Economics of Education Review 54 (2016) 113–123

Contents lists available at ScienceDirect

Economics of Education Review

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

The role of testing noise in the estimation of achievement-based peer effects☆

Hongliang Zhang

Department of Economics, Hong Kong Baptist University, Hong Kong

ARTICLE INFO

Article history:
Received 26 October 2015
Revised 19 April 2016
Accepted 22 April 2016
Available online 30 May 2016

JEL Classification:
I20
I21
C13

Keywords:
Peer effects
Student achievement
Measurement error
Mean reversion

ABSTRACT

I demonstrate that in the value-added estimation of peer effects using lagged peer achievement, testing noise may generate another bias in addition to the well-known attenuation bias. Such a bias, which I refer to as the "reversion bias," may arise when some of a student's current peers happen to be his/her former peers whose performances in the baseline test were subject to the same common testing noise as the student's own. I propose a solution to overcome this problem by exploiting only the variation in the new peers' portion of the overall peer quality. Using real-world data, I illustrate the existence of this bias and demonstrate the proposed solution.

1. Introduction

Peer influences have been investigated for various outcomes in different settings,[1] among which the effects of classroom/school peers on a student's own achievement have received the greatest attention. While empirical studies abound, they provide mixed evidence of the existence, magnitude, and even sign of peer influences among students in schools (for surveys, see Epple & Romano, 2011; Sacerdote, 2011). The failure of previous studies to arrive at a consensus partly reflects the formidable identification challenges confronted in the estimation of peer effects (e.g., Angrist, 2014; Brock & Durlauf, 2001; Moffit, 2011). In particular, the reciprocal nature of peer interactions, known as the reflection problem (Manski, 1993), hinders differentiation between endogenous and contextual effects. To circumvent this problem, prior research has often resorted to estimating the reduced-form relationship between student achievement and predetermined measures of peer composition. One strand of research focuses on

☆ An earlier version of this paper was circulated under the title "Peer effects on student achievement: An instrumental variable approach using school transition data." I thank Joshua Angrist, David Autor, Abhijit Banerjee, Jiahua Che, Weili Ding, Esther Duflo, Amy Finkelstein, Chih-Sheng Hsieh, Tao Jin, Rongzhu Ke, Lars Lefgren, Steven Lehrer, Weifeng Li, Karen Polenske, Christopher Taber, Yu Zhu, and seminar participants at Fudan University, Hong Kong University, Lingnan University, MIT, Peking University, Zhejiang University, Royal Economic Society Conference, and CUHK Symposium on the Economics of Education for their valuable comments and suggestions. I am also grateful to the co-editor and two anonymous referees for helpful comments and feedback, and to the UK Department of Education for access to the National Pupil Database. I acknowledge financial support from the Hong Kong Research Grant Council General Research Fund (No. 458610). All remaining errors are my own.

E-mail address: zhang.hongliang@gmail.com

[1] Examples include college/military academy students' GPA (e.g., Foster, 2006; Han & Li, 2009; Lyle & David, 2007; Sacerdote, 2001; Stinebrickner & Stinebrickner, 2006), juveniles' behavior (e.g., Gaviria & Raphael, 2001; Glaeser, Sacerdote, & Scheinkman, 1996; Kremer & Levy, 2008), professionals' performances in tournaments (e.g., Guryan, Kroft, & Notowidigdo, 2009), worker productivity (Mas & Moretti, 2009), movie sales (Moretti & Enrico, 2011), etc.

http://dx.doi.org/10.1016/j.econedurev.2016.04.008
0272-7757/

examining the effects of contextual peer characteristics such as race, gender, immigration status, and family background (e.g., Angrist & Lang, 2004; Gould, Lavy, & Paserman, 2009; Hoxby, 2000; Lavy & Schlosser, 2011; McEwan, 2003). Another strand of research employs value-added models to estimate the effects of lagged peer achievement (e.g., Arcidiacono & Nicholson, 2005; Hanushek, Kain, Markman, & Rivkin, 2003; Lefgren, 2004; Vigdor & Nechyba, 2007). However, both strands largely ignore the measurement problem in the peer variables that may arise from missing data and/or measurement error. To the best of my knowledge, Ammermueller and Pischke (2009), Micklewright, Schnepf, and Silva (2012), and Sojourner (2013) are the only examples of studies that consider the measurement problem in the peer variables. While Ammermueller and Pischke (2009) and Micklewright et al. (2012) deal with both missing data and measurement error related to contextual peer characteristics in the first strand of research, in the second strand of research Sojourner (2013) considers only missing data related to lagged peer achievement.

