Impacts of rural roads on household welfare in Vietnam: evidence from a replication study

Cuong Viet Nguyen
National Economics University, Hanoi, Vietnam

Received 2 March 2019
Revised 22 May 2019
Accepted 30 May 2019

Abstract

Purpose – Recently, there has been a call for replication research to validate empirical findings, especially findings that are important for development policies. Thus, the purpose of this paper is to replicate the estimation results from Mu and van de Walle (2011).
Design/methodology/approach – The author used raw data sets provided by Mu Ren and Dominique van de Walle and the same methods of Mu and van de Walle (2011). In addition to the pure replication, the author conducted the two extensions: sensitivity analysis of covariates and bandwidth selection and analysis of the effect of the road project on additional outcome variables.
Findings – Overall, the author ables to replicate most estimates from Mu and van de Walle (2011). The author find a positive effect of rural roads on local market development. The impact estimates of the road project are not sensitive to the selection of the bandwidth in kernel propensity score (PS) matching. There are no significant effects of road projects on additional outcomes, including access to credit and migration.
Practical implications – The study confirms a positive effect of rural roads on local market development. Thus, the government can provide investment in rural roads to improve the local market and its welfare.
Originality/value – This study tried to replicate and verify an important study on the impact of the rural road in Vietnam.

Keywords Vietnam, Propensity score matching, Impact evaluation, Replication, Rural roads
Paper type Research paper

1. Introduction

In recent years, there has been a remarkably increasing number of empirical socioeconomic studies. Empirical studies are important for not only researchers but also policy makers in designing socioeconomic policies. Most empirical studies rely on large-scale data sets and econometric methods to test research hypotheses. Findings from empirical studies depend heavily on the methodology selection and how data are analyzed. Even by using the same method and data sets, there can be different ways that researchers can define and select variables for model estimation, and as a result, these different ways can lead to different findings and policy recommendations. Thus, there is a call for replication research to validate empirical findings, especially important findings for development policies (Brown *et al.*, 2014). Replication research not only confirms the validity of replicated studies but also raises the importance of analyzing, documenting and keeping empirical data during the research.

The author would like to thank Mu Ren and Dominique van de Walle for generously providing me with not only the raw original data sets but also analysis do-files. Without their help, this replication work cannot be done. They also gave me useful comments on the reports. The author would also like to thank Benjamin Wood and anonymous reviewers for his help and very useful comments during this study.

Journal of Economics and Development
Vol. 21 No. 1, 2019
pp. 83-112
Emerald Publishing Limited
e-ISSN: 2632-5330
p-ISSN: 1859-0020
DOI 10.1108/JED-06-2019-0002

In this study, I tried to replicate the study of Mu and van de Walle (2011, pp. 709-34)[1]. Mu and van de Walle (2011) aim to measure the effect of rural roads on local market development in Vietnam. They test a hypothesis called "transport-induced local-market development" using data from surveys of "Vietnam Rural Transport Project I" and double differences with propensity score-matching methods. They conclude that rural roads raise local market development. By using regressions, they also find that there is heterogeneity in the impact of rural roads. The impact of rural roads tends to be higher for poorer communes, since the poorer communes have low base levels of market development.

There are several reasons for selection of this study for replication. First, rural roads play a crucial role in the socioeconomic development of rural areas (World Bank, 1994; Gannon and Liu, 1997; Lipton and Ravallion, 1995; Jalan and Ravallion, 2001). Jalan and Ravallion (2001) point out that rural roads are a necessary element for fostering rural income growth and reducing poverty. Rural roads can increase household income, including both farm and nonfarm income. Rural roads increase agricultural productivity by reducing transportation costs, increasing access to advanced technology, increasing capital and enabling the employment of labor from outside local areas. In addition, rural roads can also increase nonfarm production and nonfarm employment opportunities for local people. Mu and van de Walle (2011) provide findings on the important role of rural roads in nonfarm employment and market development. Until the end of 2013, according to the Google Scholar citation system, this paper (together with the working paper version) has been cited in 125 studies. It is important to validate its estimates and results using the original data sets.

Second, there are a large number of arguments that local market development can increase household welfare. However, there is little if anything known about the effect of public investment in transport on local market development. Most empirical studies focus on the effect of rural roads on household income and find a positive effect of rural roads on nonfarm income, e.g., Balisacan *et al.* (2002), Fan *et al.* (2002), Corral and Reardon (2001), Escobal (2001) and Nguyen (2011)[2]. Thus, Mu and van de Walle (2011) provide important evidence on the effect of rural roads on local market development. As is known, market accessibility is an important channel through which rural roads can help local people to improve nonfarm activities, income and consumption and expenditure.

Third, Vietnam is a developing country with more than two-thirds of the population living in rural areas and 95 percent of the poor living in rural areas. An important poverty reduction program in Vietnam is to improve the infrastructure for rural areas, especially those with a high poverty rate and a higher proportion of ethnic minorities. State and international agencies work continuously to improve and maintain the infrastructure, including roads[3]. In Mu and van de Walle (2011), rural roads are found to be an important factor in local market development and the effect of rural roads is higher for the poor areas. This finding is very important for policy makers in designing poverty reduction programs in Vietnam.

Fourth, the findings from Mu and van de Walle (2011) can be used for other developing countries, especially for some Asian developing countries with similar economic structures as Vietnam, such as the Philippines, Indonesia, Laos and Cambodia. Rural roads can help local market development in the short run, as a result, enhancing nonfarm employment, increasing income and reducing poverty in the long run.

In this study, I first conduct a pure replication of the study of Mu and van de Walle (2011). Mu Ren and Dominique van de Walle provided us with the raw original data sets, which allow us to replicate their published estimates. The pure replication includes the following basic steps: Reconstruct all the variables used in the study; Recalculate descriptive statistics of all the variables using the raw data; Re-estimate the results in the original study using the original specifications.

Second, I also conducted the so-called statistical replication to examine the sensitivity of the impact estimates to different sets of covariates and bandwidth used in the propensity

score (PS) matching. One of the key issues in the propensity score-matching method is to select covariates and bandwidth and there are no standard criteria for this selection. Different selections produce different comparison groups and as a result different estimates of the program impacts. Thus, it is important to investigate whether the main findings from an empirical study are robust to different model specifications.

Third, I will go beyond the outcomes that are considered in Mu and van de Walle (2011) (including market accessibility, nonfarm employment, and child education), and estimate the effect of the road project on additional outcome variables, including access to credit and migration[4]. These outcomes are important for the livelihood and nonfarm diversification of rural households, and can provide policy-relevant findings.

The report is structured into five sections. The second section describes the method and data in Mu and van de Walle (2011). The third section presents the pure replication results. The fourth section presents the results from statistical replication. Finally, the fifth section describes the conclusion.

2. Data and methods in Mu and van de Walle (2011)

Mu and van de Walle (2011) assess the impact of "the Vietnam Rural Transport Project I," which implemented the rehabilitation of 5,000 km of rural roads in communes in 18 provinces in Vietnam. The project was implemented during 1997–2001. Data used in Mu and van de Walle (2011) were collected before and after the project. This data set is called the Survey of Impacts of Rural Roads in Vietnam (SIRRV). More specifically, a panel data of 3000 households in 200 communes were conducted in 1997, 1999, 2001 and 2003. In total, 15 households were sampled from each commune. There are 100 communes in the project areas, and 100 communes from the non-project areas. Mu and van de Walle (2011) use commune data sets in 1997 (the baseline survey), 2001, and 2003 (the mid-term and endline surveys) for impact evaluation.

The endogeneity bias in the impact evaluation of "the Vietnam Rural Transport Project I" can happen because the project placement is not random. Provinces were allowed to select communes for the projects and the road links to be rehabilitated. There are several criteria for the selection of communes and road links such as cost, population density, and share of the ethnic minority population. However, these criteria are not well documented in the project documents, and it is not clear how the selection process actually happened (Mu and van de Walle, 2011). For most large-scale projects in Vietnam, it is very difficult to conduct a randomization or well-defined regression discontinuity impact evaluation (Nguyen, 2013). To solve the problem of endogeneity, Mu and van de Walle (2011) used the difference-in-difference (DD) estimator. This method controls the difference in outcomes between the treatment and control groups caused by observed variables and the time-invariant difference caused by unobserved variables. In other words, it assumes that the difference in no-project outcomes between the treatment and control groups (once observed variables are controlled for) was the same before and after the project.

Mu and van de Walle (2011) combine the DD with PS matching to estimate the effect of the rural road project on communes' market development. They estimate the average treatment effect on the treated group. According to their denotation, the estimator is expressed as follows:

$$DD = \sum_{N_P} DD_i / N_P, \tag{1}$$

where:

$$DD_i = \left(Y_{i1}^{P} - Y_{i0}^{P}\right) - \sum_{j} W_{ij}\left(Y_{j1}^{NP} - Y_{j0}^{NP}\right), \tag{2}$$

where DD_i is the estimate for the project commune i. P and NP denote the treatment (project commune) and control (non-project commune), respectively. Subscripts "1" and "0" denote

the outcome after and before the project, respectively. W indicates weights applied to the comparison communes when they are matched with the treatment communes.

Mu and van de Walle (2011) use the kernel PS matching (Heckman *et al.*, 1997) and propensity score-weighted difference-in-differences (Hirano and Imbens, 2002; Hirano *et al.*, 2003) to estimate the impact. A logit regression is used to predict the propensity score. Control variables are commune characteristics in the base year 1997. The list of control variables is presented in Tables AIII and AIV. The list of outcome variables is presented in Table II in the next section.

After estimating the effect of the rural roads on the outcomes for each commune (i.e., DD_i), Mu and van de Walle (2011) run regression of DD_i on commune characteristic variables to examine whether the effect of rural roads varies across communes of different characteristics as follows:

$$DD_i = \alpha + X_i\beta + \varepsilon_i, \tag{3}$$

where DD_i is the estimated impact on an outcome for commune i, and X_i is a vector of explanatory variables of commune i.

3. Replication results

In this section, I aim to conduct pure replication of the results from Mu and van de Walle (2011). The pure replication includes the three following basic steps: reconstruct all the variables used in the study; recalculate descriptive statistics of all the variables using the raw data; and re-estimate the results in the original study using the original specifications.

