disease Incidence and Roads

Table 5: Regressions 1, 2, 3 present the results for Equation (1) for percentage of villagers with at least one of the 14 major morbidity diseases ($diseaseIncidence_{it}$ is set to $mbAnyDisease_{it}$). Upon inclusion of $Household_{it}$ and $Village_{it}$, the effect of having a paved road in the village has no significant relationship to "mbAnyDisease." A large drop in the coefficient of "roadPaved" from 1.579 to 0.441 is observed between Table 5: Regression 1 (no controls) and Table 5: Regression 2 (controlling for $Household_{it}$ variables). Upon further analysis, this drop is primarily induced when controlling for percent of villagers that smoke tobacco and the average income of the village.[5] This suggests that disease incidence rates are primarily driven by changes in smoking habits and household income. "smokeTobacco" has a positive coefficient of 0.79 and is significant at the 1% level, even when including $Village_{it}$ controls (Table 5: Regression 3). This means a 1 percentage point decrease in percent of villagers smoking will decrease the percent of villagers with at least one of the 14 major morbiditiy diseases by 0.79 percentage points. "income" has a positive coefficient of 0.58 and is significant at the 5% level (Table 5: Regression 3). This means a 100,000 Rupee decrease in percent of villagers smoking will decrease the percent of villagers with at least one of the 14 major morbiditiy diseases by 0.58 percentage points.

Table 5: Regressions 4, 5, 6 present results for Equation (2) for incidence of any disease ($diseaseIncidence_{it}$ is set to $mbAnyDisease_{it}$). For villages without a paved road, increasing the distance to the paved road has no significant relationship with disease incidence, regardless of controls. Even among our controls, the only ones significant were percent of villagers smoking tobacco and number of private hospitals in the village. The large rise in the coefficient of "lnDistToPvdRd" from -0.835 to -0.098, in Table 5: Regression 4 and 5, is primarily due to an the inclusion of the "smokeTobacco" control variable.[4] These non-significant finds are surprising, but important nonetheless because they contradict the results of Djemai (2011) and the assessment of PMGSY by Ministry of Rural Development in India (2004). Analysis on communicable vs non-communicable diseases will help develop a more comprehensive understanding the interactions between disease incidence and road expansion.

[5]Detailed analysis can be viewed on Github at https://github.com/yesh747/RuralRoadsIndia. See "Regression: Any Disease" in Data_Analysis.Rmd file. See Appendix for more details.

Communicable Disease Incidence and Roads

Communicable and non-communicable diseases act through different channels, and Djemai (2011) and Tatem et al. (2006) have shown that roads and transportation infrastructure expansion strongly contribute to the spread of communicable disease. Table 6 details the analysis of Equations (1) and (2), on incidence of communicable diseases ($diseaseIncidence_{it}$ is set to $mbComDisease_{it}$), which are defined here to include STD/AIDS, polio, or tuberculosis. Opposite to expectations, roads in both Equations (1) and (2) have no net impact on communicable disease, regardless of controls. Even among controls, only percent of villagers with a piped water supply (`PctPipedWater`) was significant at the 5% level for Equation (2).

Non-Communicable Disease Incidence and Roads

Table 7 presents results from an analysis of Equations (1) and (2) on incidence of non-communicable diseases ($diseaseIncidence_{it}$ is set to $mbNonComDisease_{it}$). Similar to previous results thus far, road expansion appears to have no significant effect on incidence of non-communicable disease, while the only significant controls was percent of villagers that smoke tobacco (`smokeTobacco`) at the 5% level in Equation (1), where a one percentage point increase in percent of villagers that smoke tobacco leads to a 0.79 percentage point increase villagers with a non-communicable disease. Similarly to the analysis for "General Disease Incidence and Roads", a large drop in the coefficient of "roadPaved" was seen before and after controlling for $Household_{it}$ variables (Table 7: Regressions 2 and 3). Just as before, this large drop in coefficient is primarily attributed to the addition of the "smokeTobacco" variable[6], suggesting that differences in smoking habits is related to much of the variation in disease incidence. The significance of smoking tobacco on non-communicable disease is logically consistent as smoking contributes to chronic illnesses, such as cancer and asthma, both of which are captured in the non-communicable disease category.

