Social Science Research 50 (2015) 382–394

Contents lists available at ScienceDirect

Social Science Research

journal homepage: www.elsevier.com/locate/ssresearch

Instrumental variables estimates of peer effects in social networks

Weihua An

Departments of Sociology and Statistics, Indiana University Bloomington, 1020 East Kirkwood Avenue, Bloomington, IN 47405-7103, USA

ARTICLE INFO

Article history:
Received 8 November 2012
Revised 19 July 2014
Accepted 11 August 2014
Available online 27 August 2014

Keywords:
Instrumental variable
Causal inference
Peer influence
Smoking
Social Networks

ABSTRACT

Estimating peer effects with observational data is very difficult because of contextual confounding, peer selection, simultaneity bias, and measurement error, etc. In this paper, I show that instrumental variables (IVs) can help to address these problems in order to provide causal estimates of peer effects. Based on data collected from over 4000 students in six middle schools in China, I use the IV methods to estimate peer effects on smoking. My design-based IV approach differs from previous ones in that it helps to construct potentially strong IVs and to directly test possible violation of exogeneity of the IVs. I show that measurement error in smoking can lead to both under- and imprecise estimations of peer effects. Based on a refined measure of smoking, I find consistent evidence for peer effects on smoking. If a student's best friend smoked within the past 30 days, the student was about one fifth (as indicated by the OLS estimate) or 40 percentage points (as indicated by the IV estimate) more likely to smoke in the same time period. The findings are robust to a variety of robustness checks. I also show that sharing cigarettes may be a mechanism for peer effects on smoking. A 10% increase in the number of cigarettes smoked by a student's best friend is associated with about 4% increase in the number of cigarettes smoked by the student in the same time period.

1. Introduction

Sociology has been interested in studying peer effects for a long time. Early examples can be traced back to Coleman's studies on diffusion of innovations (Coleman et al., 1957) and adolescent society (Coleman, 1960) and Duncan, Haller, and Portes' study on peer aspirations (Duncan et al., 1968). Ever since, there has been a large number of studies seeking to examine peer effects in a variety of contexts, including job attainment (Granovetter, 1973), spreading of health behaviors and outcomes (Christakis and Fowler, 2007; Liu et al., 2010), agreement on political preference (Yamaguchi, 2013), etc.

But it turns out to be very difficult to causally estimate peer effects because of contextual confounding, peer selection, simultaneity bias, and measurement error, to name only a few. In this paper, I point out that measurement error in the outcome can have double-detrimental effects: not only biasing the estimated peer effects toward zero but also inflating their standard errors. The net result is that it becomes more difficult to reject the null hypothesis of no peer effects. I also show how instrumental variable (IV) methods can help to address the aforementioned problems and provide causal estimates of peer effects. The basic idea is to utilize the exogenous variations in the IVs to facilitate identification. The success of the IV methods relies critically on the strength and exogeneity of the IVs. In this paper I argument the conventional IV methods with a variety of robustness checks including sensitivity analysis, falsification tests, matching analysis, and sub-sample

E-mail address: weihuaan@indiana.edu

http://dx.doi.org/10.1016/j.ssresearch.2014.08.011
0049-089X/

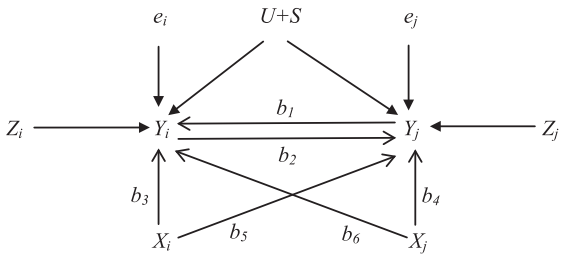

Fig. 1. Difficulties in identifying endogeneous peer effects.

analysis in order to provide cross-validating causal evidence. Perhaps more distinctly, in the survey I used to collect data for this study I have designed questions that would provide information that can be used to construct potentially strong IVs. I have also designed a question to solicit respondent's friendship preference, which enables me to evaluate in a direct fashion the exogeneity of the IVs.