3.1 Raw data sets and do-files

As mentioned, Mu and van de Walle (2011) use commune data sets in 1997 (the baseline survey), 2001, and 2003 (the mid-term and endline surveys) for impact evaluation of the rural road project. The original authors (Mu and Van de Walle) are very generous to provide me with not only the raw original data sets but also their analysis do-files (they used Stata for analysis). These data sets and do-files are used for estimation for not only the study by Mu and van de Walle (2011) but also for the study by Van de Walle and Mu (2007). The authors mentioned that they sent all the data and do-files available in their current computers. However, since the analysis was conducted by the authors a very long time ago (before 2007), do-files that are used to estimate the results of Mu and van de Walle (2011) are not fully available. It means that I cannot simply rerun the do-files sent by Mu and van de Walle to replicate their results, since some do-files are missing.

Figure 1 summarizes the data sets and do-files provided by Ren Mu and Dominique van de Walle. The Shapes 1, 2, 3 and 4 mean that data or do-files are fully available, while the "pink" shapes mean that data or do-files are just partially available. Shape 7, i.e., "Do-files to create data for analysis," is not available. Running "Do-files to estimate the impacts" (Shape 6) using "Data for impact estimation" (Shape 5) does not produce the results of Mu and van de Walle (2011), since some do-files as well as data variables are missing. I checked all the available do-files including those to create data sets and those to estimate the project impact, and find no problems.

3.2 Reconstruct all variables and recalculate descriptive statistics

In the next step, I use the raw data sets provided by the authors to create the outcome variables and the control variables that are used to estimate the project impact. Table I is replicated in Mu and van de Walle (2011). After checking the do-files, data, and questionnaires carefully, I still cannot produce the same estimates as Table I in Mu and van de Walle (2011). Table I in this study adds the column reporting the percentage difference in the outcome means between the replication and the original paper. Variables with 0 percent difference have

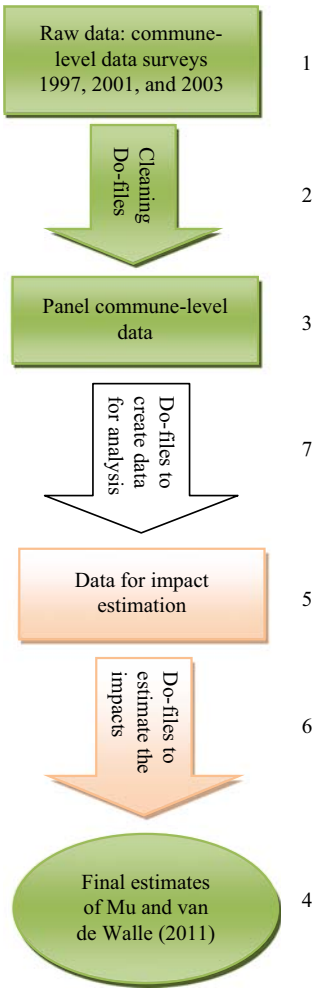

Figure 1.
Data sets and do-files

the same values as the original papers. There are 12 variables that are the same. There are four variables that differ by more than 10 percent from those from the original papers. For the remaining seven variables, the difference in the mean is less than 10 percent.

Next, I estimated the outcome variables for the years 1997, 2001 and 2003. Table AI replicates the results of Table II in Mu and van de Walle (2011). The outcomes are estimated for communes within the common support of the predicted propensity scores. In Mu and van de Walle (2011), there are 94 project and 95 non-project communes on common support. In this study, I estimated the PS using the same model specification. However, the regression results are not the same (see the next section for detailed presentation). As a result, the predicted PS is not the same, and the common support is different from Mu and van de Walle (2011). There are 85 project and 83 non-project communes on common support. The mean outcomes of project and non-project communes cannot be the same as those in Mu and van de Walle (2011) due to different common supports. However, the difference in the replicated results and the original results is not large.

Commune characteristics	Variable type	Below median (1)	Above median (2)	Difference	Difference between these and the original paper (%)
Typology: mountain	Binary	0.70	0.33	0.37***	0
Distance to the closest central market (km)	Continuous	16.09	10.46	5.63***	< 10
Share of households owning motorcycles	Continuous	6.32	10.00	−3.68***	< 10
Population density	Continuous	2.14	5.20	−3.06***	< 10
Ethnic minority share	Continuous	0.67	0.20	0.48***	0
Adult illiteracy rate	Continuous	0.11	0.03	0.07***	> 10
Flood and storm prevalence	Binary	0.60	0.64	−0.04	0
Credit availability	Binary	0.27	0.30	−0.03	> 10
North provinces	Binary	0.54	0.66	−0.12*	0
Transportation accessibility	Binary	0.23	0.31	−0.09***	0
Road density	Continuous	0.01	0.02	−0.01***	0
Market availability	Binary	0.31	0.66	−0.35***	< 10
Market frequency	Discrete	0.72	1.43	−0.71***	0
Shop	Binary	0.39	0.58	−0.19***	0
Bicycle repair shop	Binary	0.54	0.88	−0.34***	< 10
Pharmacy	Binary	0.34	0.75	−0.41***	0
Restaurant	Binary	0.23	0.44	−0.21***	0
Women's hair dressing/Men's barber	Binary	0.33	0.74	−0.41***	> 10
Men and women's tailoring	Binary	0.56	0.92	−0.36***	< 10
% farm households	Continuous	93.64	86.34	7.29***	0
% trade households	Continuous	1.17	1.70	−0.53*	0
% service sector households	Continuous	0.69	1.08	−0.39	< 10
Primary school completion (less than 15 years)	Continuous	53.78	68.89	−15.11***	> 10
Secondary school enrollment rate	Continuous	76.81	94.13	−17.32***	< 10

Notes: Table I replicates the estimates of Table I in Mu and van de Walle (2011). The definition of variables and sample is the same as the Mu and van de Walle (2011). *,**,***Significant at 10, 5 and 1 percent levels, respectively
Source: Author's estimation

Table I.
Mean baseline characteristics and outcome variables for communes classified by median household per capita consumption (log)

I found a variable of the predicted PS in the data sets sent by Mu and Van de Walle. By using this propensity score, I am able to define the common support as Mu and van de Walle (2011) (including 94 project and 95 non-project communes). Using this common support, I re-estimated the outcomes of project and non-project communes, and reported the results in Table AII. Now, there are five outcome variables (which are marked with a star *) which have the same value as the original paper.

There is a problem of the variable "Primary school completion (< 15 years)" which has very high values in 1997 but low values in 2001 and 2003. My estimates of "Primary school completion (< 15 years)" for 2001 and 2003 are close to the estimates in Mu and van de Walle (2011). However, my estimate for 1997 is substantially higher than that in Mu and van de Walle (2011). I checked the data set carefully, but cannot find the reason for this problem. A possible reason for the difference might be that the raw data sets that Mu and Van de Walle provided for me are not the same raw data sets used for Mu and van de Walle (2011). Data collectors sometimes clean and update cleaned data sets. As a result, different versions of data sets might exist.

3.3 Re-estimate the results in the original study using the original specifications

After constructing the variables and producing descriptive analysis, I estimate the impact of the rural road project on commune outcomes using the original specifications. The first step is to estimate the PS using logit regression. The logit estimation is presented in

Table II.
Impacts of road rehabilitation/building for year 2001

Outcomes	Simple DD			PS kernel matched DD			PS weighted DD		
	DD	*t*-ratio	Original estimates in Mu and van de Walle (2011)	PS kernel matched DD	*t*-ratio	Original estimates in Mu and van de Walle (2011)	PS weighted DD	*t*-ratio	Original estimates in Mu and van de Walle (2011)
Market									
Market availability	−0.01	−0.16	0.00	0.03	0.91	0.03	0.03	0.85	0.04
Market frequency	0.07	0.49	0.01	0.14	1.57	0.08	0.15	1.44	0.10
Shop	−0.05	−0.57	−0.02	−0.13	−1.23	0.01	−0.15	−1.35	0.08
Bicycle repair shop	−0.09	−1.60	−0.08*	−0.06	−1.26	−0.06	−0.06	−1.04	−0.04
Pharmacy	0.09	1.44	0.08	0.05	0.70	0.04	0.04	0.57	−0.06
Restaurant	0.11*	1.89	−0.03	0.13*	1.69	−0.01	0.14*	1.94	−0.01
Women's hair dressing/Men's barber	0.02	0.33	−0.04	0.06	0.73	−0.07	0.06	1.05	−0.07
Men and women's tailoring	0.01	0.19	0.12	0.00	0.04	0.11	0.00	0.08	0.10
Employment: % households whose main occupation is									
% farm households	−0.77	−0.47	0.04	−0.73	−0.45	0.05	−0.42	−0.29	0.03
% trade households	0.10	0.23	−0.05	−0.23	−0.34	0.03	−0.59	−0.68	0.03
% service sector households	−0.65	−1.61	−0.06	−0.18	−0.40	−1.54	0.07	0.14	−1.03
School enrollments									
Primary school completion (< 15 years)	−3.71	−0.65	0.00	1.82	0.27	0.15**	4.08	0.65	0.25**
Secondary school enrollment rate	−0.52	−0.16	0.06	1.03	0.33	0.10	0.56	0.19	0.25

Notes: Table II replicates the estimates of Table III in Mu and van de Walle (2011); the sample consists of the 85 project and 83 non-project communes on common support as determined by propensity score matching. *t*-Ratio of kernel matching is obtained from bootstrapping (100 repetitions); standard errors of weighted DD estimations are robust to heteroskedasticity and serial correlation of communes within the same district. *,**Significant at 10 and 5 percent levels, respectively
Source: Author's estimation

Van de Walle and Mu (2007, pp. 667–685). I am not able to produce the same logit result as Van de Walle and Mu (2007). The summary statistics of the explanatory variables (covariates) in the logit regression is presented in Table AIII. In Van de Walle and Mu (2007), the number of observations is 200. The number of observations in this logit regression is 198. There are missing values in some variables, and I do not know how these missing values are treated in Van de Walle and Mu (2007). In this replication study, I dropped two observations with missing values. It means that these dropped two communes are not used for impact estimation. In the logit regression (Table AIV), most explanatory variables have the same sign and close point estimates as the original paper of Van de Walle and Mu (2007). Since the logit regression results are different, the predicted propensity scores are also different from the original paper.