STD/AIDS Incidence and Roads

In order to make a direct comparison to Djemai's results (the positive correlation between distance to paved road and HIV Incidence), I applied Equations (1) and (2) to STD/AIDS incidence. Table 8 details these results. Similar to Djemai, I controlled for education (literacy rate) and income, while additionally controlling for number of health facilities in the village, immunization campaigns, and sanitary conditions (owning a toilet in the home). The models in this paper lack controls for marriage, gender, HIV/knowledge,

[6]Detailed analysis can be viewed on Github at https://github.com/yesh747/RuralRoadsIndia. See "Regression: Non-Communicable Disease" in Data_Analysis.Rmd file. See Appendix for more details.

and age, primarily because our unit of observation is at the village level, while Djemai's is at the individual level. Due to lack of GPS coordinate, I am unable to replicate Djemai's instrumental variables methods using terrain ruggedness and gradient, but our analysis uses time-fixed and village-fixed effects, whereas Djemai could only use regional-fixed effects. Both analysis have the drawbacks and strengths, but the results in Table 8 are robust in showing that roads have no effect on STD/HIV incidence, in disagreement with Djemai's results.

V. Robustness Checks

This analysis examined 14 diseases collected in the IHDS and found no significant relationship between road expansion and incidence of general, communicable, or non-communicable diseases at the village level, using Equations (1) and (2). Applying Equation (1) to each of the 14 individual diseases confirms the relationship that roads have no significant impact on any of the individual health outcomes (Tables 9, 10, and 11). While road expansion has no effect on general disease incidence according to our analysis, it may impact the number of days incapacitated due to one or more of the 14 diseases. A variation of Equations (1) and (2) are used to examine the average number of days villagers are incapacitated due to one or more of the diseases surveyed (mbDaysIncapacitated), the following models, respectively:

$$mbDaysIncapacitated_{it} = \beta RoadPaved_{it} + \delta_1 Household_{it} + \delta_2 Village_{it} + \gamma_{it} + \varepsilon_{it} \tag{3}$$

$$mbDaysIncapacitated_{it} = \beta lnDistToPvdRd_{it} + \delta_1 Household_{it} + \delta_2 Village_{it} + \gamma_{it} + \varepsilon_{it} \tag{4}$$

Table 12 details the result of this analysis and supports the findings that roads have no net impact on any of the 14 major morbidity diseases collected by IHDS.

Additionally, non-random road placement may appear to be a concern. Villages with higher incomes and larger populations could be more likely to receive roads. Similarly, wealthier villages may have better health facilities. Djemai instrumented road distance on terrain ruggedness and slope to account for non-random raod placement. However, PMGSY program is run at the national, and not state or district, level, and is less prone to influence of local village-level politics. Therefore, the assumption of non-random road placement can be made. Nonetheless, these models introduce controls to ensure robustness. Time and village-fixed effects control for several potentials areas of omitted variable bias. Differences between villages in healthcare technology and access and sanitation infrastructure may arise non-randomly as villages near urban areas may benefit from spillover effects and earlier development. However, this is adjusted for by village-fixed effects. Similarly, healthcare technology and sanitation infrastructure may improve from 2005 to 2012 in villages, but, again, these are accounted for in time-fixed effects. Addtionally, Equation (1) can be interpreted as a difference-in-differences design, where the treatment group is the 20.03% of villages in our sample that recieve a paved road between 2005 (before treatment) and 2012 (after treatment). $Household$ and $Village$ variables control for any village-level changes between 2005 and 2012.

An issue with the Equations (1) and (2) is their weak fit to the data. Equation (1) is a better fit, measured by Adjusted R^2, than Equation (2) for the data[7] and, thus, is the preferred model. Adjusted R^2 for Equation (1) ranges from -0.976 to -1.375, while Adjusted R^2 for Equation (2) ranges from -3.009 to -4.975. Equations (3) and (4) have similar values as Equations (1) and (2), respectively, for Adjusted R^2. These values show that these models poorly fit the data and explain essentially none of the variance in the observations. The low Adjusted R^2 implies that other unobserved factors may have stronger influences on the variations in health outcomes across villages.

One potential caveat in this analysis was the granularity of the unit of observation: the village. Road data was collected at the village level, so individual health data was aggregated using the mean to the village level, which results in a loss of data resolution for individuals' variations. This means our model is unable to account for individual changes in behavior that may result in different health outcomes. This is an important drawback of the analysis because the mechanisms through which road expansion could alter health outcomes occurs at the individual level. Individuals, not villages, make the decision to see physicians and travel to hospitals. Djemai (2011) was fortunate to have conducted his analysis at the individual level, and this difference in data resolution may account for the non-significant findings.

The IHDS data itself may cause bias in the results. As it was a questionairre, underreporting of health information, especially conditions that carry significant stigma like mental illness, may bias coefficients downwards. Not all the villages in the IHDS dataset were used because of missing variables for some observations. Of the 1503 villages surveyed each year, only 1217 (80.9713906%) villages were used in the analysis. This may cause an unobserved bias as villages that were missing information may be the ones that had the poorest road access, and, thus, the survey conductors were unable to reach the village.

