

Impacts of Rural Road on Household Welfare in Vietnam: Evidence from a Replication Study MS1945

Summary

The manuscript primary aims to quantify the effect of rural roads on local market development in Vietnam by replicating the estimation results based on Mu and Van de Walle (2011). The main motivation of the manuscript is derived from the reproducibility principle of scientific methods to validate and confirm previous empirical findings. The manuscript attempts to reconstruct the variables from the original raw data and re-estimate the results by applying difference-in-difference with propensity score matching while considering the same model specification as Mu and Van de Walle (2011). The author endeavors to extend the analysis by proposing an alternative model specification, varying the bandwidth parameter with regard to kernel matching and estimating the effect of rural roads on two additional outcomes, namely access to credits and migration. The main findings of the manuscript are that the estimation results of Mu and Van de Walle (2011) are robust and the improvements of rural roads do not significantly affect access to credits and migration.

General comments

In my point of view, the manuscript is discussing a policy relevant topic and stresses the importance of replication. However, my main concern about the manuscript is that it lacks focus on both the appropriate application of the chosen statistical technique and the implication of a suitable economic theory with respect to the extension of the additional outcomes.

With regard to the former, several authors underline that the assessment of the key assumptions of propensity score matching and the performance

of the chosen matching method are essentially important to yield unbiased results (Rosenbaum and Rubin, 1985; Stuart, 2010; Caliendo and Kopeinig, 2008; Austin, 2011; Lampach and Morawetz, 2016). A main concern with propensity score matching is not the variable selection of the model specification but rather to ensure the resulting balance (balancing property) between the treated and control group (Augurzky and Schmidt, 2001; Stuart, 2010). Furthermore, it is crucial to assess the ignorability assumption given the set of covariates and based on available methods (plausibility test, sensitivity analysis).

With regard to the latter, it is hard to understand why the manuscript focuses on access to credits and migration without discussing more specifically the economic theory and detailing the hypothesis to be tested deduced from the theory. Mu and Van de Walle (2011) seek to test the hypothesis of *transport-induced local-market development*. I wonder whether this hypothesis can be attributed to the additional outcomes presented in the manuscript. On the other hand, the reason why both outcomes are chosen is barely mentioned in the manuscript (*These outcomes are important for livelihood and non-farm diversification of rural households, and can provide policy-relevant findings*). The original study explains that both types of improvements, namely rehabilitated roads and new roads have been commonly considered in the empirical analysis. I wonder whether data is available to estimate the effects of both types separately on the market development. This finding could add value to the original study and would imply policy-relevant implication in the context of the government's investment decision, whether it might be more reasonable to spend public funds for the improvement of existing roads or whether new roads are necessary to link isolated and arable regions with local markets of higher activity.

The empirical results states that rural road has no-significant effect on both outcome measures. But what justifies this no-significant effect? Is it primarily driven by a potential violation of the unconfoundedness assumption or unbalanced covariates? Or which other reasons (e.g. variable creation, the original survey was not implicitly construct to measure the effect of rural road on access to credits and migration, missing data) might lead to this non-significant effect? In particular, the manuscript is lacking a substantial discussion on this matter.

To enhance the justification of the chosen covariates regarding the first extension, it might also be useful to consider the economic theory and previous findings to build up the model (Caliendo and Kopeinig, 2008). In this way, the choice of the covariates could be additional justified and helps later on to link the estimated results with the theory.

A general suggestion to improving the quality of the paper is to provide more information about the necessary assumption checks of the propensity score matching method by using available guidelines from the literature. Furthermore, the extension of the empirical analysis should be also motivated by the economic theory. Especially, a comprehensive discussion about policy relevant implications deduced from the findings would add value to the manuscript. I wonder whether similar findings could be expected for different Asiatic developing countries with similar local market structures and whether the establishment of a transnational cooperation could be beneficial to create new opportunities to increase local market development.

Furthermore, I would like to answer the following two questions:

i *Is the replication done to a high standard of professional competency?*

In my opinion the replication is basically done competently and the manuscript describes very detailed and extensive the undertaken steps. In general, the manuscript stresses problems with the data and documentation of the do-files and tries to deal with it in the best possible way. However, one specific point in the pure replication is difficult to understand: why have been both variables (*Women's hair dressing* and *Men's barber*) generated in the manuscript as an aggregated covariate, while the original study has treated them separately (as two different covariates).

ii *Do the robustness checks/extensions add value to understanding the original study?*

The manuscript deals with several robustness checks, but however, without assessing the key assumptions of the propensity score matching method it is hard to argue that the estimated results with the alternative model specification and the variation of the bandwidth parameter related to kernel matching yield robust and unbiased results. The manuscript makes a modest contribution and is lacking focus on essentially important aspects of the application of the statistical technique. Furthermore, the extension of two additional outcomes were only briefly sketched out in the manuscript and do not add value to the original study. In particular, the result of no-significant effect of rural road on access to credits and migration has not been properly discussed and explained.

In my view, the current manuscript covers a crucial topic and demonstrates that the replication of the original results is hampered by the poor documentation of the variable creation. However, the manuscript is lacking substantial robustness checks of propensity score matching assumptions and is missing specifically an added value in the context of a replication study. Throughout the review, several issues have been identified which would need substantial improvements and which make the manuscript not suitable for publication in the journal given its current format.

References

- Augurzký, B. and C. Schmidt (2001). The propensity score: A means to an end. IZA Discussion Papers 271, Institute for the Study of Labor (IZA).
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research* 46(3), 399–424. PMID: 21818162.
- Caliendo, M. and S. Kopeinig (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* 22(1), 31–72.
- Lampach, N. and U. B. Morawetz (2016). Credibility of propensity score matching estimates. an example from fair trade certification of coffee producers. *Applied Economics* 48(44), 4227–4237.
- Mu, R. and D. Van de Walle (2011). Rural roads and local market development in vietnam. *The Journal of Development Studies* 47(5), 709–734.
- Rosenbaum, P. R. and D. B. Rubin (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 39(1), pp. 33–38.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science* 25(1), 1–21.