Figure A1 presents the predicted PS for the treatment (project communes) and control groups (non-project communes). There are 85 project and 83 non-project communes on common support. This is different from Mu and van de Walle (2011), in which there are 94 project and 95 non-project communes on common support.

Tables II and III present the impact estimation of the rural road project using the original specifications and methods (these estimates replicate Table III in Mu and van de Walle, 2011). In Stata, I used the command "psmatch2" like Mu and van de Walle, 2011. Mu and van de Walle (2011) used the default bandwidth which is 0.06 in the kernel PS matching. The original estimates in Mu and van de Walle (2011) are also reported in Tables II and III for comparison. The replicated estimates are not the same as the original paper, since the predicted PS as well as the common support are different. However, most of the impact estimates for 2003 have the same sign as the impact estimates in the original paper.

As mentioned, I found a variable of the predicted PS in the data sets sent by Mu and Van de Walle. I used this predicted PS variable to estimate the effect of the project on the five outcome variables that have the same value as the original paper. Table IV presents the results of this analysis. I cannot replicate the impact estimates for the year 2001. However, for the year 2003, I am able to replicate the same impact estimates as the original paper. It means that the difference between the replicated results and the original results lies in the construction of variables, not in the methodology.

An interesting analysis in Mu and van de Walle (2011) is to examine the determinants of heterogeneous impacts of the rural road project. More specifically, after estimating the effect of the rural roads on the outcomes for each commune, Mu and van de Walle (2011) run ordinary least-square (OLS) regressions of these specific impact estimates on commune characteristic variables to examine whether the effect of rural roads varies across communes of different characteristics. Overall, they find that there is some evidence on heterogeneity in the impact of rural roads. The impact of rural roads tends to be higher for the poorer communes, since the poorer communes have low base levels of market development.

In this study, I also run regressions of the predicted impact of the rural project on explanatory variables using commune-level data. The regression results are presented in Tables from AV to AX. None of our estimates are the same as Mu and van de Walle (2011), since their common supports are different, and some of the control variables are also different. However, most of the replicated estimates have the same sign as the point estimates in Mu and van de Walle (2011).

4. Statistical replication

After conducting pure replication, I conducted the so-called statistical replication. In the statistical replication, I conduct the two extensions: sensitivity analysis of covariates and bandwidth selection, and analysis of the effect of the road project on additional outcome variables.

Table III.
Impacts of road rehabilitation/building for year 2003

Outcomes	Simple DD			PS kernel matched DD			PS weighted DD		
	DD	*t*-ratio	Original estimates in Mu and van de Walle (2011)	PS kernel matched DD	*t*-ratio	Original estimates in Mu and van de Walle (2011)	PS weighted DD	*t*-ratio	Original estimates in Mu and van de Walle (2011)
Market									
Market availability	0.07	1.27	0.09*	0.08**	2.28	0.08*	0.08**	2.00	0.09**
Market frequency	0.16	1.02	0.19	0.18	1.60	0.23*	0.18	1.28	0.25**
Shop	−0.05	−0.71	0.03	−0.14	−1.52	0.08	−0.17*	−1.70	0.14
Bicycle repair shop	−0.05	−0.94	−0.04	−0.05	−0.73	0.02	−0.05	−0.92	0.03
Pharmacy	0.14*	1.93	0.14*	0.16*	1.74	0.12	0.14	1.54	0.16
Restaurant	0.08	0.83	0.05	0.04	0.47	0.01	0.04	0.36	0.05
Women's hair dressing/Men's barber	0.05	0.95	0.14*	0.08	1.04	0.18**	0.08	1.31	0.20**
Men and women's tailoring	0.03	0.56	0.09	0.03	0.42	0.10	0.02	0.36	0.12*
Employment: % households whose main occupation is									
% farm households	−2.10	−1.35	−1.99	−2.49	−1.56	−2.04*	−2.81**	−2.11	−2.06**
% trade households	0.70	1.41	0.57	0.80	1.47	0.36	0.70	1.22	0.58
% service sector households	0.75**	2.40	1.01*	1.09**	2.16	1.68**	1.31*	2.04	1.72**
School enrollments									
Primary school completion (< 15 years)	2.52	0.37	0.04	10.13	1.45	0.17**	9.89	1.35	0.30**
Secondary school enrollment rate	−0.92	−0.31	0.10**	0.58	0.20	0.05	0.35	0.13	0.07*

Notes: Table III replicates the estimates of Table III in Mu and van de Walle (2011); The sample consists of the 85 project and 83 non-project communes on common support as determined by propensity score matching. *t*-Ratio of kernel matching is obtained from bootstrapping (100 repetitions). Standard errors of weighted DD estimations are robust to heteroskedasticity and serial correlation of communes within the same district. *,**Significant at 10 and 5 percent levels, respectively
Source: Author's estimation

Table IV.
Impacts of road rehabilitation/building on market access for years 2001 and 2003

Outcomes	Simple DD			PS kernel matched DD			PS weighted DD		
	DD	*t*-ratio	Original estimates in Mu and van de Walle (2011)	PS kernel matched DD	*t*-ratio	Original estimates in Mu and van de Walle (2011)	PS weighted DD	*t*-ratio	Original estimates in Mu and van de Walle (2011)
Impacts in 2001									
Market availability	−0.00	−0.09	0.00	0.04*	1.90	0.03	0.04	1.06	0.04
Bicycle repair shop	−0.08*	−1.76	−0.08*	0.01	0.26	−0.06	−0.04	−0.76	−0.04
% farm households	−0.28	−0.18	0.04	−1.02	−0.62	0.05	1.31	0.79	0.03
% trade households	−0.06	−0.14	−0.05	0.18	0.16	0.03	−1.03	−0.94	0.03
% service sector households	−0.68	−1.60	−0.06	0.84*	2.05	−1.54	0.10	0.26	−1.03
Impacts in 2003									
Market availability	0.09*	1.83	0.09*	0.08*	1.85	0.08*	0.09**	2.19	0.09**
Bicycle repair shop	−0.04	−0.89	−0.04	0.02	0.37	0.02	0.03	0.58	0.03
% farm households	−1.99	−1.25	−1.99	−2.04*	−1.67	−2.04*	−2.06*	−1.87	−2.06**
% trade households	0.57	1.26	0.57	0.36	0.71	0.36	0.58	1.35	0.58
% service sector households	1.01**	2.52	1.01*	1.68***	2.43	1.68**	1.72***	3.10	1.72**

Notes: Table IV replicates the estimates of Table III in Mu and van de Walle (2011); The sample consists of the 94 project and 95 non-project communes on common support as determined by the propensity score obtained from the original paper. *t*-Ratio of kernel matching is obtained from bootstrapping (100 repetitions) Standard errors of weighted DD estimations are robust to heteroskedasticity and serial correlation of communes within the same district. *,**Significant at 10 and 5 percent levels, respectively

Source: Author's estimation

4.1 Sensitivity analysis of covariates and bandwidth selection

Analysis methods. The main advantage of PS matching is that it does not rely on assumptions of functional forms of outcomes. However, the point estimates as well as the standard errors of the propensity score-matching estimators can be sensitive to the selection of control variables used in the logit (or probit) model to estimate the propensity score. The estimates might also be sensitive to the magnitude of the bandwidth in kernel matching. Thus, in the replication study, I also examine the sensitivity of the impact estimates to different bandwidths used in kernel matching.

The list of control variables (covariates) used in Mu and van de Walle (2011) is presented in Tables AIII and AIV. Variables that affect outcomes and program selection should be controlled in PS estimation. Obviously, variables which affect both the program participation and outcomes should be included in the PS model (e.g., Ravallion, 2001; Caliendo and Kopeinig, 2008). Bryson *et al.* (2002) argue that inclusion of irrelevant variables can increase the standard error of estimates. Zhao (2008) finds that overspecification of the model of the PS can bias impact estimates. However, using simulation, Nguyen (2013) shows that efficiency in the estimation of the average treatment effect on the treated group can be gained if all the variables in the outcome equation are included in the estimation of propensity scores.

A challenge in measuring the impact of "Vietnam Rural Transport Project I" is that the project selection is not fully observed. Although there are several criteria for the selection of communes and road links such as cost, population density, and share of the ethnic minority population, the actual selection of the project communes is not clear and documented (Mu and van de Walle, 2011). In addition, there are a number of outcomes, and different outcomes can be affected by different explanatory variables. Thus, Mu and van de Walle (2011) control variables that are important for program selection and other variables that can affect the program selection and outcomes. The control variables are listed in Tables AIII and AIV.

In the replication study, I can examine the sensitivity of the program impact to two additional sets of control variables as follows:

(1) Add pretreatment outcomes to the logit regression of the program selection. The pretreatment outcome can be used as control in the regression of the PS to reduce the difference in outcomes between the treatment and control groups in the baseline (Dehejia and Wahba, 1998; Smith and Todd, 2005).

(2) Limit the covariates to those that are statistically significant in the logit regression of the program selection. Several control variables are statistically significant in Mu and van de Walle (2011). They can be dropped, since these variables might affect the quality of matching of the key variables (Bryson *et al.*, 2002; Zhao, 2008).

I can also examine the sensitivity of the program impact estimates to the selection of bandwidth. Mu and van de Walle (2011) used the default bandwidth which is 0.06 in the kernel matching. In the study, I can use other bandwidths, e.g., 0.01, 0.03 and 0.09 for robust analysis. In addition, I can use a cross-validation method – a widely used selection method of bandwidth in PS matching (Frolich, 2004; Galdo *et al.*, 2010). This method selects the bandwidth as follows:

$$h^{CV} = \arg\min_h \left(\frac{1}{n_0} \sum_{j=1}^{n_0} \left(y_{0j} - \hat{m}_{-j}(p_j, h) \right)^2 \right), \tag{4}$$

where n_0 is the number of control units, y_{0j} is the outcome of the control unit j, and $\hat{m}_{-j}(p_j, h)$ is the estimated conditional mean for the control unit j at the PS p_j using all the control units

within the bandwidth but with the exception of unit j. The bandwidth that has the smallest value of h^{CV} will be selected.