[7] Adjusted R^2 for Regressions 1, 2, and 3 (Equation 1) is higher than that for Regressions 4, 5, and 6 (Equation 2), in Tables 5, 6, 7, and 8.

VI. Conclusion

Using data on individuals, households, and villages from the Indian Human Development Survey (IHDS), this paper analyzes the effect of road expansion on disease incidence in villages in rural India. This analysis examines how disease incidence is affected by whether a village has a paved road or not and, if it has no paved road, the distance to the paved road. Disease incidence is categorized by the percent of villagers in each village that have at least one of the 14 major morbidity (i.e. long-term) diseases collected by IHDS: cataracts, tuberculosis, high blood pressure, heart disease, diabetes, leprosy, cancer, asthma, polio, paralysis, epilepsy, mental illness, STD/AIDS, and an other long-term disease. Using a OLS regression model with time and village-fixed effects and with villages as the unit of observation, the results suggest that road expansion has no significant impact on major morbidity incidence in villages in rural India.

These results contribute to the limited, but growing literature on rural infrastructure and health outcomes. Bell and van Dillen (2014) found no effect of paved roads on health outcomes in upland Orissa, India. While their paper was limited to cross-sectional data from a relatively small sample of 30 villages in a specific region of the country, this paper examined 1217 villages from throughout India in both 2005 and 2012 with time and village-fixed effects. The null findings in this paper, however, support Bell and van Dillen's claims using a more robust dataset and model. Djemai (2011) demonstrated a robust, positive relationship between HIV incidence and distance to paved roads in six African countries and attributed this effect to the increased movement of mobile groups with high HIV-risk overpowering the benefits of increased medical access and knowledge associated with road expansion. My analysis contradicted Djemai's findings as roads had no impact on STD/AIDS incidence or communicable diseases in general (IHDS combined STD's and AIDS into one survey question). This difference may be attributable to a difference in cultures (e.g. high-risk mobile groups may not have as large a presence in rural India) or differences in disease environments (HIV is not as prevalent in India). To better understand how roads play a role in health outcomes, future analysis of the IHDS dataset can be conducted to explore three potential channels/questions:

1. Do paved roads increase the number of visits to local hospitals and/or decrease travel time significantly?
2. Do paved roads improve levels of education on health-related matters, like HIV risks or practices of personal hygiene by increasing access to health centers and markets?
3. Do paved roads improve economic development of a village, and as a result, the quality of care provided at local health centers?

Nonetheless, these findings have important implications for Indian domestic rural policy and the validity of internal reviews. The Ministry of Rural Development (2004) claimed an overwhelmingly positive impact of road expansion on rural health outcomes when conducting an assessment of PMGSY, the nation's rural road expansion program, which started in 2000 and continues to this day. The Ministry referenced survey data that itself collected to showed decreased travel times to health centers, higher levels of patients satisfaction, and increased hospital visits in several districts. The assessment was severely limited by a small sample size, no control groups, no control variables, or any measure of tangible improvements in the direct health of individuals. My analysis shows that simply building infrastructure alone is not enough to improve health outcomes, and internal reviews for assessing impact, like that conducted by the Ministry of Rural Development, must be held to a higher standard of rigor. Additionally, these findings shed light on a more general topic of rural road infrastructure in developing countries and health outcomes. While road expansion is vital and important to the long-term economic and social development, it alone cannot induce improvements in population health. Road infrastructure is just one component of a larger integrated population health plan, which includes health facility quality, physician access, mental health security, etc., to improving health outcomes.

VII. References

[1] Bell, Clive and Susanne van Dillen. 2014. "How Does India's Rural Road Program Affect the Grassroots? Findings from a Survey in Upland Orissa." *Land Economics*, 90(2): 372-394. https://www.researchgate.net/publication/254456306_How_Does_India's_Rural_Roads_Program_Affect_the_Grassroots_Findings_from_a_Survey_in_Upland_Orissa.

[2] Desai, Sonalde, and Reeve Vanneman. 2012. *India Human Development Survey-II (IHDS-II)*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2016-08-01. http://doi.org/10.3886/ICPSR36151.v5.

[3] Desai, Sonalde, Reeve Vanneman, and National Council of Applied Economic Research, New Delhi. 2005. *India Human Development Survey (IHDS)*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2016-02-16. http://doi.org/10.3886/ICPSR22626.v11.

[4] Djemai, Elodie. 2011. "HIV and Access to Road." University of Michigan Working Paper. http://www.ipc.umich.edu/pdf/Djema-HIV-Access-to-Road.pdf.

