

Date: 10 October 2016

There has been an increasing concern about data sets and program codes in the empirical studies in social science research. There are few journals require authors publish data and program codes. In addition, for a large number of articles published in these journals, the published data are not original ones. Thus it's important for researchers to keep their original data and program codes for replication. This paper raises the importance issues of keeping data and program codes in social science.

Overall the paper is well written. The research question is policy relevant. I would like to recommend this paper to be revised and resubmitted for publication.

Here are my comments:

First, to replicate the original research, one should follow exactly the same estimation method which the paper successfully did, specifically the paper should replicate the same model specifications, plug in all variables that the original paper used and follow the same way of variable calculation/construction, and run the models for the same sample. In addition, the author needs to check if any extreme observations were dropped in the original paper. In PSM estimation should use the same bandwidth as the original paper used. This is to make sure we compare like with the like. The difference either large or small may be due to either any of the above disparities or due to (dis)honest/reliability of the original study. This is to avoid that the original paper's authors may come back and claim that the difference is resulted from any of the differences in model specification, variable construction or sample.

Second, extension of the replicate paper is the contribution. The paper discussed inclusion of additional variables that the author believed relevant and exclusion of some controlling variables or covariates that are believed irrelevant based on the PSM literature (the common support and unconfoundedness assumptions).

The author may look at these papers to support his choice of inclusion or exclusion of covariates. Caliendo & Kopeinig (2008); Imbens (2004); Crump, Hotz, Imbens, and Mitnik (2009); Rosenbaum & Rubin (1983); Rubin & Thomas (1996); Bryson, Dorsett, & Purdon, (2002); Dehejia and Wahba (1999) and Heckman, Ichimura and Todd (1997).

Third, apart from the existing modelling extension as a contribution, is it possible for the author to re-define road access e.g. access by trucks/vans. I think the accessibility by small trucks/vans may be much more meaningful for economic growth and household welfare than by individual transportation vehicles such as bike or motorbikes.

Finally, the paper should provide more policy recommendations. For example, should governments continue to invest in rural road? Although the program improves the market access, does it improve the rural welfare of households such as income and consumption? It's very useful if the authors can look at the aggregate welfare outcomes such as income or consumption.

References:

Bryson, A., Dorsett, R., & Purdon, S. (2002). *The use of propensity score matching in the evaluation of active labor market policies*. London, England: University of Westminster, Policy Studies Institute and National Centre for Social Research.

Caleindo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31-72.

- Crump, R., Hotz, V., Imbens, G., & Mitnik, O. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96(1), 187-199.
- Dehejia, R., & Wahba, S. (1999). Casual effects in non-experimental studies: Re-evaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448), 1053-62.
- Heckman, J., Ichimura, H., & Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training program. *Review of Economic Studies*, 64(4), 605-654.
- Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*, 86(1), 4-29.
- Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.
- Rubin, D., & Thomas, N. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics*, 52(1), 249-264.