Empirical results. Table V presents the impact estimates of the road project using difference-in-differences with the PS kernel-matching method. It replicates the PS kernel-matched DD estimates in Tables II and III. The difference between the estimation method in Table V and the estimation method in Tables II and III is that the propensity scores used in Table V are estimated by using not only the covariates but also the baseline outcome variable (variable in 1997). For each outcome, the corresponding baseline variable is added to the logit regression. Thus, the logit model differs for different outcomes. Although the results are not the same as those of Mu and van de Walle (2011), most impact estimates have the same sign as those of Mu and van de Walle (2011). Similar to Mu and van de Walle (2011), the effect of the project on the market and the percentage of farming households is statistically significant.

In Table VI, the propensity scores are estimated using the logit regressions in which only covariates significant at the 10% level are kept. The results show that most estimates have the same sign as those in Mu and van de Walle (2011). However, the effect is not significant for almost all outcomes.

As mentioned, Mu and van de Walle (2011) used the default bandwidth, which is 0.06 in the kernel matching. There are no standard criteria to select the bandwidth. Using a large bandwidth results in a larger number of matched controls. This reduces the standard error, but increases potential bias, since I can match a participant with a very different nonparticipant. On the contrary, using a small bandwidth can reduce the bias but increase the standard error of the impact estimates. I can vary the bandwidth to examine whether the impact estimates are sensitive to different bandwidths. In Tables from AXI to AXIII, I used other bandwidths, e.g., 0.01, 0.03 and 0.09 for robust analysis. Three bandwidth schemes produce the same sign of the

	2001			2003		
Outcomes	PS kernel matched DD	*t*-ratio	Original estimates in Mu and van de Walle (2011)	PS kernel matched DD	*t*-ratio	Original estimates in Mu and van de Walle (2011)
Market availability	0.029	0.771	0.03	0.084**	2.260	0.08*
Market frequency	0.119	1.298	0.08	0.199*	1.803	0.23*
Shop	−0.080	−0.618	0.01	−0.115	−0.905	0.08
Bicycle repair shop	−0.012	−0.273	−0.06	0.020	0.438	0.02
Pharmacy	0.035	0.377	0.04	0.098	0.789	0.12
Restaurant	0.103	1.546	−0.01	0.003	0.029	0.01
Women's hair dressing/ Men's barber	0.071	1.038	−0.07	0.078	1.184	0.18**
Men and women's tailoring	0.026	0.523	0.11	0.039	0.674	0.10
% farm households	−0.263	−0.182	0.05	−3.293*	−1.872	−2.04*
% trade households	−1.575	−1.596	0.03	0.514	1.130	0.36
% service sector households	0.524	0.950	−1.54	2.273	2.562	1.68**
Primary school completion (< 15 years)	9.670*	1.777	0.15**	12.483**	1.992	0.17**
Secondary school enrollment rate	0.594	0.115	0.10	1.245	0.276	0.05

Notes: The sample consists of project and non-project communes on common support as determined by propensity score matching. *t*-Ratio of kernel matching is obtained from bootstrapping (100 repetitions). The propensity scores are estimated using logit models, which include covariates as Table AII and also outcome variables. *,**Significant at 10 and 5 percent levels, respectively

Source: Author's estimation

Table V.
Estimated impact of the road project using PS kernel matched DD – baseline outcome variable is controlled in estimating propensity scores

Outcomes	2001 PS kernel matched DD	2001 *t*-ratio	2001 Original estimates in Mu and van de Walle (2011)	2003 PS kernel matched DD	2003 *t*-ratio	2003 Original estimates in Mu and van de Walle (2011)
Market availability	0.000	0.004	0.03	0.064	1.198	0.08*
Market frequency	0.049	0.336	0.08	0.154	1.016	0.23*
Shop	0.001	0.014	0.01	−0.027	−0.316	0.08
Bicycle repair shop	−0.036	−0.703	−0.06	−0.013	−0.241	0.02
Pharmacy	0.044	0.554	0.04	0.063	0.732	0.12
Restaurant	0.100*	1.679	−0.01	0.050	0.492	0.01
Women's hair dressing/ Men's barber	0.045	0.639	−0.07	0.038	0.514	0.18**
Men and women's tailoring	0.040	0.790	0.11	0.022	0.361	0.10
% farm households	0.138	0.092	0.05	−1.349	−0.883	−2.04*
% trade households	−0.409	−0.703	0.03	0.317	0.677	0.36
% service sector households	−0.271	−0.736	−1.54	1.194**	1.976	1.68**
Primary school completion (< 15 years)	2.530	0.411	0.15**	6.056	1.169	0.17**
Secondary school enrollment rate	1.610	0.458	0.10	2.680	0.869	0.05

Notes: The sample consists of project and non-project communes on common support as determined by propensity score matching. The propensity scores are estimated using logit models in Table AIII. *t*-Ratio of kernel matching is obtained from bootstrapping (100 repetitions). *,**Significant at 10 and 5 percent levels, respectively
Source: Author's estimation

Table VI.
PS kernel matched DD – only covariates and baseline outcome variables, which are significant at the 10 percentlevel are controlled in estimating propensity scores

effect estimates of the project in 2003. However, the significance is slightly different between the three bandwidth schemes. For example, the effect of the road project on market availability is not significant, using a bandwidth of 0.01, while the effect of the road project on market availability is significant, using bandwidths of 0.03 and 0.09.

Finally, Table VII presents the estimates when an optimal bandwidth is used (Frolich, 2004; Galdo *et al.*, 2010). For each outcome, a bandwidth is estimated so that the difference in baseline outcomes between the treatment and control communes is minimized. The results are quite similar to those estimated using other bandwidths.

4.2 Additional outcome variables

Mu and van de Walle (2011) focus on the effect of the road project on market development, employment and education. Roads are very important for the rural economy. Thus, in this study, I examine the effect of the road project on additional outcome variables, by using the same method and data used by Mu and van de Walle (2011). The surveys contain very detailed data on commune living standards. The outcome variables are selected based on the data availability. The road project is also expected to have a significant effect on these outcomes.

The first outcome is the access to credit. The distance to banks and a credit institution is negatively correlated with the access to credit in Vietnam (Nguyen, 2008). Rural roads are expected to reduce the distance to lenders and increase the credit access of households. The second outcome is migration, out-migration and in-migration. Roads can reduce the cost of mobility and increase migration (Lucas, 2001).

Tables VIII and IX present the impact estimates of the project on credit and migration, using the same three methods as those by Mu and van de Walle (2011). Overall, there are no significant effects of the road project on credit access and migration of households in project communes.

Outcomes	2001 PS kernel matched DD	2001 *t*-ratio	2001 Original estimates in Mu and van de Walle (2011)	2003 PS kernel matched DD	2003 *t*-ratio	2003 Original estimates in Mu and van de Walle (2011)
Market availability	0.026	0.692	0.03	0.081**	2.201	0.08*
Market frequency	0.116	1.269	0.08	0.194*	1.782	0.23*
Shop	−0.058	−0.645	0.01	−0.083	−0.955	0.08
Bicycle repair shop	−0.050	−0.726	−0.06	−0.025	−0.306	0.02
Pharmacy	0.068	1.126	0.04	0.108*	1.727	0.12
Restaurant	0.087	1.542	−0.01	0.058	0.725	0.01
Women's hair dressing/ Men's barber	0.040	0.677	−0.07	0.048	0.828	0.18**
Men and women's tailoring	0.016	0.324	0.11	0.020	0.380	0.10
% farm households	−0.677	−0.440	0.05	−3.623	−1.935	−2.04*
% trade households	−0.066	−0.168	0.03	0.436	0.979	0.36
% service sector households	0.593	0.926	−1.54	2.447**	2.505	1.68**
Primary school completion (< 15 years)	4.230	0.805	0.15**	9.605	1.628	0.17**
Secondary school enrollment rate	2.480	0.614	0.10	1.632	0.488	0.05

Notes: The sample consists of 85 project and 83 non-project communes on common support as determined by propensity score matching. The propensity score is estimated by the logit model in Table AII. *t*-Ratio of kernel matching is obtained from bootstrapping (100 repetitions). *,**Significant at 10 and 5 percent levels, respectively
Source: Author's estimation

Table VII.
PS kernel matched DD – Optimal bandwidth

	Simple DD Estimates	Simple DD *t*-ratio	PS kernel matched DD Estimates	PS kernel matched DD *t*-ratio	PS weighted DD Estimates	PS weighted DD *t*-ratio
Number of credit sources available in communes	−0.050	−0.330	−0.090	−0.410	−0.148	−0.841
There is a branch of Agricultural Bank in commune	0.082	1.501	0.055	0.739	0.071	1.317
Number of households borrowing from a credit source	192.8**	1.997	139.1	1.098	95.05	0.676
% households in commune who borrowing from a credit source	8.171	1.367	6.992	1.109	5.393	0.723
Loan size per borrowing household (million VND)	−0.722	−1.093	−0.455	−0.815	−0.426	−0.521
There are private lenders in commune	−6.166	−0.671	1.685*	0.187	2.704	0.260
Percentage of people leaving commune temporarily	0.100	0.230	−0.096	−0.163	−0.191	−0.348
Percentage of men leaving commune temporarily	−0.041	−0.062	−0.255	−0.298	−0.349	−0.411
Percentage of women leaving commune temporarily	0.210	0.857	0.032	0.094	−0.057	−0.201
Percentage of households having member permanently leaving	1.015	0.906	1.789	1.069	2.115	1.189
Percentage of people coming to commune temporarily	0.006	0.018	−0.218	−0.885	−0.368	−1.384
Percentage of households coming to commune permanently	0.005	1.349	0.004	1.160	0.003	0.961

Notes: The sample consists of 85 project and 83 non-project communes on common support as determined by propensity score matching. The propensity score is estimated by the logit model in Table AII. *t*-Ratio of kernel matching is obtained from bootstrapping (100 repetitions). *,**Significant at 10 and 5 percent levels, respectively
Source: Author's estimation