[5] Ministry of Health and Family Welfare, Government of India. 2016. Indian Public Health Standards. http://nrhm.gov.in/nhm/nrhm/guidelines/indian-public-health-standards.html.

[6] Ministry of Rural Development, Government of India. 2004. Impact Assessment of Pradhan Mantri Gram

Sadak Yojana. http://pmgsy.nic.in/pmg122.asp.

[7] World Bank. 2010. Project Appraisal Document. Report No. 570810-IN. http://documents.worldbank.org/curated/en/879581468041646928/pdf/570810PAD0P1241e0only1910BOX353794B.pdf.

VIII. Tables

Table 1: 2005 Summary Statistics: Disease Incidence (Pct of village) 2005

Statistic	N	Mean	St. Dev.	Min	Max
mbCataract	1,411	0.517	1.112	0.000	12.500
mbTuberculosis	1,411	0.279	0.699	0.000	6.667
mbHighBP	1,411	1.071	2.110	0.000	20.253
mbHeartDisease	1,411	0.380	0.839	0.000	9.091
mbDiabetes	1,411	0.548	1.441	0.000	17.241
mbLeprosy	1,411	0.052	0.321	0.000	7.692
mbCancer	1,411	0.061	0.395	0.000	9.524
mbAsthma	1,411	0.625	1.181	0.000	13.158
mbPolio	1,411	0.119	0.408	0.000	4.286
mbParalysis	1,411	0.155	0.548	0.000	7.595
mbEpilepsy	1,411	0.115	0.418	0.000	5.000
mbMentalIllness	1,411	0.128	0.407	0.000	5.000
mbSTDorAIDS	1,411	0.053	0.353	0.000	6.154
mbOtherLongTerm	1,411	1.970	2.517	0.000	19.048
mbAnyDisease	1,411	1.970	2.517	0.000	19.048
mbComDisease	1,411	0.279	0.699	0.000	6.667
mbNonComDisease	1,411	0.128	0.407	0.000	5.000

Table 2: 2012 Summary Statistics: Disease Incidence (Pct of village) 2012

Statistic	N	Mean	St. Dev.	Min	Max
mbCataract	1,279	0.922	1.540	0.000	16.667
mbTuberculosis	1,279	0.291	0.707	0.000	8.696
mbHighBP	1,279	2.457	3.130	0.000	20.000
mbHeartDisease	1,279	0.632	1.211	0.000	11.538
mbDiabetes	1,279	1.089	2.075	0.000	20.548
mbLeprosy	1,279	0.061	0.374	0.000	6.250
mbCancer	1,279	0.060	0.346	0.000	6.250
mbAsthma	1,279	1.013	1.549	0.000	25.000
mbPolio	1,279	0.096	0.343	0.000	3.448
mbParalysis	1,279	0.354	1.223	0.000	17.568
mbEpilepsy	1,279	0.237	0.792	0.000	12.500
mbMentalIllness	1,279	0.332	0.881	0.000	14.286
mbSTDorAIDS	1,279	0.033	0.258	0.000	5.556
mbOtherLongTerm	1,279	3.936	3.856	0.000	32.000
mbAnyDisease	1,279	3.936	3.856	0.000	32.000
mbComDisease	1,279	0.291	0.707	0.000	8.696
mbNonComDisease	1,279	0.332	0.881	0.000	14.286

Table 3: Road Paved and Incidence of Disease Types: OLS Base Specification

	Dependent variable:			
	mbAnyDisease	mbComDisease	mbNonComDisease	mbSTDorAIDS
	(1)	(2)	(3)	(4)
roadPaved	0.622***	0.018	0.110***	−0.025
	(0.147)	(0.032)	(0.026)	(0.017)
Constant	2.456***	0.272***	0.145***	0.063***
	(0.123)	(0.028)	(0.020)	(0.016)
Observations	2,545	2,545	2,545	2,545
R^2	0.006	0.0001	0.005	0.001
Adjusted R^2	0.006	−0.0003	0.004	0.001
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 4: Road Distance and Incidence of Disease Types: OLS Base Specification

	Dependent variable:			
	mbAnyDisease	mbComDisease	mbNonComDisease	mbSTDorAIDS
	(1)	(2)	(3)	(4)
lnDistToPvdRd	−0.024	−0.031	−0.055**	−0.018
	(0.154)	(0.034)	(0.025)	(0.017)
Constant	2.483***	0.306***	0.207***	0.083***
	(0.202)	(0.047)	(0.041)	(0.027)
Observations	599	599	599	599
R^2	0.00005	0.002	0.010	0.002
Adjusted R^2	−0.002	−0.0001	0.008	−0.0001
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 5: Road Expansion and Incidence of Any Disease: OLS with Village and Time-Fixed Effects

