Critique Papers on Causal Inferences

Tony Tan

Centre for Educational Measurement

University of Oslo

MAE4051: Selected Topics in Educational Measurement

Prof Jan-Eric Gustafsson

9 April 2021

Instrumental Variables

Hanandita, W., & Tampubolon, G. (2014). Does poverty reduce mental health? An instrumental variable analysis. *Social Science & Medicine*, *113*, 59–67. https://doi.org/10.1016/j.socscimed.2014.05.005

Summary

Hanandita and Tampubolon (2014)

Causal Question

Validity

Construct Validity

Internal Validity

External Validity

Statistical Conclusion Validity

Appropriateness of Methods

Conclusion

Fixed-effect Regression

White, M. P., Alcock, I., Wheeler, B. W., & Depledge, M. H. (2013). Would you be happier living in a greener urban area? A fixed-effects analysis of panel data. *Psychological Science*, *24*(6), 920–928. https://doi.org/10.1177/0956797612464659

Summary

White et al. (2013)

Causal Question

Validity

Construct Validity

Internal Validity

External Validity

Statistical Conclusion Validity

Appropriateness of Methods

Conclusion

Propensity Scores

Sullivan, A. L., & Field, S. (2013). Do preschool special education services make a difference in kindergarten reading and mathematics skills?: A propensity score weighting analysis. *Journal of School Psychology*, *51*(2), 243–260. https://doi.org/10.1016/j.jsp.2012.12.004

Summary

Sullivan and Field (2013) attempted to investigate the marginal benefit created by a special education program mandated by the US *Individuals with Disabilities Education Act* (IDEA). The 1986 amendments to the IDEA legislation imposed legal requirement on the states to create a sequence of intervention programs targeting three stages of development for children with needs: 1) from birth to 2-year-old, 2) 3-year to kindergarten entry, and 3) kindergarten to 21-year-and-11-month-old. This paper focused on the middle segment of the intervention sequence named preschool special education services and followed children longitudinally through the delivery period. The study began by measuring children's probability of being admitted into the special education program using a logit regression, and at the end of the intervention period, compared the average performance scores in maths and reading of the children who actually received the preschool special education treatment against the average scores of those who have not. Using the propensity score weighting technique, the authors interrogated the counterfactual question of "what if" the treatment were never applied and reached an conclusion that children in treatment group would have been better off academically had they not received any intervention at all. This disappointing result lent itself to an existing literature of similar negative findings evaluating special education effectiveness, cautioning the rosy objectives of preschool special education services originally marketed by policy makers.

Causal Question

Would the children who received special education services have been better off academically, on average, had they not received such services?

Validity

Internal and External Validity

Historically, internal validity referred to inferences about whether "the experimental treatments make a difference in this specific experimental instance" while external validity asked "to what populations, settings, treatment variables, and measurement variables can this effect be generalized" (Campbell & Stanley, 1963, p. 5). Cook and Campbell (1979) advanced the idea of internal validity to the question whether the covariation observed between the independent and dependent variables were resulted from a *causal* relationship, whereas external validity further asks whether such cause-effect relationship holds over certain variation in persons, settings, treatment variables, and measurement variables.

In order to support an inference that the observed covariation between A and B reflects a causal relationship, Shadish et al. (2002) prescribed a trifecta that 1) A preceded B in time, 2) A covaries with B, and 3) no other explanations for the relationship are plausible. It is too often the third strand that undermines the internal validity of inference making—the relationship between A and B is not causal because it could have occurred even in the absence of the treatment and that it could have led to the same outcomes that were observed for the treatment. Amongst the list of potential threats to internal validity (Shadish et al., 2002, p. 54–61), maturation presents the strongest challenge to Sullivan and Field (2013). Children who have been identified as "in need" so early in life ("early onsetters") can be reasonably believed to be in possession of different developmental profiles from children who showed needs later in life ("late onsetters"). As participants in the treatment group mature, gaps in academic performance may well emerge out of such delayed developmental trajectories with or without education services. It is therefore not preschool interventions that "caused" lower academic scores but the two-tier growth profiles that did. Failing to rule out such alternative, and rather plausible, explanation weakens the internal validity of inference made by the authors.

Weak external validity has also been acknowledged by the authors in Section 4.2 of the paper. Inferences can only been drawn, first of all, over children with mild to moderate impairments resultant from the sample deletion procedure; while it were the children with the most severe impairment that policy makers wished to monitor and retain ("interaction of the causal relationship with units" by Shadish et al. (2002)). Secondly, Sullivan and Field (2013)

evaluated only the academic performance of young children to the exclusion of other developmental markers such as motor-behavioural and social-affective skills—all vital policy objectives along with reading and maths scores, if not more important, for kindergarten-entry age children ("interaction of the causal relationship with outcomes"). Lastly, the averaging procedure in calculating ATT washed out important differences across race and socio-economic groups, important factors reported by prior literatures as non-ignorable ("context-dependent mediation").

Construct Validity

Statistical Conclusion Validity

By Cook and Campbell (1979), statistical conclusion validity refers to the appropriateness of statistical techniques employed by the researcher for the purposes of inferring whether the presumed independent and dependent variables indeed covary.

Appropriateness of Methods

Conclusion

References

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Rand McNally.

Hanandita, W., & Tampubolon, G. (2014). Does poverty reduce mental health? An instrumental variable analysis. *Social Science & Medicine*, *113*, 59–67. https://doi.org/10.1016/j.socscimed.2014.05.005

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth Cengage Learning.

Sullivan, A. L., & Field, S. (2013). Do preschool special education services make a difference in kindergarten reading and mathematics skills?: A propensity score weighting analysis. *Journal of School Psychology*, *51*(2), 243–260. https://doi.org/10.1016/j.jsp.2012.12.004

White, M. P., Alcock, I., Wheeler, B. W., & Depledge, M. H. (2013). Would you be happier living in a greener urban area? A fixed-effects analysis of panel data. *Psychological Science*, *24*(6), 920–928. https://doi.org/10.1177/0956797612464659