Table VIII.
Impact of the road project on credit and migration in 2001

	Simple DD		PS kernel matched DD		PS weighted DD	
	Estimates	*t*-ratio	Estimates	*t*-ratio	Estimates	*t*-ratio
Number of credit sources available in communes	0.230	1.495	0.196	0.712	0.109	0.487
There is a branch of Agricultural Bank in commune	−0.036	−0.692	−0.013	−0.216	−0.001	−0.009
Number of households borrowing from a credit source	262.8*	1.909	236.5	1.590	192.4	1.125
% households in commune who borrowing from a credit source	10.400	1.613	9.307	1.267	7.416	0.887
Loan size per borrowing household (million VND)	41.243	1.010	0.975	0.876	41.167	1.009
There are private lenders in commune	−9.639	−0.920	−1.566	−0.143	−3.774	−0.388
Percentage of people leaving commune temporarily	−0.087	−0.218	−0.403	−0.818	−0.562	−1.265
Percentage of men leaving commune temporarily	−0.337	−0.611	−0.693	−1.067	−0.895	−1.535
Percentage of women leaving commune temporarily	0.174	0.588	−0.111	−0.288	−0.219	−0.630
Percentage of households having member permanently leaving	1.461	1.445	2.011	1.285	2.233	1.263
Percentage of people coming to commune temporarily	−0.437	−0.883	−0.989*	−1.645	−1.156	−1.560
Percentage of households coming to commune permanently	0.002	1.060	0.001	1.208	0.001	0.815

Notes: The sample consists of 85 project and 83 non-project communes on common support as determined by propensity score matching. The propensity score is estimated by the logit model in Table AII. *t*-Ratio of kernel matching is obtained from bootstrapping (100 repetitions). *,**Significant at 10 and 5 percent levels, respectively
Source:f Author's estimation

Table IX.
Impact of the road project on credit and migration in 2003

5. Conclusions

Rural roads are one of the key factors for rural development. Mu and van de Walle (2011) is an influential study, which finds a positive effect of rural roads on local market development in Vietnam. In this study, I tried to replicate the estimates of Mu and van de Walle (2011) using the raw data sets provided by the authors. I am able to produce quite similar results as those of the original paper. However, several estimates are not the same as those from the original paper. A possible reason for the difference is that the raw data sets that Mu and Van de Walle provided for me might not be the same raw data sets used for Mu and van de Walle (2011). Data collectors sometimes clean and update cleaned data sets. As a result, different versions of data sets might exist.

In addition to the pure replication, I conducted a so-called statistical replication. In the statistical replication, I conducted two extensions: Sensitivity analysis of covariates and bandwidth selection, and analysis of the effect of the road project on additional outcome variables. I find that the impact estimates of the road project are not sensitive to the selection of the bandwidth in kernel PS matching. However, using only covariates that are significant in the logit regression tends to reduce the statistical significance of the impact estimates. Finally, there are no significant effects of the road project on credit access and migration of households in project communes.

Overall, I find similar findings on the impact of the rural road project as those of Mu and van de Walle (2011). It indicates that there is a positive effect of rural roads on local market development. Thus, the government can provide investment in rural roads to improve the local market and its welfare.

Notes

1. Two-related papers of this article are Van de Walle and Mu (2007) and Mu and van de Walle (2007).
2. A review on empirical studies of the impact of rural roads can be found in Ali and Pernia (2003).
3. According to Donnges *et al.* (2007), Vietnam had a rural road network consisting of approximately 175,000 km in 2007. Around 73 percent of rural villages can be accessed by a good road (tar on gravel) (according to VietNam Household Living Standard Survey in 2010).
4. There are no data on consumption expenditure in the data set.

References

Ali, I. and Pernia, E.M. (2003), "Infrastructure and poverty reduction. What is the connection?", ERD Policy Brief No. 13, Asian Development Bank, Manila.

Balisacan, A.M., Pernia, E.M. and Asra, A. (2002), "Revisiting growth and poverty reduction in Indonesia: what do subnational data show?", ERD Working Paper Series No. 25, Economics and Research Department – Asian Development Bank, Manila.

Brown, A., Cameron, B. and Wood, B. (2014), "Quality evidence for policymaking: I'll believe it when I see the replication", *Journal of Development Effectiveness*, Vol. 6 No. 3, pp. 215-235.

Bryson, A., Dorsett, R. and Purdon, S. (2002), "The use of propensity score matching in the evaluation of labour market policies", Working Paper No. 4, Department for Work and Pensions, London School of Economics and Political Science, London.

Caliendo, M. and Kopeinig, S. (2008), "Some practical guidance for the implementation of propensity score matching", *Journal of Economic Surveys*, Vol. 22 No. 1, pp. 31-72.

Corral, L. and Reardon, T. (2001), "Rural nonfarm incomes in Nicaragua", *World Development*, Vol. 29 No. 3, pp. 427-442.

Dehejia, R.H. and Wahba, S. (1998), "Propensity score matching methods for non-experimental causal studies", NBER Working Paper No. 6829, National Bureau of Economic Research, Cambridge, MA.

Donnges, Ch, Edmonds, G. and Johannessen, B. (2007), *Rural Road Maintenance – Sustaining the Benefits of Improved Access*, International Labour Office, Bangkok.

Escobal, J. (2001), "The determinants of nonfarm income diversification in rural Peru", *World Development*, Vol. 29 No. 3, pp. 497-508.

Fan, S., Zhang, L.X. and Zhang, X.B. (2002), "Growth, inequality, and poverty in rural China: the role of public investments", Research Report No. 125, International Food Policy Research Institute, Washington, DC.

Frolich, M. (2004), "Finite-Sample properties of propensity-score matching and weighting estimators", *Review of Economics and Statistics*, Vol. 86 No. 1, pp. 77-90.

Galdo, J., Smith, J. and Black, D. (2010), "Bandwidth selection and the estimation of treatment effects with unbalanced data", *Annals of Economics and Statistics*, Vols 91/92, July-December, pp. 189-216.

Gannon, C. and Liu, Z. (1997), "Poverty and transport", TWU discussion papers – TWU-30, World Bank, Washington, DC.

Heckman, J., Ichimura, H. and Todd, P. (1997), "Matching as an econometric evaluation estimator: evidence from evaluating a job training program", *Review of Economic Studies*, Vol. 64 No. 4, pp. 605-654.

Hirano, K. and Imbens, G. (2002), "Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization", *Health Services and Outcomes Research Methodology*, Vol. 2 Nos 3-4, pp. 259-278.

Hirano, K., Imbens, G. and Ridder, G. (2003), "Efficient estimation of average treatment effects using the estimated propensity score", *Econometrica*, Vol. 71 No. 4, pp. 1161-1189.

Jalan, J. and Ravallion, M. (2001), "Geographic poverty traps? A micro econometric model of consumption growth in rural China", *Journal of Applied Econometrics*, Vol. 17 No. 4, pp. 329-346.

Lipton, M. and Ravallion, M. (1995), "Poverty and policy", in Behrman, J. and Srinivasan, T.N. (Eds), *Handbook of Development Economics vol 3B*, Elsevier, Amsterdam, pp. 2551-2657.

Lucas, R.E.B. (2001), "The effects of proximity and transportation on developing country population migrations", *Journal of Economic Geography*, Vol. 1 No. 3, pp. 323-339.

Mu, R. and van de Walle, D. (2007), "Rural roads and local market development in Vietnam", Policy Research Working Paper No. 4340, Development Research Group, World Bank, Washington, DC.

Mu, R. and van de Walle, D. (2011), "Rural roads and local market development in Vietnam", *The Journal of Development Studies*, Vol. 47 No. 5, pp. 709-734.

Nguyen, C. (2008), "Is a governmental micro-credit program for the poor really pro-poor: evidence from Vietnam", *The Developing Economies*, Vol. 46 No. 2, pp. 151-187.

Nguyen, C. (2011), "Estimation of the impact of rural roads on household welfare in Viet Nam", *Asia-Pacific Development Journal*, Vol. 18 No. 2, pp. 105-135.

Nguyen, C. (2013), "Which covariates should be controlled in propensity score matching? Evidence from a simulation study", *Statistica Neerlandica*, Vol. 67 No. 2, pp. 169-180.

Ravallion, M. (2001), "The mystery of the vanishing benefits: an introduction to impact evaluation", *The World Bank Economic Review*, Vol. 15 No. 1, pp. 115-140.

Smith, J. and Todd, P. (2005), "Does matching overcome LaLonde's critique of nonexperimental estimators?", *Journal of Econometrics*, Vol. 125 Nos 1-2, pp. 305-353.

Van de Walle, D. and Mu, R. (2007), "Fungibility and the flypaper effect of project aid: micro-evidence for Vietnam", *Journal of Development Economics*, Vol. 84 No. 2, pp. 667-685.

World Bank (1994), *World Development Report 1994: Infrastructure for Development*, Oxford University Press, New York, NY.

Zhao, Z. (2008), "Sensitivity of propensity score methods to the specifications", *Economics Letters*, Vol. 98 No. 2008, pp. 309-319.

(The Appendix follows overleaf.)

Appendix

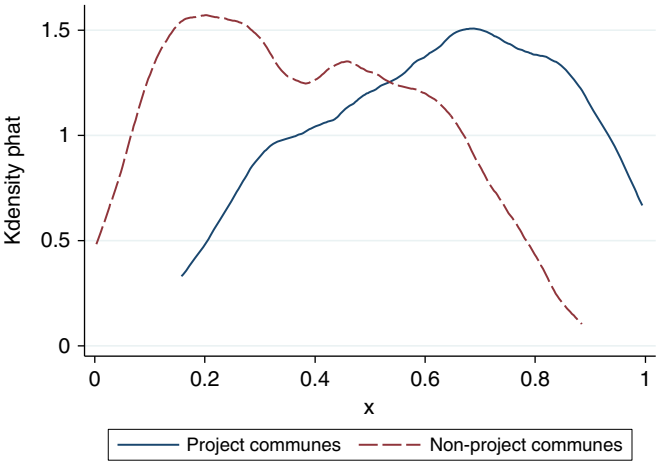

Figure A1.
Predicted propensity score

Source: Author's estimation

Table AI.
Outcome variable means using the same propensity score estimated from the replication study