	Dependent variable:					
	mbAnyDisease					
	(1)	(2)	(3)	(4)	(5)	(6)
roadPaved	1.579***	0.441	0.405			
	(0.275)	(0.276)	(0.269)			
income		0.599***	0.580**		−1.033	−1.033
		(0.219)	(0.226)		(1.244)	(1.145)
literate		2.757*	2.659*		−1.503	−0.250
		(1.496)	(1.477)		(4.519)	(4.367)
smokeTobacco		0.800***	0.790***		2.321***	2.180***
		(0.188)	(0.187)		(0.522)	(0.496)
ownToilet		0.018***	0.018***		0.021	0.017
		(0.007)	(0.007)		(0.022)	(0.023)
electricity		0.007	0.007		0.010	0.009
		(0.006)	(0.006)		(0.016)	(0.016)
seenDoctor		0.299	0.225		0.017	−0.175
		(1.131)	(1.127)		(2.440)	(2.432)
ImmuniCamps			0.037*			0.096
			(0.021)			(0.082)
PctPipedWater			−0.004			0.039*
			(0.005)			(0.023)
healthSubCntr			0.065			−0.729
			(0.208)			(0.683)
primaryHealthCntr			−0.584			−0.595
			(0.400)			(2.740)
commHealthCntr			−0.161			−0.470
			(0.588)			(1.578)
pvtClinic			0.169			0.175
			(0.168)			(0.380)
pvtHospital			0.064			5.439***
			(0.283)			(1.572)
lnDistToPvdRd				−0.835*	−0.098	−0.174
				(0.442)	(0.462)	(0.449)
Observations	2,545	2,545	2,545	599	599	599
R^2	0.030	0.163	0.171	0.022	0.336	0.382
Adjusted R^2	−1.283	−0.980	−0.976	−4.570	−3.009	−3.020
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 6: Road Expansion and Incidence of Communicable Disease: OLS with Village and Time-Fixed Effects

	Dependent variable:					
	mbComDisease					
	(1)	(2)	(3)	(4)	(5)	(6)
roadPaved	0.011	0.006	0.003			
	(0.048)	(0.054)	(0.054)			
income		−0.039	−0.035		−0.187	−0.264
		(0.055)	(0.056)		(0.199)	(0.252)
literate		−0.358	−0.371		0.288	0.021
		(0.313)	(0.313)		(0.698)	(0.672)
smokeTobacco		−0.020	−0.016		0.163	0.200
		(0.034)	(0.034)		(0.113)	(0.130)
ownToilet		0.002*	0.003*		−0.003	−0.002
		(0.001)	(0.001)		(0.005)	(0.005)
electricity		0.001	0.001		−0.002	−0.003
		(0.001)	(0.001)		(0.003)	(0.004)
seenDoctor		−0.527*	−0.531*		−0.181	−0.161
		(0.291)	(0.293)		(0.496)	(0.455)
ImmuniCamps			−0.005			0.021
			(0.005)			(0.014)
PctPipedWater			−0.001			−0.005**
			(0.001)			(0.002)
healthSubCntr			0.031			−0.061
			(0.037)			(0.146)
primaryHealthCntr			−0.051			0.464
			(0.085)			(0.327)
commHealthCntr			0.054			0.485
			(0.115)			(0.477)
pvtClinic			−0.006			−0.052
			(0.015)			(0.179)
pvtHospital			−0.054			0.439
			(0.054)			(0.359)
lnDistToPvdRd				−0.089	−0.047	−0.026
				(0.055)	(0.068)	(0.064)
Observations	2,545	2,545	2,545	599	599	599
R^2	0.00004	0.008	0.011	0.010	0.050	0.091
Adjusted R^2	−1.353	−1.348	−1.356	−4.638	−4.738	−4.912
Note:				*p<0.1; **p<0.05; ***p<0.01		

Table 7: Road Expansion and Non-Communicable Disease Incidence: OLS with Village and Time-Fixed Effects