In this paper, I extend the investigation of the role of measurement error to the estimation of achievement-based peer effects using lagged peer achievement. I point out that test scores of students from the same peer group (i.e., school or classroom) are subject to common testing noise arising from group-specific common influences having only transitory effects on test scores, e.g., a dog barking on the playground on the test day, a local flu pandemic, the coincidental overlap between the test and instruction contents, etc. While the existence of such common testing noise has been well documented in the school accountability literature (e.g., Betts & Danenberg, 2002; Kane & Staiger, 2002), in which conventional evaluation approaches are demonstrated to yield misleading assessments (Chay, McEwan, & Urquiola, 2005),[2] it has been underappreciated and largely ignored in the peer effects literature. To the best of my knowledge, this paper is the first to illustrate and highlight the relevance of common testing noise to the estimation of achievement-based peer effects. Specifically, I show that the conventional specifications using lagged peer achievement yield biased estimates of peer effects when (as is often the case) a student's current peer group consists of some former peers whose lagged performances were subject to the same transitory influences as experienced by the student him/herself. Given the existence of common testing noise in lagged performances, the continuing presence of a student's former peers in his/her current peer group leads to a spurious positive correlation between the student's own lagged achievement and mean peer lagged achievement, the workhorse variable in the linear-in-means model considered in this paper. That is, a higher mean peer lagged achievement implies more favorable testing noise in a student's own lagged achievement, which further indicates poorer achievement progress due to mean reversion, leading to a negative bias in the estimate of the coefficient on mean peer lagged achievement. Since this bias is the result of the mean-reversion property of test scores, I refer to it as the "reversion bias."[3]

Compared with the attenuation bias caused by missing data or classical measurement error considered in prior research, the reversion bias considered herein poses a more substantial challenge to the estimation of peer effects as it could even reverse the sign of the estimator. In addition to explicating the existence of the negative reversion bias, I propose a solution to overcome this problem by exploiting the variation in mean peer lagged achievement caused exclusively by new peers. Specifically, I partition mean peer lagged achievement into the old and the new peers' portions and use only the variation in the latter component (i.e., the product of the new peers' share and mean lagged achievement). I show that under some plausible mean independence conditions regarding the testing noise, mean ability of old peers, and unobserved determinants of learning, the estimated coefficient on the new peers' portion of mean peer lagged achievement is immune from the reversion bias.

To illustrate the existence of the reversion bias in the conventional estimators and demonstrate the proposed solution, I analyze achievement-based peer effects in England's secondary schools using the National Pupil Database (NPD) collected by the UK's Department of Education. The NPD contains students' test scores on both the Key Stage 2 (KS2) national exam taken at the end of primary school (sixth grade) and the Key Stage 3 (KS3) national exam taken in ninth grade at secondary school. As the same test is taken by all students of the same cohort, coincidental overlap between the test and instruction contents alone would lead to common testing noise in the scores of students from the same school. Using four cohorts of students in the NPD who finished ninth grade between 2005 and 2008, the standard school fixed-effect estimations show large, negative, and significant coefficients on mean peer lagged achievement, suggesting not only the existence of the reversion bias but also its dominance over the true peer effects (if the latter exist). The NPD data set also includes information on the primary school where students took their KS2 national exam, thus allowing a distinction to be made between old and new peers in secondary school. Performing my proposed estimation to overcome the reversion-bias problem yields modest, positive, and significant coefficients on the new peers' portion of mean lagged peer achievement, indicating that positive

[2] Chay et al. (2005) show that a difference-in-differences assessment of a school intervention program in Chile targeting low-performing schools overstates the program effect because of mean reversion in testing noise.

[3] Different from the reversion bias in the coefficient on mean peer lagged achievement considered here, Fruehwirth (2014) illustrates biases in the estimated coefficients on contextual peer characteristics conditional on lagged peer achievement. That is, when contextual peer characteristics and lagged peer achievement are both included in the estimation, the estimated coefficients on contextual peer characteristics are biased toward 0 or even take counterintuitive signs. The intuition is as follows: conditional on lagged peer achievement, more (less) favorable peer contextual characteristics partially capture a lower (higher) level of the unobservable peer quality that is not fully accounted for by lagged peer achievement, thus biasing the estimated coefficients on contextual peer characteristics toward 0 or even negative.