Variable	1997 Project	1997 Non-project	1997 Difference between these and the original paper (%)	2001 Project	2001 Non-project	2001 Difference between these and the original paper (%)	2003 Project	2003 Non-project	2003 Difference between these and the original paper (%)
Local market development									
Market availability[a]	0.51	0.45	< 10	0.57	0.52	< 10	0.61	0.48	< 10
Market frequency	1.09	0.98	< 10	1.35	1.17	< 10	1.39	1.11	< 10
Shop	0.53	0.46	< 10	0.76	0.75	< 10	0.74	0.72	< 10
Bicycle repair shop[a]	0.75	0.65	< 10	0.80	0.78	< 10	0.86	0.81	< 10
Pharmacy	0.52	0.52	< 10	0.68	0.59	< 10	0.66	0.52	< 10
Restaurant	0.32	0.35	< 10	0.46	0.39	< 10	0.49	0.45	< 10
Women's hair dressing/Men's barber	0.54	0.52	> 10	0.74	0.70	> 10	0.76	0.69	> 10
Men and women's tailoring	0.76	0.71	> 10	0.82	0.76	< 10	0.84	0.75	< 10
Employment: % households whose main occupation is									
% farm households[a]	90.31	90.85	< 10	90.18	91.50	< 10	87.57	90.22	< 10
% trade households[a]	1.18	1.34	< 10	1.62	1.69	< 10	3.13	2.59	< 10
% service sector households[a]	0.97	0.52	< 10	1.36	1.55	< 10	2.80	1.61	< 10
School enrollments (%)									
Primary school completion (< 15 years)	62.19	60.70	> 10	29.77	31.98	> 10	39.00	34.99	> 10
Secondary school enrollment rate	86.53	84.30	> 10	93.58	91.87	> 10	94.53	93.21	> 10

Notes: Table AI replicates the estimates of Table II in Mu and van de Walle (2011); the sample consists of the 94 project and 95 non-project communes on common support as determined by propensity score matching. Many outcome variables are dichotomous referring to whether the outcome is present in the commune. The exceptions are: market frequency which takes the values 0 for no market, 1 for once per week or less, 2 for more than once a week and 3 for permanent market; the percentage of households in various occupations refers to their main source of income; the primary completion rate is defined as the share of children aged 15 years and under who completed primary school; the secondary school enrollment rate is the share of children who graduated from primary school in the previous year who are enrolled in secondary school. [a]Outcomes have the same value as in Table II in Mu and van de Walle (2011)
Source: Author's estimation

Table AII.
Outcome variable means using the same propensity score variable

Variable	1997 Project	1997 Non-project	1997 Difference between these and the original paper (%)	2001 Project	2001 Non-project	2001 Difference between these and the original paper (%)	2003 Project	2003 Non-project	2003 Difference between these and the original paper (%)
Local market development									
Market availability	0.51	0.44	0	0.57	0.51	0	0.62	0.46	0
Market frequency	1.07	1.00	< 10	1.30	1.17	< 10	1.38	1.08	< 10
Shop	0.54	0.44	< 10	0.79	0.73	< 10	0.76	0.71	< 10
Bicycle repair shop	0.76	0.65	0	0.80	0.78	0	0.87	0.81	0
Pharmacy	0.55	0.53	< 10	0.70	0.62	< 10	0.66	0.52	0
Restaurant	0.33	0.33	< 10	0.48	0.39	< 10	0.49	0.43	< 10
Women's hair dressing/Men's barber	0.53	0.51	> 10	0.74	0.69	> 10	0.77	0.68	> 10
Men and women's tailoring	0.76	0.72	> 10	0.82	0.75	< 10	0.82	0.75	< 10
Employment: % households whose main occupation is									
% farm households	89.53	90.67	0	89.65	91.07	0	87.02	90.15	0
% trade households	1.45	1.41	0	1.73	1.75	0	3.17	2.56	0
% service sector households	1.12	0.54	0	1.42	1.51	0	3.20	1.60	0
School enrollments (%)									
Primary school completion (< 15 years)	62.93	60.20	> 10	31.22	31.81	> 10	38.55	34.85	> 10
Secondary school enrollment rate	86.64	84.89	> 10	93.20	92.14	> 10	94.52	93.41	> 10

Notes: Table AII replicates the estimates of Table II in Mu and van de Walle (2011). The sample consists of the 85 project and 83 non-project communes on common support as determined by propensity score matching. Many outcome variables are dichotomous referring to whether the outcome is present in the commune. The exceptions are: market frequency which takes the values 0 for no market, 1 for once per week or less, 2 for more than once a week and 3 for permanent market; the percentage of households in various occupations refers to their main source of income; the primary completion rate is defined as the share of children aged 15 years and under who completed primary school; the secondary school enrollment rate is the share of children who graduated from primary school in the previous year who are enrolled in secondary school

Source: Author's estimation, Mu and van de Walle (2011)

Explanatory variables	Obs.	Mean	SD	Min.	Max.
Terrain: coast					
Mountains	200	0.5150	0.5010	0	1
Uplands	200	0.1800	0.3852	0	1
Plains	200	0.2550	0.4370	0	1
Province: Tra Vinh					
Lao Cai	200	0.1500	0.3580	0	1
Thai Nguyen	200	0.2000	0.4010	0	1
Nghe An	200	0.2500	0.4341	0	1
Binh Thuan	200	0.1250	0.3315	0	1
Kon Tum	200	0.1250	0.3315	0	1
Population (log)	199	8.5394	0.7088	6.86	10.15
Population density (log)	199	0.6083	1.3208	−2.51	3.00
Minority population share	199	0.4338	0.3974	0	1
National road passes through commune	200	0.3700	0.4840	0	1
Railway passes through commune without stop	200	0.1350	0.3426	0	1
Waterway passes through commune	200	0.2200	0.4153	0	1
Distance to province center (km) (log)	200	48.823	37.627	2	160
Commune has a passenger transport service	200	0.6150	0.4878	0	1
Share of households engaged in non-agricultural activities	200	0.0506	0.1226	0	1.00
Share of population working in government	199	0.0027	0.0049	0	0.04
Share of population working in private enterprises	199	0.0028	0.0165	0	0.19
Share of population working in state enterprises	199	0.0006	0.0024	0	0.02
Share of crop land	198	0.3191	0.2715	0.003	0.87
Share of perennial crop land	198	0.0544	0.0800	0	0.39
Land rental market exists in commune	200	0.4300	0.4963	0	1
Number of production organizations	200	1.2450	2.2383	0	14
Commune has a radio broadcasting station	200	0.2000	0.4010	0	1
Commune has a market	200	0.4850	0.5010	0	1
Agricultural crop land adversely affected by natural disaster (1996)	200	0.6200	0.4866	0	1
Commune has an agricultural bank	200	0.1300	0.3371	0	1
Number of official credit sources	200	2.2950	1.2270	0	5
Enrollment rate for children age 6 to 15	200	85.435	19.237	0	100
Commune has a lower secondary school	200	0.7350	0.4424	0	1
Predicted consumption per capita (log)	200	7.6354	0.2766	6.91	8.14
Share of households owning motorcycles	200	8.1613	8.3419	0	49.70
Road density (commune and district level roads)	199	0.0178	0.0235	0	0.16
Share of earth and car impassable roads in total road km	200	0.3752	0.3032	0	1

Source: Author's estimation

Table AIII. Summary statistics of explanatory variables in Logit regression of commune participation in the project

Explanatory variables	Coeff.	SE	Same sign as Van de Walle, D. and Mu, R. (2007)
Terrain: Coast	Reference		
Mountains	−0.331	1.194	Yes
Uplands	0.029	0.962	Yes
Plains	−0.834	1.047	Yes
Province: Tra Vinh	Reference		
Lao Cai	0.762	1.244	Yes
Thai Nguyen	0.699	1.162	Yes
Nghe An	1.296	1.211	Yes
Binh Thuan	1.226	1.079	Yes
Kon Tum	3.007***	1.046	Yes
Population (log)	0.814*	0.424	Yes
Population density (log)	0.536	0.411	Yes
Minority population share	2.608**	1.139	Yes
National road passes through commune	−1.827***	0.559	Yes
Railway passes through commune without stop	1.492*	0.772	Yes
Waterway passes through commune	0.343	0.551	Yes
Distance to province center (km) (log)	−0.006	0.0097	Yes
Commune has a passenger transport service	0.396	0.426	No
Share of households engaged in non-agricultural activities	0.371	1.407	No
Share of population working in government	−0.639*	0.365	Yes
Share of population working in private enterprises	−0.265*	0.155	Yes
Share of population working in state enterprises	0.711	0.741	Yes
Share of crop land	1.145	2.187	Yes
Share of perennial crop land	−1.899	3.552	No
Land rental market exists in commune	0.333	0.455	Yes
Number of production organizations	0.012	0.083	Yes
Commune has a radio broadcasting station	−1.079**	0.452	Yes
Commune has a market	0.338	0.431	Yes
Agricultural crop land adversely affected by natural disaster (1996)	0.202	0.448	Yes
Commune has an agricultural bank	0.977**	0.431	No
Number of official credit sources	−0.407***	0.152	Yes
Enrollment rate for children age 6 to 15	−0.012	0.018	Yes
Commune has a lower secondary school	0.167	0.626	Yes
Predicted consumption per capita (log)	1.030	1.159	Yes
Share of households owning motorcycles	0.076**	0.036	No
Road density (commune and district level roads)	−12.21	11.40	Yes
Share of earth and car impassable roads in total road km	1.102	0.712	Yes
Constant	−15.96*	9.418	Yes
Observations	198		
Pseudo R^2	0.204		

Source: Author's estimation

Table AIV. Logit regression of commune participation in the project

Table AV.
Impact heterogeneity: market and market frequency

Explanatory variables	Market			Market frequency		
	Model 1	Model 2	Same sign as the original paper	Model 1	Model 2	Same sign as the original paper
1997 value	−0.236** (−3.07)	−0.234** (−4.36)	Yes	−0.265** (−3.22)	−0.283** (−3.86)	Yes
Distance to central district	0.006 (1.57)	0.003 (0.87)	Yes	0.008 (0.53)		No
North province	−0.011 (−0.16)		Yes	−0.208 (−1.07)	−0.202 (−1.15)	Yes
Typology: mountain	0.038 (0.27)		Yes	0.229 (0.54)		Yes
Flood and storm prevalence	0.123** (2.04)	0.133** (2.58)	No	0.553** (2.90)	0.612** (3.74)	No
Population density	−0.098 (−0.09)		No	0.72 (0.18)		No
Ethnic minority share	−0.082 (−0.55)		Yes	−0.131 (−0.30)		Yes
Adult illiteracy rate	0.018 (0.060)		Yes	0.049 (0.07)		Yes
Share of households owning motorcycles	1.057** (2.10)	1.363** (2.90)	Yes	2.143 (1.43)	2.210** (1.99)	Yes
Credit availability	0.305* (1.74)	0.328 (1.60)	Yes	1.018 (1.47)	0.974* (1.70)	Yes
Length of road rehabilitated/100	−0.014 (−1.52)		Yes	−0.032 (−1.16)	−0.017** (−2.19)	Yes
Length squared/10,000	0.01 (0.50)		Yes	0.019 (0.31)		Yes
Month since project completion/100	0.044 (1.63)	0.018 (0.96)	Yes	0.165** (2.34)	0.172** (2.72)	Yes
Month squared/10,000	−0.045* (−1.71)	−0.02 (−1.10)	Yes	−0.174** (−2.51)	−0.183** (−2.92)	No
Constant	−0.976 (−1.52)	−0.505 (−1.03)	Yes	−3.689** (−2.01)	−3.792** (−2.51)	Yes
R^2	0.42	0.39		0.41	0.39	