	Dependent variable:					
	mbNonComDisease					
	(1)	(2)	(3)	(4)	(5)	(6)
roadPaved	0.152***	0.033	0.031			
	(0.045)	(0.052)	(0.052)			
income		0.060	0.057		0.090	−0.013
		(0.054)	(0.054)		(0.136)	(0.199)
literate		0.420	0.451		0.295	0.350
		(0.336)	(0.340)		(0.416)	(0.402)
smokeTobacco		0.086**	0.086**		0.004	0.032
		(0.038)	(0.038)		(0.083)	(0.083)
ownToilet		0.0001	0.0002		0.001	0.002
		(0.001)	(0.001)		(0.002)	(0.003)
electricity		0.0003	0.0003		0.001	0.001
		(0.001)	(0.001)		(0.001)	(0.001)
seenDoctor		−0.203	−0.218		−0.104	−0.137
		(0.284)	(0.280)		(0.536)	(0.548)
ImmuniCamps			−0.007			−0.013
			(0.004)			(0.014)
PctPipedWater			0.001			0.002
			(0.001)			(0.004)
healthSubCntr			0.039			−0.046
			(0.033)			(0.080)
primaryHealthCntr			0.114			0.497
			(0.080)			(0.539)
commHealthCntr			0.036			−0.354
			(0.073)			(0.225)
pvtClinic			−0.003			0.092
			(0.011)			(0.075)
pvtHospital			−0.033			−0.012
			(0.038)			(0.177)
lnDistToPvdRd				−0.068	−0.046	−0.024
				(0.047)	(0.045)	(0.057)
Observations	2,545	2,545	2,545	599	599	599
R^2	0.007	0.041	0.046	0.009	0.033	0.081
Adjusted R^2	−1.336	−1.269	−1.273	−4.645	−4.843	−4.975
Note:					*p<0.1; **p<0.05; ***p<0.01	

Table 8: Road Expansion and Incidence of STD/AIDS: OLS with Village and Time-Fixed Effects

	Dependent variable:					
	mbSTDorAIDS					
	(1)	(2)	(3)	(4)	(5)	(6)
roadPaved	−0.037	−0.028	−0.023			
	(0.027)	(0.029)	(0.029)			
income		0.002	0.003		−0.281*	−0.185
		(0.017)	(0.017)		(0.165)	(0.121)
literate		−0.181	−0.195		−0.891	−1.043
		(0.153)	(0.154)		(0.744)	(0.653)
smokeTobacco		−0.007	−0.006		0.108	0.100
		(0.015)	(0.015)		(0.067)	(0.068)
ownToilet		−0.00002	−0.00003		0.002	0.001
		(0.001)	(0.001)		(0.004)	(0.002)
electricity		0.0002	0.0002		0.001	0.001
		(0.0004)	(0.0004)		(0.001)	(0.001)
seenDoctor		0.0003	0.014		−0.082	−0.006
		(0.117)	(0.117)		(0.284)	(0.262)
ImmuniCamps			0.002			−0.004
			(0.002)			(0.007)
PctPipedWater			−0.0001			−0.005
			(0.001)			(0.003)
healthSubCntr			−0.005			−0.026
			(0.014)			(0.156)
primaryHealthCntr			−0.032			−0.521
			(0.045)			(0.491)
commHealthCntr			−0.088			0.388*
			(0.080)			(0.220)
pvtClinic			−0.012*			−0.031
			(0.007)			(0.030)
pvtHospital			0.004			0.225
			(0.015)			(0.185)
lnDistToPvdRd				−0.003	0.009	−0.012
				(0.047)	(0.045)	(0.031)
Observations	2,545	2,545	2,545	599	599	599
R^2	0.002	0.005	0.011	0.00002	0.073	0.149
Adjusted R^2	−1.348	−1.355	−1.356	−4.695	−4.598	−4.528
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 9: Road Paved and Disease Incidence: OLS with Village and Time-Fixed Effects

	Dependent variable:			
	mbCataract	mbTuberculosis	mbHighBP	mbHeartDisease
	(1)	(2)	(3)	(4)
roadPaved	0.122	0.003	−0.020	0.031
	(0.103)	(0.058)	(0.194)	(0.084)
income	0.119	−0.035	1.218***	0.152**
	(0.087)	(0.049)	(0.163)	(0.071)
literate	−0.516	−0.371	−0.002	0.576
	(0.517)	(0.291)	(0.976)	(0.425)
smokeTobacco	0.131**	−0.016	0.570***	0.175***
	(0.061)	(0.034)	(0.115)	(0.050)
ownToilet	0.005**	0.003*	0.005	0.0002
	(0.002)	(0.001)	(0.004)	(0.002)
electricity	0.003	0.001	−0.003	−0.004**
	(0.002)	(0.001)	(0.004)	(0.002)
seenDoctor	−0.197	−0.531*	1.831*	0.727*
	(0.503)	(0.283)	(0.948)	(0.413)
ImmuniCamps	−0.015	−0.005	0.019	−0.003
	(0.010)	(0.006)	(0.019)	(0.008)
PctPipedWater	0.001	−0.001	0.007*	−0.001
	(0.002)	(0.001)	(0.004)	(0.002)
healthSubCntr	0.050	0.031	−0.033	−0.004
	(0.070)	(0.040)	(0.133)	(0.058)
primaryHealthCntr	0.135	−0.051	−0.281	−0.030
	(0.141)	(0.079)	(0.265)	(0.116)
commHealthCntr	0.284	0.054	0.986**	0.333*
	(0.238)	(0.134)	(0.448)	(0.195)
pvtClinic	0.075**	−0.006	0.007	0.016
	(0.031)	(0.018)	(0.059)	(0.026)
pvtHospital	−0.048	−0.054	−0.013	0.149*
	(0.105)	(0.059)	(0.198)	(0.086)
Observations	2,545	2,545	2,545	2,545
R^2	0.062	0.011	0.205	0.052
Adjusted R^2	−1.234	−1.356	−0.893	−1.259
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 10: Road Paved and Disease Incidence: OLS with Village and Time-Fixed Effects