Specifically, I applied the IV methods to estimating peer effects on smoking based on data collected from over 4000 students in six middle schools in China. I find that correcting measurement error is crucial for both accurate and precise estimates of peer effects. The estimated peer effects are both larger and statistically more significant when a refined measure of smoking is used. The OLS estimate accounting for peer selection shows a peer effect of about one fifth. The IV estimate is larger, indicating that if a student's best friend has smoked within the past 30 days, that student was about 40 percentage points more likely to smoke in the same time period. I also show that sharing cigarettes may be a mechanism for peer effects on smoking. The number of cigarettes smoked by a student would increase by about 4% in the past 30 days if the number of cigarettes smoked by his or her best friend increased by 10% in the same time period.

This paper proceeds as follows. I first review past research on IV methods, with close attention paid to their applications in studying peer effects. I also outline the motivations for using IV methods to address the various difficulties in estimating peer effects. Then I introduce the data and statistical methods. The fourth section presents the results, where I also provide a variety of analysis for checking the robustness of the findings. Last, I conclude and discuss possible limitations of this study.

2. Difficulties in Identifying endogenous peer effects

Given there is a strong correlation in peers' outcomes, the correlation can come from four possible sources. First is from peer influence (or so-called endogenous peer effects), namely, an individual's outcome is affected by peer's outcomes. Second is from exogenous peer effects, namely, an individual's outcome is affected by peer's exogenous characteristics (e.g., gender, age, family background) (Ammermueller and Pischke, 2009). Third is from contextual effects, namely, peers' outcomes are simultaneously affected by their shared social and environmental factors (e.g., neighborhood, school policy). And last is from selection of peers based on factors related to the outcome.[1]

Much of the research on peer effects has been focused on how to appropriately estimate endogenous peer effects while teasing out other confounding effects (Coleman et al., 1957; Bulte and Lilien, 2001; Christakis and Fowler, 2007, 2013; Cohen-Cole and Fletcher, 2008a, 2008b; Fowler and Christakis, 2008; An, 2011a; VanderWeele and An, 2013). In particular, as to whether smoking is socially contagious, despite many studies (e.g., Ennett and Baumann, 1993; De Vries et al., 2003; Pollard et al., 2010) have found a strong correlation in peer's smoking behaviors, others (e.g., Ennett and Baumann, 1994; De Vries et al., 2006; Mercken et al., 2009) show that peer selection has predominately produced or significantly contributed to the correlation.

Besides difficulties in teasing out competing causes, two other problems, which have by and large been ignored in the literature, may complicate the estimations of peer effects. One is simultaneity, namely, an individual may affect peers in the same time being affected by them. As a consequence, the individual's outcome may appear also as a predictor in the regression predicting peer's outcomes and vice versa. It has been shown that under simultaneity, estimates by conventional Ordinary Least Squares (OLS) are inconsistent, because peer's outcomes are correlated with the error terms by construction (Wooldridge, 2009: 207).

The other problem is measurement error. In the case of simple regressions with just one outcome variable and one explanatory variable, measurement error in the outcome alone will reduce the precision of the estimates while measurement error in the explanatory variable alone will bias its estimated coefficient toward zero (Wooldridge, 2010: 70–75).[2] Measurement error in models of peer effects has a unique feature. Measurement error in the outcome can transfer into

[1] Different peer effects are relevant for different policy interests. If the policy interest is to achieve changes through affecting the group composition, then exogenous peer effects may be more relevant (Graham et al., 2010). If the interest is to target the behavior of a group, leaving group composition intact, then endogenous peer effects are more relevant.

[2] The bias direction may be unpredictable if there is measurement error in multiple variables. But the general consequences of measurement error may still carry on in many empirical contexts.