Notes: Table AV replicates the estimates of Table IV in Mu and van de Walle (2011). The dependent variables are the 85 estimated commune specific impacts for 2003. Standard errors are clustered at the district level of which there are 29. Market is a zero/one dummy for whether a market exists in the commune. Market frequency takes the value 0 for no market; 1 for once a week or less; 2 for more than once a week and 3 for permanent market. *t*-Statistics are given in parentheses. *,**Significant at 10 and 5 percent levels, respectively

Source: Author's estimation

Table AVI.
Impact heterogeneity: shop and bicycle repair shop

Explanatory variables	Shop			Repair		
	Model 1	Model 2	Same sign as the original paper	Model 1	Model 2	Same sign as the original paper
1997 value	−0.962** (−7.01)	−0.969** (−8.03)	Yes	−0.738** (−6.27)	−0.729** (−6.48)	Yes
Distance to central district	0.004 (0.52)		Yes	−0.003 (−0.83)		Yes
North province	−0.084 (−0.67)		Yes	−0.012 (−0.18)		Yes
Typology: mountain	0.033 (0.17)		Yes	−0.016 (−0.28)		No
Flood and storm prevalence	−0.264** (−2.37)	−0.218** (−2.23)	Yes	0.111 (1.54)	0.106* (1.68)	Yes
Population density	2.100 (1.11)	1.381 (1.00)	Yes	0.242 (0.29)		Yes
Ethnic minority share	0.451** (2.12)	0.483** (3.22)	Yes	−0.047 (−0.37)		Yes
Adult illiteracy rate	−1.196** (−2.23)	−1.207** (−2.48)	Yes	−0.477 (−1.16)	−0.589 (−1.49)	Yes
Share of households owning motorcycles	−0.819 (−0.92)		No	0.716* (1.72)	0.714* (1.80)	Yes
Credit availability	0.983** (2.60)	0.894** (2.32)	No	−0.053 (−0.28)		Yes
Commune has a market in 1997	0.161 (1.18)	0.123 (1.15)	Yes	0.115** (2.16)	0.132** (2.35)	Yes
Length of road rehabilitated/100	−0.009 (−0.53)		Yes	−0.005 (−0.39)	−0.010** (−3.34)	Yes
Length squared/10,000	0.015 (0.33)		Yes	−0.006 (−0.19)		No
Month since project completion/100	0.068* (1.69)	0.057 (1.34)	Yes	0.063** (2.17)	0.062** (2.55)	Yes
Month squared/10,000	−0.064 (−1.60)	−0.054 (−1.29)	Yes	−0.063** (−2.26)	−0.061** (−2.65)	Yes
Constant	−1.681 (−1.63)	−1.448 (−1.34)	No	−0.957 (−1.29)	−1.008 (−1.57)	No
R^2	0.58	0.57		0.62	0.61	

Notes: Table AVI replicates the estimates of Table V in Mu and van de Walle (2011). The dependent variables are the 85 estimated commune specific impacts for 2003. Standard errors are clustered at the district level of which there are 29. All outcomes refer to availability in the commune. *t*-Statistics are given in parentheses; *,**Significant at 10 and 5 percent levels, respectively

Source: Author's estimation

Table AVII.
Impact heterogeneity – pharmacy and restaurant

Explanatory variables	Pharmacy Model 1	Pharmacy Model 2	Same sign as the original paper	Restaurant Model 1	Restaurant Model 2	Same sign as the original paper
1997 value	−0.656** (−4.61)	−0.660** (−5.38)	Yes	−0.614** (−4.59)	−0.570** (−5.82)	Yes
Distance to central district	−0.002 (−0.36)		Yes	−0.006 (−0.83)	−0.003 (−0.44)	Yes
North province	0.095 (0.84)		Yes	0.171 (1.21)		Yes
Typology: mountain	−0.094 (−0.61)		No	0.019 (0.10)		Yes
Flood and storm prevalence	−0.095 (−0.73)		Yes	0.023 (0.18)		No
Population density	0.858 (0.57)		Yes	−1.017 (−0.37)		Yes
Ethnic minority share	0.043 (0.21)		No	0.068 (0.36)		Yes
Adult illiteracy rate	−0.788 (−1.51)	−0.910** (−2.34)	Yes	−0.376 (−0.54)		Yes
Share of households owning motorcycles	0.369 (0.36)	0.483 (0.77)	Yes	−0.454 (−0.57)	−0.826 (−1.25)	No
Credit availability	0.295 (0.80)		Yes	−0.022 (−0.05)		Yes
Commune has a market in 1997	0.304** (2.53)	0.348** (3.07)	Yes	0.242** (2.58)	0.258** (2.72)	No
Length of road rehabilitated/100	−0.009 (−0.66)	−0.004 (−1.03)	Yes	0.009 (0.60)		Yes
Length squared/10,000	0.010 (0.30)		Yes	−0.012 (−0.35)		Yes
Month since project completion/100	0.055 (1.33)	0.042 (1.14)	Yes	0.035 (0.76)	0.015** (2.95)	No
Month squared/10,000	−0.055 (−1.37)	−0.042 (−1.17)	Yes	−0.022 (−0.47)		No
Constant	−0.881 (−0.88)	−0.605 (−0.69)	No	−1.110 (−1.02)	−0.565* (−1.73)	Yes
R^2	0.50	0.44		0.44	0.39	

Notes: Table A7 replicates the estimates of Table V in Mu and van de Walle (2011). The dependent variables are the 85 estimated commune specific impacts for 2003. Standard errors are clustered at the district level of which there are 29. All outcomes refer to availability in the commune. *t*-statistics are given in parentheses. *,**Significant at 10 and 5 percent levels, respectively

Source: Author's estimation

Table AVIII.
Impact heterogeneity – service availability

Explanatory variables	Pharmacy			Restaurant		
	Model 1	Model 2	Same sign as the original paper	Model 1	Model 2	Same sign as the original paper
1997 value	−0.857** (−8.13)	−0.818** (−8.81)	Yes	−0.853** (−6.28)	−0.849** (−7.03)	Yes
Distance to central district	−0.002 (−0.35)	−0.000 (−0.10)	Yes	0.002 (0.32)		No
North province	−0.213* (−1.73)	−0.154* (−1.94)	Yes	−0.011 (−0.14)		Yes
Typology: mountain	0.110 (0.78)		Yes	−0.076 (−0.96)	−0.092 (−1.13)	No
Flood and storm prevalence	0.037 (0.33)		Yes	−0.063 (−0.85)		Yes
Population density	2.711* (1.72)	2.415** (2.15)	Yes	−0.080 (−0.08)		No
Ethnic minority share	−0.156 (−0.89)		Yes	−0.212 (−1.50)	−0.203 (−1.54)	Yes
Adult illiteracy rate	−0.671 (−1.17)	−0.615 (−1.36)	Yes	−1.078** (−2.21)	−1.011** (−2.37)	Yes
Share of households owning motorcycles	0.993* (1.69)	1.015** (2.75)	Yes	0.470 (1.24)	0.613* (1.72)	Yes
Credit availability	0.224 (0.78)		Yes	0.344 (1.57)	0.312* (1.80)	Yes
Commune has a market in 1997	0.092 (0.94)	0.093 (1.12)	Yes	0.055 (0.79)		Yes
Length of road rehabilitated/100	0.003 (0.21)	−0.005 (−1.14)	Yes	−0.005 (−0.42)	−0.006** (−2.01)	No
Length squared/10,000	−0.011 (−0.31)		Yes	−0.001 (−0.02)		Yes
Month since project completion/100	0.000 (0.00)		Yes	0.077** (2.60)	0.080** (2.26)	Yes
Month squared/10,000	−0.001 (−0.05)		Yes	−0.077** (−2.74)	−0.080** (−2.38)	Yes
Constant	0.495 (0.67)	0.514** (3.54)	Yes	−1.041 (−1.35)	−1.078 (−1.26)	No
R^2	0.58	0.55		0.63	0.62	

Notes: Table AVIII replicates the estimates of Table VI in Mu and van de Walle (2011). The dependent variables are the 85 estimated commune specific impacts for 2003. Standard errors are clustered at the district level of which there are 29. All outcomes refer to availability in the commune. *t*-Statistics are given in parentheses. *,**Significant at 10 and 5 percent levels, respectively