	Dependent variable:				
	mbDiabetes	mbLeprosy	mbCancer	mbAsthma	mbPolio
	(1)	(2)	(3)	(4)	(5)
roadPaved	0.202*	0.015	−0.006	−0.099	0.038
	(0.106)	(0.033)	(0.035)	(0.100)	(0.030)
income	0.813***	−0.039	−0.014	0.153*	−0.055**
	(0.089)	(0.028)	(0.030)	(0.084)	(0.025)
literate	−0.315	0.022	0.092	−0.172	0.160
	(0.534)	(0.167)	(0.178)	(0.504)	(0.149)
smokeTobacco	−0.045	0.014	−0.014	0.296***	0.014
	(0.063)	(0.020)	(0.021)	(0.059)	(0.018)
ownToilet	0.001	0.0003	0.001	−0.001	0.0004
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
electricity	0.001	0.001	−0.0001	−0.0002	−0.001*
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
seenDoctor	0.240	−0.015	−0.012	0.435	0.221
	(0.519)	(0.163)	(0.173)	(0.490)	(0.145)
ImmuniCamps	0.008	0.002	0.001	0.001	0.004
	(0.010)	(0.003)	(0.003)	(0.010)	(0.003)
PctPipedWater	0.004**	−0.0002	0.0001	−0.001	−0.001*
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
healthSubCntr	−0.021	0.013	0.043*	−0.111	−0.018
	(0.073)	(0.023)	(0.024)	(0.069)	(0.020)
primaryHealthCntr	−0.309**	0.042	−0.026	0.043	0.005
	(0.145)	(0.046)	(0.049)	(0.137)	(0.041)
commHealthCntr	0.567**	−0.047	0.020	0.354	0.065
	(0.245)	(0.077)	(0.082)	(0.232)	(0.069)
pvtClinic	0.070**	0.005	0.011	0.048	0.0001
	(0.032)	(0.010)	(0.011)	(0.031)	(0.009)
pvtHospital	−0.066	−0.036	−0.031	−0.124	−0.0004
	(0.109)	(0.034)	(0.036)	(0.103)	(0.030)
Observations	2,545	2,545	2,545	2,545	2,545
R^2	0.161	0.006	0.007	0.064	0.021
Adjusted R^2	−0.999	−1.367	−1.365	−1.230	−1.333
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 11: Road Paved and Disease Incidence: OLS with Village and Time-Fixed Effects

	Dependent variable:				
	mbParalysis	mbEpilepsy	mbMentalIllness	mbSTDorAIDS	mbOtherLongTerm
	(1)	(2)	(3)	(4)	(5)
roadPaved	0.078	−0.019	0.031	−0.023	0.405
	(0.080)	(0.053)	(0.057)	(0.026)	(0.274)
income	0.150**	0.058	0.057	0.003	0.580**
	(0.067)	(0.045)	(0.048)	(0.022)	(0.231)
literate	−0.827**	−0.114	0.451	−0.195	2.659*
	(0.402)	(0.268)	(0.289)	(0.130)	(1.380)
smokeTobacco	0.082*	0.039	0.086**	−0.006	0.790***
	(0.047)	(0.032)	(0.034)	(0.015)	(0.162)
ownToilet	−0.001	0.001	0.0002	−0.00003	0.018***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.006)
electricity	0.002	−0.0001	0.0003	0.0002	0.007
	(0.002)	(0.001)	(0.001)	(0.001)	(0.005)
seenDoctor	−0.029	−0.573**	−0.218	0.014	0.225
	(0.391)	(0.261)	(0.281)	(0.127)	(1.341)
ImmuniCamps	0.008	−0.001	−0.007	0.002	0.037
	(0.008)	(0.005)	(0.006)	(0.002)	(0.026)
PctPipedWater	0.0003	−0.002*	0.001	−0.0001	−0.004
	(0.002)	(0.001)	(0.001)	(0.001)	(0.005)
healthSubCntr	0.230***	−0.013	0.039	−0.005	0.065
	(0.055)	(0.037)	(0.039)	(0.018)	(0.188)
primaryHealthCntr	−0.220**	0.110	0.114	−0.032	−0.584
	(0.109)	(0.073)	(0.079)	(0.035)	(0.375)
commHealthCntr	0.737***	0.185	0.036	−0.088	−0.161
	(0.185)	(0.123)	(0.133)	(0.060)	(0.634)
pvtClinic	0.049**	−0.010	−0.003	−0.012	0.169**
	(0.024)	(0.016)	(0.018)	(0.008)	(0.084)
pvtHospital	−0.043	−0.035	−0.033	0.004	0.064
	(0.082)	(0.055)	(0.059)	(0.027)	(0.281)
Observations	2,545	2,545	2,545	2,545	2,545
R^2	0.075	0.024	0.046	0.011	0.171
Adjusted R^2	−1.203	−1.325	−1.273	−1.356	−0.976
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 12: Road and Days Incapacitated Due To Illness: OLS with Village and Time-fixed Effects

	Dependent variable:					
	mbDaysIncapacitated					
	(1)	(2)	(3)	(4)	(5)	(6)
roadPaved	0.699	0.118	0.103			
	(0.437)	(0.468)	(0.463)			
income		0.799*	0.852*		−1.451	−1.016
		(0.440)	(0.451)		(1.314)	(1.258)
literate		−3.587	−4.063*		−9.137	−8.992
		(2.211)	(2.234)		(6.519)	(6.358)
smokeTobacco		0.189	0.204		0.651	0.373
		(0.259)	(0.261)		(0.626)	(0.654)
ownToilet		0.007	0.009		0.030	0.024
		(0.011)	(0.011)		(0.026)	(0.021)
electricity		0.012*	0.012*		0.019	0.022
		(0.007)	(0.007)		(0.015)	(0.017)
seenDoctor		−2.152	−2.087		−2.204	−1.880
		(1.533)	(1.517)		(3.016)	(3.037)
ImmuniCamps			0.061**			0.124
			(0.030)			(0.154)
PctPipedWater			−0.016*			−0.002
			(0.009)			(0.034)
healthSubCntr			0.028			−2.427**
			(0.328)			(1.013)
primaryHealthCntr			−0.762			−1.349
			(0.619)			(3.351)
commHealthCntr			−0.440			−2.288
			(0.902)			(2.667)
pvtClinic			0.055			−0.610
			(0.150)			(0.857)
pvtHospital			−0.119			4.466**
			(0.477)			(2.011)
lnDistToPvdRd				−0.487	−0.384	−0.596
				(0.602)	(0.678)	(0.671)
Observations	2,545	2,545	2,545	599	599	599
R^2	0.003	0.023	0.030	0.005	0.064	0.124
Adjusted R^2	−1.347	−1.312	−1.311	−4.664	−4.656	−4.697
Note:				*p<0.1; **p<0.05; ***p<0.01		

IX. Appendix

R Code on Github

All prelimary data cleaning code and exploratory data analysis can be viewed at https://github.com/yesh747/RuralRoadsIndia. Instructions on navigating these files on be found directly on the github.

Data cleaning, data analysis and creating the paper itself was done using R Markdown in R Studio. Graphs were generated using ggplot2 package (Hadley and Chang, 2016). Panel regressions were performed using the plm package (Croissant and Millo, 2008). Tables and graphs were generated using the stargazer package (Hlavac, 2015). Data Manipulation was done using dplyr (Hadley and Francois, 2016). Robust, heteroskedastic consistent standard errors were calculated using the sandwich package (Zeileis, 2004).

R Package References

[1] Achim Zeileis. 2004. "Econometric Computing with HC and HAC Covariance Matrix Estimators." Journal of Statistical Software, 11(10), 1-17. http://www.jstatsoft.org/v11/i10/.

[2] Croissant, Yves, and Giovanni Millo. 2008. "Panel Data Econometrics in R: The plm Package." Journal of Statistical Software, 27(2). http://www.jstatsoft.org/v27/i02/.

[3] Hadley, Wickham, and Romain Francois. 2016. "dplyr: A Grammar of Data Manipulation." R package version 5.0. http://CRAN.R-project.org/package=dplyr.

[4] Hadley, Wickham, and Winston Chang. 2016. "ggplot2: Elegant Graphics for Data Analysis." Springer-Verlag New York, 2009. R package version 5.2. http://CRAN.R-project.org/package=ggplot2.

[5] Hlavac, Marek. 2015. "stargazer: Well-Formatted Regression and Summary Statistics Tables." R package version 5.2. http://CRAN.R-project.org/package=stargazer.