Source: Author's estimation

Table AIX.
Impact heterogeneity: employment

Explanatory variables	Farming		Same sign as the original paper	Services		Same sign as the original paper	Trade		Same sign as the original paper
	Model 1	Model 2		Model 1	Model 2		Model 1	Model 2	
1997 value	−0.118 (−1.55)	−0.118* (−1.70)	Yes	−0.308 (−0.86)	−0.235 (−0.67)	Yes	−0.315 (−0.92)	−0.198 (−0.64)	Yes
Distance to central district	−0.010 (−0.08)		Yes	−0.090 (−1.07)	−0.099 (−1.37)	Yes	−0.014 (−0.24)		Yes
North province	−2.474 (−1.25)	−2.712 (−1.51)	Yes	1.172 (0.97)	1.732 (1.25)	Yes	−1.985** (−1.96)	−1.170 (−1.29)	Yes
Typology: mountain	−2.086 (−0.66)		Yes	−1.744 (−1.08)	−2.829** (−2.22)	Yes	0.820 (0.44)		Yes
Flood and storm prevalence	−2.534 (−1.17)	−3.189* (−1.91)	Yes	2.401* (1.79)	2.713* (1.65)	Yes	−0.140 (−0.13)		No
Population density	−37.572 (−0.87)	−21.625 (−0.80)	Yes	28.058 (0.79)		Yes	28.536 (0.92)		Yes
Ethnic minority share	0.313 (0.08)		Yes	0.132 (0.07)		Yes	0.906 (0.55)		No
Adult illiteracy rate	3.369 (0.41)		Yes	5.120 (0.90)	5.318 (1.14)	No	−3.947 (−0.88)	−4.938* (−1.67)	No
Share of households owning motorcycles	−12.164 (−0.79)	−11.355 (−0.72)	Yes	23.844** (2.83)	23.250** (3.16)	Yes	12.545 (1.64)	11.471* (1.73)	Yes
Credit availability	2.259 (0.39)	3.202 (0.51)	Yes	−8.094** (−2.16)	−9.260** (−2.59)	Yes	−4.298 (−1.56)	−4.783* (−1.76)	Yes
Commune has a market in 1997	−2.765 (−1.52)	−2.543* (−1.69)	Yes	−0.476 (−0.33)		Yes	1.612* (1.76)	1.486* (1.93)	Yes
Length of road rehabilitated/100	−0.006 (−0.03)		Yes	0.038 (0.20)		No	0.002 (0.02)		Yes
Length squared/10,000	−0.078 (−0.21)		No	−0.030 (−0.08)		No	−0.020 (−0.11)		Yes
Month since project completion/100	0.324 (0.41)	−0.015 (−0.22)	Yes	0.532 (1.22)	0.569* (1.65)	Yes	0.410 (1.04)	0.295 (0.96)	Yes
Month squared/10,000	−0.324 (−0.41)		Yes	−0.662 (−1.59)	−0.686** (−1.98)	Yes	−0.439 (−1.15)	−0.317 (−1.10)	Yes
Constant	7.292 (0.34)	14.134* (1.91)	Yes	−8.280 (−0.74)	−7.830 (−0.93)	Yes	−8.649 (−0.87)	−4.857 (−0.61)	Yes
R^2	0.21	0.19		0.29	0.27		0.19	0.16	

Notes: Table AIX replicates the estimates of Table VII in Mu and van de Walle (2011). The dependent variables are the 85 estimated commune specific impacts for 2003. Standard errors are clustered at the district level of which there are 29. *t*-Statistics are given in parentheses. All outcomes refer to availability in the commune. *,**Significant at 10 and 5 percent levels, respectively

Source: Author's estimation

Table AX.
Impact heterogeneity – schooling

Explanatory variables	Secondary school enrollment			Primary school completion		
	Model 1	Model 2	Same sign as the original paper	Model 1	Model 2	Same sign as the original paper
1997 value	−0.915** (−8.96)	−0.961** (−14.47)	Yes	−0.999** (−9.83)	−0.932** (−8.91)	Yes
Distance to central district	0.068 (0.39)		No	−0.190 (−0.40)		Yes
North province	2.052 (0.79)	3.268 (1.58)	No	7.322 (0.76)	5.700 (0.95)	Yes
Typology: mountain	1.061 (0.27)		No	−6.896 (−0.66)		Yes
Flood and storm prevalence	1.703 (0.66)		No	16.182** (2.49)	16.717** (2.87)	Yes
Population density	7.611 (0.22)		No	58.566 (0.46)		Yes
Ethnic minority share	−5.906 (−1.50)	−6.363* (−1.88)	Yes	3.152 (0.26)		Yes
Adult illiteracy rate	7.168 (0.46)		No	43.797 (1.22)	21.340 (0.93)	No
Share of households owning motorcycles	−8.500 (−0.55)	−8.096 (−0.50)	No	97.884* (1.66)	94.383* (1.94)	Yes
Credit availability	3.814 (0.60)	5.519 (0.97)	Yes	4.141 (0.22)		No
Commune has a market in 1997	1.940 (0.86)	1.720 (0.98)	Yes	9.817 (1.29)	9.344 (1.42)	Yes
Length of road rehabilitated/100	−0.024 (−0.10)	−0.173* (−1.85)	Yes	−0.633 (−0.73)	−0.368 (−1.35)	No
Length squared/10,000	−0.286 (−0.45)		No	0.327 (0.16)		No
Month since project completion/100	−0.192 (−0.25)	0.084 (0.82)	No	0.188 (0.07)		No
Month squared/10,000	0.274 (0.37)		No	−0.422 (−0.16)		No
Constant	80.464** (3.93)	81.035** (9.67)	Yes	52.136 (0.76)	45.989** (3.50)	Yes
R^2	0.87	0.86		0.71	0.67	

Notes: Table AX replicates the estimates of Table VIII in Mu and van de Walle (2011). The dependent variables are the 85 estimated commune specific impacts for 2003. Standard errors are clustered at the district level of which there are 29. All outcomes refer to availability in the commune. *t*-statistics are given in parentheses. *,**Significant at 10 and 5 percent levels, respectively

Source: Author's estimation

Outcomes	2001 PS kernel matched DD	2001 *t*-ratio	2001 Original estimates in Mu and van de Walle (2011)	2003 PS kernel matched DD	2003 *t*-ratio	2003 Original estimates in Mu and van de Walle (2011)
Market availability	0.023	0.537	0.03	0.068	1.380	0.08*
Market frequency	0.124	0.941	0.08	0.137	0.930	0.23*
Shop	−0.203	−1.617	0.01	−0.194*	−1.827	0.08
Bicycle repair shop	−0.057	−1.027	−0.06	−0.044	−0.626	0.02
Pharmacy	0.096	1.337	0.04	0.260**	2.367	0.12
Restaurant	0.145**	2.007	−0.01	0.089	0.829	0.01
Women's hair dressing/ Men's barber	0.077	1.032	−0.07	0.102	1.373	0.18**
Men and women's tailoring	0.012	0.248	0.11	0.034	0.585	0.10
% farm households	−1.961	−0.943	0.05	−3.035	−1.418	−2.04*
% trade households	0.064	0.083	0.03	1.218	1.582	0.36
% service sector households	−0.044	−0.086	−1.54	1.353**	2.306	1.68**
Primary school completion (< 15 years)	7.150	0.850	0.15**	13.848**	1.943	0.17**
Secondary school enrollment rate	2.948	0.834	0.10	0.837	0.290	0.05

Notes: The sample consists of 85 project and 83 non-project communes on common support as determined by propensity score matching. The propensity score is estimated by the logit model in Table AII. *t*-Ratio of kernel matching is obtained from bootstrapping (100 repetitions). *,**Significant at 10 and 5 percent levels, respectively

Table AXI. PS kernel matched DD: bandwidth = 0.01

Outcomes	2001 PS kernel matched DD	2001 *t*-ratio	2001 Original estimates in Mu and van de Walle (2011)	2003 PS kernel matched DD	2003 *t*-ratio	2003 Original estimates in Mu and van de Walle (2011)
Market availability	0.028	0.776	0.03	0.079**	2.003	0.08*
Market frequency	0.137	1.398	0.08	0.171	1.477	0.23*
Shop	−0.173	−1.553	0.01	−0.178*	−1.850	0.08
Bicycle repair shop	−0.059	−1.152	−0.06	−0.038	−0.575	0.02
Pharmacy	0.074	1.030	0.04	0.206*	1.883	0.12
Restaurant	0.139**	1.946	−0.01	0.073	0.795	0.01
Women's hair dressing/ men's barber	0.068	0.894	−0.07	0.092	1.231	0.18**
Men and women's tailoring	0.004	0.080	0.11	0.033	0.551	0.10
% farm households	−1.208	−0.686	0.05	−2.782	−1.529	−2.04*
% trade households	−0.191	−0.244	0.03	1.069	1.544	0.36
% service sector households	−0.032	−0.068	−1.54	1.330**	2.439	1.68**
Primary school completion (< 15 years)	4.141	0.551	0.15**	11.986*	1.718	0.17**
Secondary school enrollment rate	1.565	0.526	0.10	0.890	0.308	0.05

Notes: The sample consists of 85 project and 83 non-project communes on common support as determined by propensity score matching. The propensity score is estimated by the logit model in Table AII. *t*-ratio of kernel matching is obtained from bootstrapping (100 repetitions). *,**Significant at 10 and 5 percent levels, respectively

Table AXII. PS kernel matched DD: bandwidth = 0.03

Outcomes	2001 PS kernel matched DD	2001 *t*-ratio	2001 Original estimates in Mu and van de Walle (2011)	2003 PS kernel matched DD	2003 *t*-ratio	2003 Original estimates in Mu and van de Walle (2011)
Market availability	0.028	0.819	0.03	0.082**	2.196	0.08*
Market frequency	0.134	1.430	0.08	0.173	1.503	0.23*
Shop	−0.103	−1.011	0.01	−0.115	−1.272	0.08
Bicycle repair shop	−0.071	−1.373	−0.06	−0.058	−0.813	0.02
Pharmacy	0.045	0.601	0.04	0.140*	1.681	0.12
Restaurant	0.129	1.614	−0.01	0.038	0.393	0.01
Women's hair dressing/ men's barber	0.047	0.627	−0.07	0.069	0.926	0.18**
Men and women's tailoring	0.000	0.003	0.11	0.022	0.329	0.10
% farm households	−0.534	−0.341	0.05	−2.263	−1.527	−2.04*
% trade households	−0.161	−0.261	0.03	0.692	1.343	0.36
% service sector households	−0.325	−0.759	−1.54	0.877*	1.890	1.68**
Primary school completion (< 15 years)	0.552	0.086	0.15**	8.896	1.260	0.17**
Secondary school enrollment rate	0.915	0.293	0.10	0.607	0.205	0.05

Table AXIII.
PS kernel matched DD: bandwidth = 0.09

Notes: The sample consists of 85 project and 83 non-project communes on common support as determined by propensity score matching. The propensity score is estimated by the logit model in Table AII; *t*-Ratio of kernel matching is obtained from bootstrapping (100 repetitions). *,**Significant at 10 and 5 percent levels, respectively

About the author

Cuong Viet Nguyen holds PhD and MSc degrees in development economics from Wageningen University, the Netherlands. Dr Cuong has extensive experience in impact evaluation, poverty analysis, ethnic minority issues, education and health issues. Dr Cuong recent studies have been published in well-respected journals such as *the American Political Science Review*, *World Bank Economic Review*, *the Journal of Comparative Economics the Journal of Health Economics*, *World Development*, *the Journal of Development Studies*, etc. Dr Cuong Viet Nguyen can be contacted at: c_nguyenviet@yahoo.com